



Contribution ID: 2

Type: **not specified**

CVMFS for Data Federations

Tuesday 7 June 2016 14:20 (20 minutes)

A data federation is a cooperating set of storage resources transparently accessible across a wide area network via a common namespace. These are often implemented through a redirector hierarchy - clients query a centralized endpoint for a given file; this redirector locates an available storage resource, then redirects the client to the remote resource.

Data federations are an increasingly used as a way to distribute large-volumes of physics data. For example, the Compact Muon Solenoid (CMS) experiment has approximately 20PB of analysis data available through it's "Any Data, Any Time, Anywhere" (AAA) federation.

However, the namespace of AAA is extremely limited - it is equivalent to just a HTTP GET. There are no directory listings, authoritative size or checksum information - despite the fact this information is known to CMS and available in the underlying storage systems and across several services; it is user-hostile for data discovery.

In this presentation, we will discuss a series of improvements made to the CVMFS core to marry a user-friendly, CVMFS-based POSIX namespace with data federation. We will demonstrate a set of CVMFS repositories of increasing complexity that utilize these new CVMFS features. These repositories serve as frontends for data federations for OSG, LIGO, and CMS.

Finally, we will discuss plans to grow this work - in terms of scale (data volume), efficiency, and features used in production.

Summary

An effort to utilize CVMFS's scalable namespace features to provide a POSIX interface for data federations.

Author: BOCKELMAN, Brian Paul (University of Nebraska (US))

Co-authors: DYKSTRA, Dave (Fermi National Accelerator Lab. (US)); WEITZEL, Derek John (University of Nebraska (US))

Presenter: BOCKELMAN, Brian Paul (University of Nebraska (US))

Session Classification: CernVM Related Activities II