A distributed R&D storage platform implementing quality of service

Friday, 28 August 2020 11:20 (30 minutes)

Optimization of computing resources, in particular storage, the costliest one, is a tremendous challenge for the High Luminosity LHC (HL-LHC) program. Several venues are being investigated to address the storage issues foreseen for HL-LHC. Our expectation is that savings can be achieved in two primary areas: optimization of the use of various storage types and reduction of the required manpower to operate the storage.

We will describe our work, done in the context of the WLCG DOMA project, to prototype, deploy and operate an at-scale research storage platform to better understand the opportunities and challenges for the HL-LHG era. Our multi-VO platform includes several storage technologies, from highly performant SSDs to low end disk storage and tape archives, all coordinated by the use of dCache. It is distributed over several major sites in the US (AGLT2, BNL, FNAL & MWT2) which are several tens of msec RTT apart with one extreme leg over the Atlantic in DESY to test extreme latencies. As a common definition of attributes for QoS characterizing storage systems in HEP has not yet been defined, we are using this research platform to experiment on several of them, e.g., number of copies, availability, reliability, throughput, iops and latency.

The platform provides a unique tool to explore the technical boundaries of the 'data-lake' concept and its potential savings in storage and operations costs.

We will conclude with a summary of our lessons learned and where we intend to go with our next steps.

Primary authors: LITVINTSEV, Dmitry (FNAL); Dr LANCON, Eric (BNL); ITO, Hironori (BNL); Mrs LIU, Jane (BNL); Dr FUHRMANN, Patrick (DESY/dCache.org); Dr MILLAR, Paul (DESY); Dr GARDNER, Robert William (UniChicago); MCKEE, Shawn (UMICH); Mr MKRTCHYAN, Tigran (DESY)

Presenter: Dr FUHRMANN, Patrick (DESY/dCache.org)

Session Classification: Data Management Session

Track Classification: Data Management & Big Data