

How the INDIGO-DataCloud computing platform aims at helping scientific communities

Wednesday, 6 April 2016 14:30 (20 minutes)

Scientific workloads require customized computing power adapted to the hardware, software and configuration requirements of the applications. Providing users with the ability to deploy customized virtual infrastructures easily and execute jobs and services from an integrated system is the goal of the INDIGO-DataCloud's computing platform. This platform aims at providing an integrated IaaS + PaaS layer to support the computing requirements that arise from multiple scientific user communities, with a special focus on hybrid and federated cloud environments.

In particular the PaaS layer aims at supporting geographic brokering and deployments across multiple sites. The PaaS computing core includes services for the deployment and management of jobs, long-running services and virtual infrastructures across multiple Cloud sites based on popular open-source Cloud Management Platforms (CMPs) (i.e. OpenNebula and OpenStack). These services will be integrated with the virtualized storage, federated AAI and enhanced networking.

A key aspect of the platform is that individual operation at the site-level is preserved to leverage scalability and interoperability. Indeed, the deployment of customized virtual infrastructures is achieved at the level of each site by means of the IM [1], in the case of OpenNebula, and Heat, in the case of OpenStack. A common language to define the deployments is employed by adopting and extending the TOSCA Simple Profile in YAML version 1.0 standard [2]. The Orchestrator service, entry point for the INDIGO-DataCloud PaaS, receives a TOSCA description which flows through the PaaS layer, interacting with other services (Monitoring, Brokering, QoS/SLA, etc.), to end up being processed by either IM or Heat at the level of the site to enact the required virtual infrastructure. The life-cycle of the resources is, therefore, managed by the IaaS Cloud sites but controlled by the PaaS. The INDIGO-DataCloud PaaS is composed by multiple services exposed as Docker Images, stored in Docker Hub, which are automatically created out of the corresponding open-source repositories available in GitHub. The services running on Docker containers are managed by a Kubernetes cluster, to support a microservices-oriented deployment. This introduces benefits such as rolling updates which fosters a CI/CD approach for software development, the possibility to implement automatic scalability of the services instances, etc.

The deployment of the user application and services, are implemented via a dynamically instantiated Mesos cluster.

At the level of the Cloud site, the computing platform provided by INDIGO-DataCloud is enhancing the IaaS layers with additional features that are currently missing. Firstly, introducing TOSCA support for the CMPs enables infrastructure orchestration at the level of the site, based on a common standardized language. Secondly, the adoption of Docker containers as first-class resources in the CMPs enables lightweight isolation among computing resources and easy integration with repositories of images (e.g. Docker Hub). Indeed, developments such as OneDock [3], which introduces Docker support in OpenNebula, have already been released with significant outreach in the OpenNebula community. The scheduling algorithms for both cloud management frameworks are being improved, adding the support for preemptible instances (making possible to evacuate workloads when higher priority workloads are needed) and queuing of requests (enabling users to perform HTC on cloud resources). These two facts will provide a better experience for the end-users and a more efficient utilization of the computational resources from the resource provider standpoint. The usage of two-level orchestration (at the PaaS level and within each IaaS Cloud) will provide a scalable approach to provision customized computing resources across Cloud sites to support the computational requirements identified by the user communities. The INDIGO-DataCloud computing platform is under active development in GitHub [4], where all the developments are available under the Apache 2.0 License.

[1] Infrastructure Manager (IM). <http://www.grycap.upv.es/im>

[2] TOSCA Simple Profile in YAML Version 1.0. <http://docs.oasis-open.org/tosca/TOSCA-Simple-Profile-YAML/v1.0/csprd01/TOSCA-Simple-Profile-YAML-v1.0-csprd01.html>

[3] OneDock. <https://github.com/indigo-dc/onedock>

[4] INDIGO-DataCloud's GitHub. <https://github.com/indigo-dc>

(UPV); Dr DONVITO, Giacinto (INFN); Dr BLANQUER, Ignacio (UPVLC); Dr CAMPOS, Isabel (CSIC); Dr DUTKA, Lukasz (CYFRONET); FUHRMANN, Patrick (DESY)

Session Classification: Federated cloud PaaS and SaaS