Cloud infrastructure automation for scientific workflows

Bartosz Balis, Michał Orzechowski, Krystian Pawlik, Maciej Pawlik, Maciej Malawski
AGH Universtity of Science and Technology, Krakow, Poland
{balis,malawski}@agh.edu.pl

We present a solution for cloud infrastructure automation for scientific workflows. Unlike existing approaches, our solution is based on widely adopted tools, such as Terraform, and achieves a strict separation of two concerns: infrastructure description and provisioning vs. workflow description. At the same time it enables a comprehensive integration with a given cloud infrastructure, i.e. such wherein workflow execution can be managed by the cloud. The solution is integrated with our HyperFlow workflow management system and evaluated by demonstrating its use in experiments related to auto-scaling of scientific workflows in two types of cloud infrastructures: containerized IaaS and FaaS. Experimental evaluation involves deployment and execution of a test workflow in Amazon ECS/Docker cluster and on a hybrid of Amazon ECS and AWS Lambda. The results show that our solution not only helps in the creation of repeatable infrastructures for scientific computing but also greatly facilitates automation of research experiments related to the execution of scientific workflows on advanced computing infrastructures.

Keywords: scientific workflows, infrastructure automation, autoscaling.