

Going FaaSter: Cost-performance optimizations of serverless on Kubernetes

Erwin van Eyk, Platform9

Abstract

Serverless promises on-demand, optimal performance for a fixed cost. Yet, we see that the current serverless platforms do not always hold up this promise in practice; serverless applications can suffer from platform overhead, unreliable performance, “cold starts”, and more. In this talk we review optimizations used in popular FaaS platforms, and recent research findings that aim to optimize the trade-off between cost and performance. We will review function reuse, resource pooling, function locality, and predictive scheduling. To illustrate, we will use the open source, Kubernetes-based Fission FaaS platform to demonstrate how you can achieve specific goals around latency, throughput, resource utilization and cost. Finally, we take a look at the horizon; what are the current performance challenges and opportunities to make FaaS even faster?