Contribution ID: **59**   Type: **not specified**

IntegratingHadoop Distributed File System to Logistical Storage

Monday 25 March 2019 16:10 (25 minutes)

Logistical Storage (LStore) provides a flexible logistical networking storage framework for distributed and scalable access to data in both an HPC and WAN environment. LStore uses commodity hard drives to provide unlimited storage with user controllable fault tolerance and reliability. In this talk, we will briefly discuss LStore's features and discuss the newly developed native LStore plugin for the Apache Hadoop ecosystem. The Hadoop Distributed File System (HDFS) will directly access LStore using this plugin allowing users to create Hadoop clusters on the fly in an HPC environment. The primary benefit of the plugin is that it avoids the need for data redundancy across a traditional Hadoop and HPC cluster. Moreover, the on the fly Hadoop clusters created in the HPC environment can be scaled as needed and tune the hardware requirements to the analysis - large memory needs, GPU, etc.

We will show several empirical results using the plugin in both a traditional HPC environment and utilizing a high-latency WAN connection. The proposed plugin is compared with two current LStore interfaces: LStore command line interface and LStore FUSE mounted client interface.

Primary author: Dr BAO, Shunxing (Vanderbilt University)

Co-authors: Dr TACKETT, Alan (Vanderbilt University);   MELO, Andrew Malone (Vanderbilt University (US))

Presenter: Dr BAO, Shunxing (Vanderbilt University)

Session Classification: Storage & Filesystems

Track Classification: Storage & Filesystems