

The future of supercomputing storage

Andrey Kudryavtsev, Senior Principal Product Manager
HPE Cray Supercomputing Storage Systems K3000 (DAOS)

HPE Cray Supercomputing Storage Systems

HPE offers the highest capacity, highest performance data storage systems in the world.

690PB Capacity | >25TB Per Second | 1B IOPS

HPE Storage empowers HPC and AI customers with best-in-class data services.

POSIX | KV Record | Fast Object & S3 | Python & PyTorch | Spark | HDF5 | Block

E2000 and K3000 address critical infrastructure requirements from top-of-rack to exascale.

lustre

CORE

HPE Cray Supercomputing
Storage Systems

E2000
(Lustre)

Throughput-optimized
Full POSIX compliance
Scaled metadata servers
Lustre kernel filesystem
Gateways for S3 and NFS
Encryption and compression
Tiering between flash & HDD
Declustered RAID and hot swap
Fully HA with rolling upgrades
Fast RDMA networks & multipath
GUI and redfish API
management



NEW

HPE Cray Supercomputing
Storage Systems

K3000
(DAOS)

IOPS-optimized
KV, object, POSIX APIs
Distributed metadata transactions
100% userspace w/ bypass
Layered middleware options
Multitenancy certificates and ACLs
Asynchronous and lockless for flash
Object-level erasure and replication
Self-healing and SW-defined scale-out
Infiniband and Slingshot RDMA
CLI and Redfish API managed



HPE Cray Supercomputing Storage Systems K3000 (“K3000”)

General availability planned for Q1-CY2026

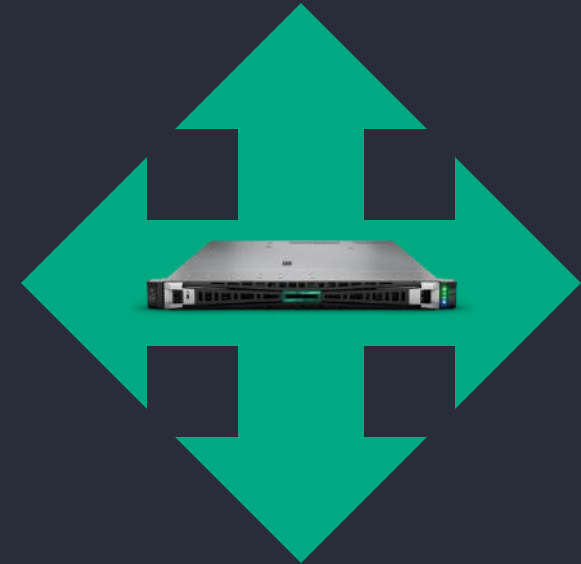
- **DAOS Management Node** (1 needed per file system)
 - 1U HPE ProLiant Compute DL325 Gen11
- **DAOS Server Nodes** (4 to 1,000s per file system)
 - 1U HPE ProLiant Compute DL360 Gen12
 - Up to 20x 3.84/7.68 /15.36 E3.S 1 DWPD NVMe SSDs per node
 - 2x 48 Core CPUs (96 total cores per node)
 - 512GB/1024GB/2048GB DRAM per node based on drive size
 - Up to 85 GB/sec Read/Write throughput per 1 U node
 - Up to 2 million random 4 KiB IOPS per 1U node
 - 4 node minimum configuration
 - Scale out in units of 1 or more node(s)
- **Management Network**
 - Single 1 Gbps Ethernet management network
 - 10 Gbps Ethernet management node connection
- **Fabric Support**
 - InfiniBand NDR
 - 400 Gbps Ethernet
 - HPE Slingshot interconnect 200 or 400
- **Multiple Rack and Rear Door Heat Exchanger Options**



1
DAOS management node



**Add as many DAOS server nodes
as necessary to meet your
requirements**



How to configure a “K3000” storage system

- **Minimum system configuration**

- 1 Management network switch
- 1 High speed network switch
(HPE Slingshot 200/400, IB NDR, 400gE)
- 1 DAOS Management Node
(HPE ProLiant Compute DL325 Gen11)
- 4 DAOS Storage Nodes
(HPE ProLiant Compute DL360 Gen12)

- **Optional:**

- Additional Management Nodes (up to 3 per system)
- Additional Storage Nodes* (up to 1,000s)
- Additional management and high speed network switches
- NFS/SMB/S3 gateways (In development)



→ **Four DAOS Storage Node densities**

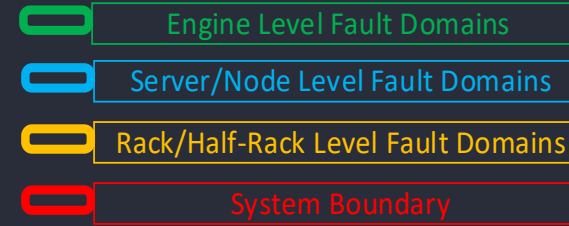
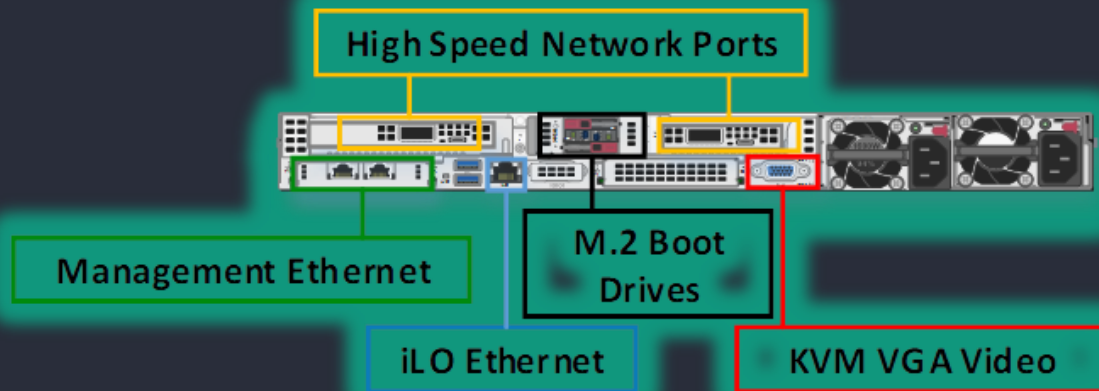
- To optimize performance and capacity needs
 - HPE Cray SC K3000 8EDSFF SSU Controller
 - HPE Cray SC K3000 12EDSFF SSU Controller
 - HPE Cray SC K3000 16EDSFF SSU Controller
 - HPE Cray SC K3000 20EDSFF SSU Controller

Three SSD capacities

- HPE Cray SC 3.84TB NVMe RI E3.S SSD
- HPE Cray SC 7.68TB NVMe RI E3.S SSD
- HPE Cray SC 15.4TB NVMe RI E3.S SSD

*All DAOS Storage Nodes must be uniform across the system

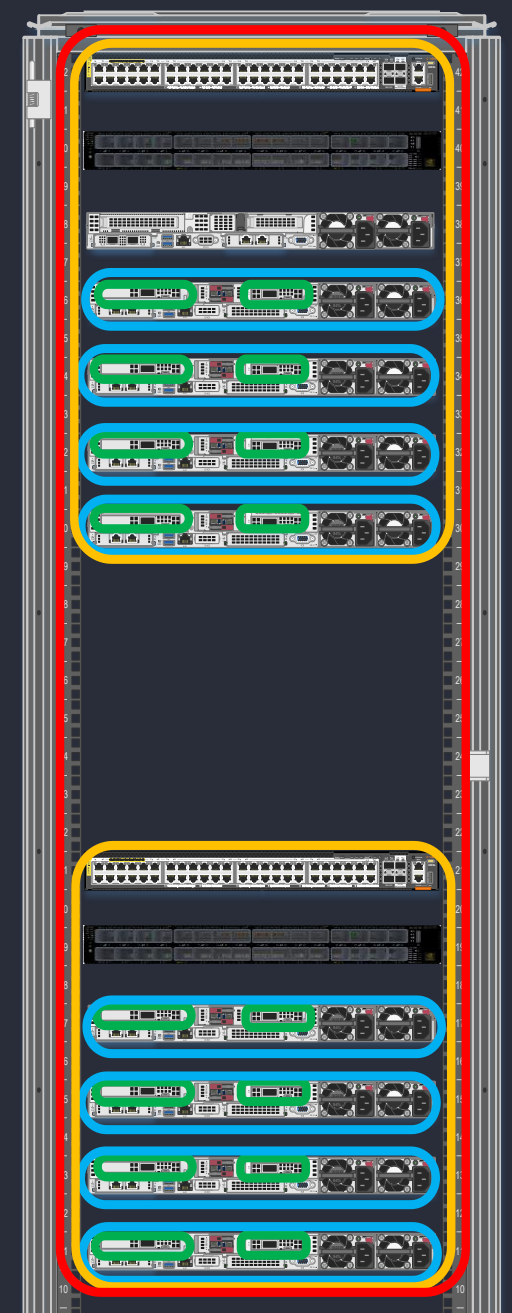
K3000 Fault Domains



Fault Domain Contains

- 1 Management Switch
- 1 High Speed Network Switch
- 1-18 DAOS Nodes

HSN Switches are 1:1 node to Uplink



HPE K3000 Value Add

• Technology

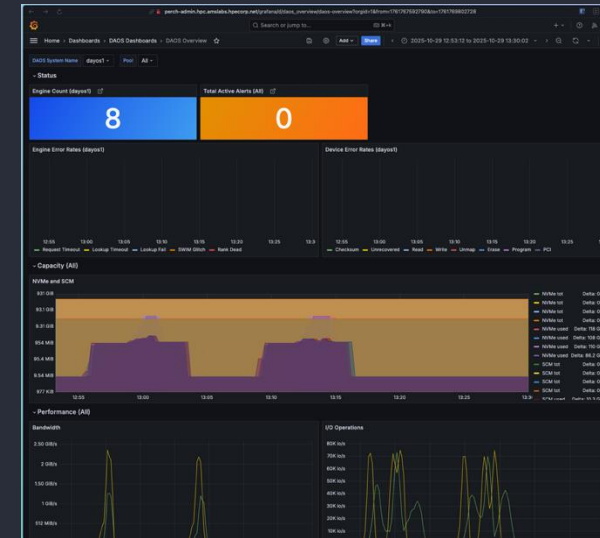
- K3000 1.0 release (DAOS 2.8 release and HPE cluster management)
- Full platform integration with HPE ProLiant DL360Gen 12 family
- Automatic HW discovery, provisioning, and configuration
- Simplified deployment and control interface with CSC (Cray Storage Command)
- Admin node for cluster monitoring and telemetry aggregation
- API RESTful interface for integration with other services, including UI
- New Dashboard for visualization

• Integration

- K3000 is fully integrated appliance up to 4000 Storage nodes
- Rack or Rackless delivery
- Infrastructure planning and optimizations

• Support

- Hardware and software installation services
- Commercial support offering with 3- and 5-year
- On-site support
- Special offerings for PoC configuration support



System setup

K3000

- Provision from bare metal
- Automatically detect and configures NVMe devices, HSN nics, BIOS
- Automatic DAOS configuration: Numa-aware, network-aware, customized fault domains
- Quickly reconfigure nodes into different DAOS systems
- Automatically manages certificate deployment and rotation

```
# csc daos system create mysys --nodecount 4
# csc daos system list
# csc daos pool create --faulttol 1 mypool --size 1T
```

Pool Size Used Avail Capacity Status

mypool 1.0 TB 26 GB 974 GB 2.6% Ready

DAOS

```
https://docs.daos.io/v2.6/QSG/setup_rhel/#start-the-daos-servers
pdsh -w $ALL_NODES 'sudo wget -O /etc/yum.repos.d/daos-packages.repo
https://packages.daos.io/v2.6/EL8/packages/x86_64/daos_packages.repo'
pdsh -w $ALL_NODES 'sudo rpm --import https://packages.daos.io/RPM-GPG-KEY'
pdsh -w $ALL_NODES 'sudo yum install -y epel-release'
pdsh -w $ADMIN_NODES 'sudo yum install -y daos-admin'
pdsh -w $SERVER_NODES 'sudo yum install -y daos-server'
daos_server nvme scan
cd /tmp /usr/lib64/daos/certgen/gen_certificates.sh
pdsh -S -w $ALL_NODES -x $(hostname -s) scp -r $(hostname -s):/tmp/daosCA /tmp
pdsh -S -w $SERVER_NODES sudo cp /tmp/daosCA/certs/daosCA.crt /etc/daos/certs/
pdsh -S -w $SERVER_NODES sudo cp /tmp/daosCA/certs/server.crt /etc/daos/certs/
pdsh -S -w $SERVER_NODES sudo cp /tmp/daosCA/certs/server.key /etc/daos/certs/
pdsh -S -w $SERVER_NODES sudo cp /tmp/daosCA/certs/agent.crt
/etc/daos/certs/clients/agent.crt
pdsh -S -w $SERVER_NODES sudo chown daos_server:daos_server
/etc/daos/certs/daosCA.crt pdsh -S -w $SERVER_NODES sudo chown
daos_server:daos_server /etc/daos/certs/server.*
pdsh -S -w $SERVER_NODES sudo chown daos_server:daos_server
/etc/daos/certs/clients/agent.crt
pdsh -S -w $SERVER_NODES sudo chown daos_server:daos_server /etc/daos/certs/clients
pdsh -S -w $SERVER_NODES sudo lspci | grep -i nvme
Create a server configuration file by modifying the default /etc/daos/daos_server.yml
Create an agent configuration file by modifying the default /etc/daos/daos_agent.yml
Create a dmg configuration file by modifying the default /etc/daos/daos_control.yml
pdsh -S -w $SERVER_NODES "sudo systemctl daemon-reload"
pdsh -S -w $SERVER_NODES "sudo systemctl start daos_server"
dmg storage format -l $SERVER_NODES # can use --force if needed
dmg pool create --size 50GB tank --properties=svc_rf:1
```

Software upgrade

K3000

- Automatically stop services, upgrade software, restart across all servers
- Different systems can be upgraded independently for TDS vs Production

```
# csc cluster software install --  
file=/root/CSSS-DAOS-0.0.7-x86_64.iso  
  
# csc cluster software deploy -i daos-  
storage-0.0.7 --nodespec 1-4
```

DAOS

```
dmg system stop --force  
pdsh -S -w $SERVER_NODES "sudo systemctl  
stop daos_server"  
pdsh -w $SERVER_NODES 'sudo wget -O  
/etc/yum.repos.d/daos-packages.repo  
https://packages.daos.io/v2.6/EL8/packag  
es/x86_64/daos_packages.repo'  
pdsh -w $SERVER_NODES 'sudo yum install  
-y daos-admin'  
pdsh -w $SERVER_NODES 'sudo yum install  
-y daos-server'  
pdsh -S -w $SERVER_NODES "sudo systemctl  
start daos_server"  
dmg system start
```

HPE K3000 Next Release Capabilities

- **Baseline**
 - DAOS 3.0
 - Migrate to newer OS (RHEL, Rocky 9.6)
 - Client OS list expanded
- **Platform capabilities**
 - Expand Client HW validation (AMD and ARM compute)
 - Metadata-on-SSD Phase 2 support
 - Expanded HW options on DL360 Gen12 for reduced memory configs
 - Introduce second vendor SSD / Larger capacity TLC SKU (30TB)
 - SED SSD SKUs support (HPE iLO managed)
 - Integrated Solutions with Juniper and Aruba switches
- **Data Integrity**
 - Incremental Rebuild / Reintegration
 - DAOS Pool migration/ transition MD-on-SSD Phase 1 to Phase 2
 - Client version interoperability
 - Rolling upgrade preparation (Pause client I/O and restart services)
 - Fault Domain beyond server support
- **Data Services**
 - Improved POSIX support
 - S3 Gateway
- **Administration**
 - GUI support
 - Expanded documentation

**HPE CRAY
SUPERCOMPUTING
STORAGE SYSTEMS
K3000**



The **industry's first** successor to an exascale machine

Discovery will replace Frontier that was first to break the exascale barrier

- HPE Supercomputing Management Software
- > 100 GX5000 racks
- HPE Cray SC GX350a accelerated blades with AMD Venice CPU and AMD MI430X GPUs
- Next generation HPE Slingshot
- ~ 300 PB Lustre and DAOS storage from HPE



[Link to announcement](#)

K3000/DAOS Related Events at SC25

16-Nov (9:00am-1:00pm): **9th DAOS User Group (DUG)**

- <https://daos.io/event/9th-daos-user-group-dug-at-sc25>
- Open to all SC25 attendees

16-Nov (08:30am-5:00pm): **Tutorial**

Modern High Performance I/O: Leveraging Object Stores

- <https://sc25.conference-program.com/presentation/?id=tut155&sess=sess270>
- Requires an SC25 tutorial pass and registration

16-Nov (11:10am-11:30am): *SmartNIC offload* presentation at **REDIS Workshop**

- https://sc25.conference-program.com/presentation/?id=ws_redis101&sess=sess219
- Requires an SC25 workshop pass

17-Nov (9:00am-5:30pm): **Parallel Data Systems Workshop (PDSW)**

- <https://www.pdsw.org/index.shtml>
- Requires an SC25 workshop pass

20-Nov (12:15pm-1:15pm): **IO500 BOF**

- <https://sc25.conference-program.com/presentation/?id=bof135&sess=sess473>
- Open to all SC25 attendees

17-Nov to 20-Nov (Exhibitor hours): **DAOS and K3000 at HPE booth #3704**

17-Nov to 20-Nov (Marriott): Private **HPE Customer Briefings**

Thank you

