CS425 MP1 Report

Group Members: Sarthak Chakraborty, Raunak Shah

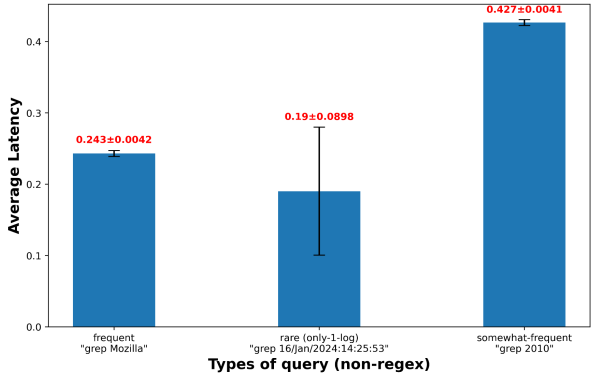
Design (Algorithm):

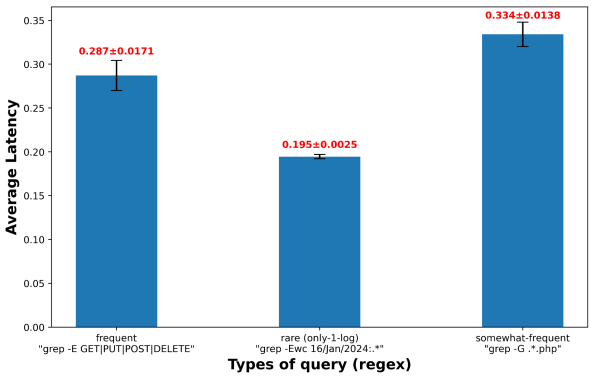
Our design includes a server and client that can be run on any VM. We start the server on each VM with a different id representing the machine number. The client can then be started on any VM, and a grep request can be sent. The client sends the grep message via sockets to each of the servers (including the server on its own VM). The server receives the grep request and runs the grep command locally. The relevant output is then sent back on the same socket to the querying client. The output for all servers is processed by the client sequentially. The client prints the line counts corresponding to each VM and the total number of matching lines across all VMs. This approach considers fault tolerance as well because if one of the servers shuts down, that particular client-server socket will not connect, and no data will be sent to the server. The client will send the grep message to the next VM in the loop.

Unit Tests:

We have queried frequent, rare (which is present in only one log), and somewhat frequent query patterns in the given apache log files. We have also queried frequent, infrequent and somewhat frequent regex patterns.

Experimental Results (with first 4 VMs):





Both the plots show the same trend where querying frequent patterns takes more time than rare patterns. However, somewhat frequent patterns experience more latency than the other two, which is against our intuition. The standard deviation for each query pattern is also not very large.