

CLOUD DATA BASES

MILESTONE 3

Report

YACOUBA CISSE, LUCA CORBUCCI, FABIAN DANISCH



TECHNICAL UNIVERSITY OF MUNICH

November 2019

Report

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

Overview of the task

The most important task from the third milestone was the implementation of the external configuration service(ECS). This service controls the KVStores, rebalances the hash ring and initiates the transfer of key ranges between the Servers when someone leaves or joins the scalable storage service.

Data analysis

In order to measure the performance of the scalable storage service, we have used the Enron Email data set to populate the storage. We have considered all the folders “*all_documents*” from all the sub-directory of the dataset. We have inserted 128.103 key,value pairs into the storage service and we have changed the number of clients and servers to understand the variation of the completion time.

1 Server – Variable number of Clients

In the first test we made, we had exactly 1 server and a variable number of client up to maximum 16. As visible in ?, the less clients we have the longer it takes to insert the Enron email dataset. With 1 client it takes up to 900 seconds while using 16 clients takes only around 250 seconds.

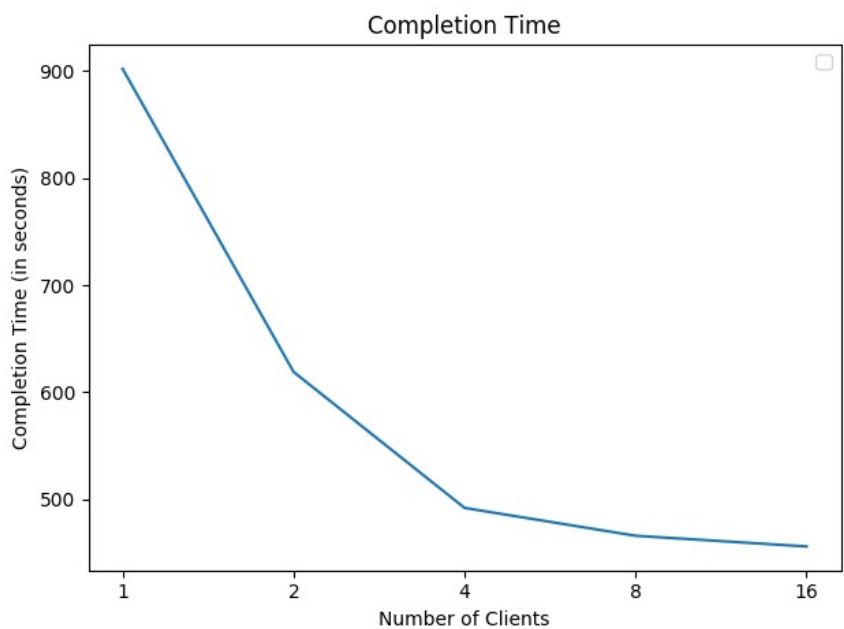


Figure 1: Completion time from 1 server and multiple clients

1 Client – Variable number of Servers

The second test takes the same dataset and inserts it into the scalable storage service, the only difference this time is, that we use a variable number of up to 16 servers. This test has a completely

different output with respect to the first test. The more servers are in use, the longer it takes to insert the dataset. This is probably caused by the communication overhead that we have when a client contacts a server which is not responsible for the key that it wants to insert in the database. Every server here is responsible for a specific part of the hash ring. If a server is not responsible for the given key value pair, then the responsible server needs to be contacted which obviously takes time if you have a huge queue of requests and multiple servers.

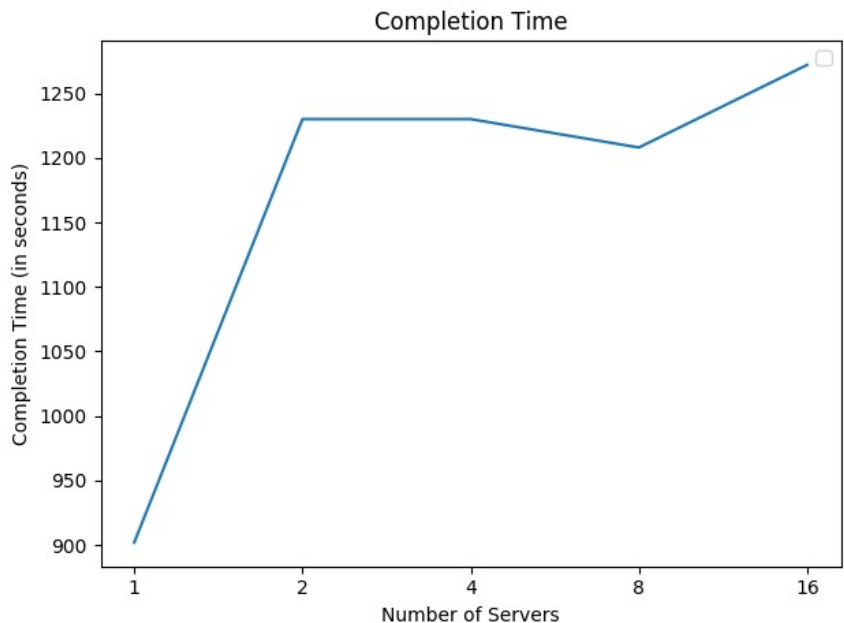


Figure 2: Completion time from 1 client and multiple server

Conclusions

We tested our approach using a laptop with 4 cores and 8 threads. To gain a better estimation of the performance it would be useful to try the database on a machine with more core or in a distributed context. For example it could be useful to launch the ECS on one machine and each server on a diffent machine so that we can also consider the communication overhead caused by the network speed.