1. Design
   1.1. Client

   The client gets the query command from the user and **distributes the query** to servers listed on its local configuration. For each server, a separate **worker thread** is designated by the client to exploit parallelism. (**Barrier** between adjacent user queries is included.) Each worker thread **sets up a socket** to communicate with the assigned server. After **sending the query**, it uses a **buffer** to collects the received result into a **temporary local file**, which secures the **integrity of each result line**. When the whole result is received, the worker thread **displays** each result line with an **identifier of the queried log file**, and **reports** the number of matched lines to client. The client does the **statistics** after all the worker threads have finished.

   1.2. Server

   The server is a **threaded TCP server**, listening to a predefined port (2333). When a connection is accepted by the server, a **handler thread** is designated to deal with the query. It **queries on the local log file** and **feedbacks** the matched lines.

   1.3. **High Configurability** & **Easy Deployment**

   We have limited the hard code as few as we can. All the variable parts are centralized in a single concise configuration file (conf.yaml).

   Code deployment on a new host can be done with ease. Keeping the common configuration file updated is enough. All the host-specific settings can be done automatically.

2. Test
   2.1. Test Logs Generation

   By generating test logs ourselves, more pattern control and test correctness is guaranteed. Three **mutual exclusive regular expression patterns** are used as patterns of high/regular/low frequencies accordingly. Additionally, three plain patterns are included in one/some/all log files accordingly. The test log file is generated and stored on each server, by referencing its local IP.

   2.2. Unit Tests

   Two unit tests are performed to test the correctness when querying patterns with **high/regular/low** frequencies and patterns that occur in **one/some/all** log files. The verification criteria is the equality between the numbers of matched lines in test and the ones of ground truth calculated from test configuration.

3. Performance

   The average query latency of 100 queries against ~100MB logs across 4 servers is **0.117s**.