Contribution ID: **6** Type: **Oral Presentation**

Serverless Computing for Data-Processing Across Public and Federated Clouds

Monday, September 23, 2019 5:15 PM (15 minutes)

Serverless computing is evolving from the initial Functions as a Service (FaaS) approach to also embrace the execution of containerised applications without the user managing the underlying computing infrastructure. Indeed, the main public cloud providers such as Amazon Web Services or Google Cloud have already started to offer services in this regard. This is the case of AWS Fargate or Google Cloud Run, mainly aimed at the deployment of microservices-based architectures. However, scientific computing can also benefit from the elastic automated management of computational infrastructure for data processing. To this aim, we developed SCAR, an open-source framework to run containers out of Docker images on AWS Lambda which defines a file-processing computing model that is triggered in response to certain events (such as file upload or a REST API invocation). This model was extended for on-premises environments through OSCAR, an open-source platform which enables the users to deploy their file-processing container-based serverless applications on a dynamically provisioned elastic Kubernetes cluster that can be deployed in multi-Clouds, integrated with the EGI Federated Cloud and the EGI Data Hub, based on Onedata.

In this work we focus on integrating a federated storage for data persistence, in particular the EGI Data Hub, with the ability to dynamically provision computational resources from a public Cloud provider to perform the data processing in response to file uploads. To this aim, we developed OneTrigger, a tool to trigger events from Onedata, that can be run as a serverless function in AWS Lambda in order to use SCAR's functionality to perform the execution of jobs in AWS Lambda, supporting thousands of concurrent executions. Longer executions, as well as those requiring specialised computing hardware, such as GPUs, are delegated to AWS Batch, a service which enables the unattended and elastic execution of batch computing workloads on the public Cloud. This allows to create hybrid data-processing serverless applications across public and federated Clouds. We demonstrate the feasibility of this approach by introducing a use case in video processing that can leverage GPU-based computing in the public Cloud to dramatically accelerate object recognition, while data persistence is still supported by the federated Cloud.

Primary authors: RISCO, Sebastián (Universitat Politècnica de València); PÉREZ GONZÁLEZ, Alfonso (UPV - GRyCAP); CABALLER, Miguel (Universitat Politècnica de València); MOLTÓ, Germán (Universitat Politècnica de València)

Presenter: RISCO, Sebastián (Universitat Politècnica de València)

Session Classification: IBERGRID Contributions

Track Classification: Development of Innovative Software Services