Contribution ID: **483**   Type: **Poster**

Using Lustre and SLURM to process Hadoop workloads and extending to the WLCG

Tuesday 10 July 2018 16:45 (15 minutes)

The Queen Mary University of London Grid site has investigated the use of its' Lustre file system to support Hadoop work flows using the newly open sourced Hadoop adaptor for Lustre. Lustre is an open source, POSIX compatible, clustered file system often used in high performance computing clusters and, is often paired with the SLURM batch system as it is at Queen Mary. Hadoop is an open-source software framework for distributed storage and processing of data normally run on dedicated hardware utilising the HDFS file system and Yarn batch system. Hadoop is an important modern tool for data analytics used by a large range of organisation including CERN. By using our existing Lustre file system and SLURM batch system we remove the need to have dedicated hardware and only have to maintain a single platform for data storage and processing. The motivation and benefits of using Lustre rather than HDFS with Hadoop are presented. The installation, benchmarks and real world performance are presented and future plans discussed. We also investigate using the standard WLCG grid middleware services (Cream CE, Storm SRM) to provide a grid enabled Hadoop service.

Primary authors: TRAYNOR, Daniel Peter (University of London (GB)); FROY, Terry (Queen Mary University of London)

Presenter: TRAYNOR, Daniel Peter (University of London (GB))

Session Classification: Posters

Track Classification: Track 4 - Data Handling