Machine Programming 1 – Distributed Log Querier

Yu Tao, Jie Yin
yutao2@illinois.edu; jiey3@illinois.edu

Design

All 10 VMs are deployed as servers. One client is chosen randomly.

On the client side, the grep command associated with desired patterns is sent to 10 servers (including the client itself) simultaneously by multithreads. The results sent back from 10 servers are then written into 10 files. Fail-stop server could be correctly identified.

On the server side, servers consistently wait to listen to the client. Once the bind is success, servers could execute the commands, i.e. grep, sent from the client. The results are then sent back to the client.

Unit Test

Two servers are deployed with pre-saved logs. There are five cases in the test: frequent pattern, somewhat frequent pattern, infrequent pattern, pattern appearing only in one log and regex expression. The first three cases only differ in the frequency that appear in logs. Querying are performed for all five cases and the result is compared with the expected value. The program runs correctly if two values match.

Performance

Configuration: four servers with 60MB log file each are queried by the client. Time latencies are recorded for querying frequent patterns, infrequent patterns and somewhat patterns respectively. Five trials are performed for each case. The figure illustrates the time latency for each trial, associated with the blue average line. The bar indicates the standard deviation.

Frequent pattern: “GET”, total lines: 2246560;

Infrequent pattern: “222”, total lines: 3112;

Somewhat frequent pattern: “images”, total lines: 1014840;

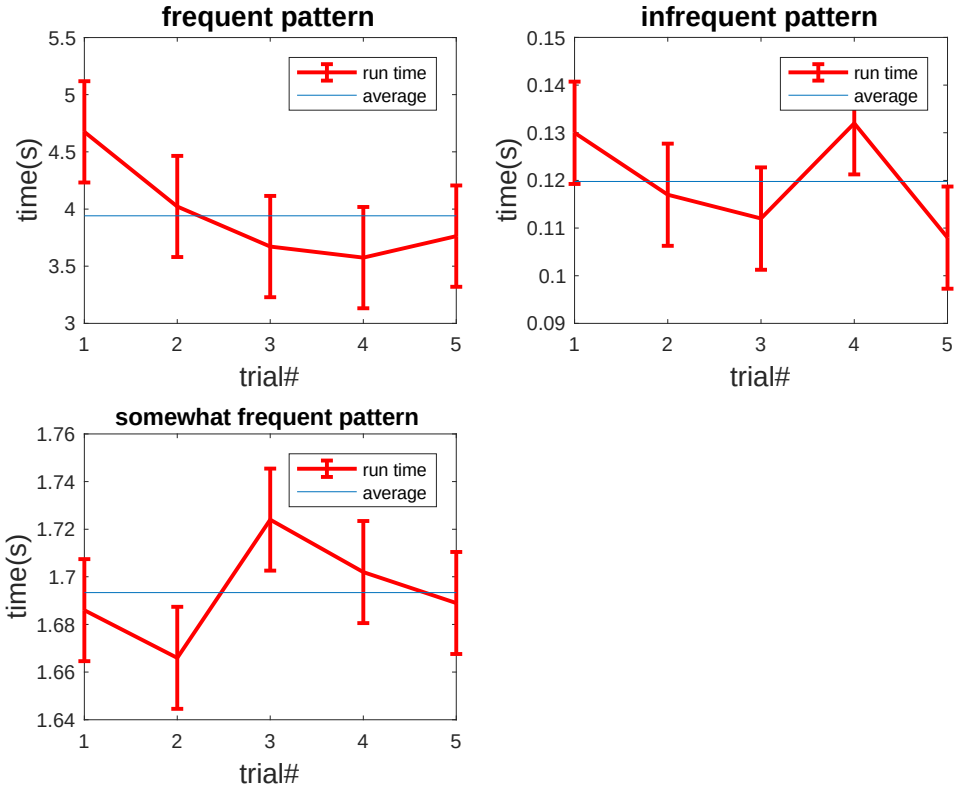

Figure 1 Performance evaluation for three cases