Project Presentation

Performance Analysis and Modeling of Software Systems, Fall 2018

MORITZ MEISTER
MORITZ.MEISTER@alumnus.upm.es

Implementation: HTML cleaning

Assumption:

HTML body is properly formed, and the text we want to analyse does not contain “<“ or “>”.

Case 1:

Text we want to analyse <html class=“…”> Text we want to analyse </html class=“…

Case 2:

ml class=”…”>Text we want to analyse <html class=“…”> Text we want to …

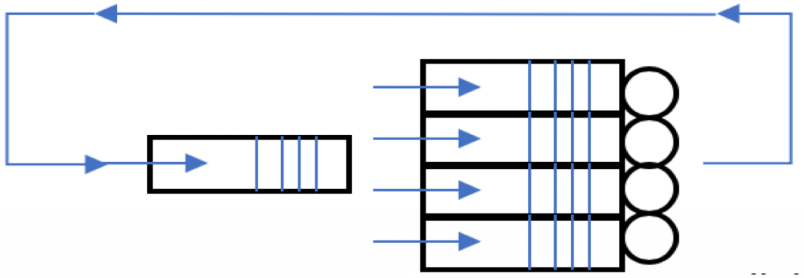
iteration pointer, inHTML = True -> if inHTML do not add character to cleaned string

Iteration over entire document, keeping track of whether the pointer is currently inside a HTML tag.

Implementation: Multi-threading

Assuming N threads for multithreading:

N x Executors.newFixedThreadPool(size=1)



N x Threads with N x Queues

+ “Network” thread with *receiveData*

Main Thread -> receiveData -> Main thread

Round robin based on client id

```
If ”$” found
 new Runnable() {
            do cleaning(doc);
            do wordcount(doc);
            do serialization(results);
 }
```

Executors.newFixedThreadPool(1)

Documents are mapped consistently to threadpools with one thread each, where each pool has its own queue.

System Stability

Experiment length:

30 (x1000) ops

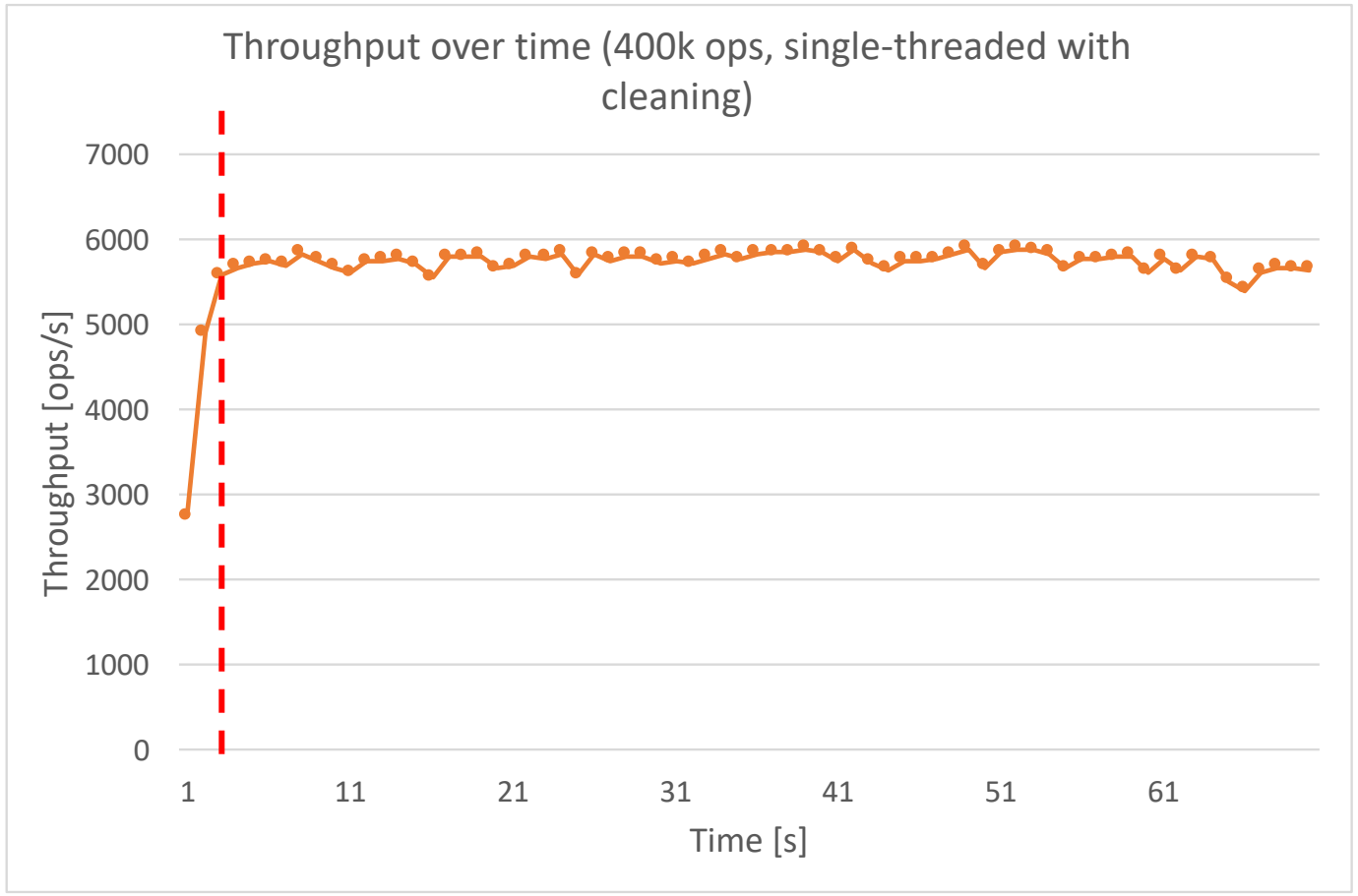
Replication factor:

1

(because of high experiment length)

Warm-up and cool-down:

2 seconds



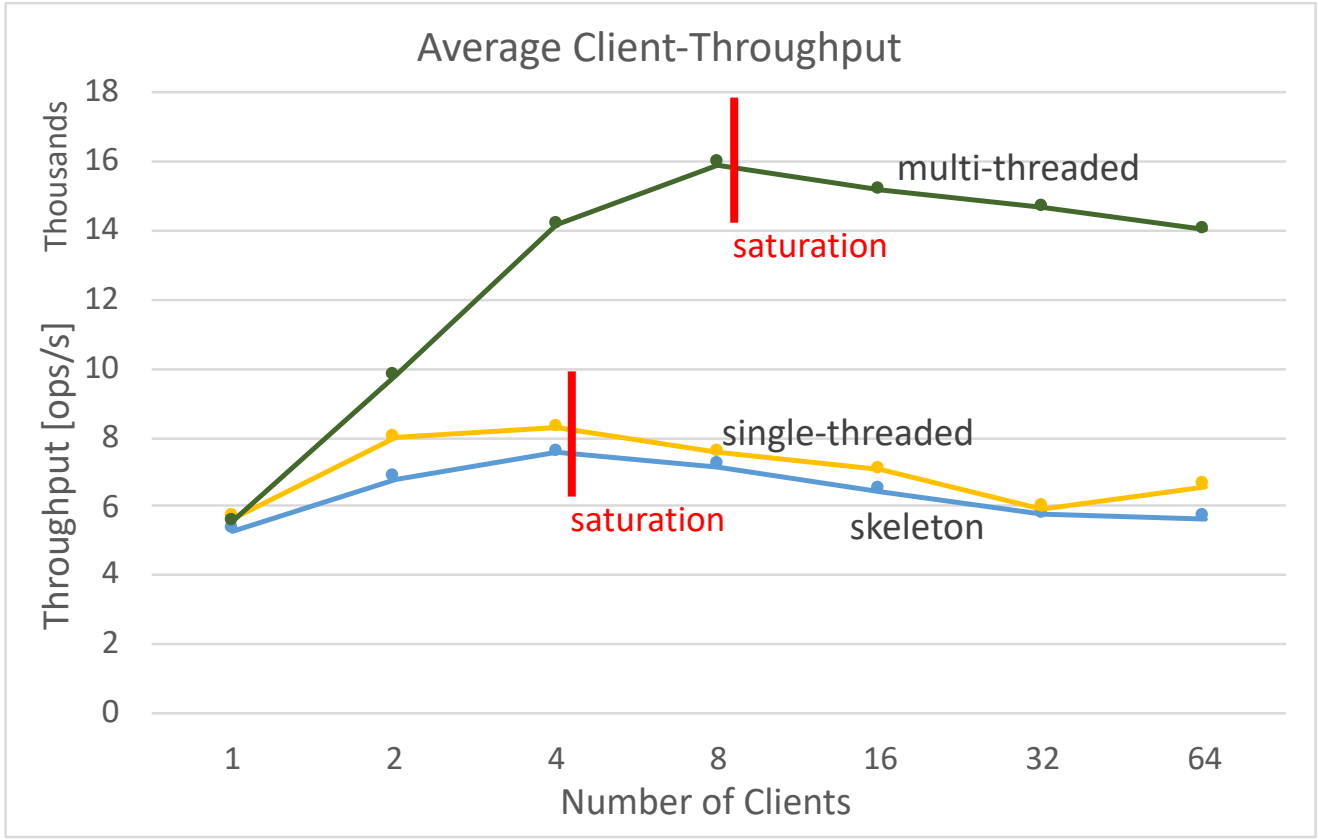
Baselines

Saturation points:

Skeleton: 4 clients

Single-threaded: 4 clients

Multi-threaded: 8 clients



Effect of Document Size

Effect:

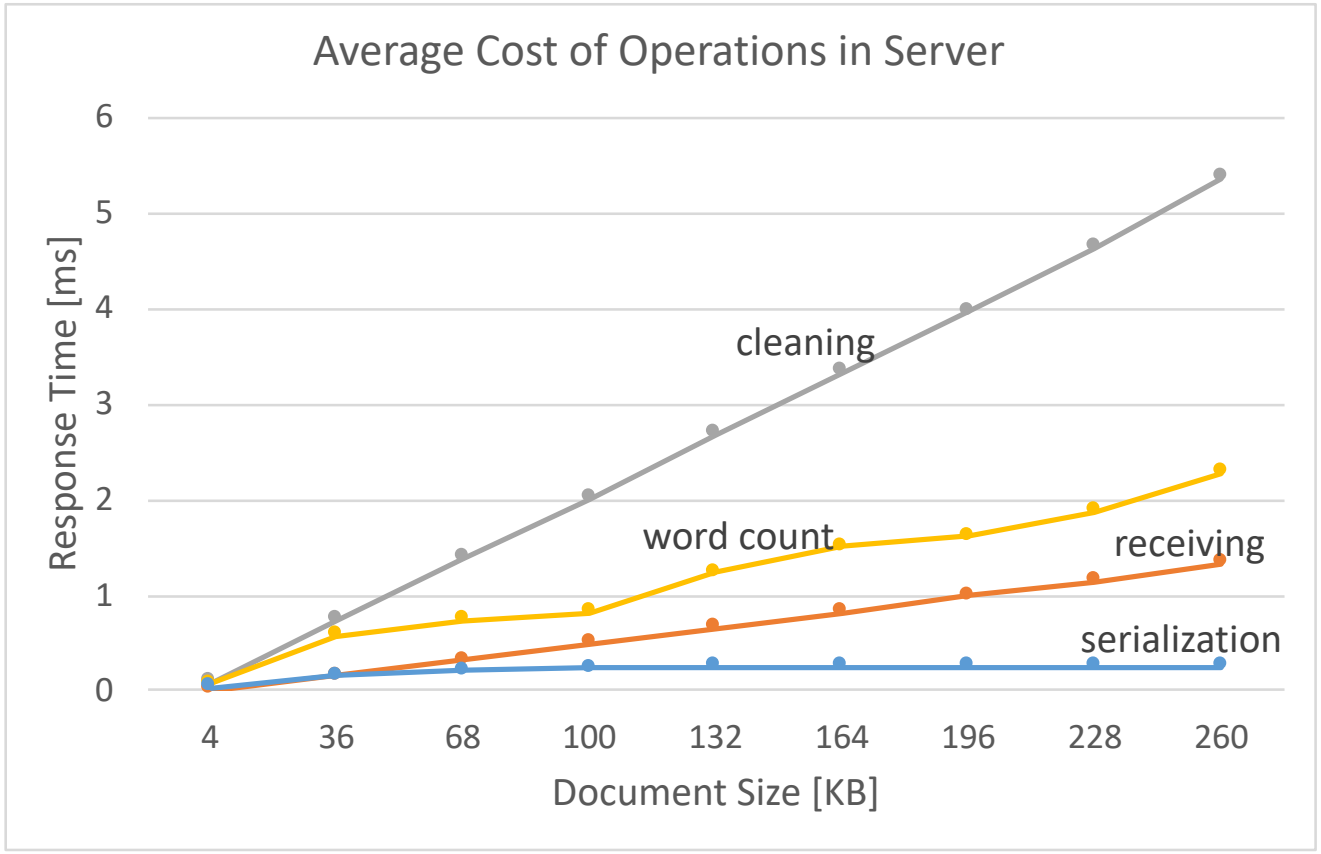
Linear relationship except for serialization (document gets repeated for > 110kb)

Bottleneck:

Document Cleaning

Why?

Expensive loop over the entire document.



Conclusion

What did you learn about the system through experiments?

- **There are a lot of things to consider for the experiments**
- **Operations which are O(n) can become a bottleneck**

What new skill did you learn while working on the project?

- **Java programming**
- **Concurrency**
- **Sockets**

What was the most difficult aspect of the project?

- **Understanding how the sockets and client connections work (Java NIO)**
- **Implementation**