CS425-MP1

Vashishth Goswami and Jorge Velez

Design

Our choice of language, Rust 🦀, was motivated by its strong support for concurrency, memory safety, and performance, all of which are essential for a distributed log search system. The server leverages TCP sockets to listen for client connections and efficiently searches log files using regular expressions. Each client connects to multiple servers concurrently, using threads to query each server in parallel and process the results. The design ensures scalability by distributing the search workload across multiple VMs, while the use of Rust's safety guarantees helps prevent common issues like data races and memory leaks.

Unit Test

The unit test validates the core functionality of the distributed log query system by running a simulated query across multiple vmi.log files from different machines. It verifies that the system correctly counts occurrences of the "GET" requests in each log, ensuring accurate results from each machine. All test scripts can be found in our directory *test*
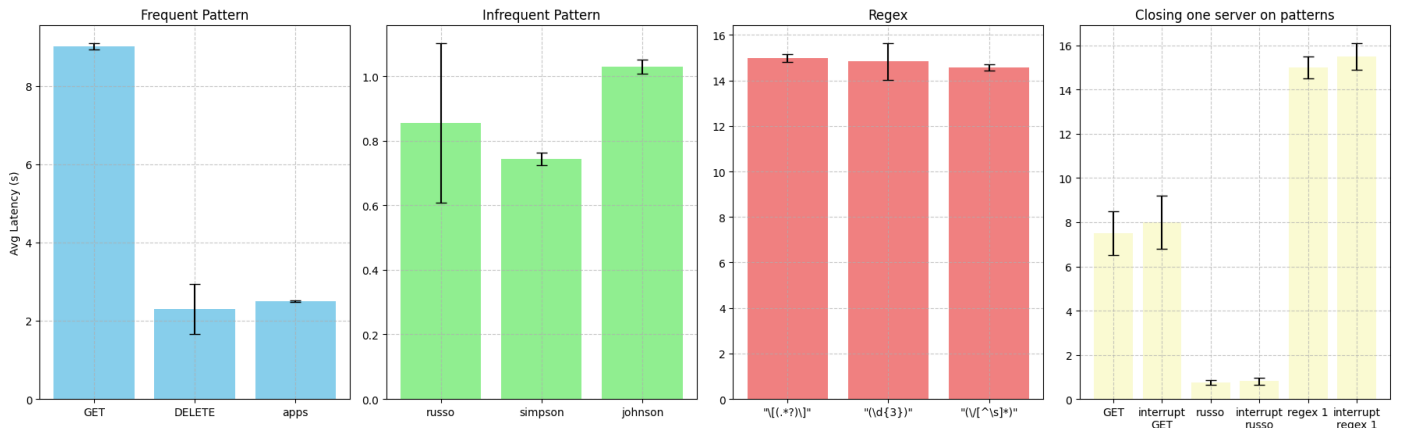
Analysis of Average Query Latency

Figure 1: Plot of the average latency with standard deviation

Frequent Patterns: These patterns have higher counts (in the 150,000s), leading to longer processing times. This is expected, as more frequent queries typically require more resources and time to handle.
Infrequent Patterns: With counts in the hundreds or low thousands, these patterns show faster processing times. This is due to fewer resources being required, allowing for quicker query handling.
Regex Patterns: These patterns have even higher counts (around 250,000), resulting in the longest processing times. Regex operations are computationally intensive, especially with large datasets, which explains the increased latency.
Server Interruption: When a server is closed during the test, latency decreases. This is because the query bypasses the closed server, reducing the overall processing time. This scenario highlights the impact of server availability on query performance.
Standard Deviation: The error bars indicate variability in the data, suggesting fluctuations in query times due to network conditions or server load.