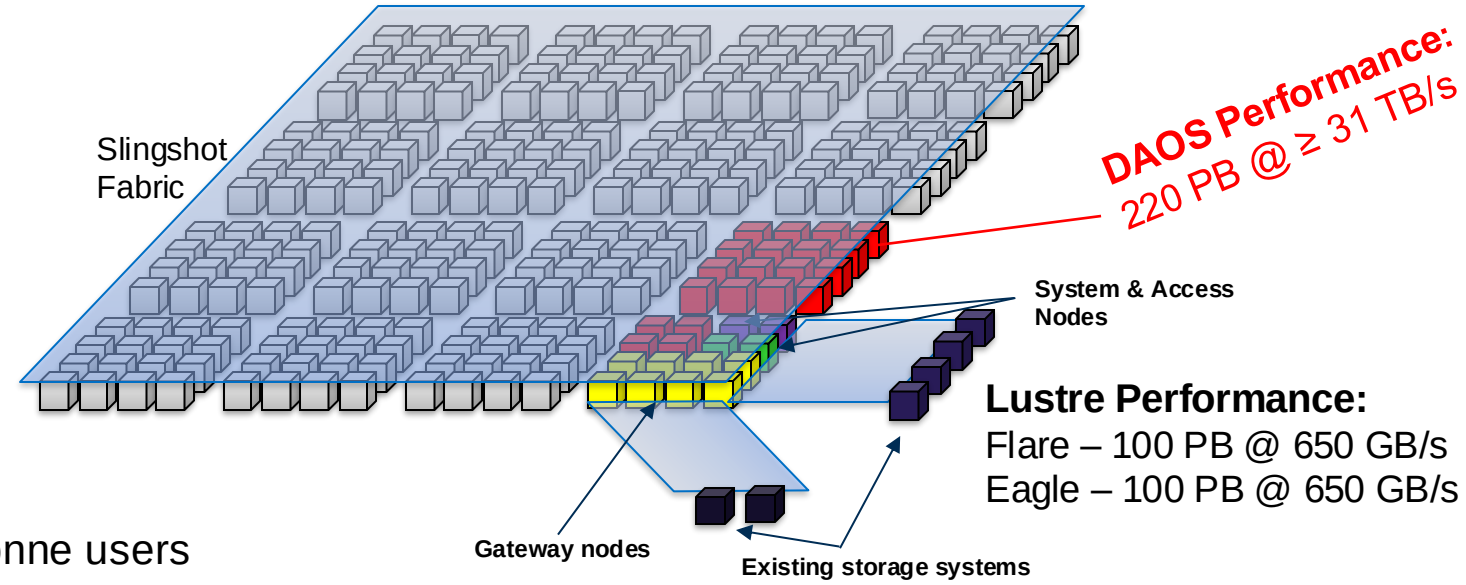
ALCF Site Update

Kevin Harms
Paul Coffman
Kaushik Velusamy
Gordon McPheeters

Aurora DAOS

- System is under acceptance now
  - daos_perf
    - 800 daos server nodes
    - Used for acceptance testing
  - daos_user
    - 128 daos server nodes
    - General access for Intel and Argonne users
- Both systems running v2.6.2-tb2 and SHS 11.0.0
- Testing up to 8000 compute nodes or 64,000 endpoints
- Core focus is on hardening and bug fixing



num CN nodes	ppn	num DAOS nodes	interception lib	daos backend mode	access	type	transfer size	block size	agg file size	Write TB/s	Read TB/s
256	16	128	libpil4dfs.so	posix	single-shared-file	independent	16MiB	2 GiB	8TiB	5.23	4.47
256	16	128	libpil4dfs.so	mpiio-daos:/	single-shared-file	independent	16MiB	2 GiB	8TiB	4.64	4.52
256	16	128	libpil4dfs.so	dfs	file-per-process	independent	16MiB	2 GiB	8TiB	5.51	4.12
256	16	128	libpil4dfs.so	dfs	file-per-process	collective	16MiB	2 GiB	8TiB	5.54	4.11
256	16	128	libpil4dfs.so	dfs	single-shared-file	independent	16MiB	2 GiB	8TiB	5.73	4.44
256	16	128	libpil4dfs.so	dfs	single-shared-file	collective	16MiB	2 GiB	8TiB	5.73	4.37

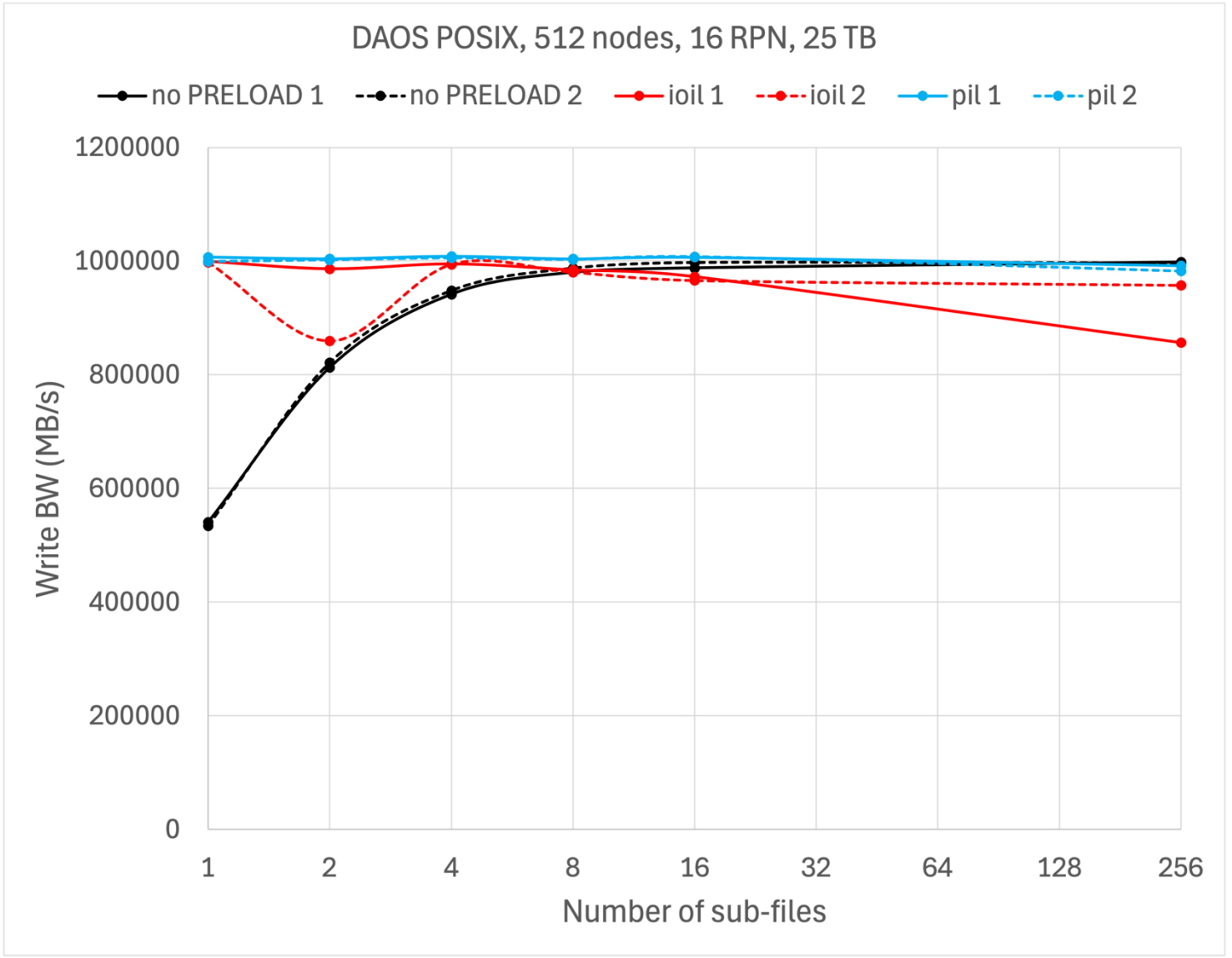
SX – no redundancy, roofline performance

PDSW’24

- **Initial Experiences With DAOS Object Storage on Aurora**
  - Rob Latham, Argonne National Laboratory
    Robert Ross, Argonne National Laboratory
    Phillip Carns, Argonne National Laboratory
    Shane Snyder, Argonne National Laboratory
    Kevin Harms, Argonne National Laboratory
    Kaushik Velusamy, Argonne National Laboratory
    Paul Coffman, Argonne National Laboratory
    Gordon McPheeters, Argonne National Laboratory
- Good overview of per-client performance and tuning
- Results on small cluster

HACC

- Running HACC Generic IO
  - HACC is Cosmology code which simulates galaxy formation
  - Generic IO is a benchmark using their application I/O code
- Running at 512 compute nodes
  - Writing 25TB checkpoint
  - Using POSIX API
- Using daos_user (20 servers)
- Code has been write optimized
- 1 TB/s with SX



DAOS at Large Scale

- Intel testing at large server counts and large client counts prior to acceptance testing
- IOR
  - 793 daos servers (2 engines each)
  - 8192 compute nodes (65,536 endpoints)
  - 1MiB EC cell size
  - EC_16P2GX
  - IO-500 ”easy” configuration using DFS
    - 16 MiB block size
    - 22,489 GiB/s write
    - 22,618 GiB/s read
  - IO-500 “hard” configuration using DFS
    - 21115360 chunk size
    - 1223 GiB/s write
    - 8729 GiB/s read
- mdtest
  - 775 daos servers (2 engines each)
  - 8192 compute nodes (64,536 endpoints)
  - DFS with files configured RP_3G1 and directories RP_3GX
  - IO-500 “easy”
    - Create: 48438 k/ops
    - Stat: 225635 k/ops
    - Remove: 31407 k/ops
  - IO-500 “hard”
    - Create: 35626 k/ops
    - Stat: 237749 k/ops
    - Remove: 35312 k/ops

Data graciously provided by Dalton Bohning, Intel DAOS