



LA-UR-24-28183

Author(s): Paul D. Karhnaek

Mentors: Dave Bonnie, Garrett Ransom

Title: MUSTANG: A Powerful Vehicle for MarFS Object Cataloging and Retrieval

Abstract:

MarFS is a POSIX-like file system which stores user data at Los Alamos National Laboratory (LANL) on a medium-term basis. MUSTANG (MarFS Underlying Storage Tree and Namespace Gatherer) is a utility to scan a user-facing MarFS instance and its contents in parallel. MUSTANG outputs a list of unique MarFS object IDs encountered during traversal which may later be used to retrieve specific objects from backup storage. Eventually, MUSTANG is intended for use as an administrative tool in Marchive (MarFS archive) hybrid POSIX-tape file systems to efficiently dictate which tapes need to be accessed to move an object or objects into a user-facing POSIX file system. MUSTANG has demonstrated promising performance at scale for various realistic datasets.

Keywords: file systems, parallel programming, threads