Content:
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.

Besides legal issues, those two worlds have little overlap in user authentication and access protocols. While traditional storage technologies, popular in HEP, are based on X.509, 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.

While OwnCloud serves as the sync-n-share entry point into dCache, the dCache team is exploring CDMI as a standard for Cloud Middlewares (e.g. OpenStack, etc) to interact with storage.

Traditionally storage systems have had well understood responsibilities and behavior, codified by the POSIX standards. More sophisticated systems (such as dCache) support additional functionality, such as storing data on media with different latencies.
From a user’s perspective, this forms a relatively simple adjunct to POSIX: providing optional quality-of-service values when writing data and optionally requesting data be staged from tape ahead of use.

The CDMI protocol provides a standard mechanism for clients to discover and use many advanced features. Such features include storing and querying metadata, searching for files matching metadata predicates, controlling a file’s quality-of-service and retention policies, providing an object store and alternative protocol discovery.

A CDMI enabled storage has the potential for greatly simplifying a more general service as some high-level functionality can be delegated to the storage system. This reduces and may remove the need to run additional services, which makes it easier for sites to support their users.

By implementing the CDMI standard, dCache can expose new features in a standards compliant fashion. Here, various scenarios are presented where CDMI provides an advantage and the road-map for CDMI support in dCache is explored.

We will describe the design of the dCache cloud support with OwnCloud and CDMI, report on several months of operations experience running it at DESY, and elucidate on the future road-map.

Primary authors: Dr. FUHRMANN, Patrick (DESY) ; Dr. MILLAR, Paul (DESY) ; Prof. HESSLING, Hermann (HTW-Berlin) ; Mrs. WESCHENFELDER, Jana (HTW-Berlin) ; Mr. VAN DER REEST, Peter (DESY) ; Mr. BUCHHOLZ, Quirin (DESY) ; Dr. BEHRMANN, Gerd (NDGF) ; Dr. ROSSI, Albert (FNAL) ; Dr. LITVINTSEV, Dmitry (FNAL) ; Mr. SCHWANK, Karsten (DESY) ; Mr. BERNARDT, Christian (DESY)

Co-authors:

Presenter: Dr. FUHRMANN, Patrick (DESY)

Session classification: Infrastructure Cloud I

Track classification: Infrastructure Clouds and Virtualisation

Type: Oral