

# Performance considerations on execution of large scale workflow applications on cloud functions

Maciej Pawlik, AGH University of Science and Technology

## Abstract

Function-as-a-Service is a novel type of cloud service used for creating distributed applications and utilizing computing resources. Application developer supplies source code of cloud functions, which are small applications or application components, while the service provider is responsible for provisioning the infrastructure, scaling and exposing a REST style API. This environment seems to be adequate for running scientific workflows, which in recent years, have become an established paradigm for implementing and preserving complex scientific processes. In this paper, we present work done on evaluating three major FaaS providers (Amazon, Google, IBM) as a platform for running scientific workflows. The experiments were performed with a dedicated benchmarking framework, which consisted of instrumented workflow execution engine. The testing load was implemented as a large scale bag-of-tasks style workflow, where task count reached 5120 running in parallel. The studied parameters include raw performance, efficiency of infrastructure provisioning, overhead introduced by the API and network layers, as well as aspects related to run time accounting. Conclusions include insights into available performance, expressed as raw GFlops values and charts depicting relation of performance to function size. The infrastructure provisioning proved to be governed by parallelism and rate limits, which can be deduced from included charts. The overhead imposed by using a REST API proved to be a significant contribution to overall run time of individual tasks, and possibly the whole workflow. The paper ends with pointing out possible future work, which includes building performance models and designing a dedicated scheduling algorithms for running scientific workflows on FaaS.