

Small Fish in a Big Data Pond: Data Storage and Distribution Strategies for new VOs.

Wednesday, 10 April 2013 16:40 (20 minutes)

Summary

While most of the data currently moved over the EGI Grid is owned by the four LHC experiments, an increasing number of non-LHC, and non-HEP VOs have emerging non-trivial storage and data transfer needs.

Most of these VOs have significantly lower staffing levels than ATLAS and CMS, and economies of scale therefore make managing their growing storage and network needs disproportionately difficult.

This talk aims to provide some background and suggested strategies for such VOs: common, non-proprietary protocols; data transfer scheduling and automation; metadata management; and so on. We will also discuss which tools and changes small VOs may need made in the infrastructure to better support their needs.

Impact

Work done leading towards this presentation has included support for the T2K, NEISS.ORG.UK, Enroller, NA62 VOs in their storage management. The presenter and co-authors are members of the GridPP Storage Group, and have extensive experience in storage and data management, and support of the same across the UK.

We expect this talk to provide a basis for improving both policy and infrastructure for storage and data management for the smaller and emerging VOs. These VOs are arguably the future of the European Grid, as the problems solved by the larger scale LHC VOs are also mostly managed by their larger staffing levels.

Feedback for this talk will influence provision of future tools, and potentially core-development, for the DPM storage system.

Description

This talk aims to provide a state of the nation overview of the policies currently adopted by smaller VOs in Europe for Storage and Data Management, with general UK focus. In our experience, the use of data by small VOs varies very widely in all dimensions (storage size, throughput, replication/resilience, metadata complexity, and efficiency).

We then explore, in tandem, both suggested movements in policy that might be adopted by VOs to improve their efficiency (use of common protocols like WebDAV/https and NFS4.1/pNFS, better use of data management tools like FTS) and improvements that could be made in the infrastructure (for example, integration of FTS with metadata catalogues like the LFCs) which would dramatically improve the user experience for small VOs.

We also discuss tools and services provided by GridPP Storage Group which may be of use to small VOs. We expect/desire feedback from smaller VOs on improvements that we could provide, especially given our almost certain involvement in the DPM Community Support model which is now evolving.

Primary author: SKIPSEY, Sam (UG)

Co-authors: Dr DAVIES, Brian (RAL); BHIMJI, Wahid (UE)

Presenter: SKIPSEY, Sam (UG)

Session Classification: Community Platforms

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