

Project Report

Computer Systems Performance, Spring 2021

YOUR NAME, YOUR-EMAIL@itu.dk

Formatting guidelines

You can use Word or Latex to write the report, but please make sure that you use at least 2.5cm wide margins (on all sides), A4 page size and font size of 11pts. The total length of the document can't exceed 8 pages (including all sections). The submission format is PDF.

1. Implementation overview

In this section you should explain how you implemented the three features listed in the project description. Make sure to elaborate on how multi-threading works in your system and what data structures are accessed by multiple threads, if any.

Feel free to include a drawing of the internal architecture if it helps to illustrate at what points of the server logic you have added instrumentation for performance measurement.

Explain shortly how you automated the experiments, how you collected the log files from both the clients and the server and how did you process them.

2. Baselines

The first section of the report establishes the baselines of performance for the word count server and explores what effect the additional functionality has when compared to the plain skeleton.

2.1 Test Environment and Experimental Design

You will run the experiments using Amazon EC2, with two t2.2xlarge machines (one for the client and one for the server).

Explain the default setup for the experiments in terms of length, repetitions, amount of warm-up and cool-down time. Explain why you chose these specific values. We recommend that you maintain the same setup for all experiments in the paper to ensure that numbers can be correctly compared.

2.2 Throughput and Response Time

Using 1 to 16 clients¹ on a single client machine and the default 16KB document size, plot the throughput and response times as measured by the clients for the following server variants: 1) skeleton, 2) single-threaded server with HTML tag cleaning, 3) multi-threaded server with tag cleaning. Repeat the experiment with input document type 1 and input document type 2.

You need to provide four plots in total, two per input type: one for throughput and one for response time, each of them with multiple lines for the different server variants. Please follow the best practices presented in class when plotting.

¹ In case the system does not become saturated with 16 clients, double the number of clients until it does. Use the same maximum number for all experiments that follow in the report.

2.3 Discussion

Discuss the behavior of the system and the overheads (if any) that your implementation introduces. Identify and quantify the cost of performing HTML cleaning and comment on how well the system parallelizes with increasing number of clients.

Please make sure that in the explanations you don't just describe what is in the graphs, but instead put the results in the context of your design decisions and implementation details.

3. Effect of Document Size

3.1 Time Spent in Server

Based on the server version with all features included (that is, variant 3), pick a number of clients C just below the saturation point of Section 2.2 and run experiments with different document sizes between 4KB and 256KB (at least 7 different sizes) both on input type 1 and 2. Plot how the response time changes at the clients, as well as inside the server (for this, show at least 2 different internal costs).

3.2 Discussion

Describe how the relative cost of different operations inside the server change and how this relates to your implementation decisions. Identify the operation that limits performance the most.

Please make sure that in the explanations you don't just describe what is in the graphs, but instead put the results in the context of your design decisions and implementation details.

4. Modeling

4.1 M/M/m

Using the experimental results in Section 2.2 variant 3) and the insights you gathered in Section 3, build a model of your system using an M/M/m queue.

Plot the predicted response time and throughput as a function of load² and compare this to the real-world results.

4.2 Discussion

Elaborate on how well the models match the real-world behavior and how these results relate to your design decisions.

Please make sure that in the explanations you don't just describe what is in the graphs, but instead put the results in the context of your design decisions and implementation details.

² Note that you will have to transform the number of clients (the parameter in your experiments) into equivalent arrival rate (the parameter in the model) to be able to compare to a model.