Report

CS425-MP1

For each VM, we have an executable server and client. We first open server on all ten VMs, and on any of VM, we execute client and the client will send grep request to all ten VMs(including itself). Those VMs will then run grep command on their local log files and return output to the VM that sends the request. We achieved this by muti-threading the client program so that client receives outputs simultaneously from all other VMs.

We have 9 unit tests. They are the combination of two groups: group that represents the frequency of the query pattern( “rare”, “somewhat frequent”, “frequent”) and group that represents how many log files the query occurs( “one”, “Some”, “All”). That’s why they are named “rareOne”, “rareSome”, “rareAll”, “somewhatFrequentOne”, “somewhatFrequentSome”, “somewhatFrequentAll”, “frequentOne”, “frequentSome”, “frequentAll”. In each log file, there are about three to five known lines followed by three to four random unknown lines. If the query occurs in log file, it occurs in known lines and the frequency depends on whether it’s a rare test, somewhat frequent test or frequent test.

Report: In the graph we generated, query latency includes printing time and each query data point is the average that was ran 5 times on 4 VMs.



Figure 1. Figure 2.

Figure 1. describes the relationship between query data points and average latency. 9 queries are selected and sorted by total line number result from grepping those queries. When total line number is less than one thousand, despite there are some small fluctuations, the average latency rises slowly overall. But the average latency increased significantly if resulted line number increased suddenly. This may due to the reason that more time is consumed in printing the result. This satisfy our expectations. Figure 2. describes the relationship between query data points and standard deviation. The result we expect is standard deviation increases as total line number rises. However, this figure does not satisfy our expectations. We agree standard deviation is dominated by network fluctuations which is unpredictable.