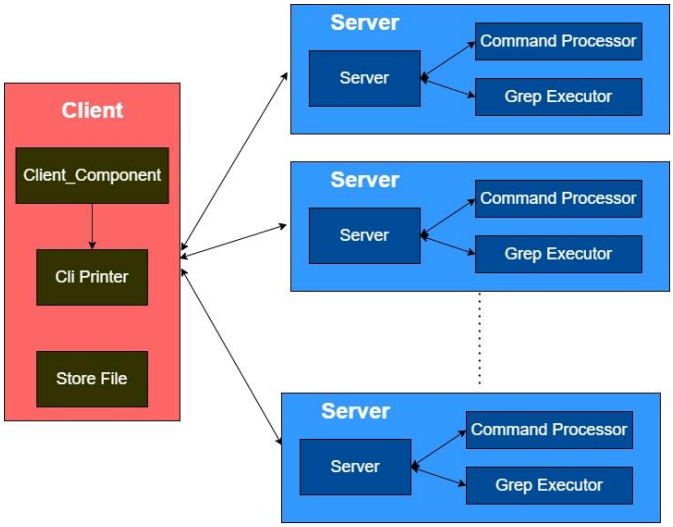
CS 425 MP1 Report
GA6 - Saurabh Darekar(sdare1) & Ankit Chavan(auc3)

Design

We have designed a system comprising two distinct processes: the server process and the client process. The program is in Java using socket programming to establish communication between the client and the server. On the client side, command is passed with command line parameters, and it initiates multiple threads to send command to each target machine, respectively. These threads connect to the respective servers that are continuously running in the background, awaiting connections. Upon receiving a connection, the server processes the command using a Command Processor, executes it via a Grep Executor, and returns the results to the client. The client then prints the no. of lines and saves the output to individual files for respective VM. All threads on Client machine operate concurrently to manage the command execution across different machines. After receiving and printing the results from all machines, the client terminates.



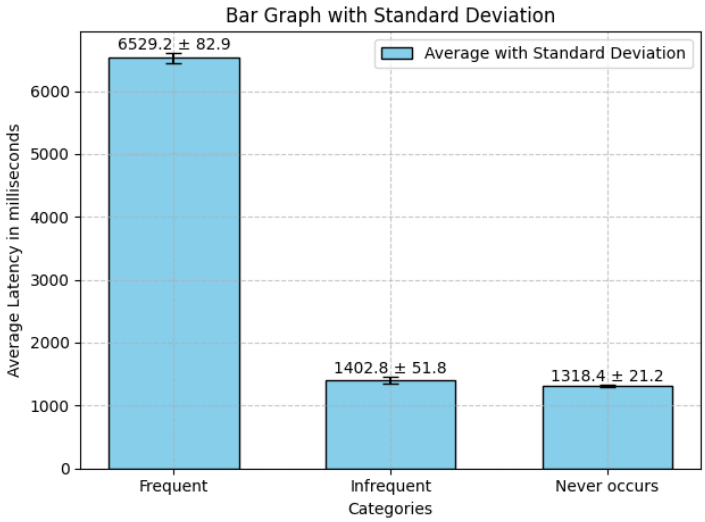
Unit Test Cases

We have added two JUnit classes named GrepTest and GrepExecutorTest. The GrepExecutorTest contains six test cases that cover various functionalities: testing option -n, option -c, option -i, a combination of options -n and -i (both -ni and -n -i), and a regular expression test. The GrepTest class contains a JUnit test that verifies the end-to-end flow of the distributed log querier.

Analysis

	Frequent	Infrequent	Never occurs	Reg Exp
Test 1	6437	1405	1324	82034
Test 2	6600	1369	1305	84574
Test 3	6613	1353	1333	85352
Test 4	6448	1400	1289	83938
Test 5	6548	1487	1341	84220
Average	6529.2	1402.8	1318.4	84023.6
Std Dev	82.88969	51.78031	21.208489	1232.071

All Time values are in Milliseconds.



Plot for Non-Regex queries.

We performed grep queries on 4 machines and calculated the average time to process the query. We found that while executing a frequent pattern it takes more time than the infrequent and a keyword never occurs. This might be due to increased time in transferring the data and saving the large file. Regular Expressions also take a large amount of time as there might be backtracking, lookahead or lookbehind going on due to which it is taking an more time than others. Thus, to avoid irregularities in the graph we have excluded Regular Expression.