Contribution ID: **47** Type: **Presentation**

An integrated IaaS and PaaS architecture for scientific computing

Wednesday, 11 November 2015 13:30 (20 minutes)

Scientific applications often require multiple computing resources deployed on a coordinated way. The deployment of multiple resources require installing and configuring special software applications which should be updated when changes in the virtual infrastructure take place. When working on hybrid and federated cloud environments, restrictions on the hypervisor or cloud management platform must be minimised to facilitate geographic-wide brokering and cross-site deployments. Moreover, preserving the individual operation at the site-level in federated clouds is also important for scalability and interoperability.

In that sense, the INDIGO-DataCloud project [1] has been designed with the objective of building up a PaaS-level cloud solution for research. One of the key multi-level components is the PaaS computing core. This part constitutes the kernel for the deployment of services and computing virtual infrastructures for the users. It is complemented with the virtualized storage, federated AAI and networking. The INDIGO-DataCloud PaaS core will be based on a microservice architecture [2]. Microservices consist of a set of narrowly focused, independently deployable services, typically implemented using container-embedded applications, exposed by RESTful interface. Microservices are designed to be highly scalable, highly available and targeted for the use in cloud environments. INDIGO's microservices will be deployed, dynamically scheduled and managed using tools such as kubernetes [3]. In cases where multi-tenancy is not yet intrinsically supported by the particular microservice, like the container manager, INDIGO-DataCloud may decide to offer multiple instances to bridge that gap.
INDIGO PaaS will offer an upper layer orchestration service for distributed applications using the TOSCA language standard [4]. It will deal with the requested service instantiation and application execution, managing the needed microservices in order, for example, to select the right end-point for the deployment. Cross-site deployments will also be possible.

This PaaS, aimed at providing a more efficient platform for scientific computing, will require additional characteristics from the underlying layers. The INDIGO PaaS will leverage an enhanced IaaS that will provide a richer set of features currently missing. The usage of TOSCA permits IaaS providers to offer infrastructure orchestration, making possible to manage the deployment and configuration of the resources that are being provided. The life-cycle of the resources is therefore managed through the APIs exposed by the IaaS end-points. The TOSCA templates will be translated into their native deployment schemas using IM [5] for OpenNebula and Heat-Translator [6] for OpenStack HEAT.
Both OpenNebula and OpenStack will incorporate drivers to support the deployment of containers as first-class resources on the IaaS. This will provide high efficiency when building up complex configurations from a repository of container images. The scheduling algorithms for both cloud management frameworks will be improved, in order to provide a better experience for the end-users and a more efficient utilization of the computational resources. The usage of two-level orchestrator (at the level of PaaS and within each IaaS instances) will enhance the capabilities of providing a dynamic and on-demand increase in cloud resources.

Summary

INDIGO-DataCloud [1] (INtegrating Distributed data Infrastructures for Global ExplOitation) is a project within the E-INFRA-1-2014 call of the Horizon 2020 framework program of the European Community. It aims at developing a data/computing platform targeting scientific communities, deployable on multiple hardware and provisioned over hybrid (private or public) e-infrastructures.

TOSCA [4], OASIS Topology and Orchestration Specification for Cloud Applications, is a vendor-neutral specification for enhancing the portability of cloud applications and services. TOSCA enables the interoperable description of applications and services and the relationships between their parts independent of the cloud provider.

Kubernetes [3] is an open source orchestration system for Docker containers created by Google that actively manages workloads and isolates applications into different levels.

Infrastructure Manager [5], is a tool to deploy complex and customized virtual infrastructures on multi-Clouds, by automating the VMI selection, deployment, configuration, software installation, monitoring and update of

Virtual Appliances, using the Resource and Application Description Language (RADL) to describe applications and virtual infrastructures.

Heat-Translator is an OpenStack project that provides a tool which takes TOSCA templates as an input and produces a Heat Orchestration Template (HOT) which can be deployed by Heat.

Primary authors: Mr LOPEZ GARCIA, Alvaro (CSIC); Dr MOLTÓ MARTÍNEZ, Germán (UPVLC); Dr DONVITO, Giacinto (INFN); Dr BLANQUER, Ignacio (UPVLC); FUHRMANN, Patrick (DESY)

Presenters: Dr MOLTÓ MARTÍNEZ, Germán (UPVLC); Dr DONVITO, Giacinto (INFN); Dr BLANQUER, Ignacio (UPVLC)

Session Classification: Exploiting the EGI Federated clouds - Paas & SaaS workshop