CS425 MP1 Report
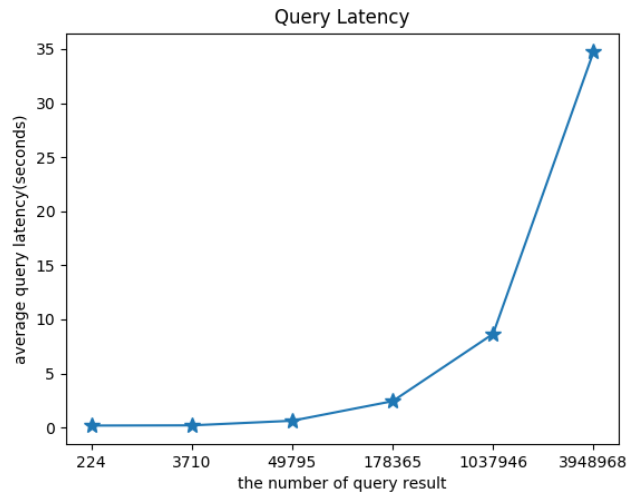## Mingrui Yin && Siti Zhang

Design

The design is based on client-server model and we implemented the distributed log querier by using Python with xmlrpc. We choose nine different VMs as servers and another one as client. The client got the user input and created several thread to communicate with the server. The server part was responsible for listening and handling the request, and then returning the results to the client. Finally, the query results directly have showed in our terminal.

Unit test

The unit test is based on python unittest. First, we created log files in every VM with some frequent and random lines. Then, we wrote several functions to test the frequent, infrequent and rare patterns. We tested both the number of the query result and the several beginning characters of results and compared them with the expected output.

Plot

We queried 4 machines and each had log file that stored more than 97MB. The plot showed the relationship between total line numbers and the corresponding query latency. As we expected, when the number of query results increases, the query latency also increases significantly. Also, we did frequent pattern and infrequent query. We also calculated the average and standard deviation of query time in the below form.

pattern	1	2	3	4	5	average	sd
10	8.96	8.47	8.55	8.65	8.69	8.664	0.186494
13596	0.184	0.19	0.176	0.168	0.203	0.1842	0.013387
10*	36.61	34.09	34.33	34.29	34.41	34.746	1.048656