

Abstract

HEC I/O Initiative - A Multi-Institutional Effort between Georgia Tech's Center for Experimental Research in Computer Systems (CERCS)

Oak Ridge National Laboratory and Sandia National Laboratories,
University of New Mexico

with additional support from CERCS member companies, including CISCO and Intel

Collaborators: Karsten Schwan (GT), Matthew Wolf (GT/ORNL), Patrick Bridges (UNM), Barney Maccabe (UNM), Ron Oldfield (Sandia), Scott Klasky (ORNL)

The difficulties faced by scientists and engineers in attaining high performance I/O for data-intensive MPP applications are exacerbated by the low level of abstraction presented by current I/O systems. The multi-institutional HEC I/O initiative aims to create higher level I/O abstractions, along with automated and dynamic methods that will ease future MPP developers' tasks. Specifically, the proposed framework models I/O as I/O Graphs that 'connect' application components with input or output mechanisms like file systems, based on metadata constructed offline by autonomous metabots. Our approach enhances the I/O functionality available to end users in several ways:

- I/O Graphs can be programmed to realize application-specific I/O functionality, such as data filtering and conversion, data remeshing, and similar tasks.
- I/O Graph management is automated, including the mapping of their logical graph nodes to underlying physical MPP and distributed machine resources.
- I/O performance in SSDS will be improved compared to that of current high performance file systems, by integrating the computational I/O actions of I/O Graphs with the backend file systems that store high volume data and with the I/O actions already taken by applications, and by moving metadata management offline into metabots.

The implementation of the HEC I/O system and its I/O Graph model will impact a substantial HPC user community, due to its planned integration with the Lightweight File System (LWFS) currently under development at Sandia National Laboratories (SNL). As a key early adoption/co-development platform, we leverage our engagement with a prominent high performance application, the gyrokinetic fusion application GTC, through our collaborators at Oak Ridge.