

CLOUD DATA BASES

MILESTONE 3

Report



Group 18: Weijian **Feng**, Jingyi **Jia**, Mingrun **Ma** (alphabet order)

TECHNICAL UNIVERSITY OF MUNICH

June 2023

Report

This report concerns itself and evaluates the performance of the implementation of the third milestone 3 from the advanced practical course “Cloud Databases” at the Technical University Of Munich.

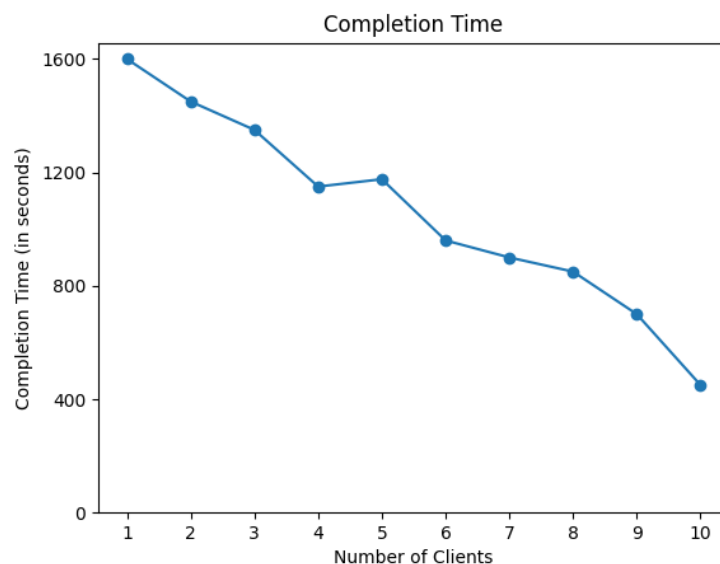
Overview of the task

The primary objective of the third milestone was to implement the external configuration service (ECS). This service plays a vital role in overseeing the KVStores, facilitating rebalancing of the hash ring, and initiating the transfer of key ranges between servers in response to node additions or departures within the scalable storage service.

Data analysis

To evaluate the efficiency of the scalable storage service, we utilized the Enron Email dataset to populate the storage. Specifically, we focused on the "*alldocuments*" folders within all sub-directories of the dataset. By inserting 128,103 key-value pairs into the storage service, we observed how the completion time varied by altering the number of clients and servers.

1 Server and Multiple Clients



For the second test, we repeat the process of inserting the same dataset into the scalable storage service. However, the key difference lies in the utilization of a variable number of servers, ranging up to 10. This test yields completely distinct results compared to the first test. Interestingly, as the number of servers increases, the

dataset insertion time also prolongs. This phenomenon is likely due to the communication overhead that occurs when a client interacts with a server that is not responsible for the specific key it intends to insert into the database. Each server is assigned a specific portion of the hash ring's responsibility. If a server is not accountable for a given key-value pair, it must redirect the request to the responsible server, resulting in additional processing time, especially when there is a significant queue of requests and multiple servers involved.

1 Client and Multiple Servers

Conclusion

We conducted our experimentation on a Mac Pro laptop equipped with 8 Intel i5-8257U. To obtain a more accurate assessment of performance, it would be beneficial to evaluate the database on a machine with a higher core count or within a distributed environment. For instance, deploying the ECS on one machine and each server on separate machines would enable us to account for the communication overhead arising from network speed.