MP1 Report

Design: Developed a coordinator less, distributed grep, client server based system where a single client VM connects to multiple server VMs through a unique *(ip_addr:port_number)* combination. Initially, the user inputs a grep command through the client to the servers. Each server then executes the grep command on its local log files, searches the pattern and returns the count of occurrences from each server. Apart from the count of occurrences(-c) the user can try out other options of grep too.

Test cases: 9 scenarios combined by 3 pattern frequencies (rare, some, frequent) and 3 types of occurrences on machines (only one machine, some machines, and all machines). We use 5 machines from 2~6 to perform unit tests. Take the rare pattern as example:

For vm2: test the only-one-machine scenario and all-machines scenario. We generated a log file with the string "rare1" in vm2, and query all machines to search for "rare1" to test the only-one machine scenario.

For vm3~5: designed to test the some-machines scenario and all-machines scenario. We generated log files with the string "rare2" in vm3~5, and query all machines to search for "rare2" when testing for some-machines scenario.

For vm6: designed only for the all-machines scenario, so the log file contains the string "rare". For the all-machines scenario, we just query all machines with the string "rare", and all log files containing either "rare", "rare1", "rare2" will be counted. Same logic applies to the other 2 pattern frequencies

Graphs:
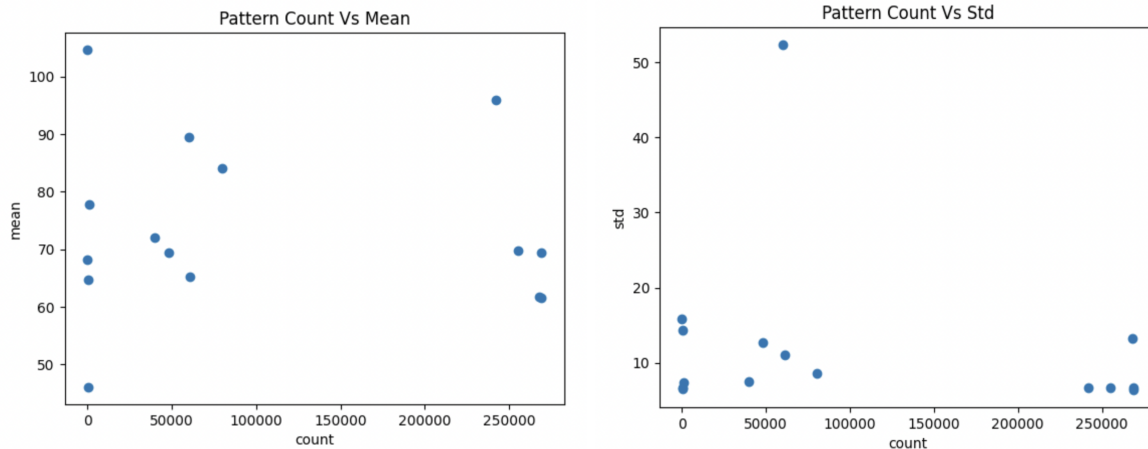
The above graphs are divided into 3 sections *viz.* The data points aligning at x=0 are (i) rare patterns, data points in the 50 - 100 thousand range are (ii) somewhat frequent patterns and the latter the (iii) frequent patterns. From the above graphs it can be seen that the mean and std for (i) remains constant, for (ii) it is decreasing over count and for (iii) it's almost stays constant