

mpiFileUtils

Parallel File Utilities for HPC

November 13, 2017

Danielle Sikich



Motivation

- What is mpiFileUtils?
 - Suite of MPI-based suite of file utilities built on a common library for handling large HPC datasets
 - Similar to current unix file commands such as cp, rm, cmp, etc.
- Why mpiFileUtils?
 - Applications that produce and consume data are highly-parallel and frequently optimized for HPC file systems
 - UNIX file management utilities are single-process solutions

Tools

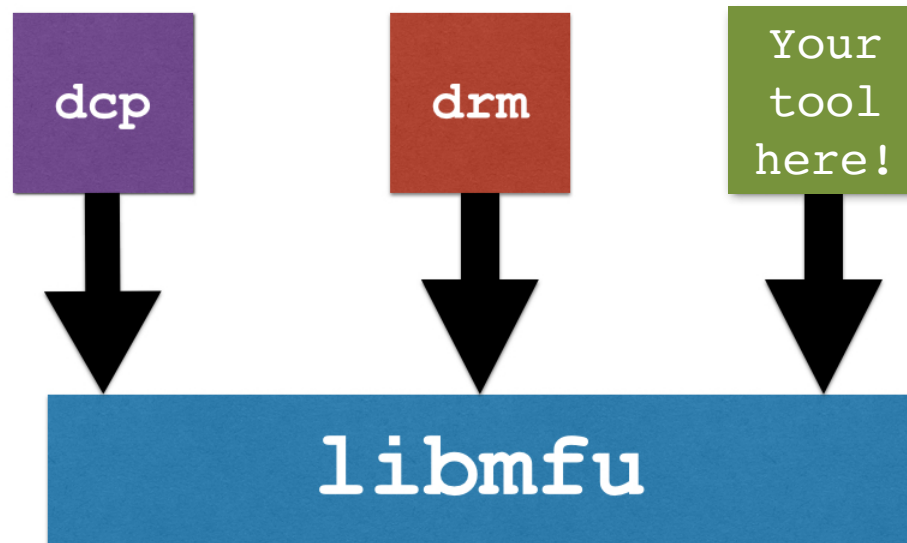
- Currently available for use:
 1. dchmod – change permissions & group access
 2. dcp – copy files
 3. dcmp – compare files
 4. drm – remove files
 5. dstripe – restripe files
 6. dwalk – list files

1 Million File Experiment

Tool	1 Proc	64 Procs	256 Procs
cp	12 hours		
dcp	1x (12hr)	37x (19m)	51x (14m)
rsync	6 hours		
dcmp	0.5x (12hr)	51x (7m)	60x (6m)
rm	15 minutes		
drm	0.5x (28m)	3x (5m)	5x (3m)

Common Library

- Easy for contributors to add new parallel file utilities based on a common API



Acknowledgements & Download

- Result of multi-institutional collaboration: LANL, LLNL, ORNL, DataDirect Networks, Red Hat, and Australian National University
- We are in Spack
 - spack install mpifileutils
- Link to GitHub site:
 - <https://github.com/hpc/mpifileutils>

