Contribution ID: **128**   Type: **Presentations**

dCache: dependable storage for new distributed communities

Wednesday, 10 April 2013 14:30 (30 minutes)

Impact

The storage of data is a common requirement for sites. Such institutes often have to support many different user communities; some users are physically located in the campus while others may be geographically distributed. These different communities will have different requirements for their storage. Examples of where this is used will be presented.

Rather than adopting different storage solutions for the different communities, dCache provides sufficient flexibility that it can support many communities. By consolidating their software, a site can spend less effort on their storage software and more on supporting their communities.

dCache also offers features not found in similar storage solutions. This allows a site to offer new features to a community without installing new software, simply by enabling the facility within dCache.

The software within dCache is strongly standards driven; preferring standards-based protocols over proprietary ones. This means that communities will be able to use the storage capacity provided by a dCache instance with little, if any, impact on their analysis chain.

There are new challenges in how users identify themselves and their group-membership. By taking an active role in this area, dCache is ensuring that it will continue to support scientific work in a world where what people do and how they identify themselves is changing.

URL

http://www.dcache.org/

Summary

The dCache storage project is a collaboration between DESY, Fermilab
and NDGF. The software is mature and stable. dCache software is deployed on many sites and currently provides more storage capacity for CERN Large Hadron Collider (LHC) experiments than any other software.

New technologies are allowing different scientific communities to generate vast amounts of data. Most of these scientists have no experience with grid computing, yet they require storage capacity of comparable scale to LHC experiments, or exceeding it. They are also often form short-lived collaborations, which is distinct from current grid practise.

Over its lifetime, the dCache project has been evolving to provide storage for different groups of users. This has resulted in a software framework that is very flexible. We present how dCache is continuing to adjust, so it can satisfy the demands from these new communities.

Description

The dCache storage project is a collaboration between DESY, Fermilab and NDGF. The software is mature and stable: the first instance went into production over ten years ago. The task then was to provide fast access to storage for local users: users who are well known to the site running dCache.

With the introduction of grid computing, as championed by CERN and the experiments using the Large Hadron Collider (LHC) facility, the focus shifted. dCache continues to support high-performance access to data for local users but, in addition, it allows access for users unknown to the site. These users identify themselves using X.509 certificates rather than username-password. dCache successfully adapted to the grid

environment and is now used by sites throughout the world which, combined, store roughly half of all LHC data.

We are now at the brink of a data revolution. New technologies allow different scientific communities to generate vast amounts of data, comparable to and potentially exceeding that of the LHC experiments. This will likely trigger ad-hoc collaborations where people from different institutes work on large datasets. Many of these users will be unknown to the site hosting the data; however, unlike with grid computing, such collaborations will likely lack the management cohesion necessary to adopt X.509 certificates.

We will describe some of the modern, alternate approaches by which users may identify themselves. Some of the challenges of these approaches will be presented, especially those that are most problematic when handling or accessing data. An overview will be provided of the projects operating in this space, including project moonshot and LSDMA.

Work is underway within the dCache team to support these new
communities by allowing access to users who authenticate via federated identity systems. We will describe the progress so far and what the future will bring, as dCache adapts to solve storage for a federated identity world.

Primary authors: FUHRMANN, Patrick (DESY); MILLAR, Paul (DESY)

Presenter: MILLAR, Paul (DESY)

Session Classification: Community Platforms

Track Classification: Community Platforms (Track Lead: P Solagna and M Drescher)