Seeing the trees for the forest: Describing HPC Filesystems with the Grand Unified File-Index (GUFI)

Abstract

Jenna Kline
Mentors: Jason Lee and Rusty Davis, LANL HPC DES
LA-UR-23-28958

High performance computing (HPC) filesystems are extremely large, complex, and difficult to manage with existing tools. It is challenging for HPC administrators to describe the current structure of their filesystems, predict how they will change over time, and the requirements for future filesystems as they continue to evolve. Previous studies of filesystem characteristics largely predate the modern HPC filesystems of the last decade. The Grand Unified File Index (GUFI) was used to collect the data used to compute the characteristics of six HPC filesystems indexes from Los Alamos National Laboratory (LANL) representing 2.8 PB of data, containing 36 million directories and 600 million files. We will present a methodology that uses GUFI to characterize the shape of HPC filesystems to help system administrators to understand their key characteristics.

Los Alamos
NATIONAL LABORATORY