Data processing pipeline automation on cloud platform - how to prepare your data processing web service for a sudden surge in demand?

Espoo, February 12, 2019

Following the recent advancements in web technologies, competition has increased in the web hosting industry, and while the leading multinational technology companies have established sufficiently large data centers around the world to inexpensive locations, it has become profitable for smaller companies to outsource their hosting services. The use of cloud services provides increased reliability, power, scalability and cost-effectiveness for software businesses. In his thesis, Sampsa Hyvämäki describes the process of transferring a public transportation journey planner service from a company's own datacenter to a cloud platform. During the process, he documented the consequential changes in different abstraction levels of the software architecture, and evaluated the methods that were used in planning and executing the transfer. He combined a container virtualization method with Amazon's container platform service and it increased the software reliability with automatic service health monitoring and scaling. These new features of the system he verified with load tests, and their results showed that the cloud platform is able to restore the normal response times to the service in five minutes after the number of users has grown drastically, by provisioning more resources for the routing algorithm.

The old server environment consisted of physical servers that formed clusters, among which the processing load generated by user requests was distributed. With a container virtualization technology called Docker, containers were created to resemble the different servers in the clusters. Therefore, every container had a different purpose, such as delivering the web page content as a frontend or calculating routes with routing algorithm.

In the load tests, Sampsa Hyvämäki simulated 350 users with Apache JMeter to make random routing requests to a one algorithm container, in a pre-defined area with a random intervals that were sampled from a Gaussian distribution. After a rise in the load of the central processing unit of a container instance, the Amazon's container platform Fargate was able deploy more containers during five minutes to recover the routing service from a 1,6 second response time back to 30-60 millisecond response time.

During the Migration project, Sampsa utilized the methods of Lean software development, by exploring different options that cloud service providers offered, and embedding the gathered knowledge to the product with decisions that optimized the whole. He also followed a reuse-oriented software development process model, where requirements for the newly developed or adapted components are refined after software discovery and evaluation phase.

More information:
Sampa Hyvämäki
svante.hyvamaki@aalto.fi