Benchmarking Effects of Erasure Scheme and MPI Configuration on MarFS Throughput

Abstract

Janya Budaraju, Paul Karhnak, Zachary Snyder
Mentors: David Bonnie, Garrett Ransom

MarFS is an open-source, medium-term campaign storage platform used in Los Alamos National Laboratory (LANL) supercomputing clusters. MarFS performs parallel file operations to read to and write from a mount point, employing multi-layer erasure coding through an Intel Intelligent Storage Acceleration Library (ISA-L) Reed-Solomon error correction implementation to create redundancy and resiliency. Currently, optimizing MarFS erasure scheme and Message Passing Interface (MPI) parameters for a specific cluster requires close hardware familiarity; we created a software suite to benchmark MarFS throughput across multiple configuration parameters, abstracting away hardware-level considerations. In addition, we identified performance patterns and validated assumptions about ISA-L erasure coding in HPC workloads. Our results offer detailed insight into ISA-L erasure coding performance on our cluster and demonstrate our tools' viability. These tools we developed provide a key starting point for optimizing MarFS performance, accommodating growing storage system performance needs at LANL.

Los Alamos
NATIONAL LABORATORY