Contribution ID: **50**   Type: **Oral Presentation**

EOS Open Storage - the CERN storage ecosystem for scientific data repositories

Thursday, 22 March 2018 11:30 (30 minutes)

EOS Open Storage platform is a software solution for central data recording, user analysis and data processing.

In 2017 the EOS system at CERN provided 250 PB of disk storage on more than 50k disks. Most of the stored data originates from the Large Hadron Collider and various other experiments at CERN.

Originally developed as a pure disk storage system, EOS has been extended with interfaces to support data life cycle management and tiered storage setups. A work flow engine allows to trigger chained work flows on predefined storage events to notify external services on arrival, retrieval or deletion of data. It is planned to connect EOS during 2018 to the CERN tape archive system (currently archiving 200 PB) to optimize data durability and costs.

Unlike many classical storage systems, EOS is also designed and used for distributed deployments in WAN environments. The CERN storage setup is distributed over two computer centers in Geneva and Budapest. Another example is a distributed setup in Australia for the CloudStor service of the australian research network AARNet.

EOS participates in the eXtreme Data Cloud project with the goal to use EOS as an overlay layer to implement a data lake concept where various non-uniform storage systems can be virtualized into a centrally managed distributed storage system. The described deployment model allows to optimize redundancy parameters on a higher level for cost reduction and availability optimization - in contrast to a simplistic file replication model where local redundancy within servers or a storage site is not taken into account.

To enable scientific collaboration and interactive data analysis EOS is used as the back-end implementation for CERNBox with file synchronization, sharing and collaborative editing capabilities - and SWAN as a service for web-based data analysis (Jupyter notebook interface). The CERN OpenData digital repository based on Invenio uses EOS as its storage back-end. The enabling functionality for these front-end services is a very versatile quota and access control system built-in to EOS and a large variety of access protocols optimized for LAN and WAN usage.

Another major development area in the EOS ecosystem is a FUSE based file system client providing low-latency/high-throughput access to EOS data via a file system interface. The third generation allows kerberos or certificate based authentication with similar performance levels of distributed file systems like AFS and allows transparent re-exporting via NFS or CIFS protocol.

Primary author: Dr PETERS, Andreas-Joachim (CERN)

Presenter: Dr PETERS, Andreas-Joachim (CERN)

Session Classification: Data Management & Big Data Session

Track Classification: Big Data & Data Management