

<center>Montenegro, Budva, Becici, 28 september - 02 october 2015</center>



Contribution ID: 12

Type: **not specified**

dCache, Sync-and-Share for Big Data

Friday 2 October 2015 11:10 (30 minutes)

The availability of cheap, easy-to-use sync-and-share cloud services has split the scientific storage world into the traditional big data management systems and the very attractive sync-and-share services. With the former, the location of data is well understood while the latter is mostly operated in the Cloud, resulting in a rather complex legal situation.

Beside legal issues, those two worlds have little overlap in user authentication and access protocols. While traditional storage technologies, popular in HEP, are based on X509, cloud services and sync-n- share software technologies are generally based on user/password authentication or mechanisms like SAML or Open ID Connect. Similarly, data access models offered by both are somewhat different, with sync-n-share services often using proprietary protocols.

As both approaches are very attractive, dCache.org developed a hybrid system, providing the best of both worlds. To avoid reinvent the wheel, dCache.org decided to embed another Open Source project: OwnCloud. This offers the required modern access capabilities but does not support the managed data functionality needed for large capacity data storage.

With this hybrid system, scientist can share files and synchronize their data with laptops or mobile devices as easy as with any other cloud storage service. On top of this, the same data can be accessed via established mechanisms, like GridFTP to serve the Globus Transfer Service or the WLCG FTS3 tool, or the data can be made available to worker nodes or HPC applications via a mounted filesystem. As dCache provides a flexible authentication module, the same user can access its storage via different authentication mechanisms; e.g., X.509 and SAML. Additionally, users can specify the desired quality of service or trigger media transitions as necessary, so tuning data access latency to the planned access profile. Such features are a natural consequence of using dCache.

Summary

DESY has been setting up a Cloud Storage System, composed of dCache and OwnCloud, serving the needs of the scientific community in terms of fast data access, as well as Sync-and-Share functionality. We will describe the design of the hybrid dCache/OwnCloud system, report on several months of operations experience running it at DESY, and elucidate on the future road-map.

Authors: Mr QUIRIN, Buchholz (DESY); Dr FUHRMANN, Patrick (DESY); Dr MILLAR, Paul (DESY); Mr VAN DER REEST, Peter (DESY); Mr MKRTCHYAN, Tigran (DESY)

Presenter: Dr FUHRMANN, Patrick (DESY)

Session Classification: Computing for Large Scale Accelerator Facilities (LHC, FAIR, NICA, etc.) and Big Data