



Contribution ID: 102

Type: Oral contribution

Speed up research by leveraging INDIGO-DataCloud solutions: containers in user-land and on-demand computing clusters

Thursday, 20 April 2017 12:20 (20 minutes)

INDIGO-DataCloud (INDIGO for short, see <https://www.indigo-datacloud.eu>) is a project started in April 2015, funded under the EC Horizon 2020 framework program. It includes 26 European partners located in 11 countries and addresses the challenge of developing open source software, deployable in the form of a data/computing platform, aimed to scientific communities and designed to be deployed on public or private Clouds and integrated with existing resources or e-infrastructures.

In this contribution we will describe briefly the architectural foundations of the project starting from its motivations, discussing technology gaps that currently prevent effective exploitation of distributed computing and storage resources by many scientific communities and we'll illustrate the main components of the INDIGO architecture in the three key areas of IaaS, PaaS and User Interfaces. The modular INDIGO components, addressing the requirements of both scientific users and cloud / data providers, are typically based upon or extend established open source solutions such as OpenStack, OpenNebula, Docker containers, Kubernetes, Apache Mesos, HTCondor, OpenID-Connect, OAuth, and leverage both de facto and de jure standards.

The INDIGO DataCloud solutions are the real driver and objective of the project and derive directly from use cases presented by its many scientific communities, covering areas such as Physics, Astrophysics, Bioinformatics, Structural and molecular biology, Climate modeling, Geophysics, Cultural heritage and others. In this contribution we will specifically highlight how the INDIGO software can be useful to tackle common use cases in the HEP world by describing two of the key solutions that the project has been working on:

- Udocker –a tool for executing simple Docker containers in user space without requiring root privileges. It does not require any type of privileges nor the deployment of services by system administrators, miming a subset of the docker capabilities with minimal functionalities:

- basic download and execution of docker containers by non-privileged users in Linux systems where docker is not available.

- access and execution of docker containers in Linux batch systems and interactive clusters that are managed by other entities such as grid infrastructures or externally managed batch or interactive systems.

- It can be downloaded and executed entirely by the end user.

- DoDAS –the Dynamic On Demand Analysis Service - a service that provides the end user with an automated system that simplifies the process of provisioning, creating, managing and accessing a pool of heterogeneous (possibly opportunistic) computing resources. In particular we'll describe the support of a batch system as a Service based on HTCondor which in turn can:

- o Seamlessly be integrated in the existing HTCondor GlobalPool of CMS.

- o Deploy a standalone, auto-scaling HTCondor batch farm, also using different - geographically distributed computing centers.

Both tools are already used by various research communities like the MasterCode collaboration (<http://cern.ch/mastercode>) which is concerned with the investigation of supersymmetric models that go beyond the current status of the Standard Model of particle physics, or the CMS experiment at CERN.

Summary

The INDIGO-Datacloud project aims to develop a Cloud-based Platform as a Service oriented to scientific computing for the exploitation of heterogeneous resources. Without the need of system-admins, researchers can lower the time to have the results by deploying their analysis workflows into resources offered by distributed data-centers offering cloud e-infrastructures, by using two of the solutions developed within the project -

udocker, tool to deploy containers in user-land and dynamic on-demand deployment of batch, HTCondor, computing clusters. The level of maturity of the tools is demonstrated by the communities that started already to use them: the MasterCode collaboration and CMS experiment at CERN.

Primary authors: SALOMONI, Davide (CNAF); MICHELOTTO, Diego (CNAF); DUMA, Doina Cristina (CNAF)

Presenter: MICHELOTTO, Diego (CNAF)

Session Classification: Sessione Nuove Tecnologie

Track Classification: Sessione Nuove Tecnologie