



**SDC<sup>18</sup>**

September 24-27, 2018  
Santa Clara, CA

[www.storagedeveloper.org](http://www.storagedeveloper.org)

# **Gen-Z Communication at the Speed of Memory**

**Kurtis Bowman**  
**President, Gen-Z Consortium**

# Disclaimer

This document is provided 'as is' with no warranties whatsoever, including any warranty of merchantability, noninfringement, fitness for any particular purpose, or any warranty otherwise arising out of any proposal, specification, or sample. Gen-Z Consortium disclaims all liability for infringement of proprietary rights, relating to use of information in this document. No license, express or implied, by estoppel or otherwise, to any intellectual property rights is granted herein.

Gen-Z is a trademark or registered trademark of the Gen-Z Consortium.  
All other product names are trademarks, registered trademarks, or servicemarks of their respective owners.

All material is subject to change at any time at the discretion of the Gen-Z Consortium

<http://genzconsortium.org/>

# About The Gen-Z Consortium

- ❑ The Gen-Z Consortium launched in October of 2016 to create an open, industry standard for a high speed, low latency, scalable, memory centric fabric
- ❑ There are currently 60+ member companies covering all of the disciplines required to create an ecosystem based on this open standard
- ❑ Gen-Z Has shown demos of memory pooling with multiple servers over the last year
- ❑ Gen-Z members have released design IP and silicon vendors have started detailed designs for Gen-Z devices
- ❑ Released and selected draft specifications are available on **[www.GenZConsortium.org](http://www.GenZConsortium.org)** for public review and comment



# Open Consortium With Broad Industry Support

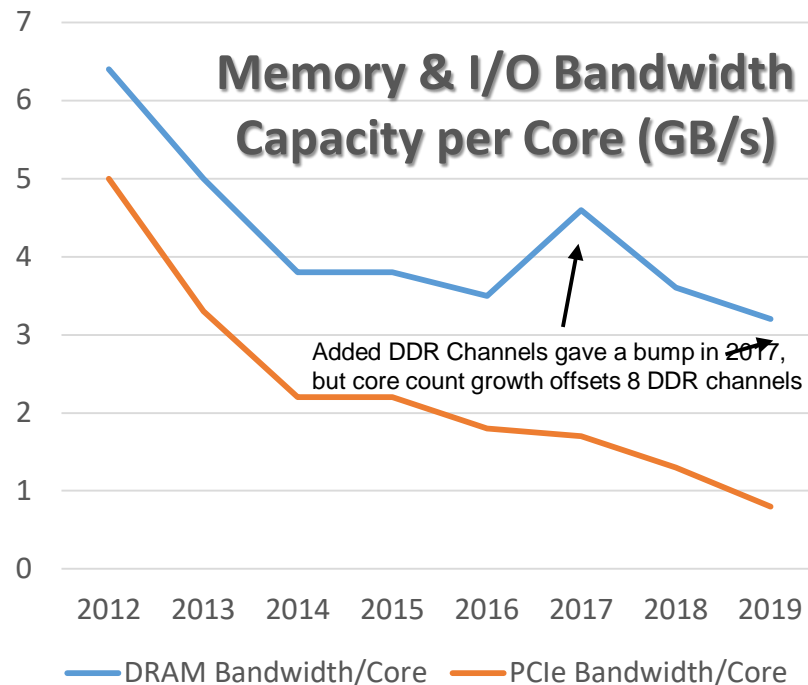
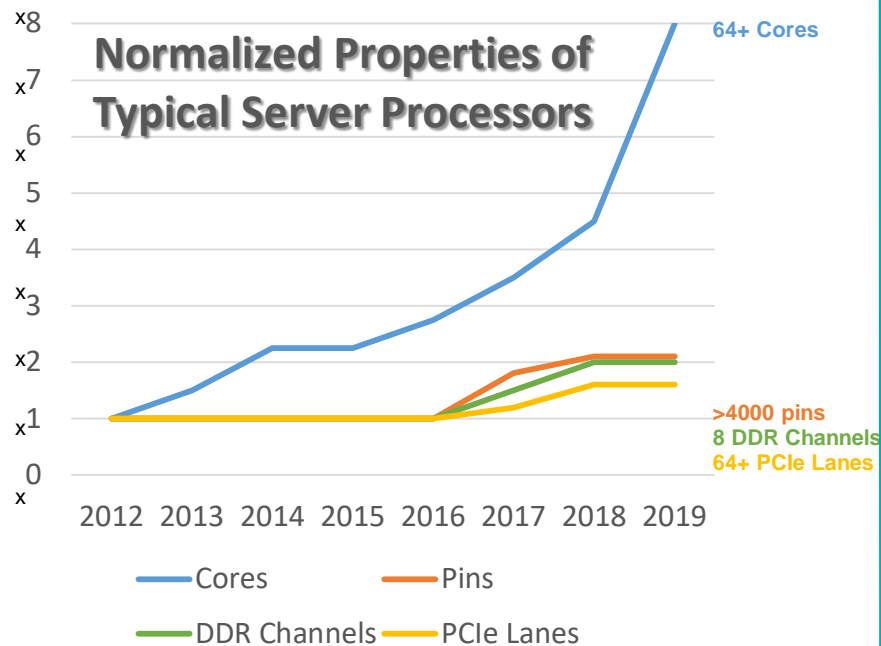


## GEN Z Consortium Members

- Allion Labs
  - Alpha Data
  - AMD
  - Amphenol
  - ARM
  - Avery Design Systems
  - Broadcom
  - Cadence
  - Cavium
  - Cisco
  - Cray
  - Dell EMC
  - Everspin
  - ETRI
  - FIT
  - Genesis Conn Solutions
  - Google
  - Hirose
  - HP
  - HPE
  - Huawei
  - IBM
  - IDT
  - IntelliProp
  - ITT Madras
  - Jess Link
  - Keysight
  - Lenovo
  - Lotes
  - Luxshare-ICT
  - Mellanox
  - Mentor
  - Micron
  - Microsemi
  - Mobiveil
  - Molex
  - NetApp
  - Nokia
  - Oak Ridge Natl Labs
  - PLDA Group
  - Qualcomm
  - Red Hat
  - Samsung
  - Samtec
  - Seagate
  - Senko Advanced Comp
  - Simula Research Lab
  - SK hynix
  - Smart Modular
  - Spin Transfer Tech
  - Teledyne LeCroy
  - TE
  - Toshiba Memory Corp
  - Univ. New Hampshire
  - VMware
  - Western Digital
  - Xilinx
  - Yadro
  - Yonsei University
  - 3M
- \*Board member  
\*Associate member



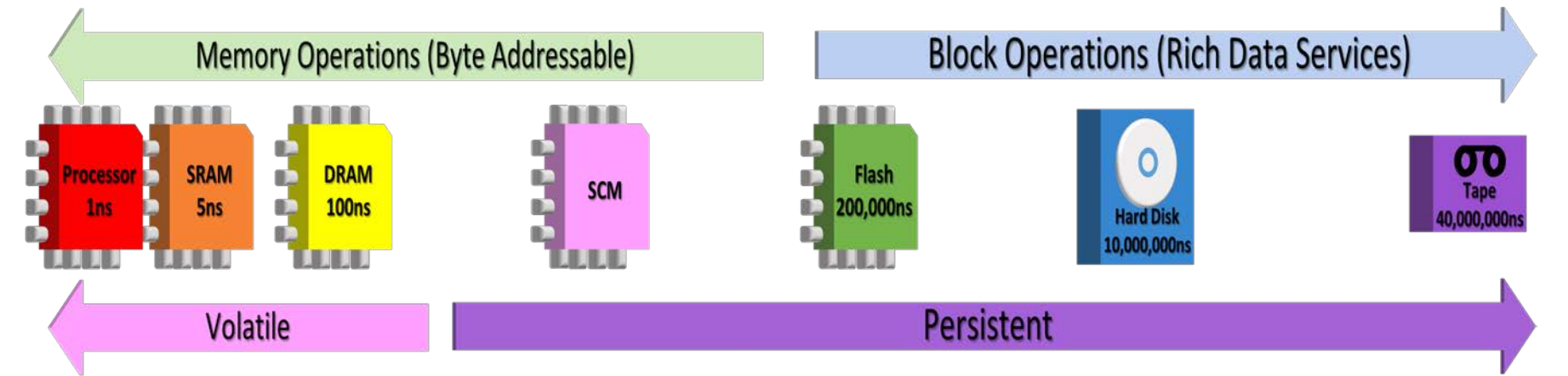
# Compute-Memory Balance is Degrading



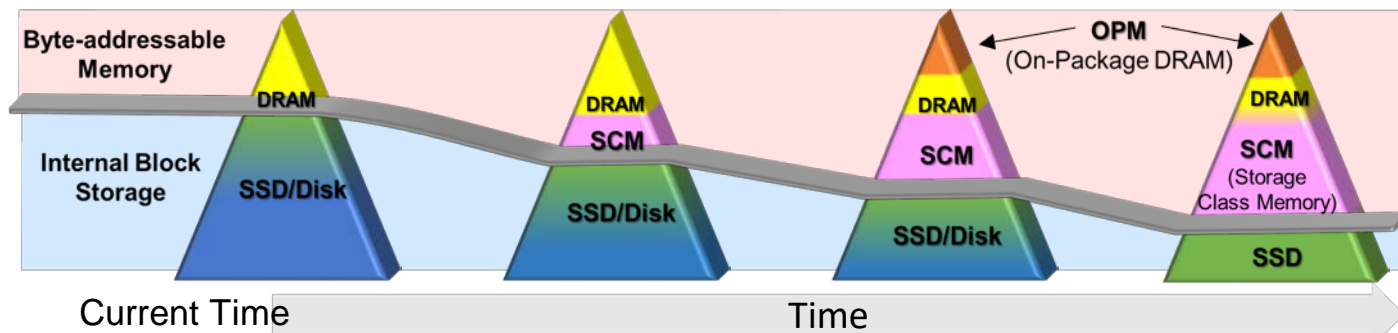
Processor memory and I/O technologies ...

... are being stretched to their limits

# Memory and Storage are Converging



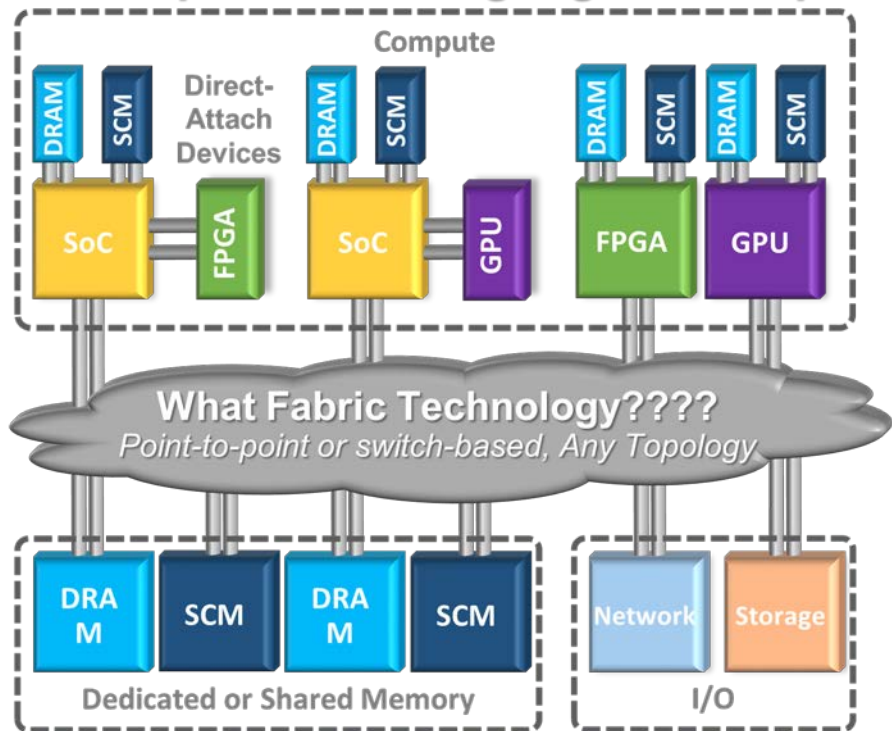
With memory/storage convergence, memory semantic operations become predominant (volatile & non-volatile)



# What's the Solution? Gen-Z!

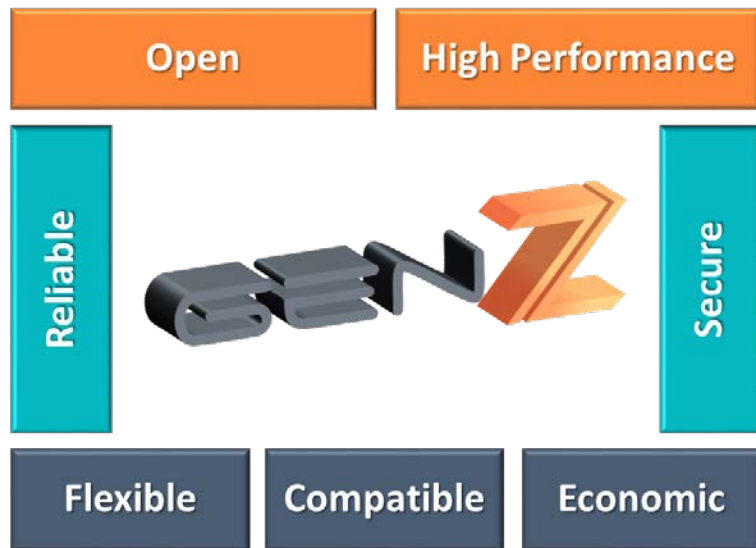
- High Performance
  - High Bandwidth, Low Latency, Scalable
  - Eliminates protocol translation cost / complexity / latency
  - Eliminates software complexity / overhead / latency
- Reliable
  - No stranded resources or single-point-of-failures
  - Transparently bypass path and component failure
  - Enables highly-resilient data (e.g., RAID / erasure codes)
- Secure
  - Provides strong hardware-enforced isolation and security
- Flexible
  - Multiple topologies, component types, etc.
  - Supports multiple use cases using simple to robust designs
  - Thorough yet easily extensible architecture
- Compatible
  - Use existing physical layers, unmodified OS support
- Economic
  - Lowers CAPEX / OPEX, unlocks / accelerates innovation

## Gen-Z speaks the language of compute

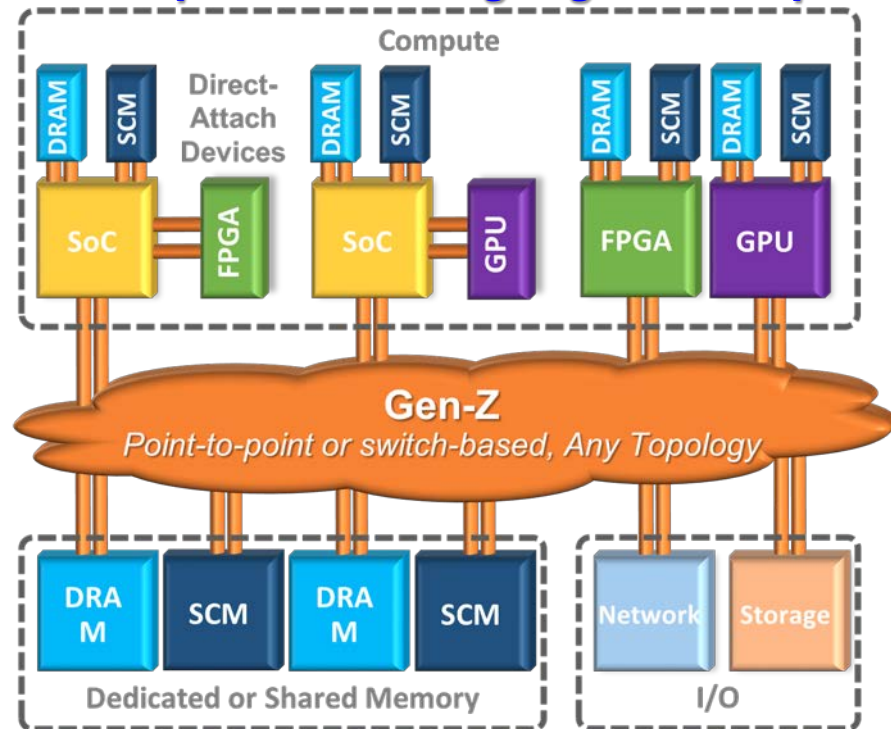




# What's the Solution? Gen-Z!



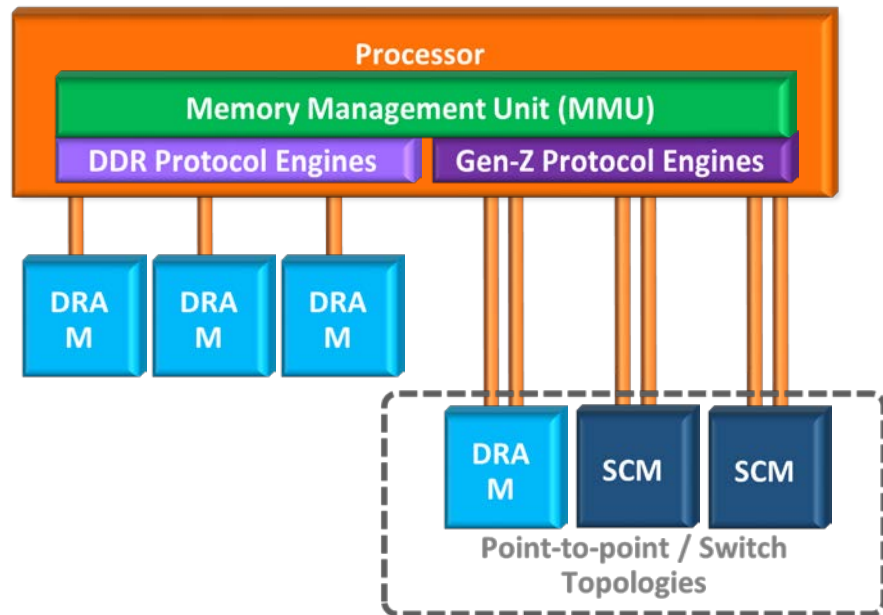
## Gen-Z speaks the language of compute





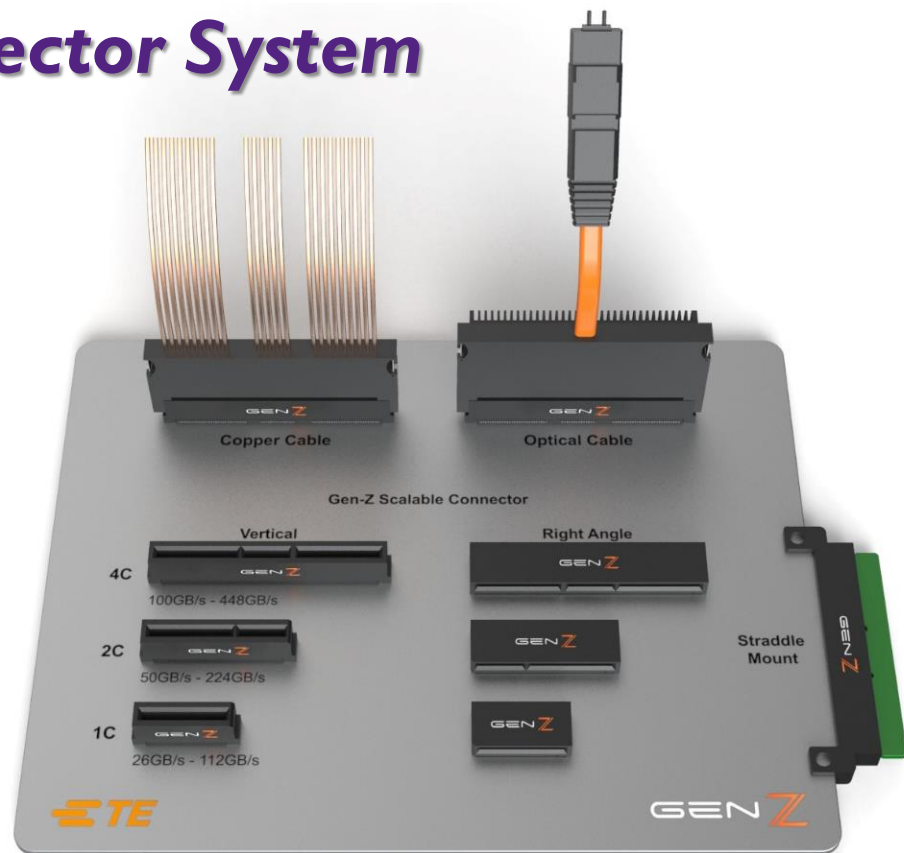
# Example Gen-Z Memory Use Case

- Seamlessly augments DDR / HBM solutions
  - Supports unmodified applications, OS, middleware
  - Load-stores transparently translated into read-writes
- Abstracts media to break processor-memory interlock
  - Accelerates solution agility
  - Creates a virtuous circle of innovation
  - Supports any mix of DRAM, SCM, and NVM media
- Very high bandwidth (16 GT/s to 112 GT/s signaling)
  - Delivers 32 GB/s to 400+ GB/s per memory module
- Supports legacy and new high-capacity form factors
  - 10s GB to multi-TB capacities
- Supports point-to-point and switch-based topologies
  - Scales from single motherboard to rack-scale
- Flattens memory / storage hierarchy



# Flexible: Universal Connector System

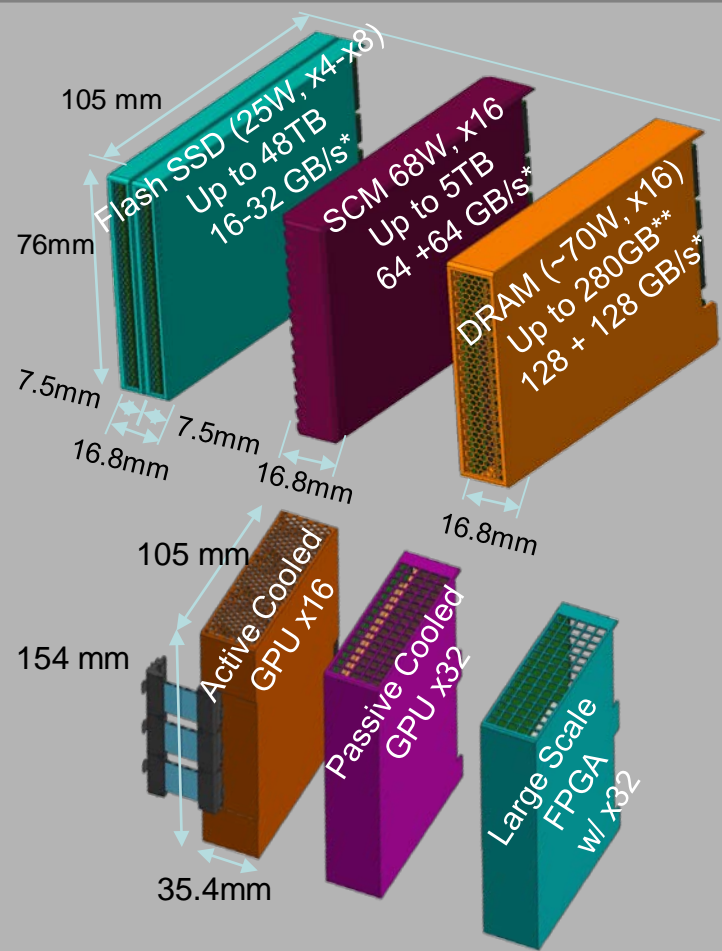
- Vertical, horizontal, right angle, straddle mount
- Same connectors for memory, I/O, storage, etc.
- Cabled solutions: for copper & optical
- Eliminates “hard choices”
  - Universal connector eliminates industry fragmentation
    - Simplifies supply chain—drives volume and lowers cost
  - Any component, any slot, any time
    - Any mix of static and hot-plug form factors
  - Multi-connector option to provide added scalability
    - 80W incremental power
    - Incremental bandwidth
  - Supports internal and external cable applications
    - Enables modular system design
    - Enables system disaggregation
    - Eliminates expensive board materials
  - Multipath—can bifurcate connector into multiple links
    - Aggregate bandwidth, resiliency, no stranded resources
    - Support multiple topologies—point-to-point, daisy-chain, mesh
- Supports multiple interconnect technologies—Gen-Z, PCIe, etc.



Gen-Z members contributed mechanical & electrical specification to SNIA—see SFF-TA-1002 Gen-Z Scalable Connector specification (final version is publicly available) covers remaining functionality.

# Flexible: Scalable Form Factor<sup>1</sup>

- Supports any component type
  - Flash, SCM, DRAM, NIC, GPU, FPGA, DSP, ASIC, etc.
- Supports multiple interconnect technologies—Gen-Z, PCIe, etc.
- Single and double-wide—scale in x-y-z directions
  - Increased media, power, performance, and thermal capacity
  - Double-wide can be inserted into pairwise single slots
- Supports 1C, 2C, and 4C scalable connectors
  - Larger modules can support multiple connectors—scale power & performance
- Scalable Form Factor Benefits:
  - Simplifies supply chain
  - Lower customer CAPEX / OPEX
  - Consistent customer experience
  - Increases solution and business agility @ lower dev cost
  - Eliminates Potential ESD Damage
    - Can safely move modules from failed / old to new enclosure
  - Eliminates SPOF or stranded resources
    - Multiple links per connector, multiple connectors per module
  - Scalable thermal plus improved airflow across components



<sup>1</sup> Draft specification publicly available—see [www.genzconsortium.org](http://www.genzconsortium.org)

\* Bandwidth calculated using 32 GT/s Signaling

\*\* DRAM module provides 3.5x the highest-capacity DDR5 DIMM

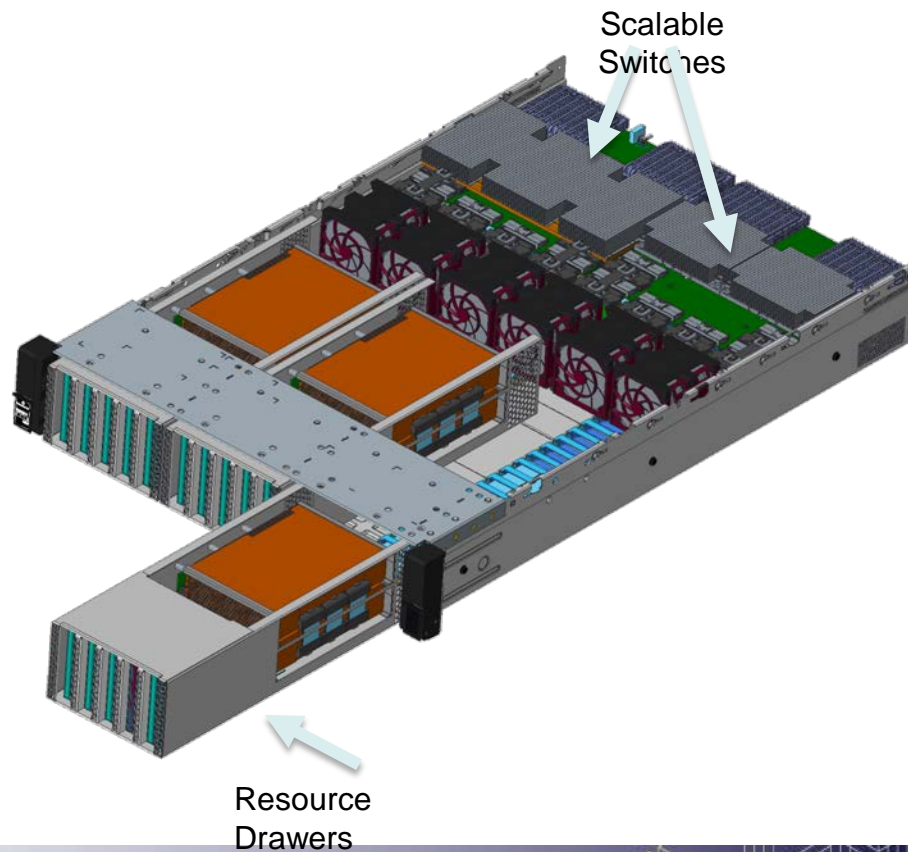
# Flexible: Scalable Resource Enclosures

- **Resource enclosures can support multiple drawers**

- Fixed or hot-plug drawers
- Supports any mix of component types
- Supports memory-centric solution architectures
- Supports data-centric architecture—any mix of media and computational technologies
- Eliminates / reduces need for TOR switches
- Supports multiple interconnect technologies

- **Benefits:**

- Easily augment any server, storage, or network enclosure
- Reduces customer CAPEX / OPEX
- Fully enable / exploit composable infrastructure
  - Fixed or dynamically compose to meet customer workload need
- Eliminates stranded media, stranded communication capacity (e.g., no stranded memory channels / I/O links), etc.
- Eliminates platform design “hard choices”
- Unlocks solution innovation and agility



# Compatible: *Unmodified OS Support*

## Block Storage: Gen-Z SSDs (NVMz)

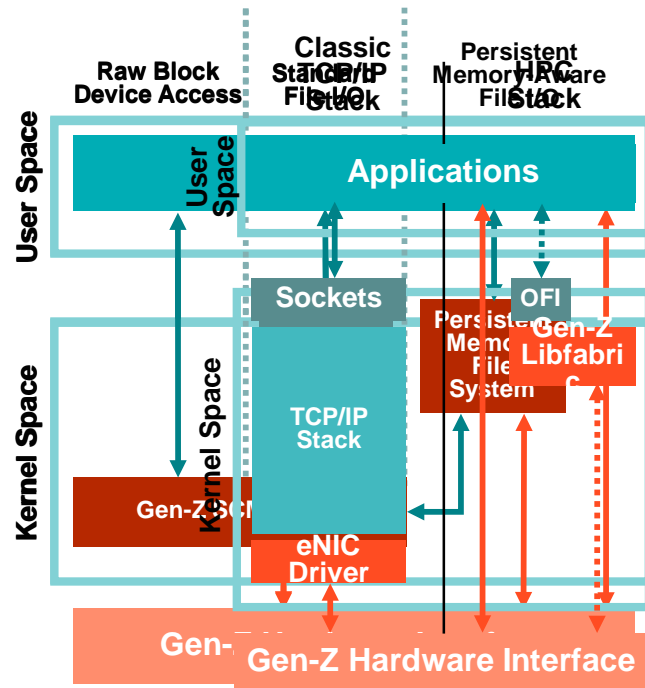
- Block driver is virtually unchanged from NVMe
- Leverage Gen-Z Logical PCIe Device (LPD) support

## DRAM and Storage Class Memory (SCM)

- DRAM appears as DRAM (just like DDR / HBM)
- SCM driver for block & persistent memory stacks
  - Use Persistent Memory File System developed for NVDIMMs

## Messaging: Ethernet & Low Latency

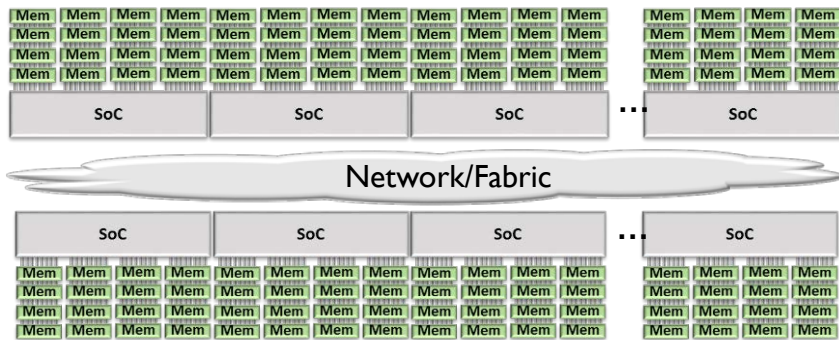
- Emulated NIC (eNIC) driver for traditional Ethernet stacks
  - Tunnels Ethernet packets over Gen-Z
- OFI/Libfabric software for HPC apps and middleware
  - Embedded Reads and Receive Tag support
  - Robust set of collectives and collective accelerator support






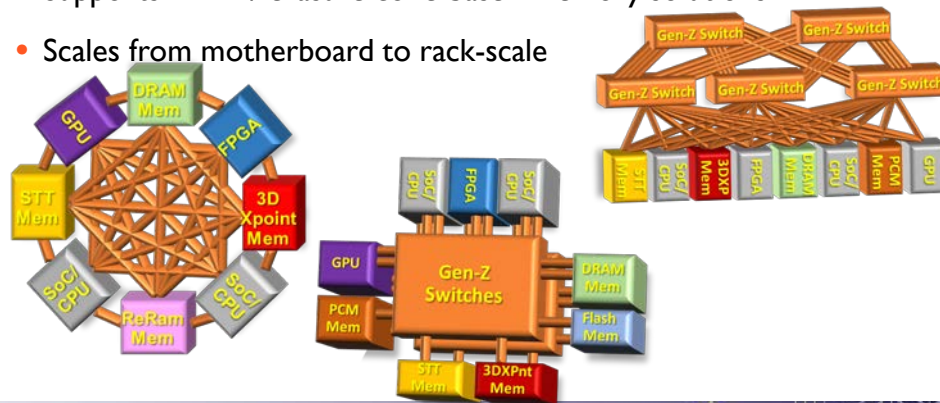
# Today

- Memory is captive of the host device (processor)
  - Stranded memory channels and memory resources
- Can't scale memory independently of processing
- All accesses must traverse host processor



## Gen-Z

- Memory and processing scale independently
  - Heterogeneous compute & memory deployments
  - Direct access to memory devices across fabric
  - Memory can be dedicated or shared by processors
  - Supports up to 64-way barber pole memory interleave
  - Supports RAID / erasure code-based memory solutions
  - Scales from motherboard to rack-scale
- 
- A small icon of a Gen-Z Switch, which is a yellow rectangular block with the text "Gen-Z Switch" in black, resting on a brown base that resembles a circuit board or a stack of components.

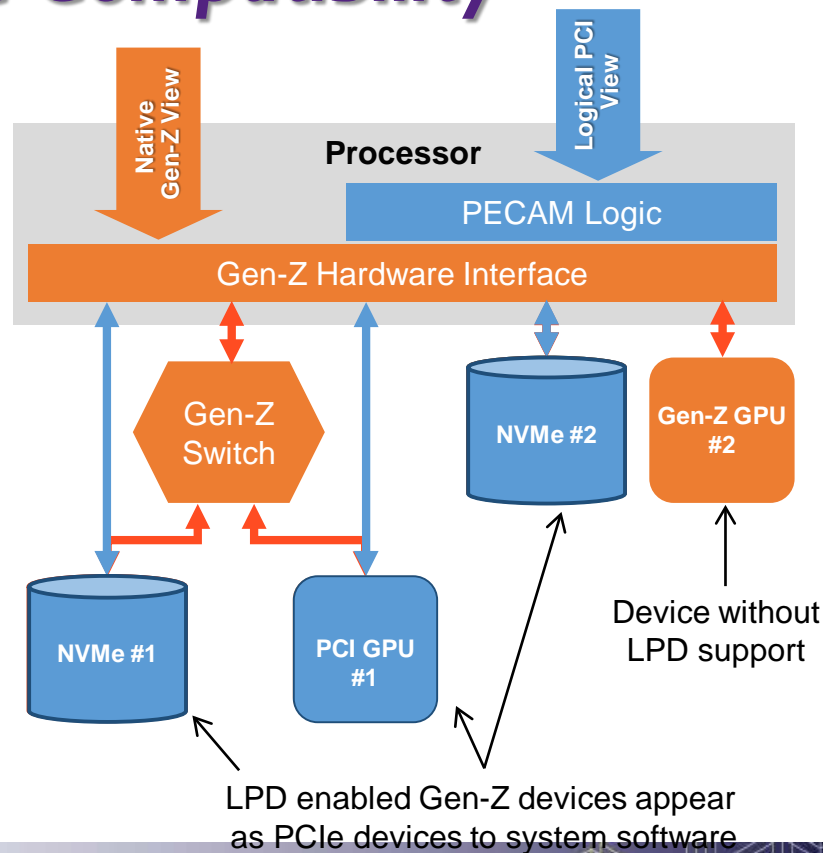




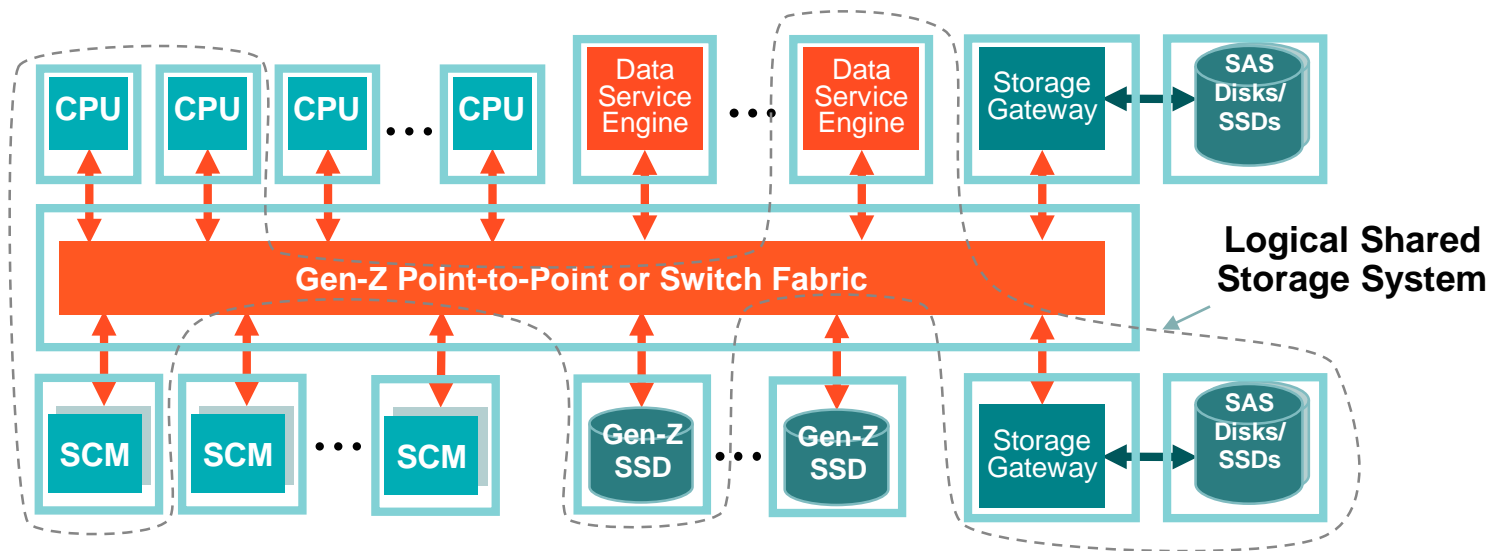
# Composable I/O with PCIe Compatibility

## Gen-Z Logical PCI Devices (LPDs)

- Gen-Z devices can be discovered/configured
  - Via standard PCI / PCIe system software
- LPDs can fully exploit Gen-Z Architecture
  - Low-latency switching
    - Gen-Z 30 ns vs. PCIe 130-150 ns translates to 200-240 ns savings per read operation
  - Memory-speed CPU-to-device communication
  - Security and fine-grain hardware-enforced isolation (any-to-any communication without compromise)
  - Supports all x86 / ARM / Power architecture Atomics
  - Simplified single and multi-host I/O virtualization and sharing capabilities
  - Multipath—aggregation / resiliency / robust topologies
  - PCIe 2.5-32 GT/s PHY and 25-112 GT/s 802.3 Electrical
  - Legacy plus New Gen-Z Scalable Connector and Scalable Form Factors
  - CPU-based data movers to enable new software paradigms
  - Scale-up and scale-out connectivity and performance
  - Simplified software—any mix coherent and non-coherent operations
  - And much more...



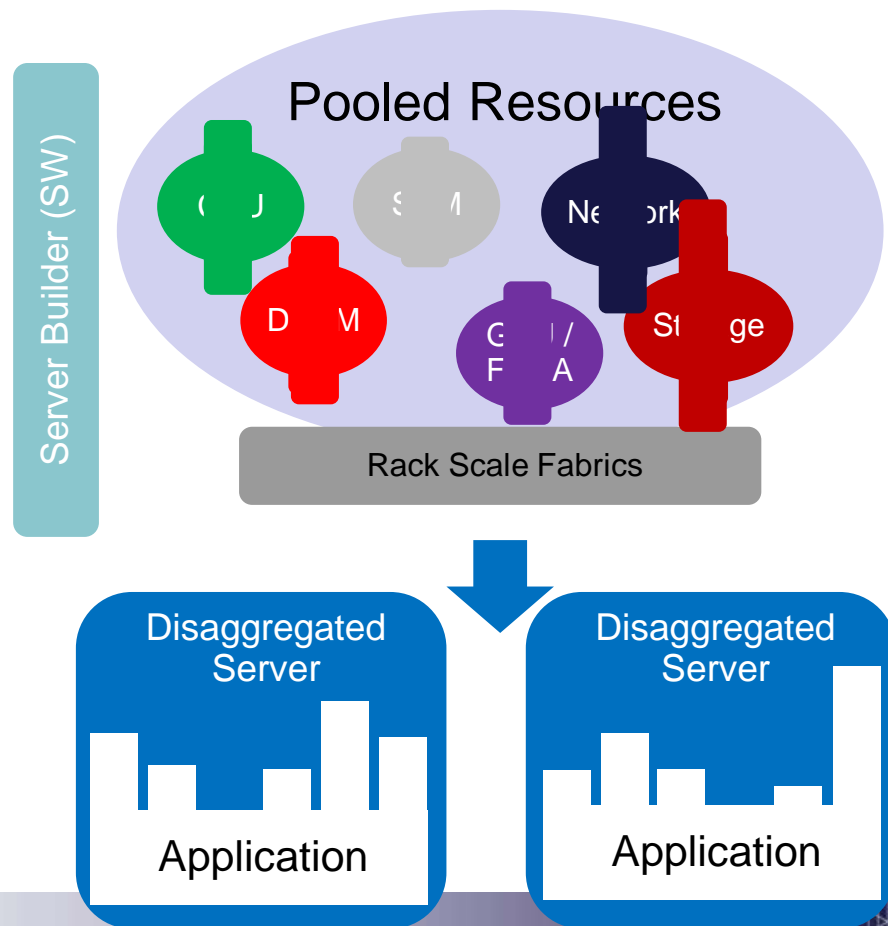
# Composable Storage



- Logical storage systems composed from components on Gen-Z fabrics
- Supports: Object storage, Key-Value Stores, Storage Arrays (small/large), etc.
- Supports Rich Data Service Accelerator components
  - RAID, dedup, replication, compression, thin provisioning, encryption, etc.

# Server Disaggregation

- All resources are collected into shared pools
- High-speed, low-latency fabric connects pools
- Management software:
  - Configures network to connect components
  - Assigns resources
- Result:
  - Disaggregated server
  - True bare-metal bootable server
  - Ready for installation of any OS and application

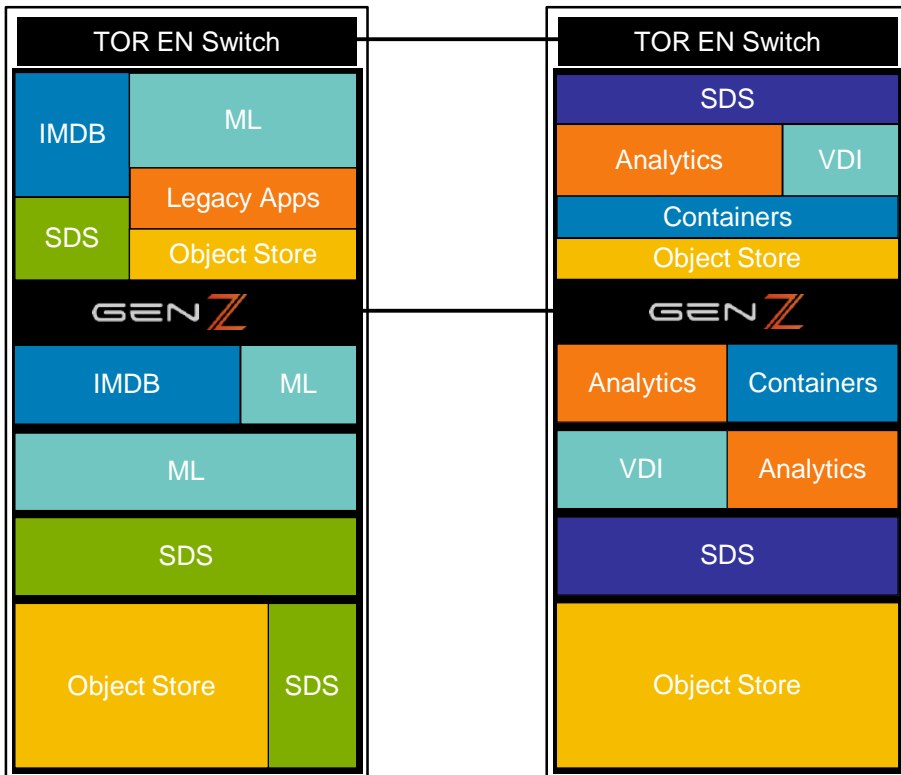


# The ideal end state with composability

For customers, finally capitalize on the promise of “pay-as-you-go” model of IT.

For end users, dynamically adjust IT resources as the business needs fluctuate.

For IT providers, efficiently orchestrate business needs without physical setup or reconfiguring of resources.



# Disaggregated infrastructure benefits



Compose servers with  
resources app requires



Unlock trapped  
resources



Avoid overprovisioning



Purchase resources  
independently

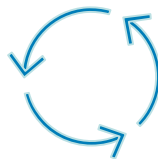
## INCREASE AGILITY

## OPERATE EFFICIENTLY

## UNLOCK VALUE



Increase RAS



Repurpose retired  
resources



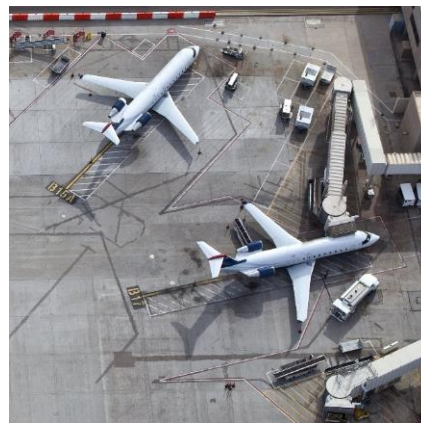
Technologies can  
evolve – and deployed  
independently

# Gen-Z Transforms Performance

Modify existing  
frameworks

New algorithms

Completely rethink



In-memory analytics

Similarity search

Large-scale  
graph inference

Financial models

**15x**  
faster

**20x**  
faster

**100x**  
faster

**8,000x**  
faster



# *Communication at the Speed of Memory*

