

## **DODAS: How to effectively exploit heterogeneous clouds for scientific computations**

*Tuesday, 20 March 2018 11:30 (30 minutes)*

Dynamic On Demand Analysis Service (DODAS) is a Platform as a Service tool built combining several solutions and products developed by the INDIGO-DataCloud H2020 project. DODAS allows on-demand generation of a container-based HTCondor batch system over cloud-based infrastructures implementing the “Batch System as a Service” paradigm. As such, it is a cloud enabler designed for scientists seeking to easily exploit distributed and heterogeneous clouds to process data.

Aiming to reduce the learning curve as well as the operational cost of managing community specific services running on distributed cloud, DODAS completely automates the process of provisioning, creating, managing and accessing a pool of heterogeneous computing and storage resources.

DODAS has a modular architecture providing the following major features: the abstraction of the underlying compute and data resources, which is crucial for the interoperability across IaaS providers; the security toolkit, a key aspect for managing community services on opportunistic resources; the automated resource provisioning and configuration, to simplify the setup of the complex experiment computing environments.

The resource abstraction and the full automation are implemented combining together the INDIGO PaaS Orchestrator and the INDIGO Infrastructure Manager (IM). IM provides the support for several types of IaaS like, for example, OpenStack, OpenNebula, Amazon AWS and Microsoft Azure, while the PaaS Orchestrator represents the DODAS entrypoint. As such it is directly exposed to the end user, who is required to provide properly configured TOSCA templates, which will be then processed by the INDIGO PaaS layer. The cluster setup and the services configuration are automated using Ansible recipes. The TOSCA and Ansible combination guarantees an easy procedure to describe complex computing infrastructures.

The Identity Access Manager (IAM) and Token Translation Service (TTS) are used to manage user authentication/authorization both to grant access to the resources and to protect experiment services such as Workload Management Systems and data storages.

Apache Mesos is the baseline solution used by DODAS to abstract CPU, RAM and storage, while Marathon is adopted as container orchestration platform on top of Mesos. Marathon takes care of setting up both HTCondor and any additional services (e.g squid proxy, proxy certificate cache) that experiments might require. Such architecture provides high scaling capabilities and self-healing support that results in a drastic reduction of time and cost, through setup and operational efficiency increase.

The high level of modularity of DODAS is a key to its generic applicability, providing the ability to easily customize the workflow depending on the community computational requirements. Although originally designed for the Compact Muon Solenoid (CMS) Experiment at LHC, DODAS has been quickly adopted by the Alpha Magnetic Spectrometer (AMS) astroparticle physics experiment mounted on the ISS as a solution to exploit opportunistic computing, nowadays an extremely important topic for research domains where computing needs constantly increase. Due to its flexibility and efficiency, DODAS was selected as one of the Thematic Services that will provide multi-disciplinary solutions in the EOSC-hub project, an integration and management system of the European Open Science Cloud starting in January 2018.

The main goals of this contribution are to provide a comprehensive overview of the overall technical implementation of DODAS, as well as to illustrate two distinct real examples of usage: the integration within the CMS Workload Management System and the extension of the AMS computing model.

**Primary author:** Dr SPIGA, Daniele (INFN-PG)

**Co-authors:** Dr DUMA, Cristina (INFN-CNAF); Dr SALOMONI, Davide (INFN-CNAF); Dr ANTONACCI, Marica (INFN-Bari); Dr DURANTI, Matteo (INFN-Perugia); Dr BOCCALI, Tommaso (INFN-Pisa); Dr FORMATO, Valerio (INFN-Perugia); Dr CIANGOTTINI, diego (INFN-Perugia); Dr DONVITO, giacinto (INFN-Bari)

**Presenter:** Dr SPIGA, Daniele (INFN-PG)

**Session Classification:** VRE

**Track Classification:** Virtual Research Environment (including Middleware, tools, services, workflow, ... etc.)