Annual Update on Flash Memory for Non-Technologists

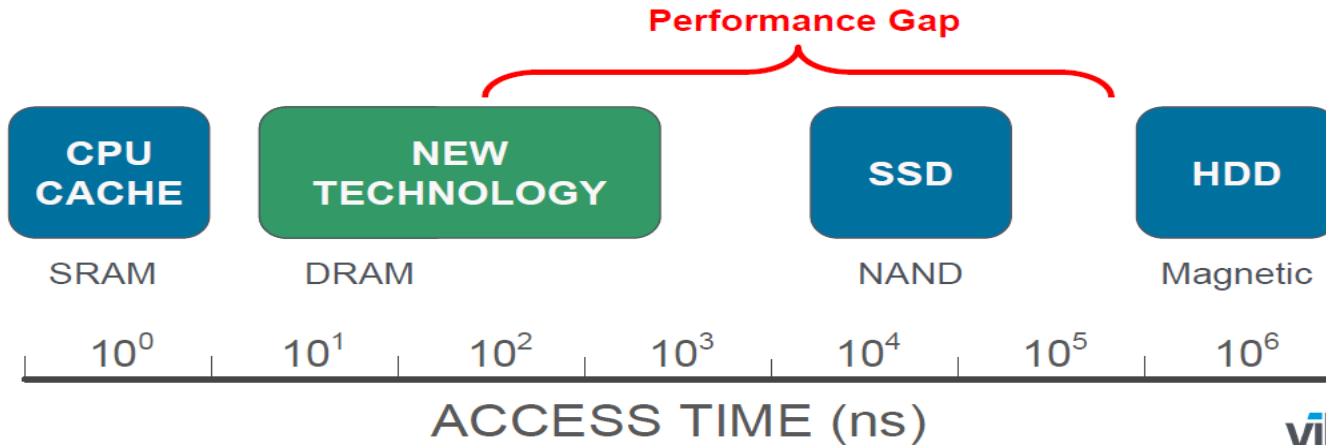Jay Kramer, Network Storage Advisors

&

George Crump, Storage Switzerland

Memory / Storage Hierarchy

- Data-Intensive Applications Need Fast Access To Storage
- Large Performance Gap Between Main Memory And HDD
- SSDs Have Narrowed The Gap, But A Gap Still Exists
- Opportunity For Innovation!

Performance Gap

CPU CACHE	NEW TECHNOLOGY	SSD	HDD
SRAM	DRAM	NAND	Magnetic

10^0 10^1 10^2 10^3 10^4 10^5 10^6

ACCESS TIME (ns)

NAND Flash Solutions

What's an SSD?

- **Solid State Disk – all electronic, no moving parts**
- **As opposed to the Hard Disk Drive (HDD) – electromechanical**
  - **SSDs typically use NAND flash for the persistence layer**
  - **SSD have evolved from 2D NAND to 3D NAND (future MRAM, RRAM, STT-RAM, PCM)**

Why SSDs?

- **Can address different Use Cases for capacity, speed, endurance, & price**
- **Have now surpassed HDDs in density (capacity per physical size)**

SSD Packaging Options

- **SSDs come in various form factors**
  - **U.2 connector 2.5" and 3.5" form factor – SATA, SAS, or NVMe**
  - **M.2 and M.3 form factor – SATA or NVMe**
  - **Add-on card form factor – PCI-Express – NVMe**

# NAND Flash Solutions

Evolution of flash form factor:

- **SLC** – single level cell – the first form of NAND
  - Nice and simple – and fast with high PE cycles
    – **but not dense**

- **MLC** – multiple level cell – aka two-bits per cell
  - **Slower than SLC and fewer PE cycles** but more dense

- **TLC** – triple level cell – three bits per cell
  - **Slower than MLC and fewer PE cycles** but more dense

- **QLC** – four bits per cell
  - **Slow to read & write and lower endurance**
    - WORM (Write Once Read Many) flash use

NAND Flash Specs
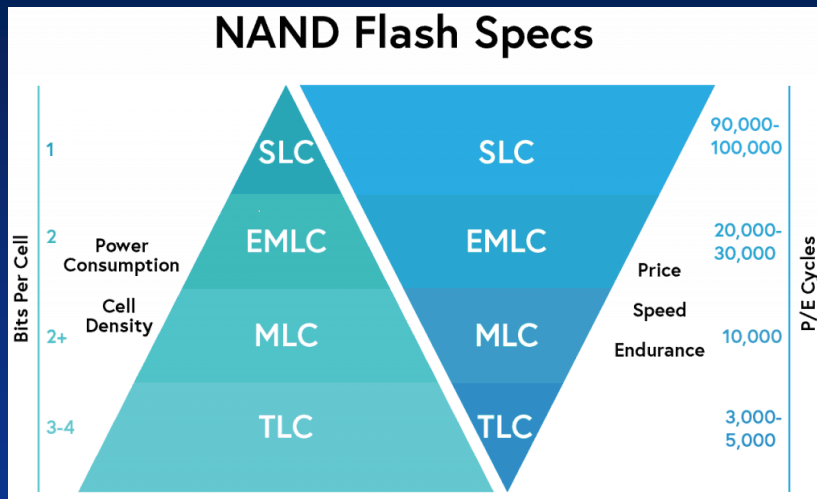
As we move from **SLC** ➡ **MLC** ➡ **TLC** ➡ **QLC**
we get **higher density** and **lower cost** but we also get **lower performance** and **shorter life**
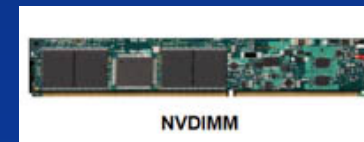
Memory Technologies

- **Flash Memory – the technology of SSDs**
  - An electronic (solid-state) non-volatile computer storage medium that can be electrically erased & reprogrammed
    - Faster, less power, less cooling, smaller footprint than spinning media
  - Still in many ways emulating the behavior of HDDs

- **NVDIMM - Non-Volatile Dual In-line Memory Module**
  - A random access memory for computers that retains data even when electrical power is removed
    - Speed, endurance, and random byte addressability of DRAM and the non-volatility of NAND
  - But higher cost than SSDs and HDDs

- **3D NAND – is a three dimensional computer logic circuit**
  - A type of flash memory in which the memory cells are stacked vertically in multiple layers.
    - Higher densities at a lower cost per bit, reduced power consumption, boost reliability, higher data write performance
  - But the layering process adds steps to the 3D NAND manufacturing process initially and adds to a higher manufacturing cost

FLASH

DRAM
Fast, Volatile, Expensive

NAND
Slow, Persistent, Cheap

3D XPOINT
Fast, Persistent, High Density

NVDIMM

Storage Interfaces

- **SAS/SATA – Serial Attached SCSI and Serial Advanced Technology Attachment**
  - Particular type of electronic interfaces developed for a hard drive and used with SSDs
    - SATA is lower cost and cost effective for entry level servers, desktops, laptops
    - SAS is higher performance enables greater scale of multiple devices to be connected simultaneously
    - Common protocols for HDDs and SSDs but restricts the highest possible flash performance

- **PCIe - Peripheral Component Interconnect Express**
  - A high-speed serial computer expansion bus standard and doubles the data transfer rates of original PCI. Provides lower latency & higher data transfer rates than parallel busses such as PCI.
    - PCIe can provide better performance as devices are not competing for bandwidth and not sharing the same bus. PCIe puts the SSD closest to the CPU memory complex for lower potential latency.
    - PCIe can provide greater scale (from one to 32 separate lanes) using a switched architecture of channels. It supports bidirectional capabilities, meaning reads and writes can occur simultaneously.
    - Many traditional storage array architectures need to evolve to support a large number of PCIe slots to achieve greater scalability and performance

Storage Interfaces

- **SAS/SATA – Serial Attached SCSI and Serial Advanced Technology Attachment**
  - Interfaces developed for a hard drive and used with SSDs (But can't achieve the highest flash performance)
    - SATA is lower cost and cost effective for entry level servers, desktops, laptops
    - SAS is higher performance enables greater scale of multiple devices to be connected simultaneously

- **PCIe - Peripheral Component Interconnect Express**
  - A high-speed serial computer expansion bus providing lower latency & higher data transfer rates than PCI.
    - PCIe can provide better performance as devices are not competing for bandwidth and not sharing the same bus. PCIe puts the SSD closest to the CPU memory for lower potential latency.
    - PCIe can provide greater scale (from one to 32 separate lanes) using a switched architecture of channels. It supports bidirectional capabilities for simultaneous reads and writes. (Legacy storage arrays don't support large # of PCIe slots)
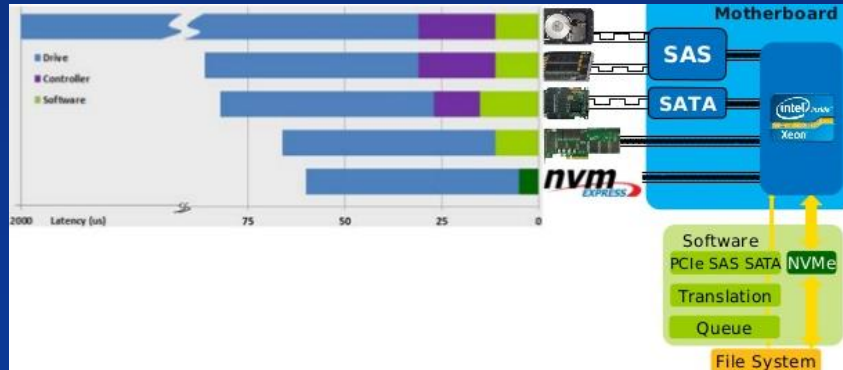
- **PCIe** removes controller latency

- **NVMe** reduces software latency

Storage Interfaces

- **NVMe – Non-Volatile Memory Express**
  - Enables a solid-state drive (SSD) to use of a variety of form factors including high-speed Peripheral Component Interconnect Express (PCIe) cards, M.2 SSDs and U.2 2.5-inch SSDs & can do the following:
    - Reduce the latency in the host software stack
    - Provide higher input/output operations per second (IOPS)
    - Potentially lower power consumption depending on the form factor and number of PCIe lanes in use

    Raises the bar on performance but at a higher cost

- **NVMe over Fabrics (NVMe-oF)**
  - Specification completed on June 5, 2016 and it's designed to extend the high-performance and low-latency benefits of NVMe using remote direct memory access (RDMA) and across network fabrics that connect servers and storage systems (e.g., Fibre Channel (FC), Ethernet and InfiniBand)

    Compatibility and interoperability of NVMe-oF with multiple vendor products and with the ability to protect existing customer investments in SAN makes this new solution very compelling for IT ROI.
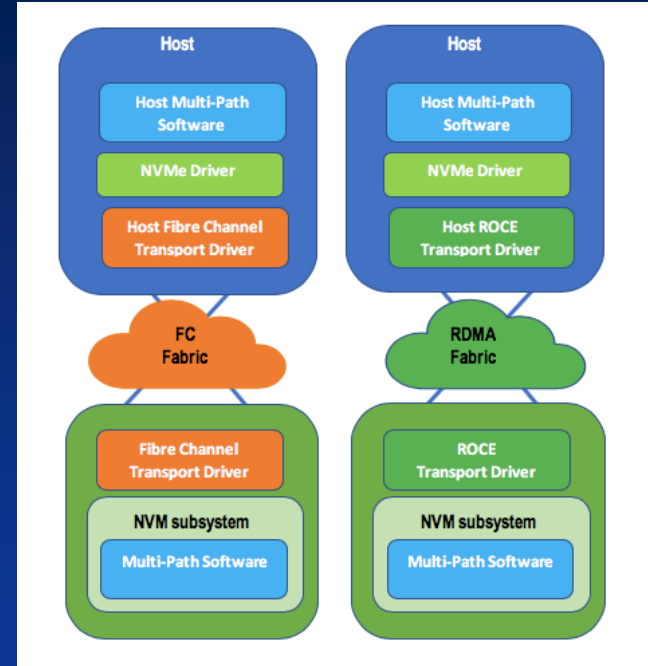
NVMe oF – Transport Protocols

RDMA – Remote Direct Memory Access

- RDMA is a host-offload, host-bypass technology that allows an application (including storage) to make data transfers directly to/from another application's memory space. It achieves accelerated performance

Multiple Transport Protocol Choices – Ethernet, InfiniBand, Fibre Channel
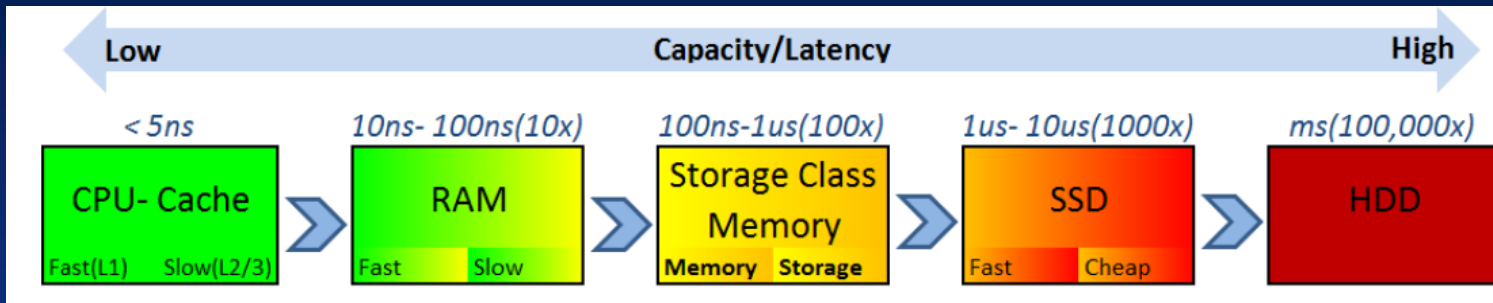
- **RoCE** - RDMA over Converged Ethernet (RoCE) is a network protocol that allows remote direct memory access (RDMA) over an Ethernet network. It is based on InfiniBand transport over Ethernet.
- **iWARP** - is another RDMA protocol alternative layered on top of TCP/IP (joint venture Intel & Carnegie Mellon U).
- **TCP** – TCP enables efficient end-to-end NVMe operations between NVMe-oF host(s) and NVMe-oF controller devices interconnected by any standard IP network with optimized performance and latency characteristics.
- **PCIe** – NVMe-oF enables PCIe as a transport for high performance when extended distances are not required.
- **FC** – Allows mapping of other protocols (SCSI, FICON) and can coexist with Gen6 FC for investment protection.

Availability

- **RoCE and iWARP** run on all major Ethernet switches (Arista, Cisco, Dell, HPE, Mellanox, etc.)
- **RoCE and iWARP** adapters from a variety of vendors (Avago, Cavium, Chelsio, Intel)
- **TCP** can leverage existing ubiquitous Ethernet infrastructure with switches & traditional Ethernet network adapters
- **Fibre Channel** supports NVMe-oF from the major SAN interconnect vendors (Broadcom, Cavium, Cisco, Mellanox)

Storage Class Memory

Low	Capacity/Latency			High
< 5ns	10ns- 100ns(10x)	100ns-1us(100x)	1us- 10us(1000x)	ms(100,000x)
CPU- Cache	RAM	Storage Class Memory	SSD	HDD
Fast(L1) Slow(L2/3)	Fast Slow	Memory Storage	Fast Cheap	

- **Storage Class Memory (aka Persistent Memory)**
  - A new class of data storage/memory devices that overcomes the high cost and volatility of DRAM and has greater performance than traditional SSDs.
    - **It's persistent, meaning that information is not lost if the server crashes or loses power.**
    - **It can provide read performance similar to DRAM and write performance that is significantly faster than SSD or HDD technology.**

    Can potentially shift the bottleneck to another part of the ecosystem like the Server CPU

Storage Class Memory

- **MRAM** **(magnetoresistive random access memory)**
- A non-volatile RAM memory technology and a method of storing data bits by using magnetic charges to store data instead of electric charges used by DRAM (dynamic random access memory). Unlike DRAM, MRAM will retain data even when the power is turned off
    - MRAM was first developed by IBM in the 1970s

- **RRAM (resistive random access memory)**
- A non-volatile RAM memory technology with higher switching speed and draws less power than NAND flash.

- **PCM (Phase-Change Memory)**
- A type of non-volatile RAM that has bit-alterability to change the state of the memory for performance optimization. Market forecasts that the global phase change memory market will grow at a CAGR of 84% during the period 2017-2021.

The higher memory density, faster read and write speeds, greater durability and lower power draw are reasons why memristor-based memory technologies are often cited as the logical replacement in apps like solid-state drives (SSDs) and nonvolatile dual in-line memory modules (NVDIMMs).

DRAM
~10ns

SCM
1~10us

3D TLC
100~200us

3D QLC
100~200us

HDD
3~5ms
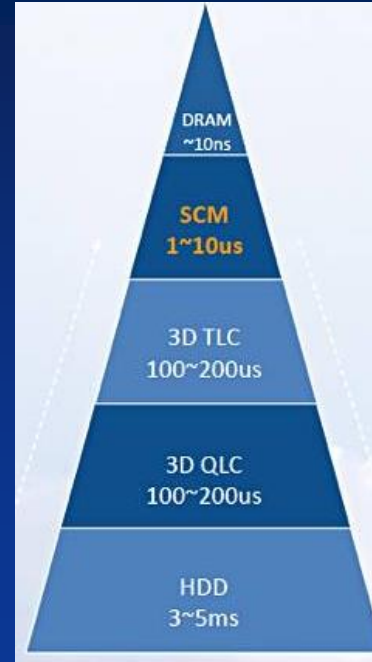
Software Defined Storage

- **Software Defined Storage**
  - An evolving concept for computer data storage software to manage policy-based provisioning and management of data storage independent of the underlying hardware.

- **Software Defined Storage (SDS) must include:**
  - **Automation – Simplified management that reduces the cost of maintaining the storage infrastructure**
  - **Standard Interfaces – APIs for the management, provisioning and maintenance of storage devices and services**
  - **Virtualized Data Path – Block, File and/or Object interfaces that support applications written to these interfaces**
  - **Scalability– Seamless ability to scale the storage infrastructure without disruption to the specified availability or performance**
  - **Transparency – The ability for storage consumers to monitor and manage their own storage consumption against available resources and costs**
  - **Fabric – The ability to move data to the cloud and across clouds**

  It is part of a larger industry trend for a Software Defined Data Center that includes software-defined networking (SDN).
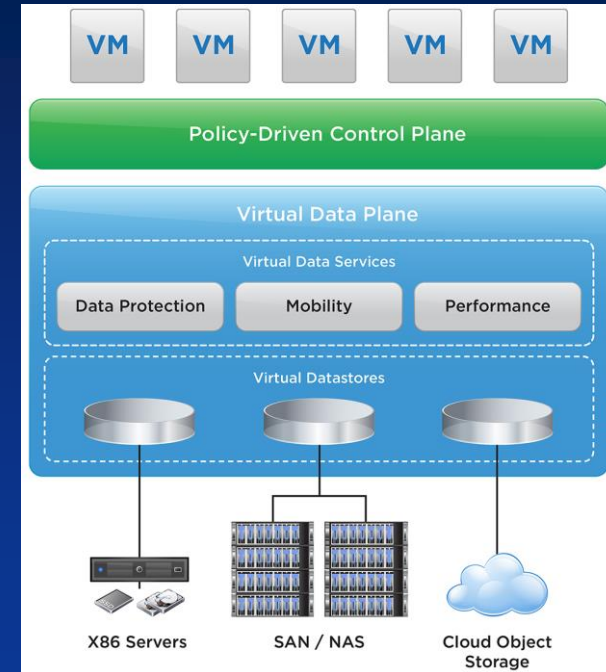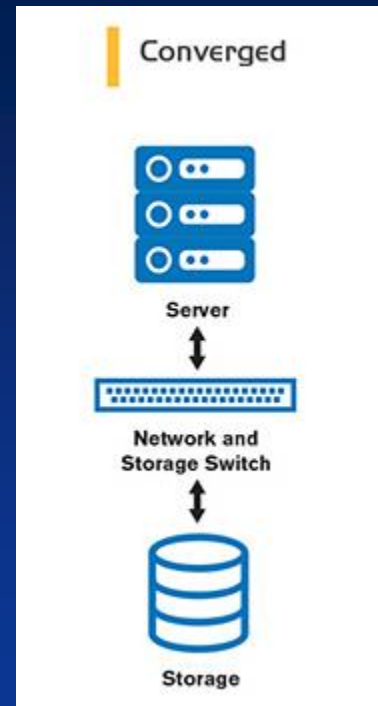
Converged and Hyperconverged

- **Converged Storage**
- Converged storage is a storage architecture that combines storage and computing resources into a single entity. This integration of storage and computing hardware and processes is designed to achieve the following:

    - **Speed delivery time**

    - **Simplicity of implementation**

    - **Optimize application performance**

    - **Minimize power, cooling, and physical space requirements**

    - **Packaged solution for virtualized and cloud-based environments**

    Proprietary packaged solution that can be considered **a lock-in IT infrastructure** or can be viewed as **a value-added single vendor supported solution.**

Converged and Hyperconverged

- **Hyperconverged Storage**
  - A software-defined approach to storage management that combines storage, compute, networking and virtualization technologies in one physical unit that is managed as a single system.
    - **Flexibility:** Leverages industry standard hardware and commoditization for flexibility of hardware vendors.
    - **Scalability:** Hyperconverged systems can be expanded through the addition of nodes to the base unit.
    - **Integrated:** Hyperconverged systems are highly integrated and can not be broken down into separate components.

    An architecture adopted by the Hyperscale Cloud Providers now available in the general marketplace but it is an architecture that is **unique and different from traditional NAS and SAN solutions.**


Hyper - Converged

Network

Storage and Server
(combined)
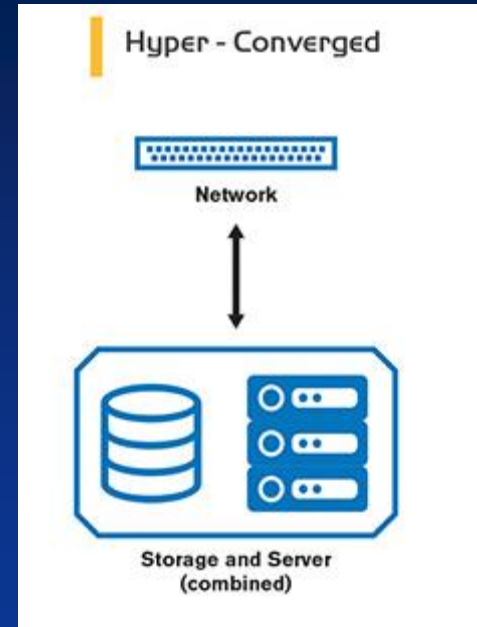
What is Computational Storage?

SNIA has Defined the Following

- **Computational Storage Drive (CSD):**
  - A component that provides persistent data storage and computational services

- **Computational Storage Processor (CSP):**
  - A component that provides computational services to a storage system without providing persistent storage

- **Computational Storage Array (CSA):**
  - A component that provides computational services to a storage system without providing persistent storage

- **Computational Storage Enabled by NVMe**
  - Low latency
  - High throughput
  - Low CPU overhead
  - Multicore awareness
 - Management at scale
  - QoS awareness
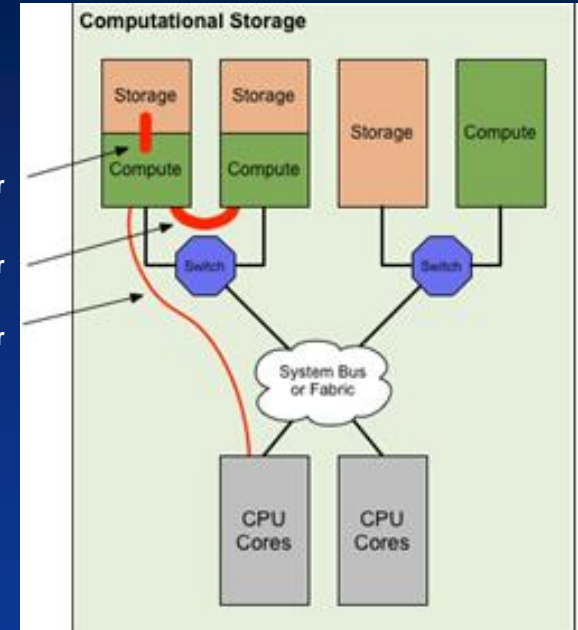
What is Composable Storage?

SNIA has Defined the Following

Composable Infrastructure

- ▪ **Composable Storage - definition**
    - ▪ Assign controllers and assign capacity on the fly to an application with the ability to release it back to the pool of resources.

    This is typically enabled by the flexibility of an NVMe architecture.

- • **Composable Storage - advantages**
    - ▪ **Flexibility:**
        - ▪ Composable storage offers you the flexibility to quickly scale up for increased storage capacity or scale out for improved performance or scale down when not needed.

        AI will play an increasing role in storage   management.
    - ▪ **Performance:**
        - ▪ Performance sensitive analytics workloads can borrow additional controller compute to manage surge requirements and then return to the pool when no longer needed.

        Composability can create a QoS environment for storage

Subsystem Aggregation

IO

Compute
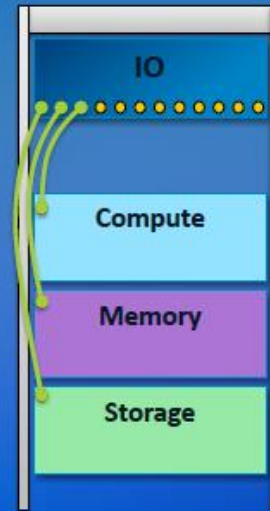
Memory

Storage

Pooled compute
Pooled storage
Pooled memory
Shared boot

References

- **The SNIA Dictionary**
  - https://www.snia.org/sites/default/files/dictionary/SNIADictionaryv2016_1.pdf

- **Wikipedia**
  - https://en.wikipedia.org/wiki/

- **Tech Target Search Solid State Storage**
  - searchsolidstatestorage.techtarget.com/definition

- **The Memory Guy – Jim Handy**
  - thememoryguy.com

- **Storage Switzerland – George Crump**
  - https://storageswiss.com/

Thank You!!!

Annual Update on Flash Memory for Non-Technologists

Jay Kramer, Network Storage Advisors

jay@networkstorageadvisors.com

&

George Crump, Storage Switzerland

georgeacrump@storageswiss.com