Changing Storage Architectures require new standards

Mark Carlson, Toshiba Memory

May 2019

What changing Architectures?

- New high speed low latency fabrics
- Composable Infrastructure
- New tiers of storage
- Computational Storage
- …

New Fabrics

- PCIe works great for direct attach high speed storage
    - NVMe over PCIe is quickly replacing SATA and SAS in new designs
    - However, this locks in a compute/storage ratio once a server is built
    - Possible to extend PCIe for a handful of hosts (OCP Lightning)
- New fabrics are emerging to enable disaggregation of systems into CPU, Memory, Storage components

GenZ

◆ An open systems Interconnect designed to provide memory-semantic access to data and devices via direct-attached, switched or fabric topologies.

◆ Bridges to PCIe/NVMe devices

◆ Leverages DMTF Redfish for management

OpenCAPI

◆ An open coherent high performance bus interface based on a new bus standard called Open Coherent Accelerator Processor Interface

◆ Emphasis on Accelerators

◆ A new class of interconnect focused on emerging acceleration applications such as machine learning, network processing, storage off-load, in-memory data base and 4G/5G wireless technology.

- ◆ A cache-coherent host-to-device interconnect, focusing on GPUs and FPGAs
- ◆ Leverages PCIe
  - ◆ Gen 6?

Composable Infrastructure

- As these high speed fabrics emerge, with disaggregated components, how can you then assemble them on the fly to create the ideal "system" for the application?
- How are the components managed?
- How is the resultant system managed?
- The components may not all be in the same chassis, so BMCs may not be in common (if needed at all)
- This all requires new standards

New tiers of storage

◆ **Persistent Memory is emerging to fill a gap between memory latencies and that of SSDs**

- ◆ The idea is to be cheaper than DRAM but faster than SSDs
- ◆ Bridging applications while the technology and standards catch up are NVDIMMs with both DRAM cache and NAND/PM (but cannot be cheaper than DRAM)

◆ **New QLC NAND increases capacity but at a cost in latency and perhaps wear**

- ◆ Bulk NAND may fit between TLC NAND and HDDs

9

Persistent Memory

- SNIA has created the NVM Programming model for how applications can take advantage of PM
- Major OS and Hypervisor vendors have added support for PM and the programming model
- The ecosystem is here and applications are moving to this new model use case by use case
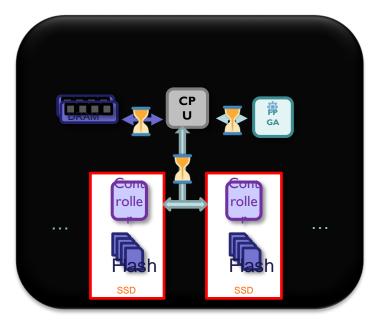 - Primarily for performance reasons

Computational Storage

◆ **The notion has been around for some time with various names**

◆ **Offload computation from the CPU**

   ◆ Specialized processing

   ◆ Move the computing to the data, don't move tons of data through the CPU
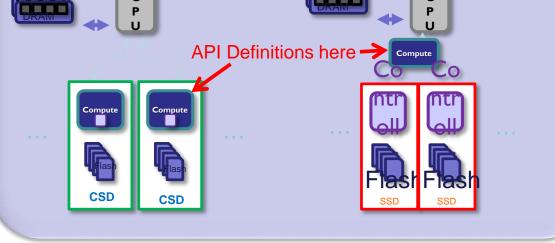
What's the Problem?

With the widespread adoption of high-speed storage and deployment of data-intensive applications, the traditional architecture…

o *Creates* CPU, Memory & Storage *Bottlenecks*

o With Offloads that are *not* in-line with Storage

o and *fails to scale* Performance with Capacity

How do we fix it?

Example Architectures

Distributed-Processing and Data-Driven

API Definitions here

Compute

CSD   CSD

Flash Flash
SSD   SSD

Computational Storage Solutions

- ✓ **Alleviate** CPU, Memory & Storage **Bottlenecks**

- ✓ with Offloads **in-line** with Storage

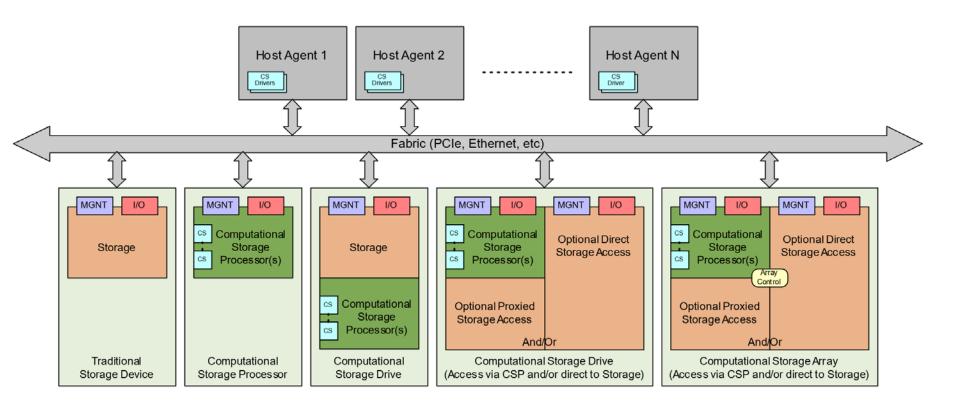- ✓ and Performance that **scales** with additional Computational Storage

Notice the missing pieces-

- • Software design- we need your help

Computational Storage

Diagram showing Host Agent 1 (CS Drivers), Host Agent 2 (CS Drivers), through Host Agent N (CS Driver), connected to Fabric (PCIe, Ethernet, etc).

Five device types connected to the fabric:

- **Traditional Storage Device**: MGNT, I/O, Storage
- **Computational Storage Processor**: MGNT, I/O, CS Computational Storage Processor(s)
- **Computational Storage Drive**: MGNT, I/O, Storage, CS Computational Storage Processor(s)
- **Computational Storage Drive (Access via CSP and/or direct to Storage)**: MGNT, I/O, CS Computational Storage Processor(s), Optional Direct Storage Access, Optional Proxied Storage Access, And/Or
- **Computational Storage Array (Access via CSP and/or direct to Storage)**: MGNT, I/O, CS Computational Storage Processor(s), Array Control, Optional Direct Storage Access, Optional Proxied Storage Access, And/Or

Synergy

- ◆ While initially Computation Storage will be realized on PCIe, you can see how this can fit with new emerging fabrics

- ◆ Need new NVMe features to enable computation

- ◆ Need to manage computation based components

- ◆ Need to assemble "Stacks" of computation and storage

Standards

◆ **Many standards for next gen fabrics – who will win?**
- Both Hyperscalers and Enterprise customers are participating in multiple orgs

◆ **Composable Infrastructure is enabled by these fabrics**
- Management standards (Redfish) need to be extended

◆ **Persistent Memory also needs new DDR and Memory fabric standards**

◆ **Computational Storage needs NVMe enhancements and new fabrics as well**

SNIA

- **The Storage Networking Industry Association has always had a key role in storage management standards**
  - SNIA Swordfish extensions to Redfish needed for these new architectures

- **SNIA has formed a new group to address Computational Storage**
  - Come join the effort
  - (See Steven's talk)

Thank You!

QUESTIONS?