Contribution ID: **103**   Type: **Talk**

Converging Storage Technologies Using a Flexible HPC Storage Framework

Wednesday, 2 December 2020 12:15 (30 minutes)

Traditionally, file systems are mostly monolithic, making it hard to experiment with new approaches and technologies. Exchanging core functionality within a file system is a burdensome task, leading to a lack of innovation in this area. However, data volumes are growing rapidly because the ability to capture and produce data is increasing at an exponential rate. Rising core counts and data volumes present challenges for contemporary storage systems, especially regarding metadata performance and data management. This makes it even more important to investigate new ways of storing and managing data efficiently.

HPC storage systems are typically designed around POSIX-compliant parallel distributed file systems that are accessed using sophisticated I/O libraries. The file system and library layers are strictly separated for portability reasons. While this allows exchanging individual layers, their complexities pose a high barrier of entry. This is especially problematic for shorter research projects and presents a significant hurdle for young researchers and students.

Within the JULEA and CoSEMoS projects, we are aiming to change this. JULEA is a flexible storage framework that can be used to prototype new ideas related to storage and file systems. It allows offering arbitrary I/O interfaces to applications and includes interfaces for object, key-value and database storage. The framework has been designed to be easy to set up and run without administrative privileges, so it can be used on a wide range of software and hardware environments. It also serves as the foundation of the CoSEMoS project, which explores the benefits of a coupled storage system for self-describing data formats, such as HDF5 and ADIOS2. This allows the storage system to manage file metadata found within these data formats and makes it possible to use structural information for selecting appropriate storage technologies. For instance, metadata can be stored in database systems that can be queried efficiently. Moreover, making use of established data formats allows running existing applications without modifications, which helps preserve past investments in software development.

CoSEMoS enables novel data management approaches via a data analysis interface that gives applications direct access to JULEA's backends, eliminating the need to sift through large volumes of data to find relevant data points. Breaking up the strict separation has additional long-term benefits, such as being able to take data migration decisions based on structural information found within the file formats.

This talk will briefly introduce the JULEA and CoSEMoS projects, and show the opportunities enabled by their novel storage system design.

Student?

No

Supervisor name

Supervisor email

Primary author: KUHN, Michael (Otto von Guericke University Magdeburg)

Presenter: KUHN, Michael (Otto von Guericke University Magdeburg)

Session Classification: HPC

Track Classification: Storage and IO