Hewlett Packard Enterprise

ISC High Performance
REINVENTING HPC
MAY 12 – 16, 2024 | HAMBURG, GERMANY

HPE DAOS Update

Torben Kling Petersen, PhD

Distinguished Technologist

Lead HPC Storage Architect

Global HPC and AI BU

HPE's DAOS Involvement

DAOS Foundation:	Founding member of the Foundation Widens the community of DAOS developers and adopters
Metadata on SSD:	Eliminates the original requirement to use persistent memory hardware Enhancements to the Memory allocator
PyTorch:	Optimize DAOS IO for PyTorch Evaluation is based on fsspec over DFS (& adding dlio_benchmark)
SlingShot-200	Enabling full SlingShot support including using the ANL "Mercury" RPC framework
Solution HW:	Proliant gen 11 servers (e.g., DL360 G11 or DL380 G11) Roughly doubles single-rack performance
POC Systems:	DAOS testbed availability for application integration and tuning Assists customers, partners, ISVs in integrating with DAOS

DAOS POC System Update

- A Single-Rack Solution with Maximums:
  - 32 DL-360 Gen11; 1TB DDR5, 5PB flash
  - Four 200Gb (Slingshot) or 400Gb (Nvidia) Switches
  - ~1.4TBps/0.7TBps raw read/write throughput
  - *~64M/32M peak read/write operations per second*
  - Double these maximums available with chilled doors
- Unbundled Repeatable Solution Delivery Method
  - Qualified hardware and software BOM
  - HPCM cluster management software
  - Light installation / configuration scripting
  - Reference doc set: for field or factory integration
  - Customer system administration skills required

- Customer accessible test system in DMZ in CF labs

Up to 2 HPE Management Servers:
- DL-325 Gen11 single-socket

Up to 4 200 or 400GbE Switches:
- HPE Slingshot 200
- Mellanox QM8700

Up to 32 HPE DAOS Server Cfg:
- DL-360 Gen11
- 1-Socket Sapphire Rapids Cfg *
- Max 10x Gen5 NVMe SSD 3.2TB+ *
- 16x DDR5 *
- 200Gb or 400Gb NIC *

* Double density possible with chilled doors

DAOS Administration Support

- Cluster Setup Process
  - Compliant HW is pre-assembled onsite or in HPE Manufacturing with firmware / BIOS leveled / configured
  - Admin node's OS/HPCM is installed & added to customer admin network
  - OS distro to be deployed to DAOS servers is added to admin node's HPCM repository
  - BMC & server OS access MACs, and BMC login info are added
- HPE Performance Cluster Manager (HPCM)
  - Server cluster management & monitoring via top-of-rack admin node
  - Can optionally manage compute nodes attached to DAOS as well
- DCM command set augments HPCM
  - Supports multiple logical DAOS systems / clusters within one physical cluster of HPE Proliant nodes
  - Programmatically sets up and tears down mini-clusters on subgroups of nodes
  - Operates / administers DAOS on each of the configured mini-clusters
  - Familiar to HPCM administrators using similar commands

Hewlett Packard Enterprise

ISC High Performance
REINVENTING HPC
MAY 12 – 16, 2024 | HAMBURG, GERMANY

Thank you

tkp@hpe.com