Contribution ID: **71**   Type: **Oral Presentation**

Facilitating the distribution of software by using CernVM File System and S3 bucket (Remote Presentation)

Thursday, 28 March 2024 16:30 (30 minutes)

The adoption of user-friendly solutions aimed at sharing data as well as software and related configuration files, among heterogeneous and distributed resources, becomes a necessity for the scientific community. By adopting and using software products dedicated to this purpose, it is possible to facilitate the distribution of software, configurations and files. To this extent, the CernVM-File System has been adopted and integrated with other technologies such as S3 object storage, Vault identity-based secrets and encryption management system and RABBitMQ open source message broker.

The CernVM File System (CVMFS) provides a scalable, reliable and low-maintace software distribution service. It was developed to assist High Energy Physics (HEP) collaborations in deploying software on the worldwide-distributed computing infratructure used for running data processing applications. It is a network file system implemented as a POSIX read-only file system in user space (a FUSE model) and it uses a standard HTTP transport, thereby avoiding most of the firewall issues. Files and directories available via CernVM-FS are hosted on standard web servers and are always mounted in the universal namespace /cvmfs.

The integration with Vault provides encryption services that are gated by authentication and authorization methods to ensure secure, auditable and restricted access to secrets and to store the CVMFS keys. On the other hand, RABBitMQ collects events used to process the creation and /or the update of new repositories.

The objective of the present work is to design and to develop a cloud-oriented service aimed at enabling the final user to require a personal or a group CVMFS repository. At a later time the user can upload data to the dedicated object storage space and access it in his personal CVMFS repository in a transparent way. The whole system will provide an abstraction layer enabling the final user to distribute data, software, libraries and related depedences from an S3 object storage to different resources, having them installed under a proper path of the file system, and having a POSIX access. Hiding the complexity of the above mentioned system, in fact, will shorten the learning curve by improving the user experience in the adoption of the service itself.

In the present work, the deployment of the different services on the INFN Cloud distributed infrastructure is presented, together with the integration process. Also, some practical examples are presented to demonstrate both the high level of reproducibility and the usability of the deployed solution suitable to be adopted by other communities.

Primary authors: COSTANTINI, Alessandro (INFN-CNAF); SPIGA, Daniele; DEL CORSO, Francesca; MALATESTA, Giada (INFN-CNAF); SGARAVATTO, Massimo; TRALDI, Sergio

Presenter: MALATESTA, Giada (INFN-CNAF)

Session Classification: Data Management & Big Data

Track Classification: Track 6: Data Management & Big Data