

# CS425, Fall 2017 Machine Programming 1

## Goup64: wc4, siyub2

### Design

The functional requirement of this project is to implement distributed grep of different log files on different VMs.

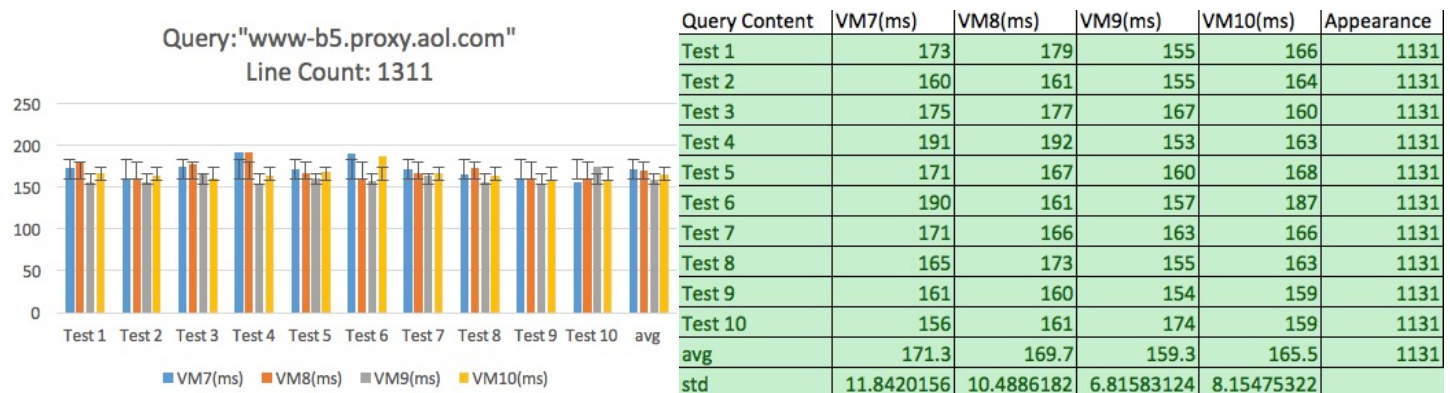
For communication between VMs, we used Client-Server architecture. The query machine will be considered as client and all others computers to be queried from will run a server. The client takes a query string from user and send it to all servers, servers run grep regarding to that query string locally and return the grep result back to the client. Specifically, we store all the matched lines locally on VM being queried and only send back the total line count of matched lines.

Additionally, both client and server is multithreaded. The server is multithreaded such that it can handle queries from multiple clients (if there is any). The client is also designed to be multiple threaded so that it can talk to 10 servers simultaneously and one or several failed servers won't affect others. It's faster and more robust than talking to each server one after one.

For implementation, we use java socket programming and invoke system grep command within server thread.

### Measurement

To measure performance, we generated a 60MB log file by reuse the logs provided on Piazza. We uploaded it to VM7-10 and run query on VM6. We run 10 trails and calculated the average speed in ms.



### Unit Tests

For unit test, we first test the functionality of SysGrep class. We test both frequent and infrequent pattern and the result is consistency what we get from running command line directly in terminal. We also test the logic of Client class given two different condition: with or without server running on machine and listening the port it send socket to.

### Reference

1. <http://wiki.jikexueyuan.com/project/java-socket/tcp.html>
2. <https://alvinalexander.com/java/edu/pj/pj010016>