Contribution ID: **95** Type: **Long Talk**

INFN-DataCloud a distributed infrastructure supporting multi purpose Scientific data analytics services

Wednesday, 2 October 2024 15:00 (20 minutes)

In Italy thanks to the investment coming from Italy's Recovery and Resilience Plan projects (mainly ICSC: https://www.supercomputing-icsc.it/, Terabit: https://www.terabit-project.it/, and others) INFN is implementing a distributed HW, SW infrastructure that will be used to support very heterogeneous scientific use cases, not only coming from the INFN Community, but also with the full Italian scientific community.
INFN DataCloud aims to create a next-generation integrated computing and network infrastructure by 2025. The primary goal is to enhance collaboration and information exchange among Italian scientific communities.
In particular both within ICSC and Terabit, INFN is collaborating with CINECA and GARR with the aim of building an integrated computing and network infrastructure to eliminate disparities in access to high-performance computing across Italy.
The ICSC project has a very large partnership of around 55 entities both public and private bodies that assure that the infrastructure we will put in operation by the end of the project has to be able to support requirements from about full Italian scientific communities.
The INFN DataCloud project must address technical and not technical challenges related to network architecture, data storage, and computational resources together with the distributed team of people working in all the projects activities from each of the INFN main sites (about 12 distributed data center in Italy).
One of the main challenges is Data Security and Privacy: Handling sensitive scientific data requires robust security measures. The project must address data encryption, access controls, and compliance with privacy regulations to protect researchers'work.
The computing and storage distributed facilities upgraded by Italy's Recovery and Resilience Plan projects that founded many fat nodes (we call those "HPC-Bubbles") with GPU, many cores CPU, large RAM Memory and SSD based storage increase the resources available to the researchers in order to support the modern requirements of AI algorithms and very large dataset needed for most of the science (Physics, bioinformatics, earth studies, climate etc).
In the contest of the Data access/management/transfer, the INFN DataCloud project is federating both posix and Object storage with a geographically distributed data lake.
In the talk we will show both technical and not technical solutions implemented to build a transparent high-level federation of both Compute and Data resources, and how the development, operation and user support activities are organized in such a heterogeneous environment.
In summary, the INFN DataCloud project faces a mix of technical, organizational, and logistical challenges. However, its potential impact on Italian scientific communities makes overcoming these hurdles worthwhile. Moreover, the INFN DataCloud project has democratized access to high-performance computing and data resources, empowering Italian researchers to accelerate their scientific endeavors.

Topic

Needs and solutions in scientific computing: National and scientific perspectives

Primary author: DONVITO, Giacinto (INFN)

Co-authors: MARTELLI, Barbara; PELLEGRINO, Carmelo (INFN-Cnaf); GRANDI, Claudio (INFN); CESINI, Daniele (INFN); SPIGA, Daniele; MICHELOTTO, Diego (INFN); GIORGIO, Emidio (INFN); CARBONE, Luca (INFN-Milano); SGARAVATTO, Massimo (INFN); FOGGETTI, Nadina; CIASCHINI, Vincenzo (INFN); STALIO, stefano

Presenter: DONVITO, Giacinto (INFN)

Session Classification: National Perspectives: EGI Member Countries' Latest Developments and Future Initiatives