TSC Update



Johann Lombardi, TSC Chair, DAOS Foundation Virtual DAOS User Group, May 2025



d a o s

https://foundation.daos.io

TSC Structure

- Voting Members
 - O Argonne: Kevin Harms
 - *Google*: corwin
 - O HPE: Lance Evans
 - O Intel: Allison Goodman
 - O Vdura: Brian Mueller
 - O TSC Chair: Johann Lombardi
- Meet weekly (public) with rotating schedule
 - O Members distributed across US, EU, China and Australia



intel VDURA





TSC Scope

- Define community roadmap (2.8+)
 - O Gather contributions from all community members
 - O Publish roadmap on https://daos.io
- Produce community releases (2.8+)
 - O Track progress, review jira tickets & test results
 - O Tag release and sign/distribute packages
 - O Provide docker images
- Organize DAOS development
 - O Simplify contributions
 - O Organize gatekeeping (members, responsibilities, process)
 - O Document contribution process





TSC Scope

- Community test infrastructure
 - O Goal: artifacts and logs available to all contributors
 - O Expand coverage
 - ARM/AMD
 - More fabrics
 - More linux distributions
 - Cloud environments
 - Focus on pmem-less mode
- Working groups
 - O Open to anyone
 - O Forums for DAOS users/administrators/contributors to exchange
 - O Rotating schedule







(CR)

- Container parking/serialization

а 0 S

DAOS 2.8

DAOS 2.8 (DAOS Foundation Release)

OS Packages:

- Leap 15.6

- RHEL/Rocky/Alma 8.10/9.4

Networking:

- DOCA-OFED support

Features:

- Optimized object placement
- Mount POSIX snapshots RO
- Client telemetry
- Incremental rebuild/reintegration
- Catastrophic recovery and
- distributed consistency checker
- Fault domains beyond servers

Tech preview:

- Non-PMem support phase 2
- Pytorch data loader
- Rolling upgrade preparation

UX Improvements:

- Reintegration of all pools
- daos pool listing





Delayed Rebuild & Incremental Reintegration

• Delayed rebuild

- O Originally developed to address cloud needs
- O Don't trigger rebuilt on failure
- O Wait for the original or a new node to be added
- O Degraded reads
- O Write supported
- Incremental reintegration
 - O Don't wipe out storage on reintegration
 - O Reuse content (if available)
 - O Reconstruct only I/Os starting from a given epoch/version
- Combining the two for better robustness
 - O Delay rebuild for some time to address temporary outage
 - O Exclude the failed node after the timeout
 - O Reintegrate failed node quickly





SuperMUC NG System

SuperMUC NG Phase 2 DAOS

- 42x Lenovo Storage nodes
 - O 2x Xeon 8352Y CPUs (ICX)
 - O 512GB DRAM
 - O 8x 3.84TB NVMe SSDs
 - O 2x HDR IB NICs
 - O 2TB Optane Persistent Memory 200
- 90x Client nodes







SuperMUC NG System Comparison

SuperMUC NG Phase 2 DAOS

- 42x Lenovo Storage nodes
 - O 2x Xeon 8352Y CPUs (ICX)
 - O 512GB DRAM
 - O 8x 3.84TB NVMe SSDs
 - O 2x HDR IB NICs
 - O 2TB Optane Persistent Memory 200
- 90x Client nodes





IRIS MSKCC WekalO

- 54x Dell Storage nodes
 - O 2x Xeon 5317 CPUs (ICX)
 - O 256GB DRAM
 - O 8x 15TB NVMe SSDs
 - O 2x HDR IB NICs
- 261x Client nodes



SuperMUC NG Performance Comparison

SuperMUC NG Phase 2 DAOS

IOR & FIND	
EASY WRITE	896.71 GiB/s
EASY READ	1,872.09 GiB/s
HARD WRITE	252.43 GiB/s
HARD READ	718.81 GiB/s
FIND	12,733.44 kIOP/s

IRIS MSKCC WekalO

IOR & FIND	
EASY WRITE	383.77 GiB/s
EASY READ	1,076.53 GiB/s
HARD WRITE	51.68 GiB/s
HARD READ	190.49 GiB/s
FIND	424.40 kIOP/s





SuperMUC NG Performance Comparison

SuperMUC NG Phase 2 **DAOS**

METADATA	
EASY WRITE	6,324.79 kIOP/s
EASY STAT	29,403.34 kIOP/s
EASY DELETE	3,442.67 kIOP/s
HARD WRITE	2,644.93 kIOP/s
HARD READ	17,023.13 kIOP/s
HARD STAT	23,242.01 kIOP/s
HARD DELETE	3,112.59 kIOP/s

IRIS MSKCC WekalO

METADATA	
EASY WRITE	1,484.48 kIOP/s
EASY STAT	15,370.21 kIOP/s
EASY DELETE	1,693.76 kIOP/s
HARD WRITE	281.11 kIOP/s
HARD READ	6,806.84 kIOP/s
HARD STAT	8,791.83 kIOP/s
HARD DELETE	324.23 kIOP/s





IO500 Per-server Performance (production list)











Resources

- Foundation website: <u>https://daos.io/</u>
- Github: <u>https://github.com/daos-stack/daos</u>
- Online doc: <u>https://docs.daos.io</u>
- Mailing list & slack: <u>https://daos.groups.io</u>
- YouTube channel: <u>http://video.daos.io</u>

		100 1000			
	Agenda		August anness Statistical anness		
į	- Unionshi - Unionshi - Unionshi		Elite, tagener constants of the second	-	Anna a bairtean ann Anna a bairtean Anna a bairtean Anna a bairtean Anna a bairtean Anna a bairtean
-		-	4 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		and the second
	-		Course of Courses		The second second





The Open-Source Storage Platform for Al

& HPC

Features



