Contribution ID: **34** Type: **Oral Presentation**

Evolution of the INDIGO-DataCloud architecture toward an advanced open source Cloud platform integrating AI-based workflows exploiting large-scale Big Data facilities

Friday, 5 April 2019 09:30 (30 minutes)

In this contribution, we describe an innovative open source Cloud platform evolving the architecture and outcomes of the successful INDIGO-DataCloud project (INDIGO, https://www.indigo-datacloud.eu) and of the two INDIGO follow-on projects DEEP-HybridDataCloud (DEEP, https://deep-hybrid-datacloud.eu) and eXtreme-DataCloud (XDC, http://www.extreme-datacloud.eu).

INDIGO developed a modular open source data and computing platform targeted at scientific communities, deployable over public or private resources. By filling many technology gaps at the PaaS and SaaS levels, INDIGO helped developers, resource providers, e-infrastructures and scientific communities to overcome key challenges in the Cloud computing, storage and network areas.
XDC is developing scalable technologies for federating storage resources and managing massive amount of data in highly distributed computing environments, as required by the most demanding, data intensive research experiments in Europe and worldwide.
DEEP is evolving intensive computing services offered via Cloud infrastructures and exploiting specialised hardware components, such as GPUs, low-latency interconnects, and others typically accessed as "bare metal" resources.

The proposed architecture described here is the evolution of these three solutions, which are either already available or planned to be made available shortly. This new platform will extend, leverage and achieve convergence of services used to access extreme large datasets over distributed environments and to exploit specialised hardware for high-level, easy to use support of AI workloads in Cloud and HPC environments.

Some of the key characteristics of the proposed architecture are:

- Integration of third-party storage services, allowing to automatically stage in data required by applications into dynamically-created clusters, running on any IaaS Cloud infrastructure.
- Support for experiment reproducibility, providing the possibility to e.g. select or playback given versions of applications or containers, PIDs for the input files, etc.
- Advanced QoS support on heterogeneous and specialized hardware, including e.g. SSD storage systems vs. other storage types. This will make it possible to build hybrid computing clusters with local dynamic data caches tailored to actual workloads.
- Provisioning of high-level interfaces for the exploitation of metadata used to find data to be processed. Based on this metadata and without having to know the details of data locations, file names, etc., users will be able to request analysis code to be automatically executed on distributed resources.
- Support for user-level workflows via standard languages such as CWL or similar, extending toward end users the work on service-oriented TOSCA templates that was started in INDIGO. Through these workflows, users will not only be able to ask for the creation and set-up of complex clusters of services, but also to co-design the execution of entire multi-step distributed analysis processes.
- Support for building overlay HTCondor-based computing clusters running on top of distributed heterogeneous resources, which can range from large Cloud or HPC allocations, down to opportunistic resources such as a single server or user desktops.

The talk will describe how we intend to drive and implement the proposed architecture, which existing components will need further development to support the features mentioned above, and eventually how services for these features can be deployed into real infrastructures.

Primary authors: Dr SALOMONI, Davide (INFN); Mr DONVITO, Giacinto (INFN)

Presenter: Dr SALOMONI, Davide (INFN)

Session Classification: Supercomputing, High Throughput Computing, Accelerator Technologies, and their Integration

Track Classification: Supercomputing, High Throughput Computing, Accelerator Technologies, and their Integration