

Part3: Performance Evaluation

Our team conducted a straightforward load test to evaluate the performance of our application. Since we had two machines handy, we utilized one machine to deploy clients ranging from 1 to 5, and these clients simultaneously sent requests to the server, which we hosted on another machine. We carefully measured the performance metrics for both Part1 and Part2 during the load test.

We have presented the results of our load test in the form of plots, where the number of clients is shown on the X-axis, and the time taken to process the requests is displayed on the Y-axis, in seconds. The plots for Part1 and Part2 show how the time taken to process requests varies with the number of clients.

The latency numbers below are for **1000 requests per client**.

Fig 1: Lookup vs Trade (Part1)

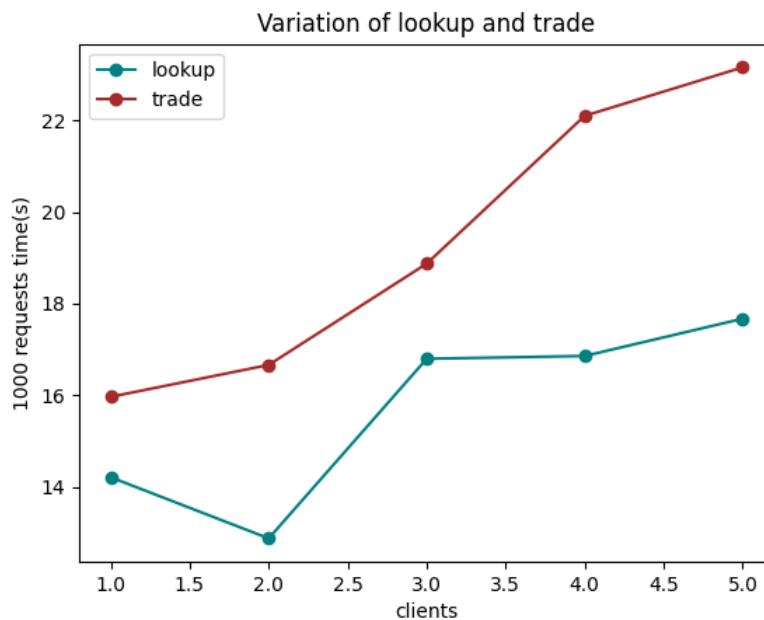


Fig 2: Lookup vs Trade (Part2 – with Docker)



Fig 3: Latency comparison between Part1 and Part2 (The requests sent are a mixture of lookup and trade requests)



Here are our observations:

1. Does the latency of the application change with and without Docker containers? Did virtualization add any overheads?

As Figure 3 indicates, the latency of the application without docker is lower than the latency of the application with Docker containers when using different number of containers. The network latency is affected when using Docker containers since Docker sets up its own network interface and IP addressing scheme for the containers. When the application is deployed in Docker containers, it can introduce some additional overhead due to virtualization. This overall increases the latency of the application with Docker containers.

2. How does the latency of the lookup requests compare to trade? Since trade requests involve all these microservices, while lookup requests only involve two microservices, does it impact the observed latency?

As Figure 1 and 2 shows, in both Part 1 and Part 2, the latency of trade requests is higher than the latency of lookup requests.

Lookup requests involve two microservices, catalog and front-end which only retrieves the information of the stock. Since there are only two microservices involved, the lookup requests can be processed relatively quickly and with lower latency.

Whereas for trade requests all three microservices, catalog, order and front-end are used. The requests are used to complete a transaction which involves more complex data processing and coordination between all the microservices. The front-end service receives an order request, it will forward the request to the order service which still needs to interact with the catalog service to complete the order. All this data processing and coordination can take time, which can result in higher latency for trade requests compared to lookup requests.

3. How does the latency change as the number of clients change? Does it change for different types of requests?

Our analysis of Part1 and Part2 shows that the latency, or the time taken to complete a request, tends to increase as the number of clients increases as shown in all the figures above. This is mainly due to the limited access of shared database and coordination between all the microservices. The application can handle concurrent requests from multiple clients but accessing the shared database can be performed by only one request at a time due to locks in place.

Even with increasing clients, the time taken for lookup requests are lower compared to the trade requests. This is because the lookup needs to access only the stock catalog to get the details of the stock, so it must go through only one read lock. Whereas the order request needs to go through one read lock and two write locks (access the stock

catalog for lookup, update stock catalog after transaction, and update transaction detail in order catalog). As the number of clients increases, the concurrent requests received increases. These overall increases the time taken by order requests increases drastically when compared to the time taken by the lookup requests to process as the number of clients increases.