



April 26-28, 2022

DoubleTree by Hilton San Jose

SmartNICsSummit.com

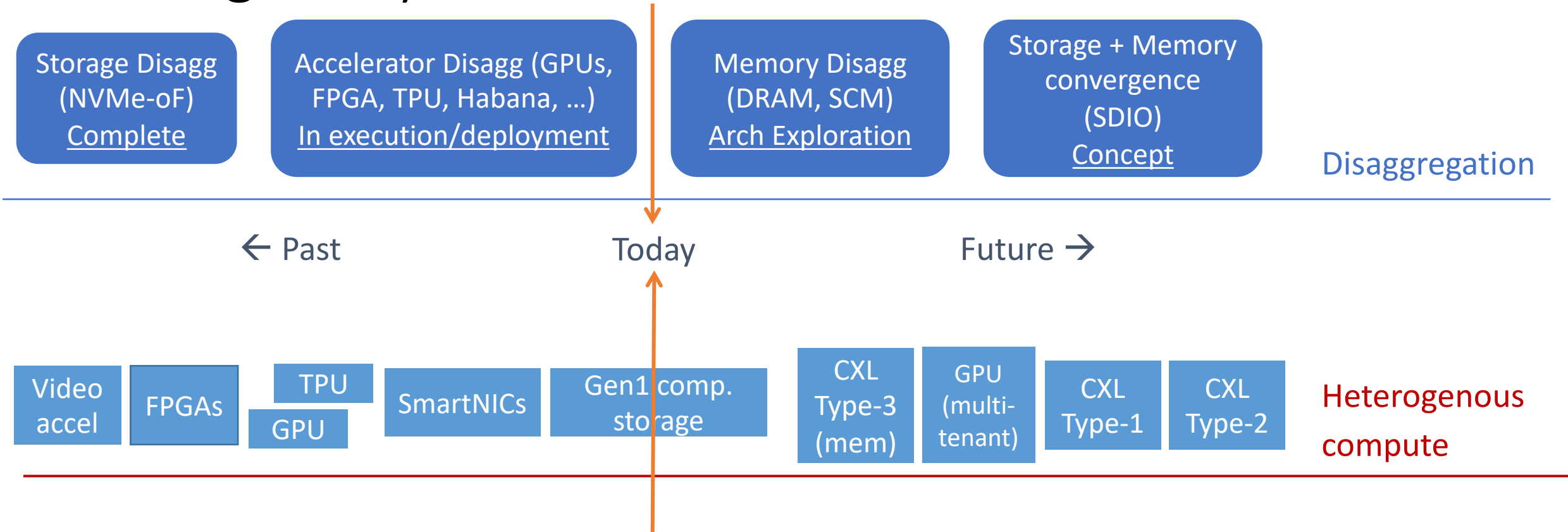
# Advantages and Use Cases for Adding the CXL interface to DPUs

Pavel (Pasha) Shamis, Sr. Principal Engineer

Kshitij Sudan, Director Storage & Accelerator Segment Marketing

Arm Inc.

# Enabling technologies for disaggregation and heterogeneity



# Drivers for CXL interface on DPUs

---

- Domain specific accelerators gaining prominence in datacenter
- Line-rate processing of network traffic is critical for certain use-cases
  - Like compression, de-dupe, encryption, streaming data-processing
- Near-storage/near-memory processing technologies advancing
  - Computational storage, in-memory processing

All these use-cases need larger DPU memory at low-cost  
CXL interface addresses this need

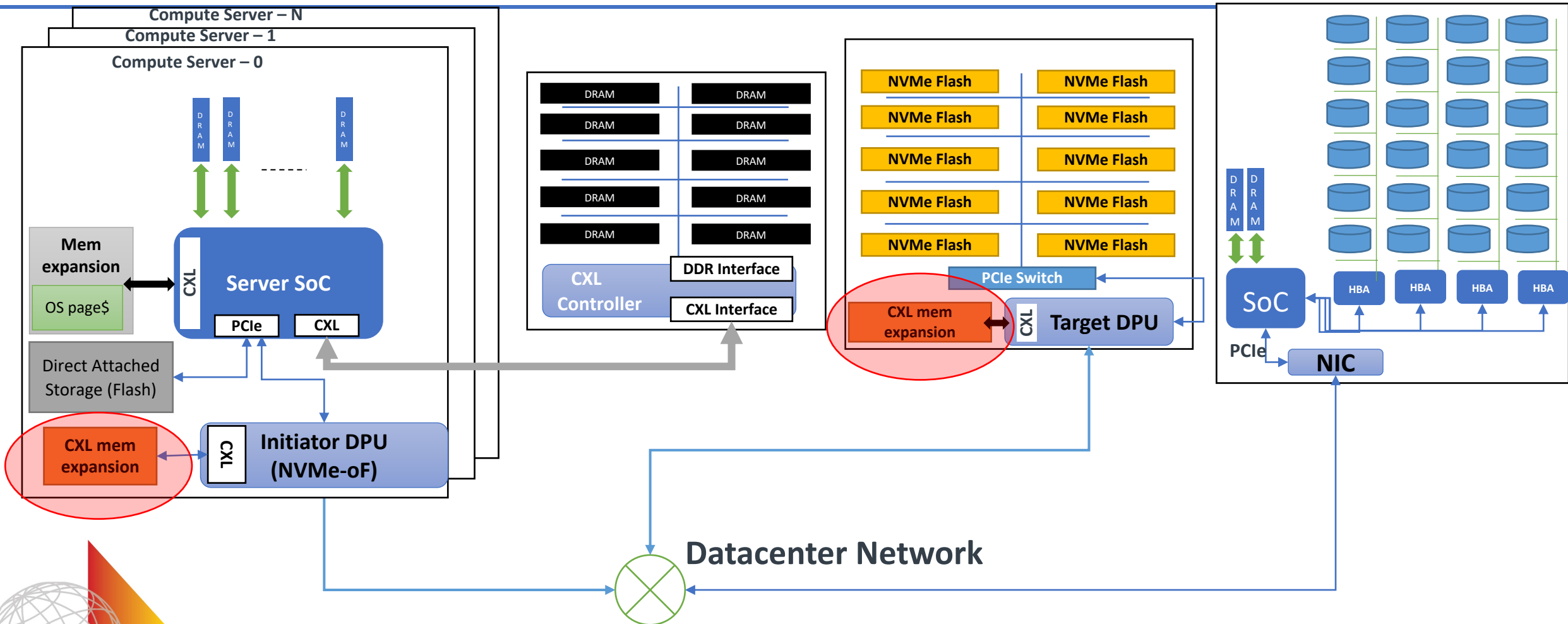
# Emerging datacenter storage & memory architecture

## General Purpose Compute

## Disaggregated DRAM Pool

## Warm caching tier

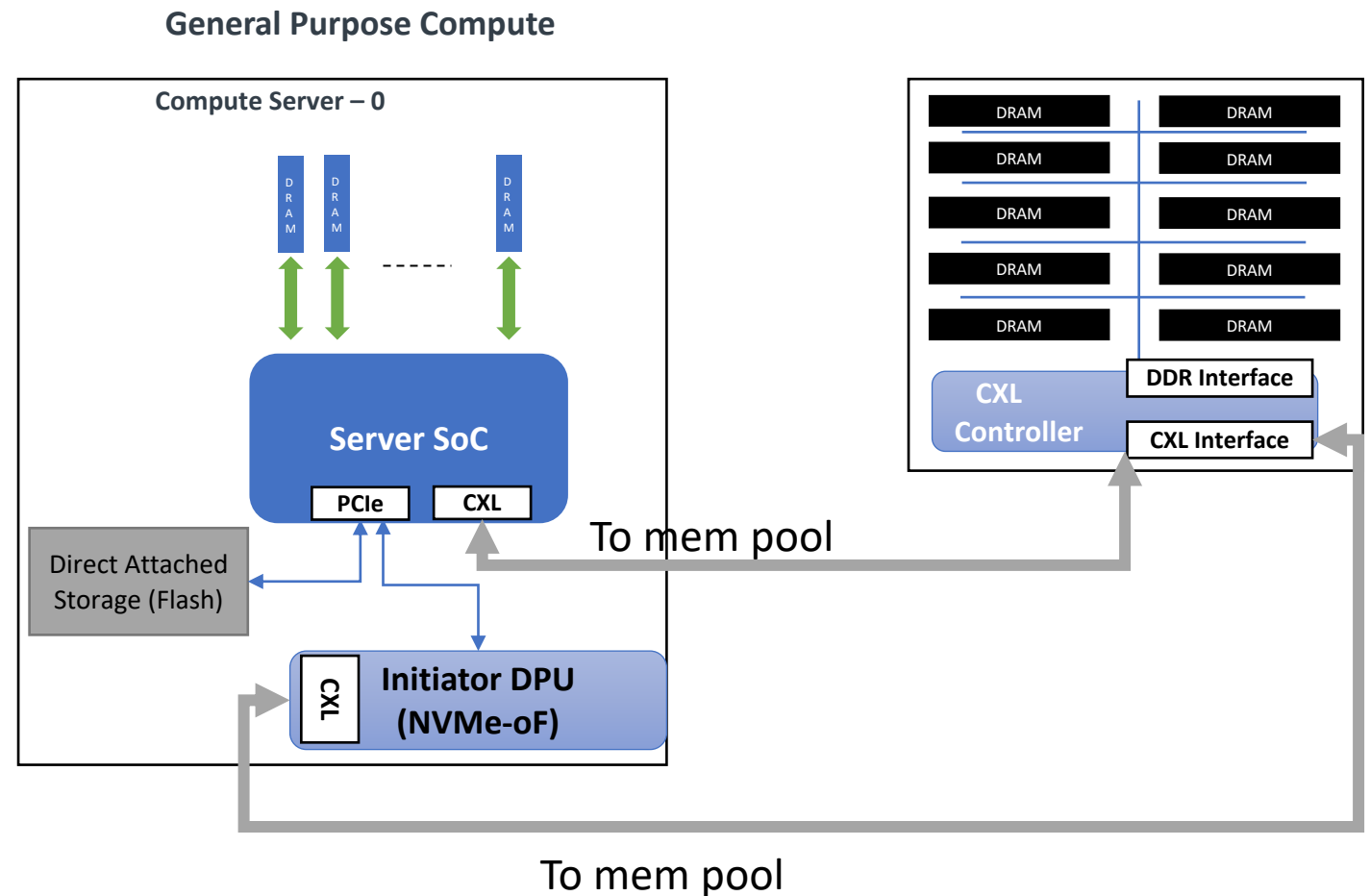
## Storage/Cold tier





# Reducing stranded-ness with memory pools

- Expand DPU's ability to leverage pooled memory
  - Resolves stranded-ness challenges
- Future CXL specifications, discussing potential to share memory at the memory pool
  - Security challenges



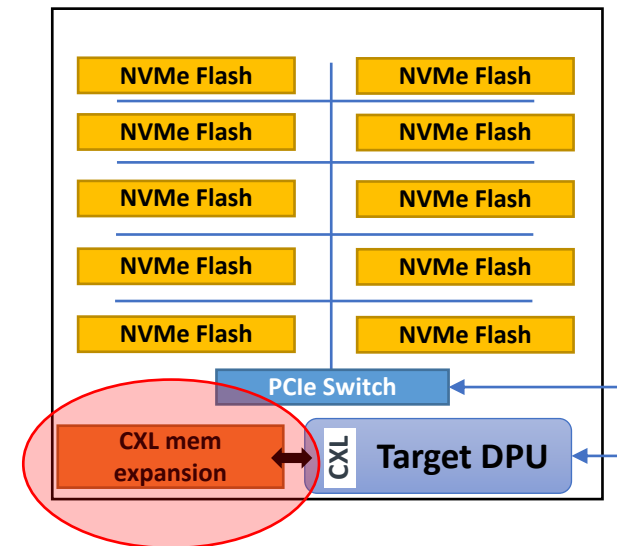
# Appliance use-case for DPUs with CXL

- Advantages

- Builds on DPU-as-target use-case
- Uses large CXL-connected memory capacity for “conditioning tier”
- Lower pin-count for DPU memory
- Lower DPU memory cost/bit

- Concerns

- Increased BOM cost with CXL interface and CXL memory expansion card



# DPU + NVDIMM Storage Appliance Study

- **Bluefield**

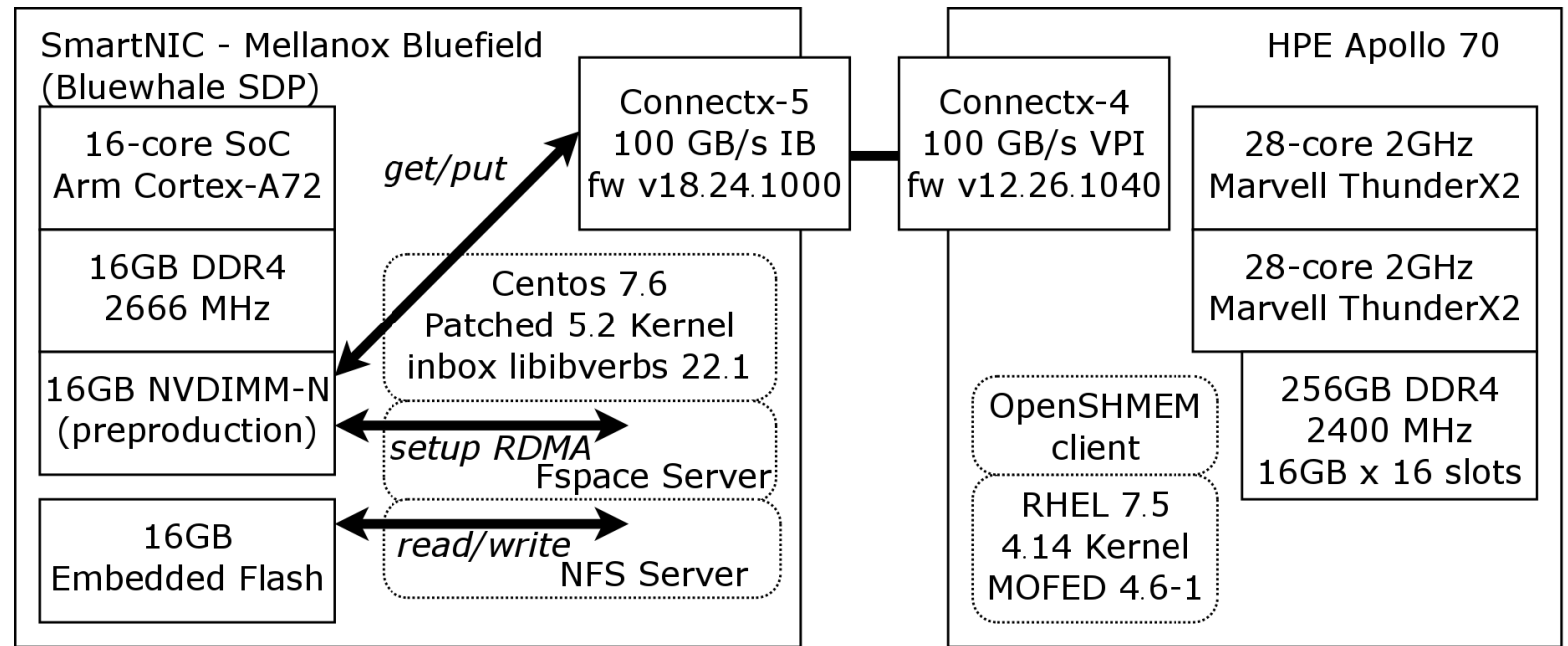
- 16 core SoC for off-chip in-memory processing
- Direct access to memory and attached storage
- Integrated into Infiniband adapter
- Runs full Linux/IB stack

- **NVDIMM-N**

- NVDIMM-N installed with battery pack and fits into DDR module slot
- Appears as unique PMEM device type to Linux
- Using Linux DAX, appears as file system but behaves as memory when files are mmap'd
- Syncs files to on-module battery backed NVM at processor request or on emergency power loss, so data is persistent.

- **HPE Apollo 70 client**

- **OpenSHMEM programming model**

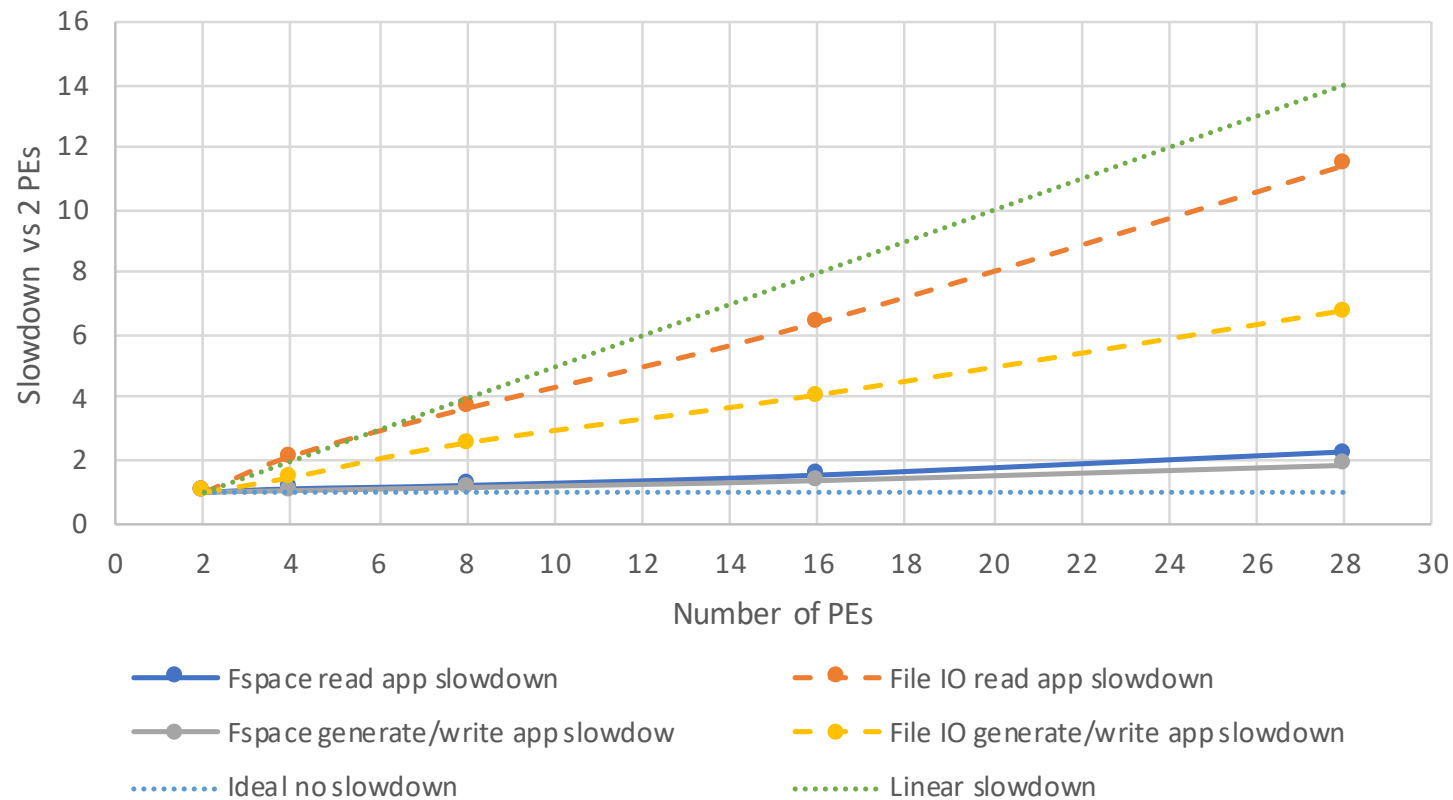


Grodowitz, Megan, Pavel Shamis, and Steve Poole. "OpenSHMEM I/O Extensions for Fine-Grained Access to Persistent Memory Storage." Smoky Mountains Computational Sciences and Engineering Conference. Springer, Cham, 2020.



# Edge Sort Total Runtime

Weak Scaling of Total Runtime for Graph Edge Decomposition



POSIX File I/O on NFS degrades linearly as number of processes increase. All access in parallel to non-overlapping file regions. Read app (App1) also writes back after sort so performance degrades worse for App1 as expected.

Fspace File I/O over same network fabric shows only small performance degradation as number of processes increases.

# Key challenges emerging in datacenter memory and storage hierarchies

Cloud vendor specific



## Programmability

Minimizing software overheads for emerging storage devices



## New memory technologies

New price/perf points and interfaces for storage (SCM, CXL)



## Increasing network costs

Need to minimize data movement



## Increasing compute heterogeneity

Improves efficiency via acceleration

Watch the companion talk in session B-201: [Extending DPUs to Enable Software-Defined I/O \(SDIO\)](#)

---

# Thank You!

