



# Lenovo Storage Solutions for IBM Spectrum Scale

Lenovo™

Michael Hennecke | Spectrum Scale Expertentage, 05-Mar-2020

# Agenda

- **Lenovo DSS-G : Integrated Solutions** based on Spectrum Scale RAID
  - DSS-G2xy ... based on „paired RGs“ with the Data Acces / Data Management Editions
  - DSS-G100 ... based on „scale-out RGs“ with the Erasure Code Edition
- **Spectrum Scale on Lenovo DE-Series Storage**
  - Design and Implementation Guidance for Lenovo Block Storage Controllers
- **Developments in the NVMe over Fabrics space**
  - Excelero NVMesh
  - NetApp EF600
  - Off-topic: DAOS

# Lenovo DSS-G

Integrated Solutions based on Spectrum Scale RAID (GNR)



# Lenovo Distributed Storage Solution for IBM Spectrum Scale

## The Hardware Components:



### SR650 Servers

### D3284 JBODs

### D1224 JBODs

## The Solution:

### DSS-G2xy



#### Lenovo ThinkSystem SR650 Server

Product Guide

Lenovo ThinkSystem SR650 is an ideal 2-socket 2U rack server for small businesses up to large enterprises that need industry-leading reliability, management, and security, as well as maximizing performance and flexibility for future growth. The SR650 server is designed to handle a wide range of workloads, such as databases, virtualization and cloud computing, virtual desktop infrastructure (VDI), enterprise applications, collaboration/email, and business analytics and big data.

Featuring the Intel Xeon Processor Scalable Family, the SR650 server offers scalable performance, storage capacity, and I/O expansion. The SR650 server supports up to two processors, up to 1.5 TB (support for up to 3 TB is planned for future) of 2880 MHz TurboDDR4 memory, up to 24 x 2.5-inch or 14 x 3.5-inch drive bays with an extensive choice of NVMe PCIe SSDs, SAS/SATA SSDs, and SAS/SATA HDDs, and flexible I/O expansion options with the LOM slot, the dedicated storage controller slot, and up to 6 PCIe slots.

The SR650 server offers basic or advanced hardware RAID protection and a wide range of networking options, including selectable LOM, ML2, and PCIe network adapters. The next-generation Lenovo XClarity Controller, which is built into the SR650 server, provides advanced service processor control, monitoring, and alerting functions.

The following figure shows the ThinkSystem SR650.




Figure 1. Lenovo ThinkSystem SR650

**Did you know?**

The SR650 server features a unique AnyBay design that allows a choice of drive interface types in the same drive bay: SAS drives, SATA drives, or U.2 NVMe PCIe drives.



The SR650 server offers onboard NVMe PCIe ports that allow direct connections to the U.2 NVMe PCIe SSDs, which frees up I/O slots and helps lower NVMe solution acquisition costs.

The SR650 server delivers impressive compute power per watt, featuring 80 PLUS Titanium and Platinum redundant power supplies that can deliver 96% (Titanium) or 94% (Platinum) efficiency at 50% load when connected to a 200-240 V AC power source.

The SR650 server is designed to meet ASHRAE A4 standards (up to 45 °C [113 °F]) in select configurations, which enable customers to lower energy costs, while still maintaining world-class reliability.

[Click here to check for updates](#)

Lenovo ThinkSystem SR650 Server 1



#### Lenovo Storage D3284 External High Density Drive Expansion Enclosure

Product Guide

The Lenovo Storage D3284 High Density Expansion Enclosure offers 12 Gbps SAS direct-attached storage expansion capabilities that are designed to provide density, speed, scalability, security, and high availability for medium to large businesses. The D3284 delivers enterprise-class storage technology in a cost-effective dense solution with flexible drive configurations of up to 84 drives in 5U rack space and RAID or JBOD (non-RAID) host connectivity.

The D3284 expansion unit is designed for a wide range of workloads, including big data and analytics, video surveillance, media streaming, private clouds, file and print serving, e-mail and collaboration, and databases. They also well-suited for software defined storage (SDS) and Windows Server solutions with Storage Spaces.

The D3284 expansion unit is also well-suited for software defined storage (SDS) and Windows Storage Spaces.




Figure 1. Lenovo Storage D3284 HD Expansion Enclosure

**Did you know?**



The D3284 expansion enclosures support 12 Gbps SAS connectivity, which doubles the data transfer rate compared to 6 Gb SAS solutions to maximize performance of storage I/O-intensive applications.

With support for daily chaining, the D3284 expansion enclosures can be scaled up to 3.36 PB for capacity-optimized configurations.

The D3284 expansion enclosures allow daily chaining with D1212 and D1224 expansion enclosures: Up to two D3284 and two D1212 or one D1224 drive enclosures is supported in a single chain.

[Click here to check for updates](#)

Lenovo Storage D3284 External High Density Drive Expansion Enclosure 1



#### Lenovo Storage D1212 and D1224 Drive Enclosures

Product Guide

The Lenovo Storage D1212 and D1224 Disk Expansion Enclosures offer 12 Gbps SAS direct-attached storage expansion capabilities that are designed to provide simplicity, speed, scalability, security, and high availability for small to large businesses. The D1212 and D1224 deliver enterprise-class storage technology in a cost-effective solution with flexible drive configurations and RAID or JBOD (non-RAID) host connectivity.

The D1212 and D1224 expansion units are designed for a wide range of workloads, including big data and analytics, video surveillance, media streaming, private clouds, file and print serving, e-mail and collaboration, and databases. They also well-suited for software defined storage (SDS) and Windows Server solutions with Storage Spaces.

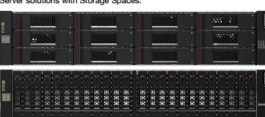


Figure 1. Lenovo Storage D1212 and D1224 Disk Expansion Enclosures

**Did you know?**

The D1212 and D1224 expansion enclosures offer flexible drive configurations with the choice of 2.5-inch and 3.5-inch drive form factors, 10K or 15K rpm SAS and 7.2K rpm NL SAS hard disk drives (HDDs) and self-encrypting drives (SEDs), and SAS solid-state drives (SSDs).

The D3284 expansion enclosures support 12 Gbps SAS connectivity, which doubles the data transfer rate compared to 6 Gb SAS solutions to maximize performance of storage I/O-intensive applications.



With support for daily chaining, the D1212 can be scaled up to 960 TB for capacity-optimized configurations with HDDs, and the D1224 can be scaled up to 192 drives for performance-optimized configurations.

The D1212 and D1224 expansion units support 12 Gbps SAS connectivity, which doubles the data transfer rate compared to 6 Gb SAS solutions to maximize performance of storage I/O-intensive applications.

[Click here to check for updates](#)

Lenovo Storage D1212 and D1224 Drive Enclosures 1





#### Lenovo Distributed Storage Solution for IBM Spectrum Scale (DSS-G) (ThinkSystem based)

Product Guide

Lenovo Distributed Storage Solution for IBM Spectrum Scale (DSS-G) is a software-defined storage (SDS) solution for dense scalable file and object storage suitable for high-performance and data-intensive environments. Enterprises or organizations running HPC, Big Data or cloud workloads will benefit the most from the DSS-G implementation.

DSS-G combines the performance of the Lenovo ThinkSystem SR650 servers, Lenovo D1224 and D3284 storage enclosures, and industry leading IBM Spectrum Scale software to offer a high performance, scalable building block approach to modern storage needs.

Lenovo DSS-G is delivered as a pre-integrated, easy-to-deploy rack-level engineered solution that dramatically reduces time-to-value and total cost of ownership (TCO). All DSS-G base offerings described in this product guide are built on Lenovo ThinkSystem SR650 servers, Lenovo Storage D1224 Drive Enclosures with high-performance 2.5-inch SAS solid-state drives, and Lenovo Storage D3284 High-Density Drive Enclosures with large capacity 3.5-inch NL SAS HDDs.

Combined with IBM Spectrum Scale formerly IBM General Parallel File System, GPFS, an industry leader in high-performance clustered file system, you have an ideal solution for the ultimate file and object storage solution for HPC and Big Data.

**Did you know?**

The DSS-G solution gives you the choice of shipping fully integrated into the Lenovo 1410 rack cabinet, or with the Lenovo Client Site Integration Kit, 7274, which allows you to have Lenovo install the solution in a rack of your own choosing. In either case, the solution is tested, configured, and ready to be plugged in and turned on; it is designed to integrate into an existing infrastructure effortlessly, to dramatically accelerate time to value and reduce infrastructure maintenance costs.

Lenovo DSS-G is licensed by the number of drives installed, rather than the number of processor cores or the number of connected clients, so there are no added licenses for other servers or clients that mount and work with the file system.

Lenovo provides a single point of entry for supporting the entire DSS-G solution, including the IBM Spectrum Scale software, for quicker problem determination and minimized downtime.

[Click here to check for updates](#)

Lenovo Distributed Storage Solution for IBM Spectrum Scale (DSS-G) (ThinkSystem based) 1




Figure 1. Lenovo DSS-G Model G280

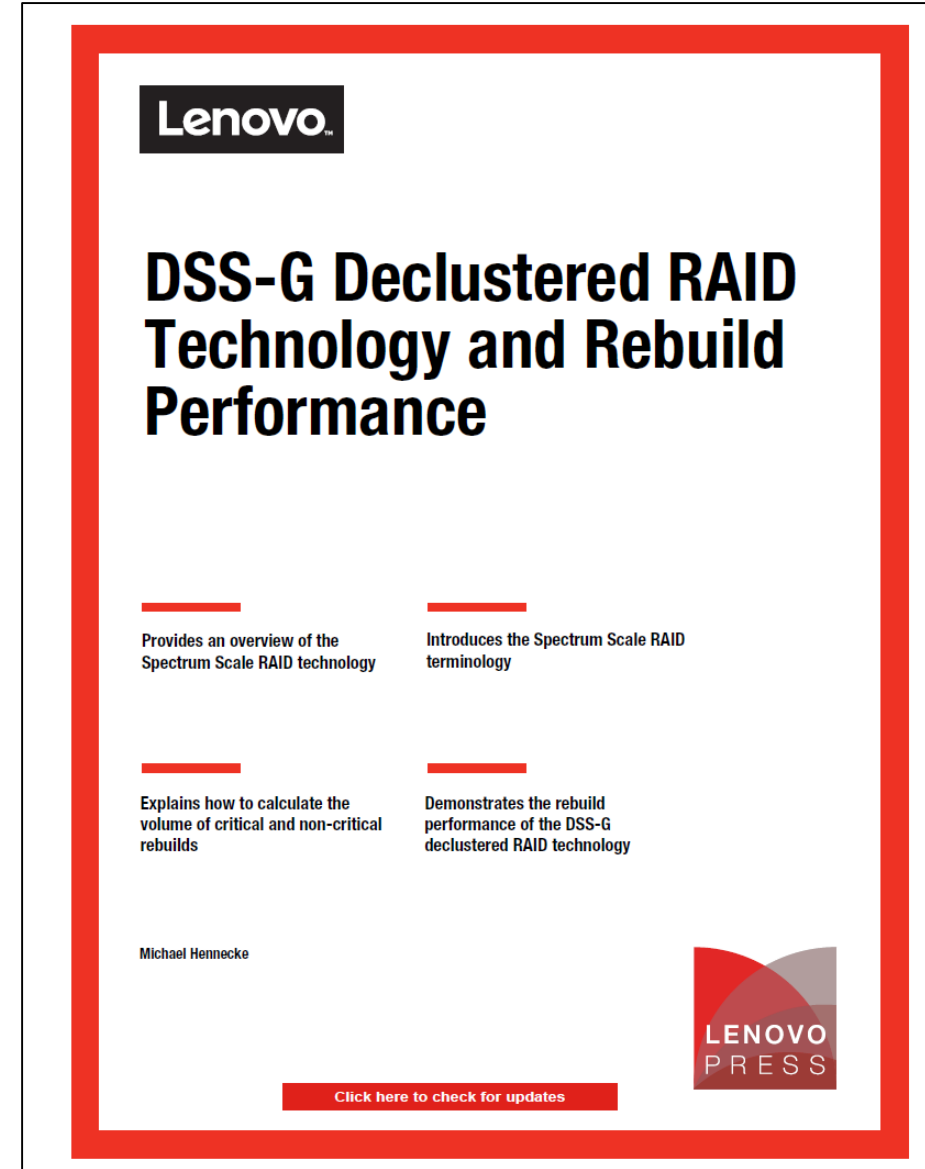
<https://lenovopress.com/lp0644-lenovo-thinksystem-sr650-server>

<https://lenovopress.com/lp0513-lenovo-storage-d3284-external-high-density-drive-expansion-enclosure>

<https://lenovopress.com/lp0512-lenovo-storage-d1212-d1224-drive-enclosures>

<https://lenovopress.com/lp0837-lenovo-dss-g-thinksystem>

# New Lenovo DSS-G Collateral



# Lenovo DSS-G100 NVMe Servers



- Intel „Performance“: Lenovo ThinkSystem SR630 (1U2S)

- 2x Intel „CascadeLake“ CPUs; 192GB
- **8x U.2 NVMe** drives (**4-lane** gen3)
- **2x 100Gbps** networking (CX-5 IB/Eth, OPA)



- Intel „Capacity“: Lenovo ThinkSystem SR650 (2U2S)

- 2x Intel „CascadeLake“ CPUs; 192GB
- **24x U.2 NVMe** drives (**2-lane** gen3)
- **2x 100Gbps** networking (CX-5 IB/Eth, OPA)

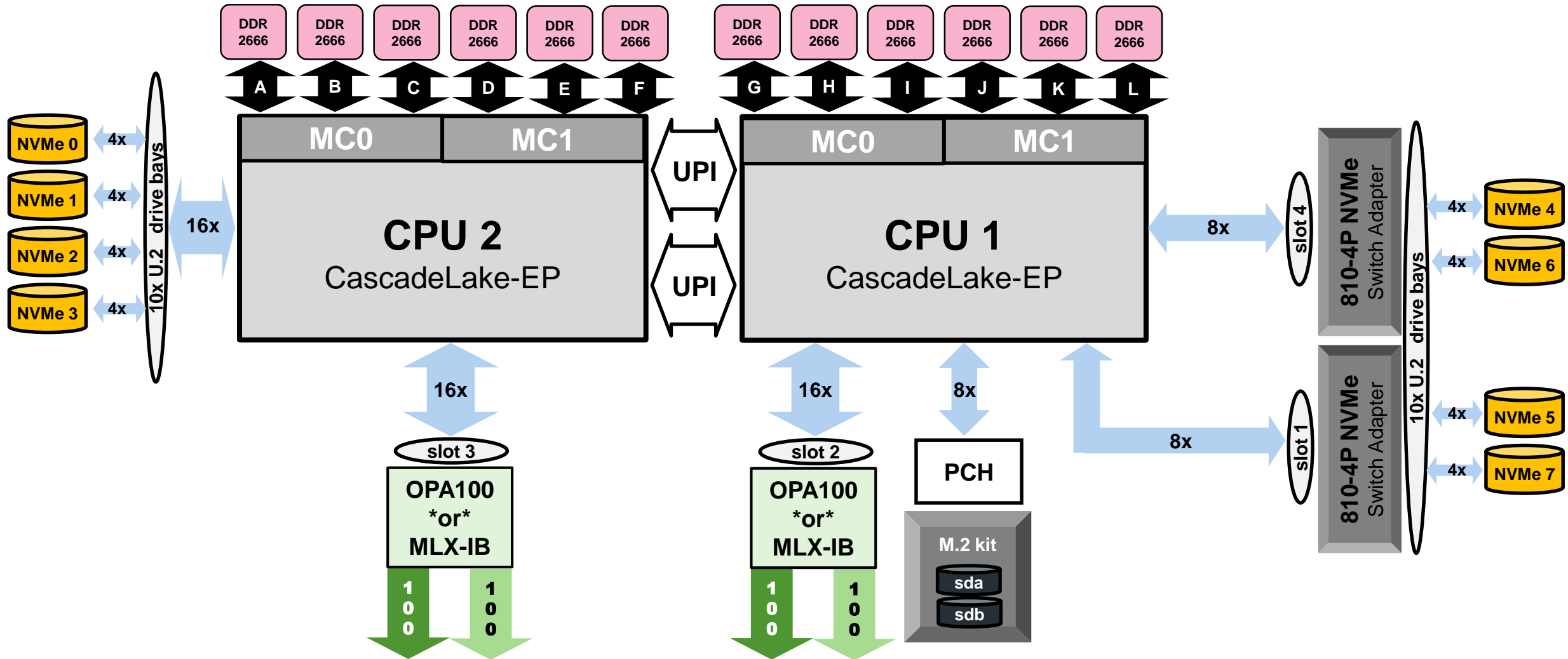


- AMD „Performance and Capacity“: Lenovo ThinkSystem SR655 (2U1S)

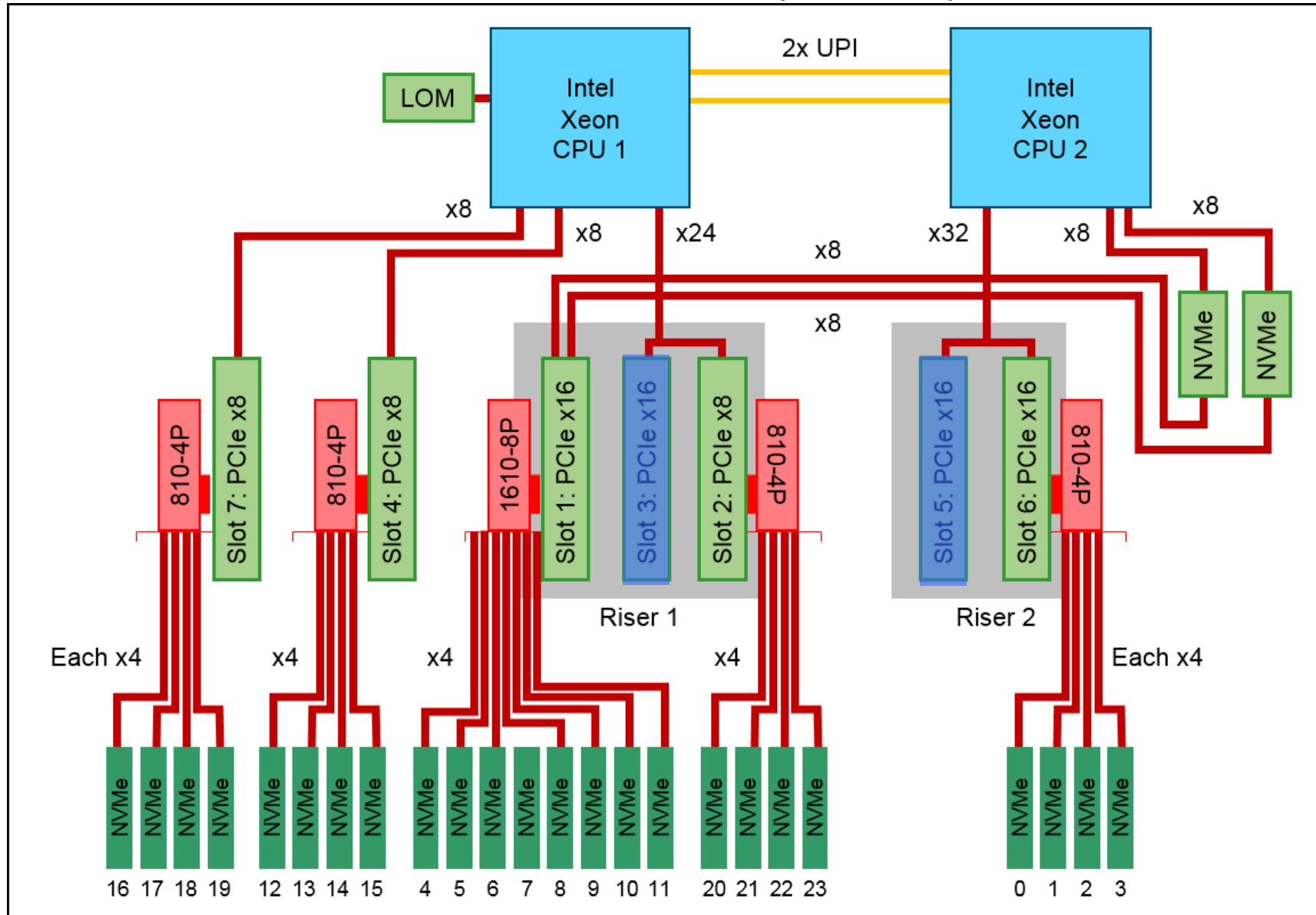
- 1x AMD EPYC „Rome“ CPU; 256GB
- **≥16x U.2 NVMe** drives (**4-lane** gen3/gen4)
- **2x 200Gbps** networking (**gen4** CX-6 IB/Eth)



# SR630 „CXL“ with 8x NVMe (4-lane) and 2x 100Gbps

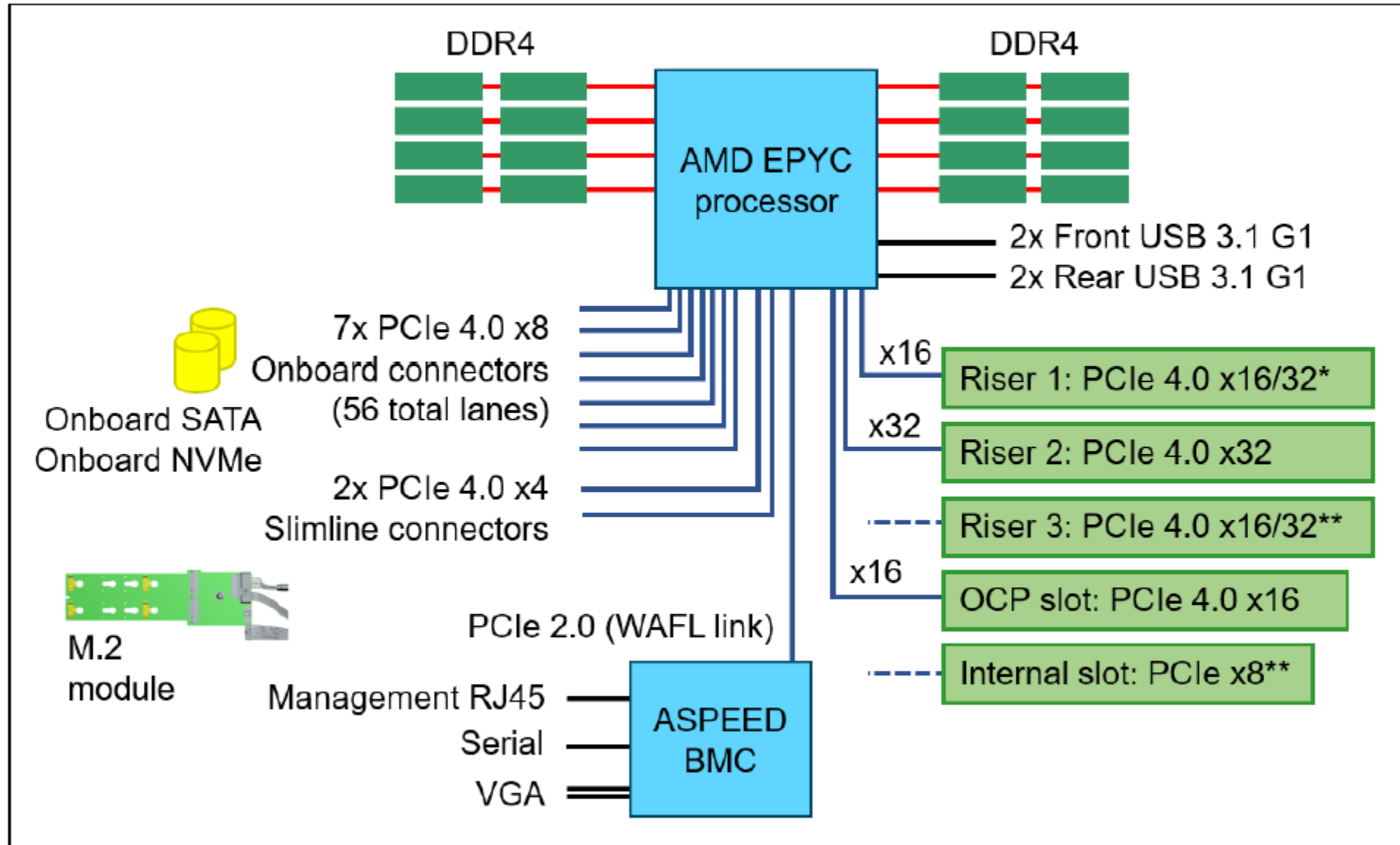


# SR650 „CXL“ with 24x NVMe (2-lane) and 2x 100Gbps





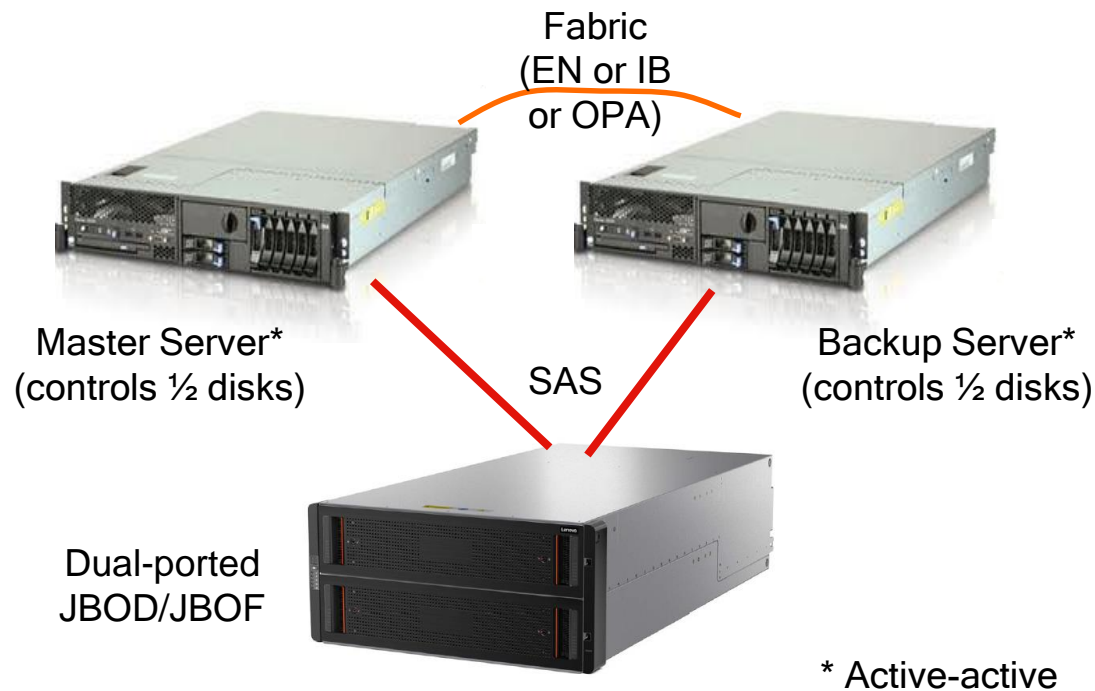
# SR655 „Rome“ with $\geq 16$ x NVMe (4-lane) and 2x 200Gbps



# Comparison of DSS-G2xx and ECE on DSS-G100

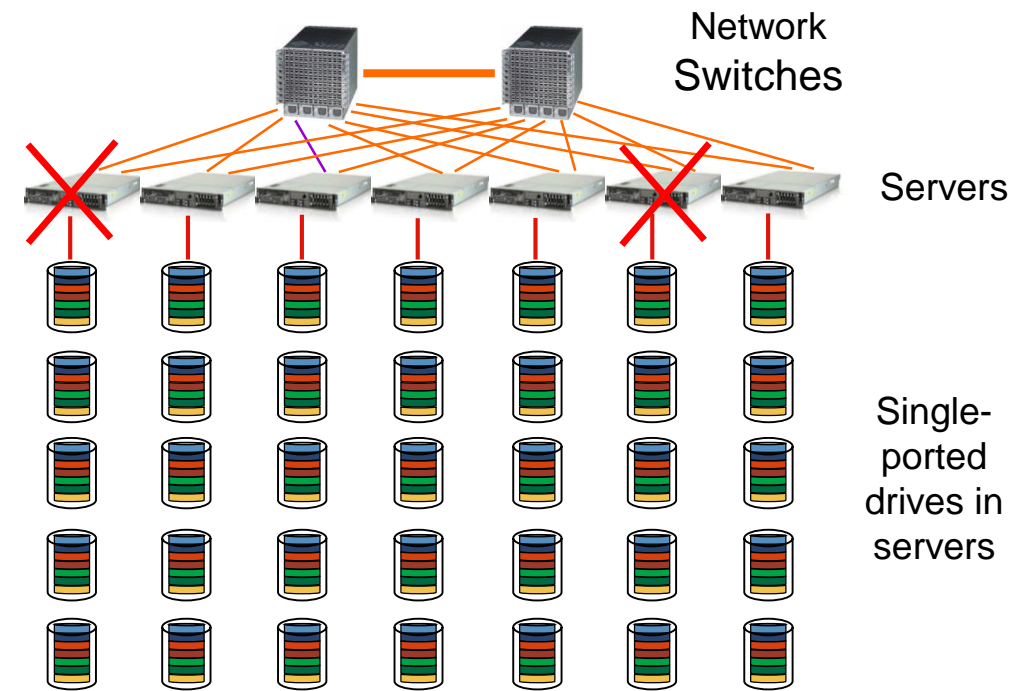
- Twin-tailed disks, dual servers – provide very high availability
- Master controls 50% of the disks and backup controls other 50%, when one server fails, the other takes over all disks

## DSS-G2xx Dual Server Building Block:



- Single-tailed drives, scalable number of servers
- Tolerates concurrent failure of two arbitrary servers (or 3 servers if 8+3p erasure code), or 2 (3) disks
- Multiple servers take over the work of failed servers

## ECE (on DSS-G100) Scale-out Configuration:



# Lenovo DE-Series

Spectrum Scale on SAN Storage Controllers



# Lenovo DE-Series Product Guides at Lenovo Press



## Lenovo ThinkSystem DE4000H Hybrid Storage Array Product Guide

Lenovo ThinkSystem DE4000H is a scalable, hybrid entry-level storage system that is designed to provide performance, simplicity, capacity, security, and high availability for medium to large businesses. It delivers enterprise-class storage management capabilities with a wide choice of host connectivity options, flexible drive configurations, and enhanced data management features. The ThinkSystem DE4000H is a perfect fit for a wide range of enterprise workloads, including big data and analytics, video surveillance, technical computing, backup and recovery, and other storage I/O-intensive applications.

ThinkSystem DE4000H models are available in a 2U rack form-factor with 24 small form-factor (2.5-inch SFF) drives (2U24 SFF), 12 large form-factor (3.5-inch LFF) drives (2U12 LFF), or a 4U rack form-factor with 60 LFF drives (4U60 LFF) and include two controllers, each with 8 GB or 32 GB cache for a system total of 16 GB or 64 GB. Universal 1/10 Gb iSCSI or 4/8/16 Gb Fibre Channel (FC) ports provide base host connectivity, and the host interface cards provide additional 1/10 Gb iSCSI or 4/8/16 Gb FC, 12 Gb SAS, 10/25 Gb iSCSI, or 8/16/32 Gb FC connections.

The ThinkSystem DE4000H Storage Array scales up to 192 drives with the attachment of Lenovo ThinkSystem DE120S 2U12, DE240S 2U24 SFF, and DE600S 4U60 LFF Expansion Enclosures. It also offers flexible drive configurations with the choice of 2.5-inch (SFF) and 3.5-inch (LFF) form factors, 10 K rpm SAS and 7.2 K rpm NL SAS hard disk drives (HDDs), and SAS solid-state drives (SSDs).



Figure 1. Lenovo ThinkSystem DE4000H 2U24 SFF (top), 2U12 LFF (middle), and 4U60 LFF (bottom)

### Did you know?

The ThinkSystem DE4000H scales up to 2.3 PB of raw storage capacity, and it offers block storage connectivity with support for 1/10 Gb iSCSI or 4/8/16 Gb FC, and 12 Gb SAS, 10/25 Gb iSCSI, or 8/16/32 Gb FC at the same time.

For the ThinkSystem DE4000H, customers can change the host port protocol from FC to iSCSI or from iSCSI to FC for the SFP+ host ports built into the controller (base host ports), or the universal SFP+ host ports on the host interface card (HIC ports), or for all SFP+ base and universal HIC ports.

[Click here to check for updates](#)

Lenovo ThinkSystem DE4000H Hybrid Storage Array

1



## Lenovo ThinkSystem DE6000H Hybrid Storage Array Product Guide

Lenovo ThinkSystem DE6000H is a scalable, hybrid mid-range storage system that is designed to provide high performance, simplicity, capacity, security, and high availability for medium to large businesses. The ThinkSystem DE6000H delivers enterprise-class storage management capabilities in a performance-optimized system with a wide choice of host connectivity options, flexible drive configurations, and enhanced data management features. The ThinkSystem DE6000H is a perfect fit for a wide range of enterprise workloads, including big data and analytics, video surveillance, technical computing, backup and recovery, and other storage I/O-intensive applications.

ThinkSystem DE6000H models are available in a 2U rack form-factor with 24 small form-factor (2.5-inch SFF) drives (2U24 SFF) or a 4U rack form-factor with 60 LFF drives (4U60 LFF) and include two controllers, each with 16 GB or 64 GB cache for a system total of 32 GB or 128 GB. Universal 10 Gb iSCSI or 4/8/16 Gb Fibre Channel (FC) ports provide base host connectivity, and the host interface cards provide additional 12 Gb SAS, 10/25 Gb iSCSI, or 8/16/32 Gb FC connections.

The ThinkSystem DE6000H Storage Array scales up to 240 (base configuration) or 480 (optional upgrade) drives with the attachment of Lenovo ThinkSystem DE240S 2U24 SFF and DE600S 4U60 LFF Expansion Enclosures. It also offers flexible drive configurations with the choice of 2.5-inch (SFF) and 3.5-inch (LFF) form factors, 10 K rpm SAS and 7.2 K rpm NL SAS hard disk drives (HDDs), and SAS solid-state drives (SSDs).

The Lenovo ThinkSystem DE6000H 2U24 SFF and 4U60 LFF enclosures are shown in the following figure.



Figure 1. Lenovo ThinkSystem DE6000H 2U24 SFF (top) and 4U60 LFF (bottom) enclosures

### Did you know?

The ThinkSystem DE6000H scales up to 2.88 PB of raw storage capacity in the base configuration or up to 5.76 PB with the optional Features on Demand upgrade.

The ThinkSystem DE6000H offers block storage connectivity with support for 10 Gb iSCSI or 4/8/16 Gb FC, and 12 Gb SAS, 10/25 Gb iSCSI, or 8/16/32 Gb FC at the same time.

For the ThinkSystem DE6000H, customers can change the host port protocol from FC to iSCSI or from iSCSI to FC for the SFP+ host ports built into the controller (base host ports).

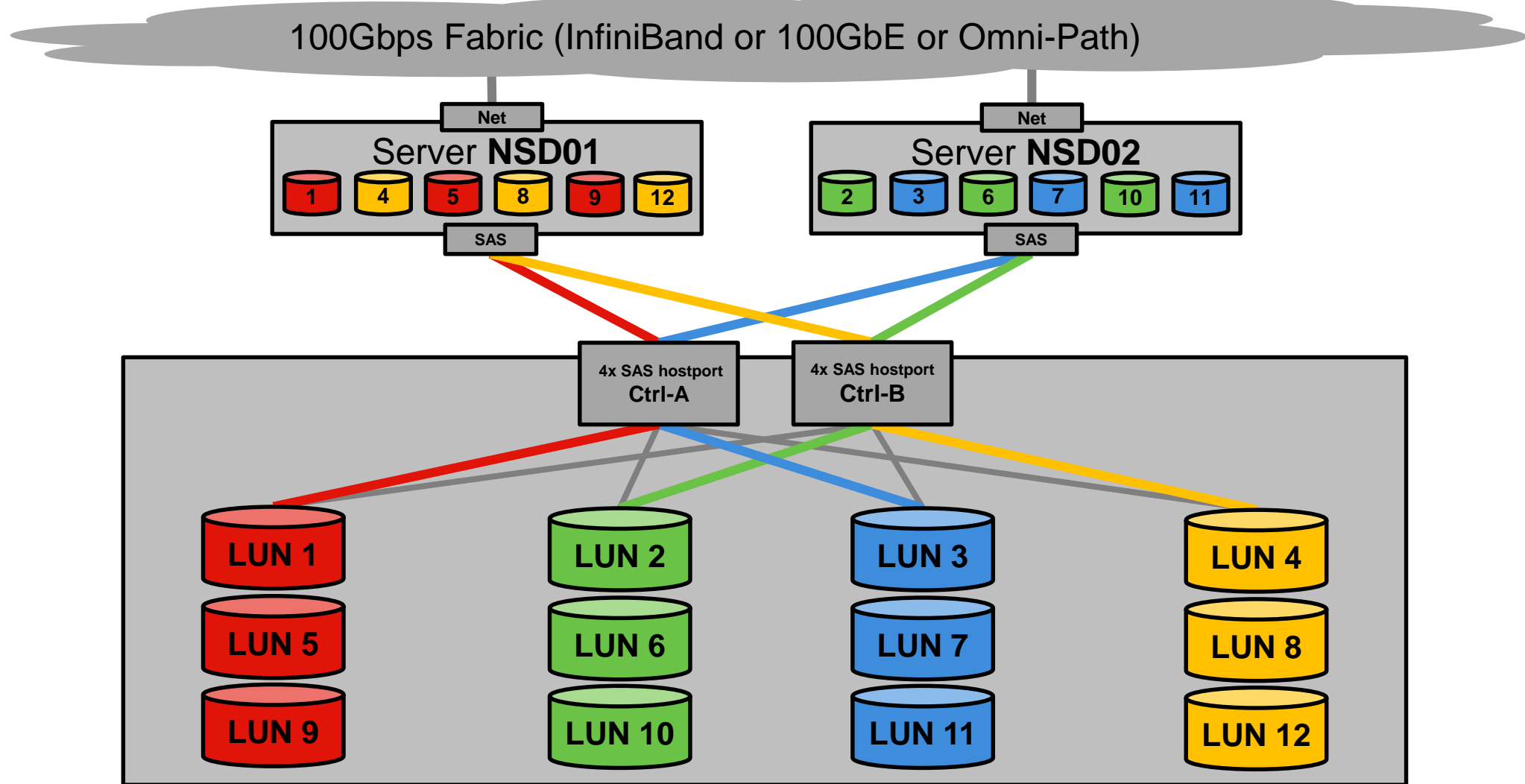
[Click here to check for updates](#)

Lenovo ThinkSystem DE6000H Hybrid Storage Array

1



# Building Block – Controller and NSD Server Ownership



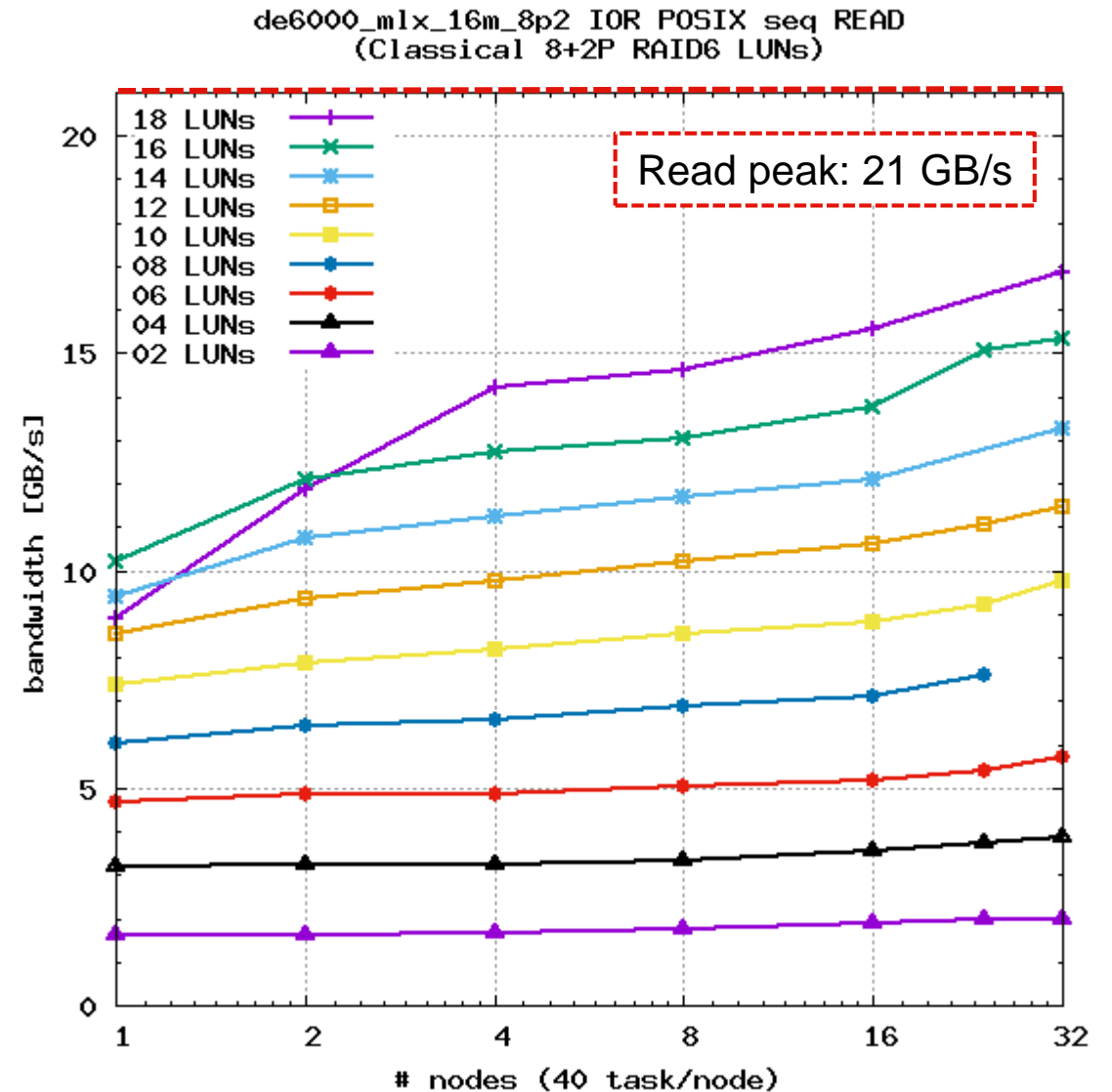
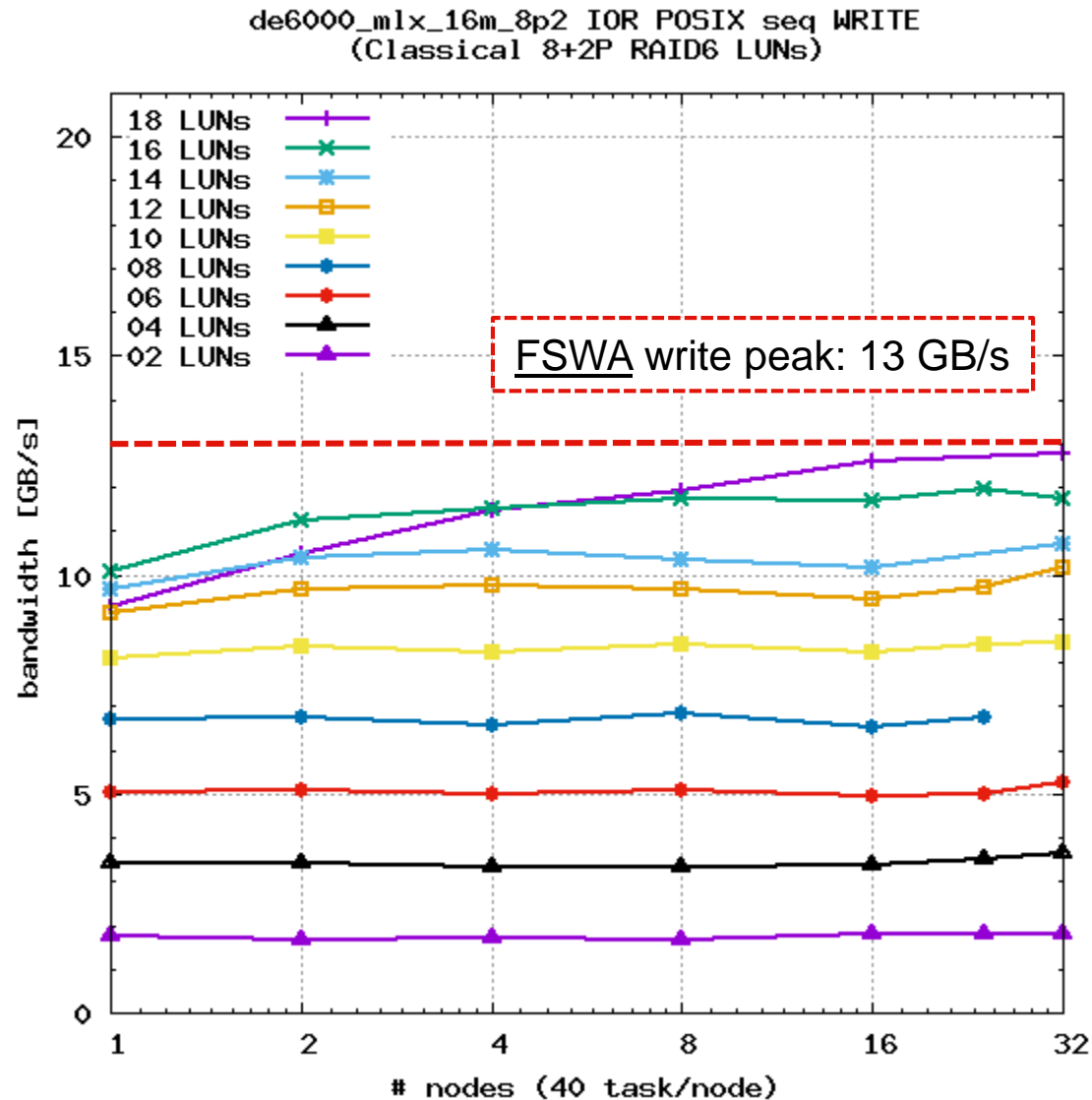
Preferred owner: **Ctrl-A ; NSD01**

**Ctrl-B ; NSD02**

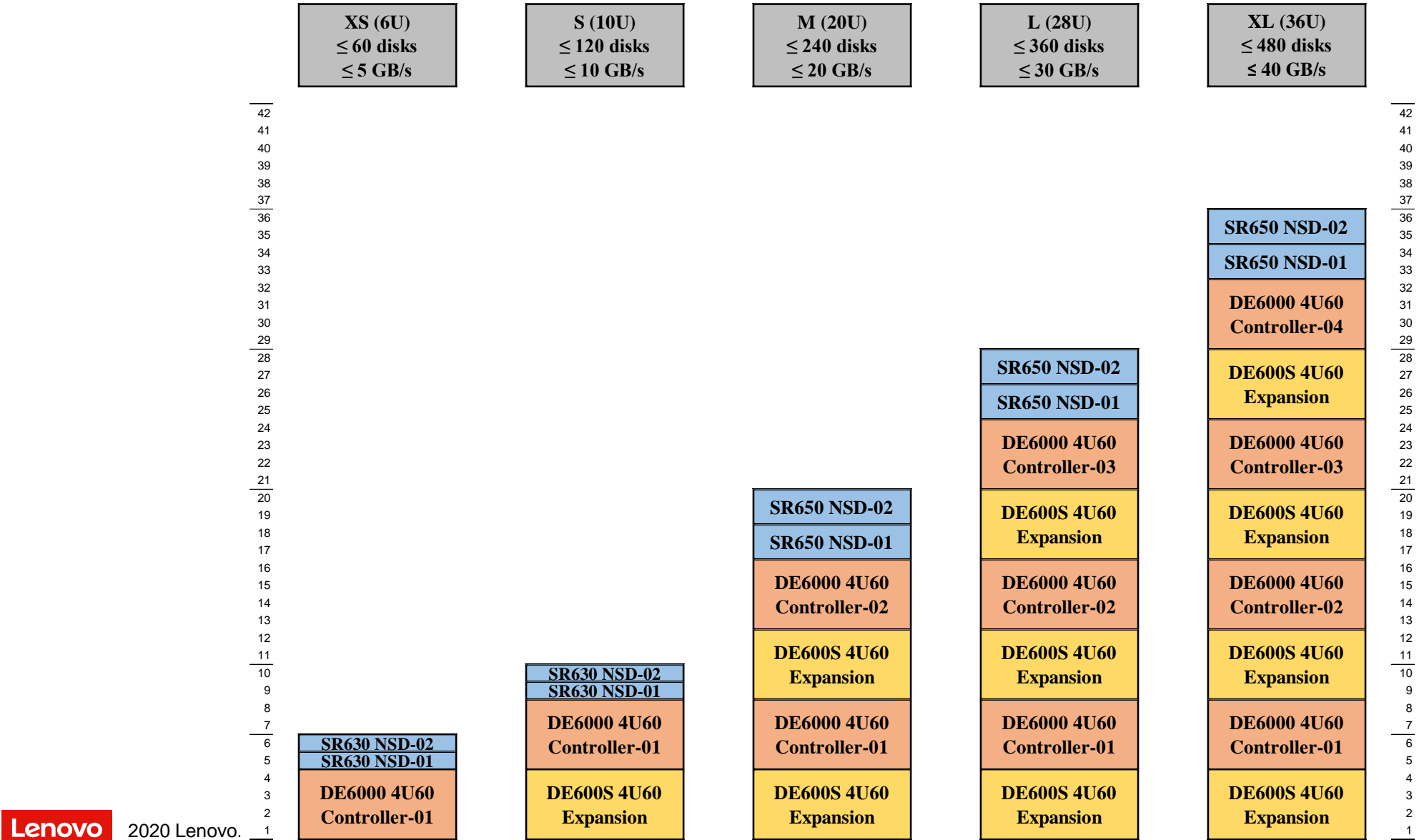
**Ctrl-A ; NSD02**

**Ctrl-B ; NSD01**

# DE6000H Spectrum Scale Bandwidth „Scale-Up“



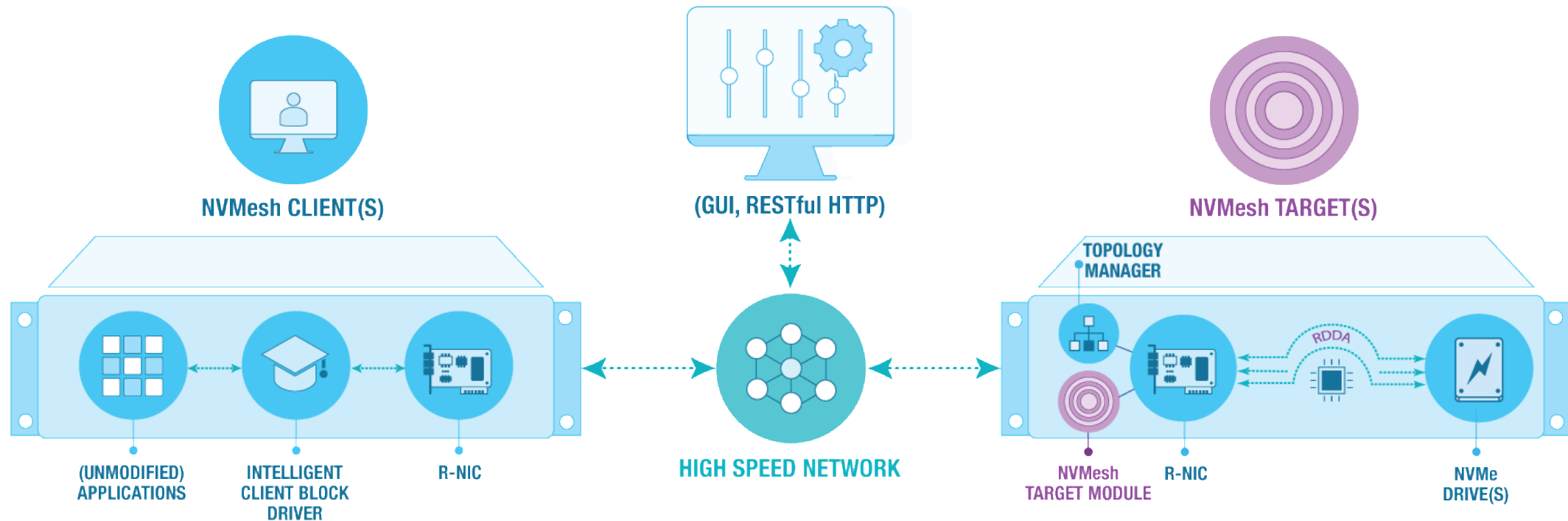
# Spectrum Scale Building Block Scale-Up (DE6000H)



# NVMe over Fabrics



# Excelero NVMesh Software and Hardware Components



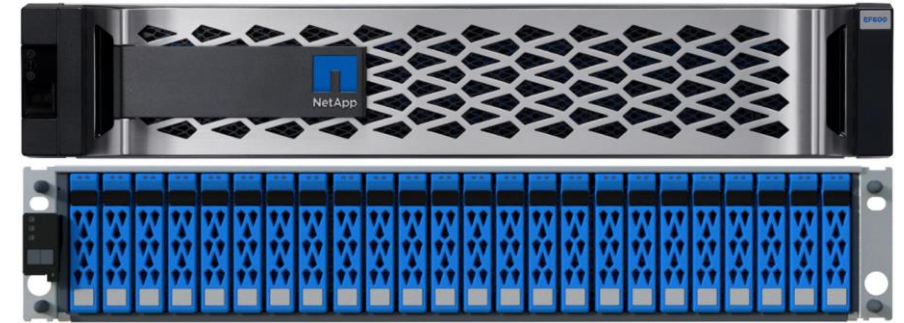
- All server hardware must support PCIe 3.0
- Supported Linux OS distributions include RHEL/CentOS, Ubuntu and SLES
- NVMe devices are supported in PCIe adapter as well as U.2 and M.2 drive form-factors

# Spectrum Scale on NetApp EF600 (2U24 NVMe)

- 24x Dual-Ported NVMe in 2U (1.92 to 15TB)
- 8x IB Host ports (4x A and 4x B controller)
- SANtricity 11.60 software on EF600
- Host Software requirements
  - Latest RHEL 7.7 patches (or latest SLES12)
  - MOFED built with NVMe-o-F support (`./mlnxofedinstall --add-kernel-support --with-nvmf`)
- Discover and connect to the EF600 host ports from the cluster nodes

```
# nvme discover -t rdma -a $ef600_a1_ip | grep subnqn # repeat for a2,a3,a4,b1,b2,b3,b4 (all 8 ports)
subnqn: nqn.1992-08.com.netapp:6000.6d039ea0003ef5100000000059729fff
# nvme connect -t rdma -n $subnqn_a1 -a $ef600_a1_ip -Q 1024 -l 3600 # repeat for b1 (connected ports)
```
- Add the Node NQNs to the EF600 host topology, and map the volumes (GUI or CLI)

```
# cat /etc/nvme/hostnqn
nqn.2014-08.org.nvmexpress:uuid:630639ed-2a0d-44a1-820c-1ce026f74710
```
- Set up DM-Multipathing on the Nodes to manage the redundant IB paths
- Use the `/dev/dm-0` etc. as NSDs **without defining NSD servers („SAN Mode“)**



# Listing the EF600 Volumes / Paths with nvme

- One NVMe **device #** per visible EF600 **host port** (A1,A2, ...B4)
- One NVMe **namespace ID** per mapped EF600 **volume (LUN)**

```
[root@c1i0801 ~]# nvme netapp smdevices
```

```
/dev/nvme4n1, Array Name de0704-ef600, volume Name vd0, NSID 1,  
Volume ID 000009 7859b331b2d039ea00003ef510, Controller A, Access State unknown, 19.15TB # A1  
/dev/nvme4n2, Array Name de0704-ef600, volume Name vd1, NSID 2,  
Volume ID 000009 4e59b33c26d039ea00003ef1fd, Controller A, Access State unknown, 19.15TB # A3  
/dev/nvme5n1, Array Name de0704-ef600, volume Name vd0, NSID 1,  
Volume ID 000009 7859b331b2d039ea00003ef510, Controller A, Access State unknown, 19.15TB # A1  
/dev/nvme5n2, Array Name de0704-ef600, volume Name vd1, NSID 2,  
Volume ID 000009 4e59b33c26d039ea00003ef1fd, Controller A, Access State unknown, 19.15TB # A3  
/dev/nvme6n1, Array Name de0704-ef600, volume Name vd0, NSID 1,  
Volume ID 000009 7859b331b2d039ea00003ef510, Controller B, Access State unknown, 19.15TB # B1  
/dev/nvme6n2, Array Name de0704-ef600, volume Name vd1, NSID 2,  
Volume ID 000009 4e59b33c26d039ea00003ef1fd, Controller B, Access State unknown, 19.15TB # B3  
/dev/nvme7n1, Array Name de0704-ef600, volume Name vd0, NSID 1,  
Volume ID 000009 7859b331b2d039ea00003ef510, Controller B, Access State unknown, 19.15TB # B1  
/dev/nvme7n2, Array Name de0704-ef600, volume Name vd1, NSID 2,  
Volume ID 000009 4e59b33c26d039ea00003ef1fd, Controller B, Access State unknown, 19.15TB # B3
```

# DM-Multipathing for the EF600

```
# yum install -y device-mapper-multipath
```

```
# cat /etc/multipath.conf
```

```
# NetApp EF600 NVMe-o-F devices:
```

```
devices {
  device {
    vendor "NVME"
    product "NetApp E-Series*"
    path_grouping_policy group_by_prio
    failback immediate
    no_path_retry 30
  }
}

# exclude locally attached NVMe drives:
blacklist {
  wwid nvme.8086-*
}
```

```
# multipath -ll
eui.0000097859b331b2d039ea00003ef510 dm-1 NVME,NetApp E-Series
size=17T features='1 queue_if_no_path' hwhandler='0' wp=rw
|+- policy='service-time 0' prio=50 status=active
| |- 4:0:1:0 nvme4n1 259:4 active ready running
| `-- 5:0:1:0 nvme5n1 259:6 active ready running
`+- policy='service-time 0' prio=10 status=enabled
  |- 6:0:1:0 nvme6n1 259:8 active ready running
  `-- 7:0:1:0 nvme7n1 259:10 active ready running
eui.0000094e59b33c26d039ea00003ef1fd dm-2 NVME,NetApp E-Series
size=17T features='1 queue_if_no_path' hwhandler='0' wp=rw
|+- policy='service-time 0' prio=50 status=active
| |- 6:0:2:0 nvme6n2 259:9 active ready running
| `-- 7:0:2:0 nvme7n2 259:11 active ready running
`+- policy='service-time 0' prio=10 status=enabled
  |- 4:0:2:0 nvme4n2 259:5 active ready running
  `-- 5:0:2:0 nvme5n2 259:7 active ready running
```

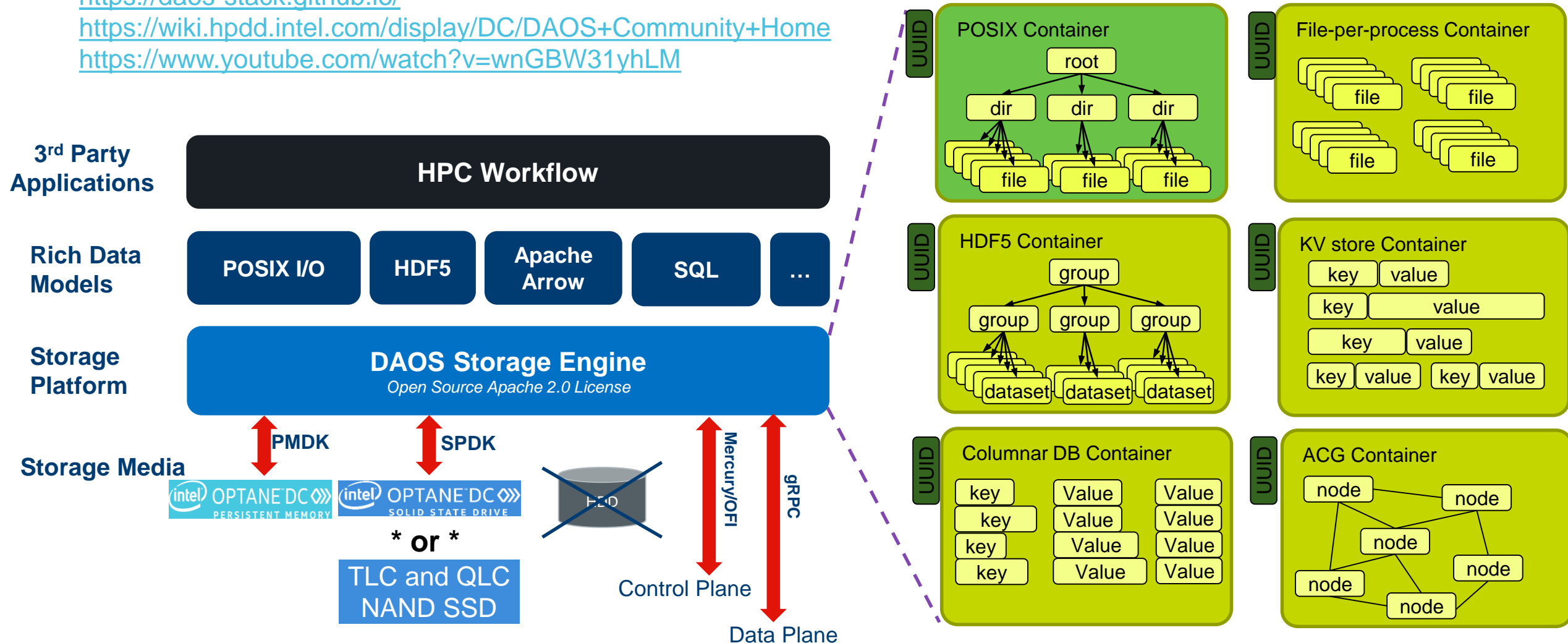


# Intel Distributed Asynchronous Object Storage

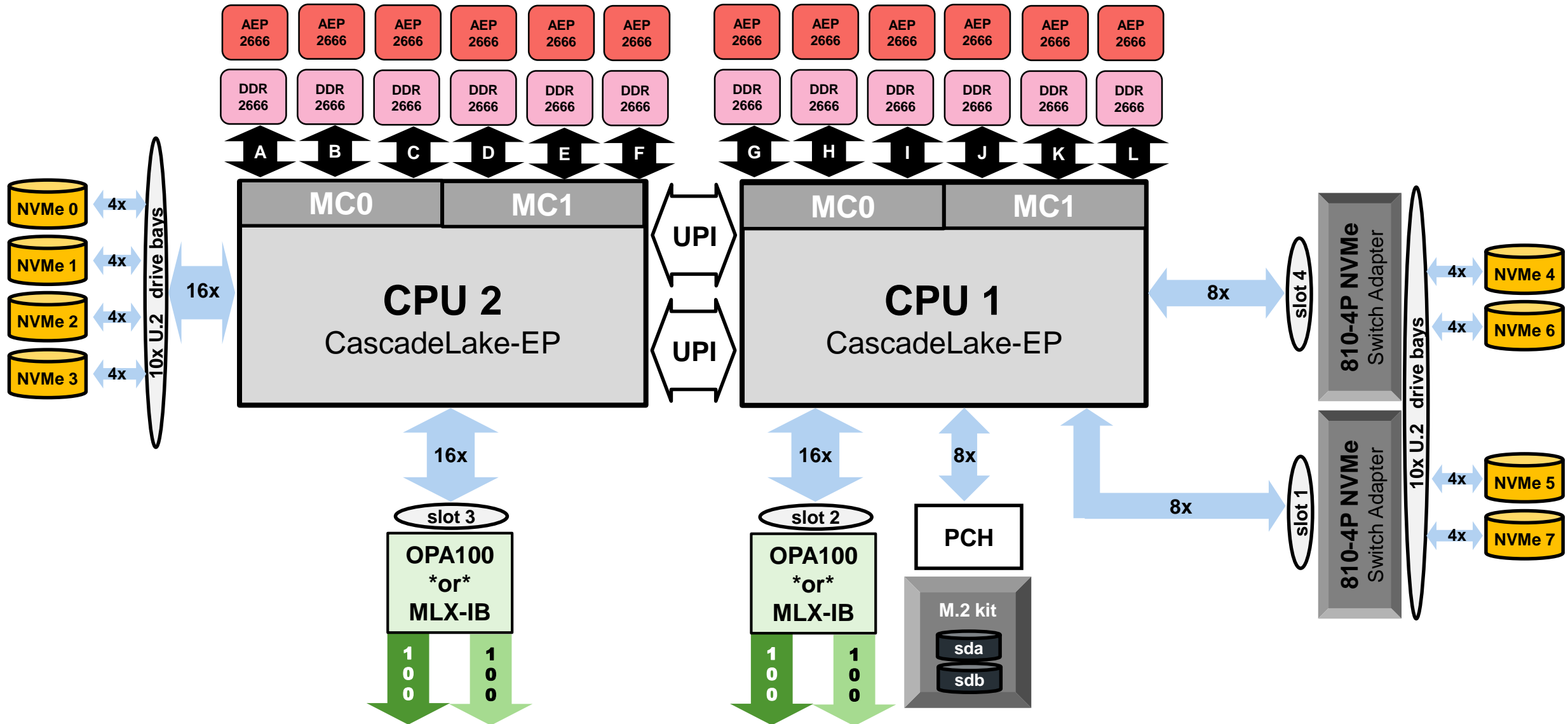
<https://daos-stack.github.io/>

<https://wiki.hpdd.intel.com/display/DC/DAOS+Community+Home>

<https://www.youtube.com/watch?v=wnGBW31yhLM>



# DAOS Server: SR630 „CXL“ with 8x NVMe and 2x 100Gbps



# thanks.

**mhennecke @ lenovo.com**

