Redefining Data Redundancy with RAID Offload

Mahinder Saluja
Director of Strategy, SSD BU
KIOXIA America, Inc.

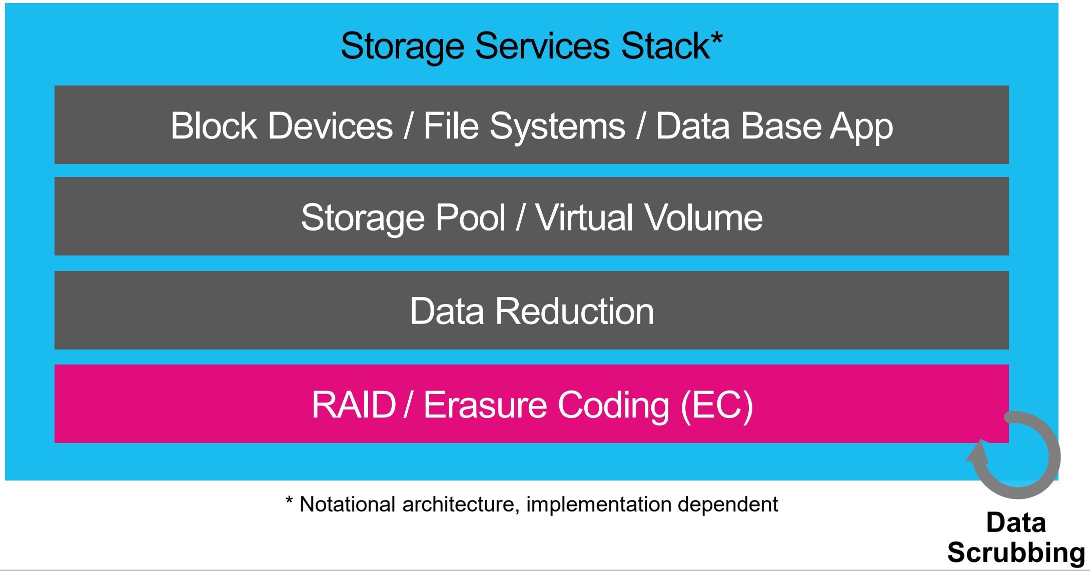
Agenda

- Storage Services Evolution
- Data Redundancy Compute Challenges
- Offload to SSD
- Why xPU Should Leverage SSD Offload

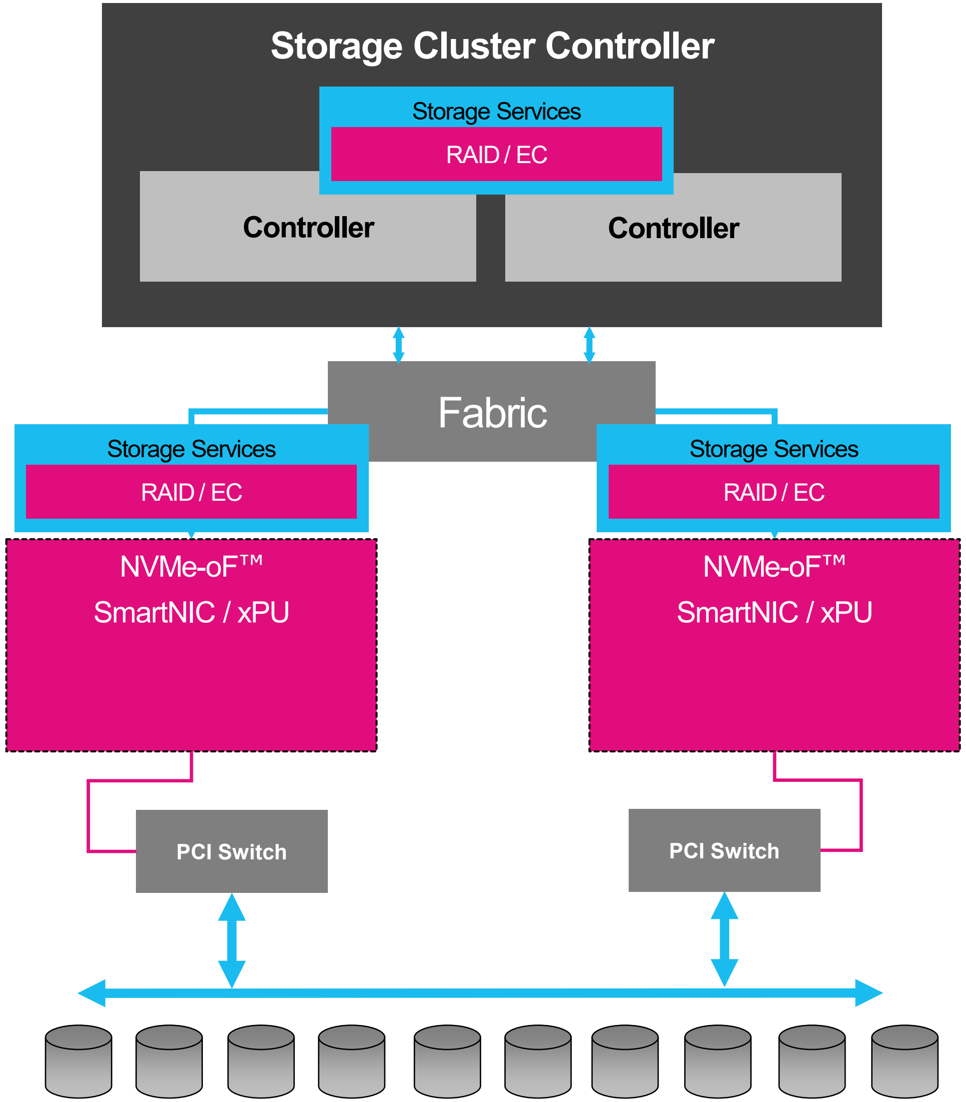
Storage Services Evolution

- Data redundancy in storage services demands high compute resources
- xPUs are making inroads to offload and accelerate storage services stack
- xPUs will be challenged for performance in future
  - NVMe™ performance continues to double with every PCIe® generation*

    *Based on the PCIe evolution as published by PCI-SIG.



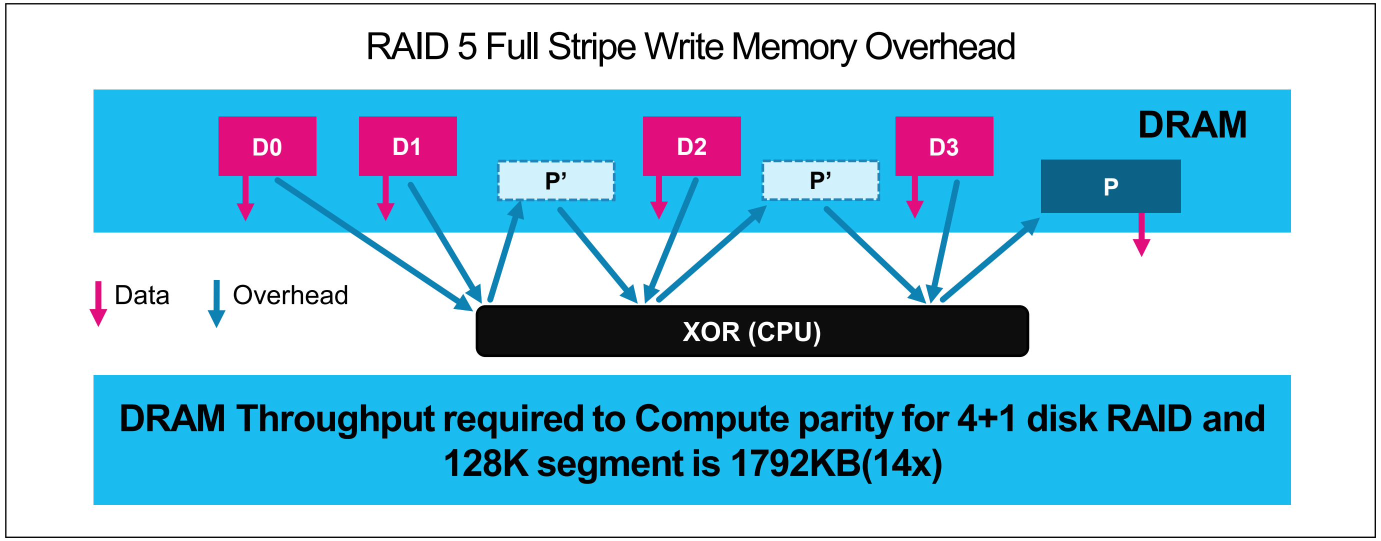
* Notational architecture, implementation dependent

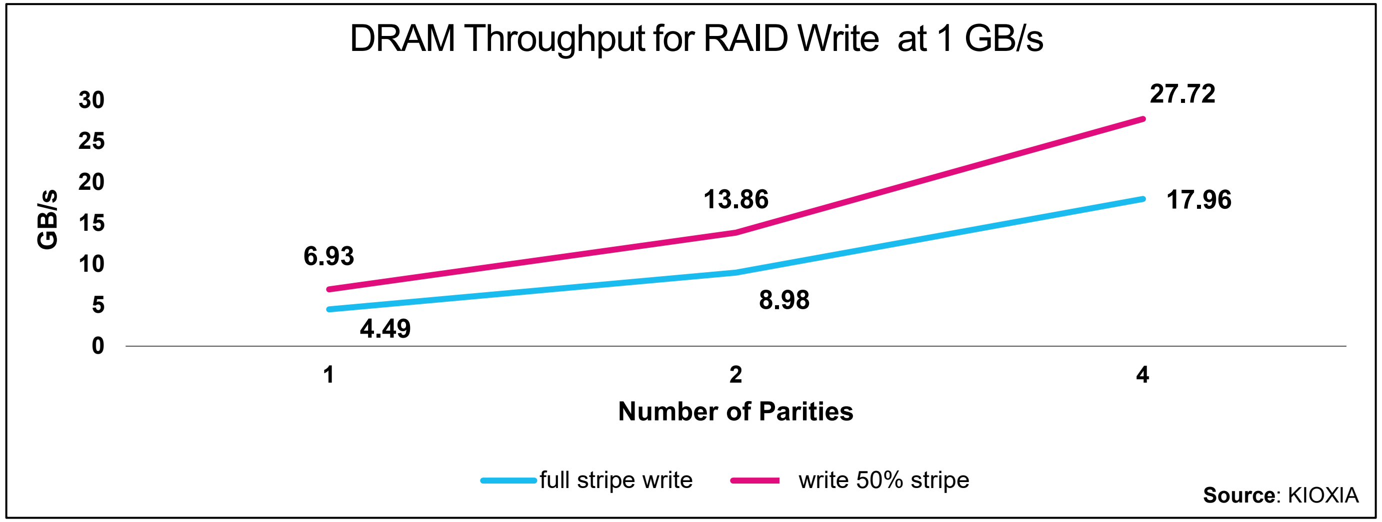


xPU represents a portfolio of architectures (i.e., CPU, GPU, FPGA and other accelerators), depending on the application.
NVMe™, and NVMe-oF™ are trademarks of NVM Express, Inc. PCIe is a registered trademark of PCI SIG.

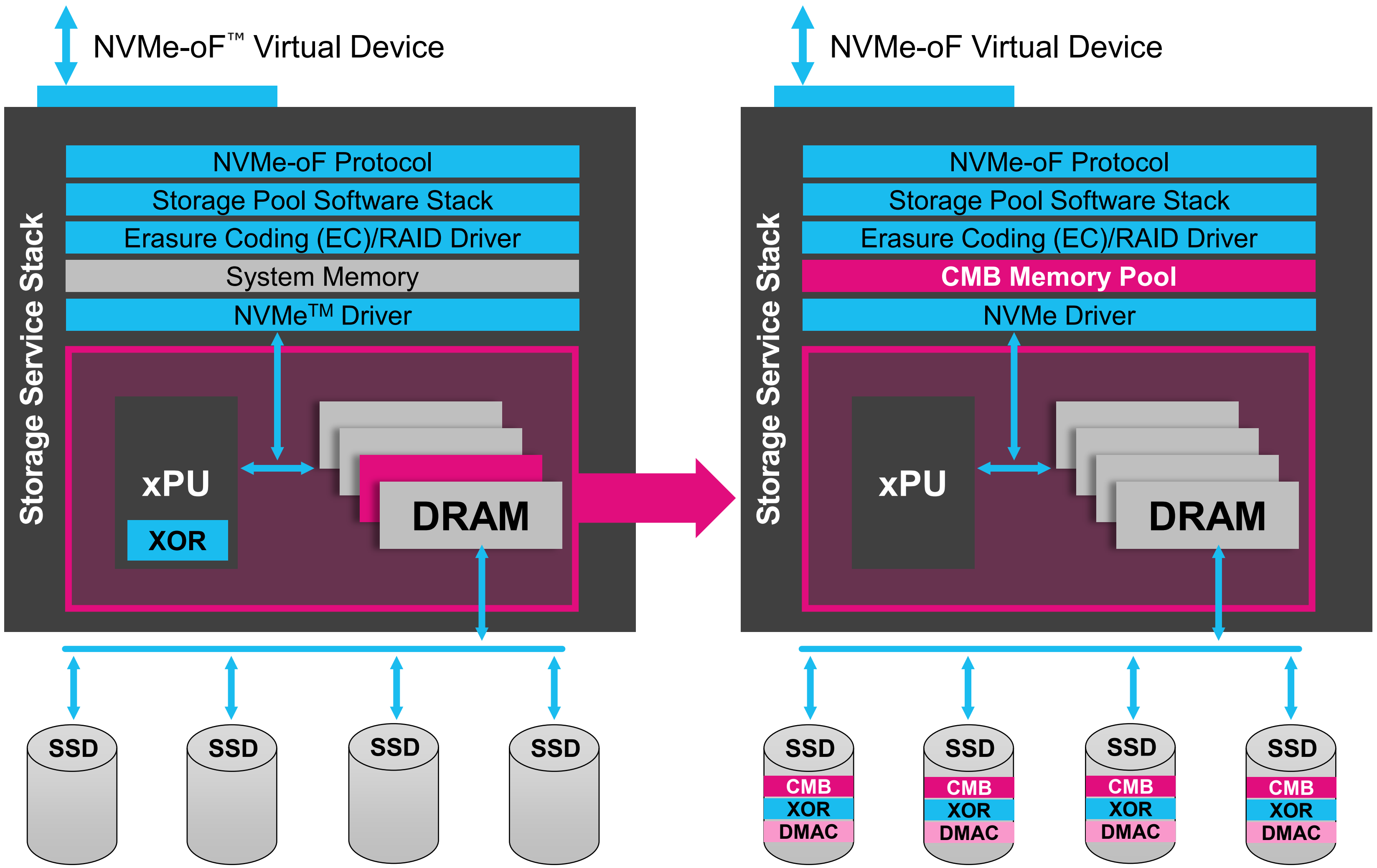
What are Data Redundancy Challenges?

- Parity compute is memory bandwidth and CPU intensive
- A RAID 5 partial stripe write requires ~10x DRAM throughput, for example a modest 4KB block RAID 5 write will consume 40KB DRAM bandwidth
- The problem worsens with RAID 6 / erasure coding (EC)
- System resources are overprovisioned to meet these demands



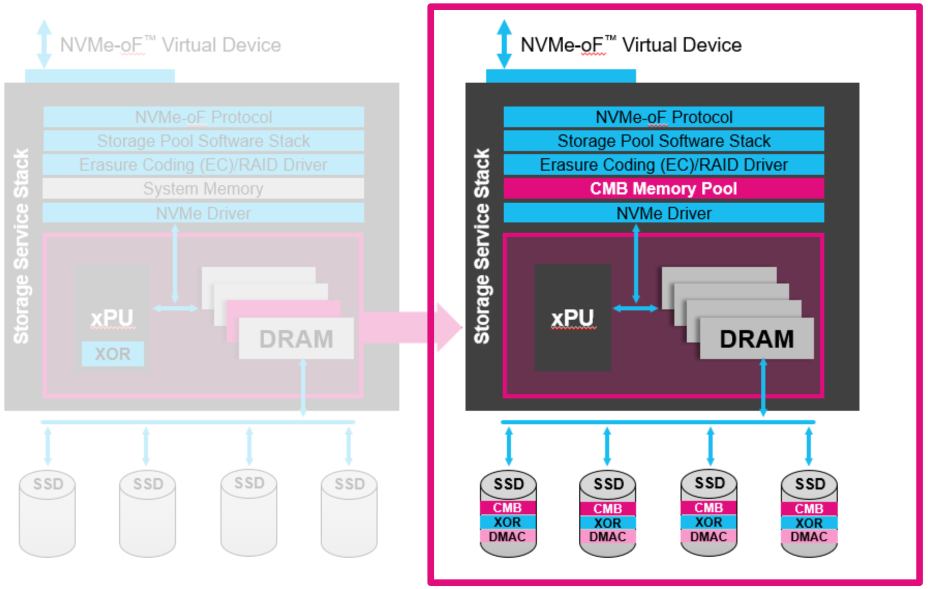


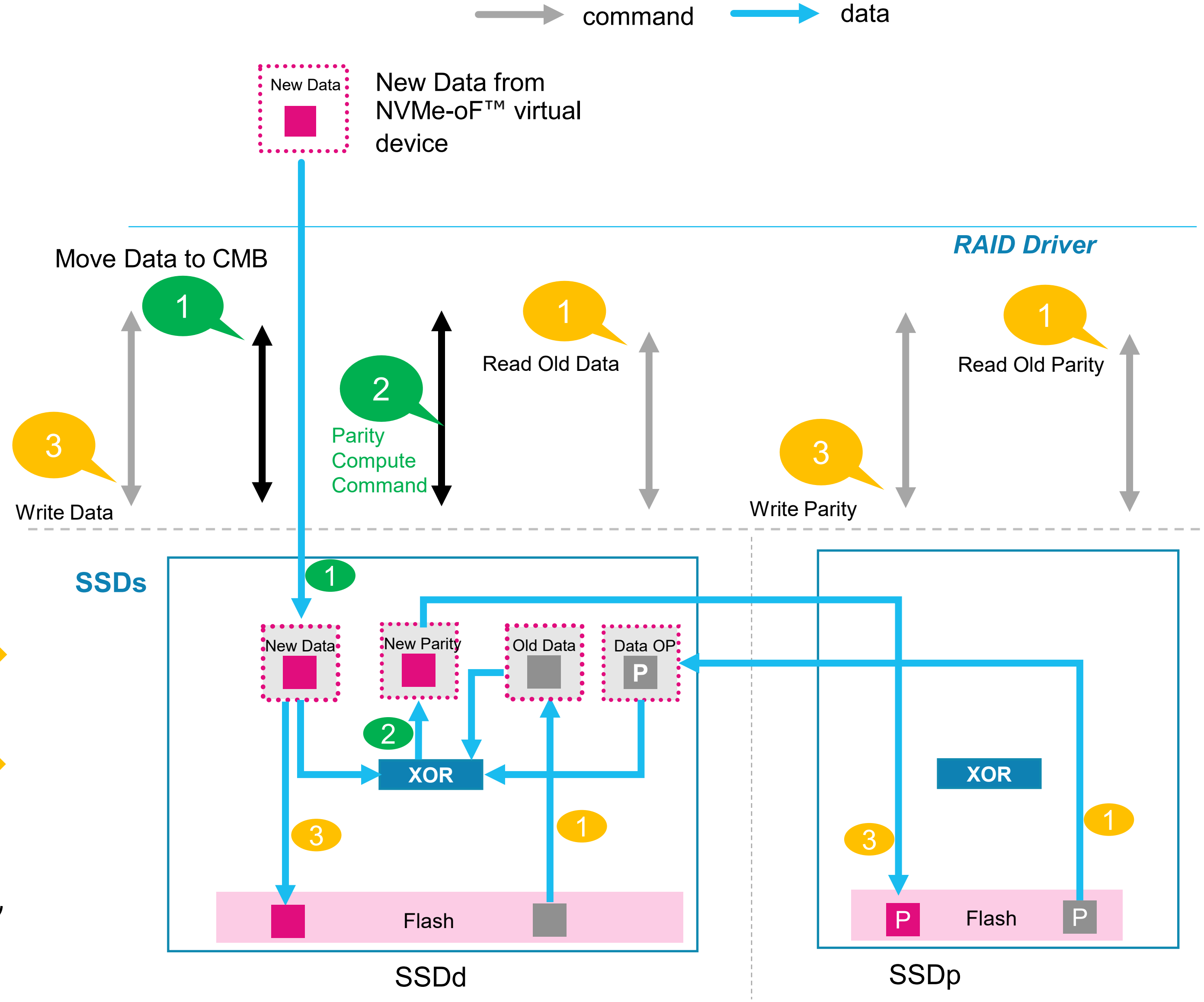
How xPUs Leverage RAID



- **KIOXIA NVMe SSD features**
  - Controller memory buffers (CMB) to offload DRAM
  - Exclusive OR (XOR) engine to compute up to 8 parities
  - Direct memory access controller (DMAC)
    - To place data in host address space (including remote CMB)
- **RAID Offload enables parallel compute and linear scaling**

How xPUs Can Leverage RAID – Sample Command Flow





Move New Data in CMB

Read Old Data in CM

Read Old Parity in CMB

Calculate New P Parity

Write New Data

Write New Parity

Step 1: Data In

Step 2: Compute

Step 3: Write

RAID 5 – Read / Modify / Write Sequence

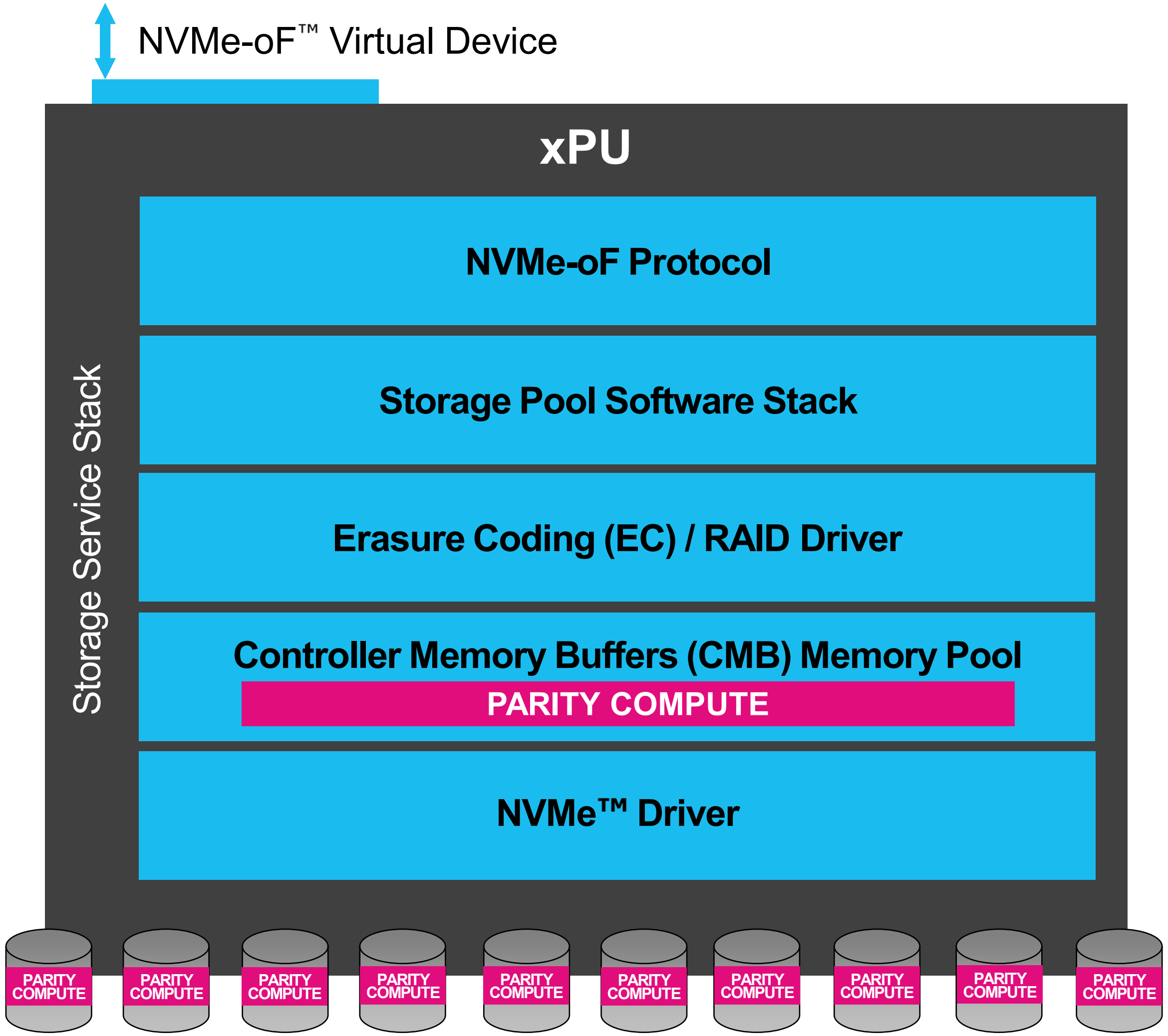
Why Should xPUs Leverage SSD RAID Offload?

xPUs can leverage their own accelerators, but why offload to SSD?

- Performance of accelerators will be limited to design time considerations
- The high memory bandwidth requirements increases the cost of xPUs
- The SSD offload can scale linearly with every SSD added to the cluster

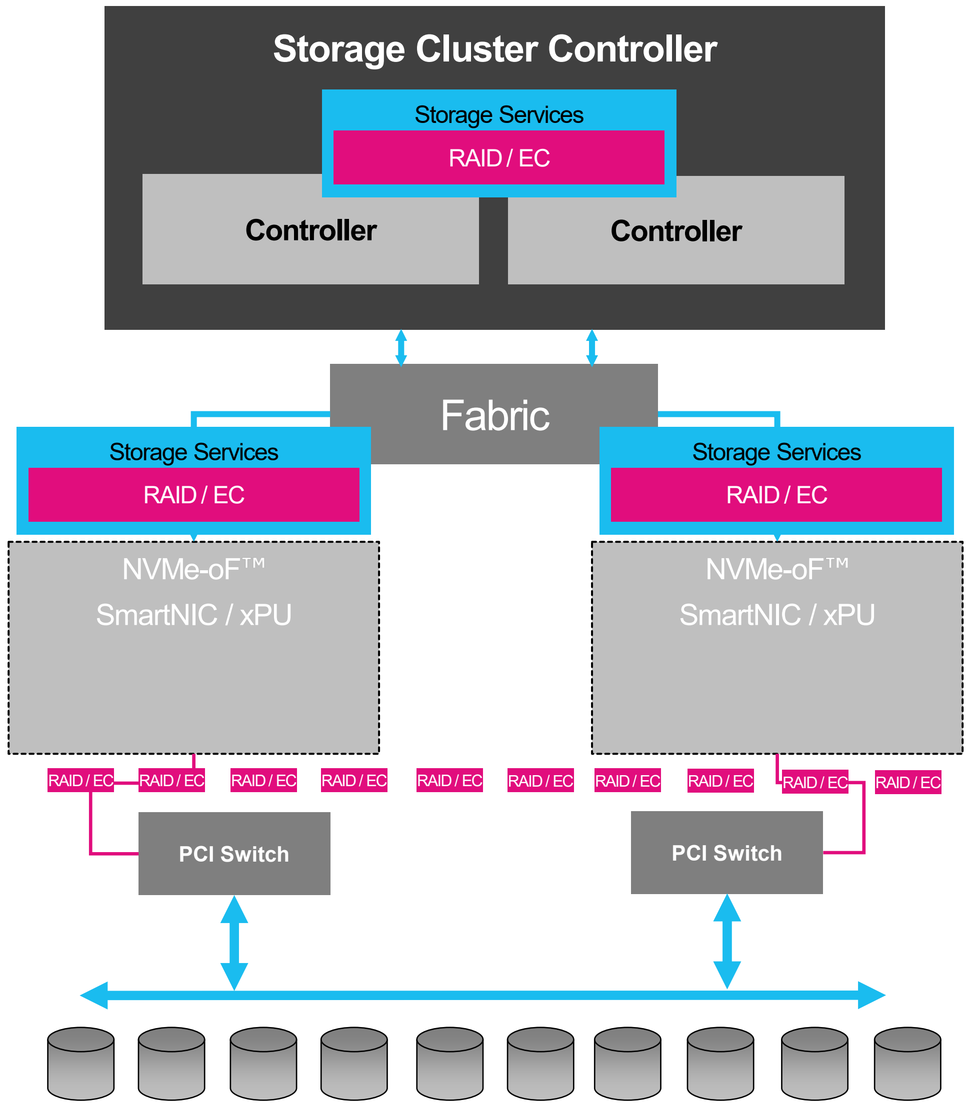
With offload…

- Save compute and memory bandwidth for value add storage functions
- Throw away operations like data scrubbing can be offloaded to SSD; 99% data movement reduction in data scrubbing operation to SSD
- Develop cost-effective data processing systems and solutions
- xPUs can scale RAID solution by leveraging its Remote Direct Memory Access (RDMA) capabilities



xPU represents a portfolio of architectures (i.e., CPU, GPU, FPGA and other accelerators), depending on the application.
NVMe™ and NVMe-oF™ are trademarks of NVM Express, Inc. PCIe is a registered trademark of PCI SIG.

RAID Offload Proof of Concept (PoC) results on CPU



RAID Offload : PoC Results (with KIOXIA CM7 and mdraid5) CPU attached

System	KIOXIA CM7 Gen4 x4 – mdRAID 5#	RAID Offload	% Benefit
CPU Utilization	42	37	12% reduction
DRAM Bandwidth (in MiB/s)	3450	340	91% reduction

System: DELL® PowerEdge™ R650xs Xeon® Gold 6338N 2.2GHz (2 Socket, 32 Cores) PCIe® 4.0 , SSDs: 5xCM7 Gen4 (1.92TB)

IO workload: FIO 512K Random Write @ 950MB/s

Data Scrubbing PoC Results CPU attached

	Offload Disabled	Offload Enabled
Scrubbing Time	129s	91s
DRAM Bandwidth	10.24 GB/s	1.43 GB/s
Total CPU Utilization	99.5%	~70%
L3 Cache Misses	14.7M	4M
Total PCIe® Write (MB/s)	3694 MB/s	159 MB/s

xPU represents a portfolio of architectures (i.e., CPU, GPU, FPGA and other accelerators), depending on the application. NVMe™ and NVMe-oF™ are trademarks of NVM Express, Inc. PCIe is a registered trademark of PCI SIG. Dell is a registered trademark of Dell Inc. PowerEdge is a trademark of Dell Inc.. Xeon is a registered trademark of Intel Corporation or its subsidiaries in the U.S. and/or other countries.

xPU and SSD can team up to build a cost-effective storage services solution.

Additionally, KIOXIA is exploring offload functions beyond RAID Offload.

For more information,
read our RAID Offload brief.

For more details,
visit KIOXIA FMS Booth #307.

KIOXIA