

Building a global file system for data access using Large Scale CVMFS and DynaFed

Thursday, 12 July 2018 14:00 (15 minutes)

CVMFS has proved an extremely effective mechanism for providing scalable, POSIX like, access to experiment software across the Grid. The normal method for file access is http downloads via squid caches from a small number of Stratum 1 servers. In the last couple of years this mechanisms has been extended to allow access of files from any storage offering http access. This has been named Large Scale CVMFS. Large Scale CVMFS has been shown to work for experiments whose entire dataset can be stored at a single site however it has not been designed for when the data is distributed across many sites or when there is more than one copy of a file available.

DynaFed can federate http storage endpoints and is able to present a huge distributed repository as if it were one. It is an ideal complement to Large Scale CVMFS as it provides a mechanism to select the most appropriate file when more than one copy exists. The dynamic nature of the federation also allows storage to be added and removed without requiring changes to CVMFS clients running on every worker node. This paper reports on the work done within GridPP to build a global file system for data access using Large Scale CVMFS and DynaFed. The data federation includes both traditional Grid storage endpoints such as DPM as well as cloud storage such as S3 and this paper also describes the differences in their setup and observed performance.

Primary authors: DEWHURST, Alastair (STFC-Rutherford Appleton Laboratory (GB)); CONDURACHE, Catalin (STFC-Rutherford Appleton Laboratory (GB)); ROY, Gareth Douglas (University of Glasgow (GB)); SKIPSEY, Samuel Cadellin; Mr JOHNSON, Ian (STFC)

Presenter: DEWHURST, Alastair (STFC-Rutherford Appleton Laboratory (GB))

Session Classification: T4 - Data handling

Track Classification: Track 4 - Data Handling