

## MP1 report - Team 05

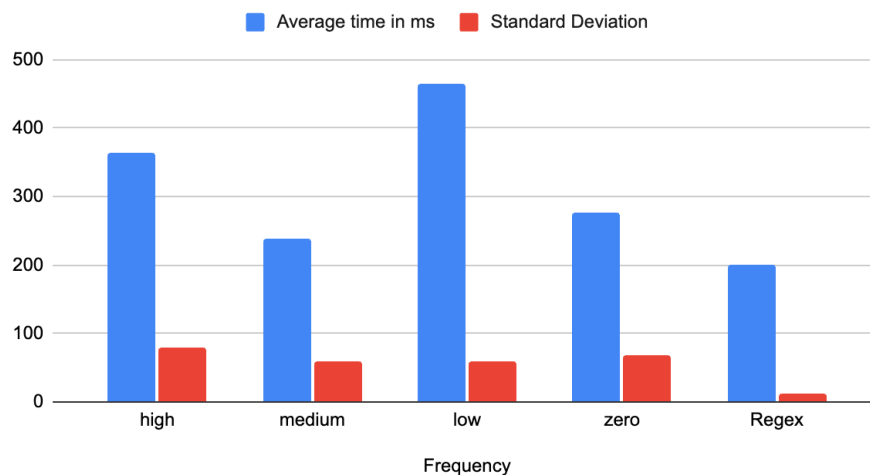
**Design:** Our log searcher system will need to have the server part running on all the machines that contain the log file. The client will create multiple sockets and connect to all the servers. So we are only creating the socket connections that are needed. It will also send the grep query to all the servers. Servers will do the grep search locally and send the result back to the client. Therefore, the search is done in parallel to get better performance and the client will only need to combine all the results. The grep search is done by running a process as a Linux grep command. The client will print out the results for the user.

**Unit tests:** We tested the accuracy of our local query results with different keyword/regex frequencies that occur in different numbers of files. We tested for 0%/10%/50%/80%/100% frequencies in one/half/all files.

**Latency data:** We use the vm2-5.log files which are about 60MB each. We expect the latency to be about the same in these cases because for any keyword, we will need to go through all the lines in the file.

Frequency	high	medium	low	zero	Regex
Time used in ms	440, 473, 328, 286, 289	181, 236, 254, 180, 341	509, 512, 493, 350, 457	371, 333, 200, 275, 205	222, 193, 196, 194, 193

Average time in ms and Standard Deviation



From the data, we can see that low and high-frequency keywords will take more time. The standard deviation is about the same for the four cases with words and the regex has a low standard deviation value, so the results are pretty consistent. The result is different from what we expected. We guess the reason might be that it takes more time to search for an infrequent pattern and it takes more time to process the matching lines for frequent patterns. So those two cases need more time.