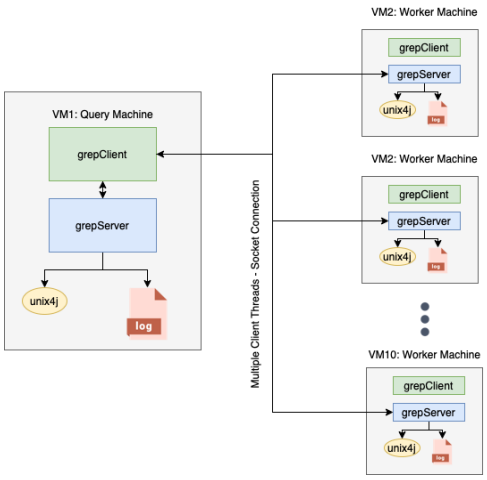
CS 425 - Distributed Systems

Machine Problem 1: Distributed Grep Query System

Group 65 : Sanjit Kumar (netid: sanjitk3) ; Aruna Parameswaran (netid: aruna2)

Design



The distributed grep system is designed with 2 major modules in each machine - a server and a client module. When querying from any machine the client module is used - it obtains the grep command, contains network configuration information of the other VMs (hostnames, ports, file paths) and spawns a thread for N servers to distribute the input command via a socket connection. The individual remote (and local) server(s) wait till a command is received and run the grep command locally with the unix4j library. The result is then sent back via the same socket as the response which the client thread prints before it terminates.The main thread waits for a volatile static variable to collect line counts from each thread (machine) to print the total line count at the end before the main thread terminates.

Unit Tests

Unit tests are executed during build time for each module, and have been written using JUnit. Each module has separate unit tests, and we have also implemented two distributed unit tests that exercise the whole end-to-end connection flow. The first unit test verifies whether the client-server connection is established, and that data is received back from the server by the client. The second unit test covers the output validation portion, and compares the grep output when run on a single machine against the consolidated output received by the client from across all machines.

	frequent	infrequent	never occurs	regex
0	1206.00	1050.00	1041.00	1300.00
1	1191.00	1049.00	1027.00	1222.00
2	1193.00	1047.00	1129.00	1204.00
3	1196.00	1051.00	1068.00	1230.00
4	1186.00	1057.00	1031.00	1425.00
mean	1194.40	1050.80	1059.20	1276.20
standard deviation	7.44	3.77	42.17	90.82

Latency Metrics

Mean query latency across different query classes ("frequent", "infrequent", "never occurs" and "regex") was measured to be 1145.15 milliseconds with an average standard deviation of 109.4 milliseconds. (See here).