

Integrated, heterogeneous data access in INFN-Cloud and beyond

Tuesday, November 3, 2020 11:45 AM (30 minutes)

INFN-Cloud integrates an object storage service as its main data backend for end user applications as well as for internal use.

The INFN-Cloud Object Storage Service is a geographically distributed OpenStack Swift instance, instantiated over the INFN-Cloud backbone, where data is replicated over two different data centers about 600km away from each other. In the current deployment different replica policies can be applied, depending on both the characteristics of specific sets of data and the requirements of their owners.

High availability, resilience, ubiquitous and authenticated access, as well as ease of use and support for multiple technologies, are the highlights of the service described in this talk.

The INFN-Cloud Object Storage service has been coupled with high-level tools and facilities by taking advantage of the OpenStack Swift and S3 APIs. Using this approach, Nextcloud, ownCloud, Minio, Duplicati, Rclone, AWS cli, S3fs and many other similar tools act as the contact points between the backend storage service and end user applications deployed on the INFN-Cloud infrastructure at the IaaS or PaaS levels. The variety of the supported storage products, each one with its distinct characteristics and different data access paradigms, allows to implement ad-hoc solutions aimed at satisfying requirements coming from different scientific communities.

Some typical requests that INFN-Cloud is addressing within specific use cases are related to scientific data archival and distribution, remote and encrypted data backup, personal and shared data storage. Besides that, the INFN-Cloud Object Storage Service is also used internally for image and software repository and data backup for its own core services.

The talk will provide details about the storage integration capabilities already implemented in INFN-Cloud, as well as future directions; in addition, some representative use cases dealing with Jupyter notebooks with persistent storage deployed on Kubernetes cluster, the integration of the entire data workflow of some physics experiments and the use of sync&share solutions for scientific data management will be described, with the goal of highlighting the role and impact of the INFN-Cloud Object Storage Service on the solutions provided to scientific user communities.

Primary authors: STALIO, Stefano (INFN); SALOMONI, Davide (INFN); DONVITO, Giacinto (INFN); DUMA, Doina Cristina (INFN); GIORGIO, Emidio (INFN); SPINOSO, Vincenzo (INFN); CIASCHINI, Vincenzo (INFN); SGARAVATTO, Massimo (INFN); ANTONACCI, Marica (INFN); SPIGA, Daniele (INFN)

Presenter: STALIO, Stefano (INFN)

Session Classification: Data management solutions - Part 2