



HYPERION RESEARCH

Considerations for Storage and Interconnects

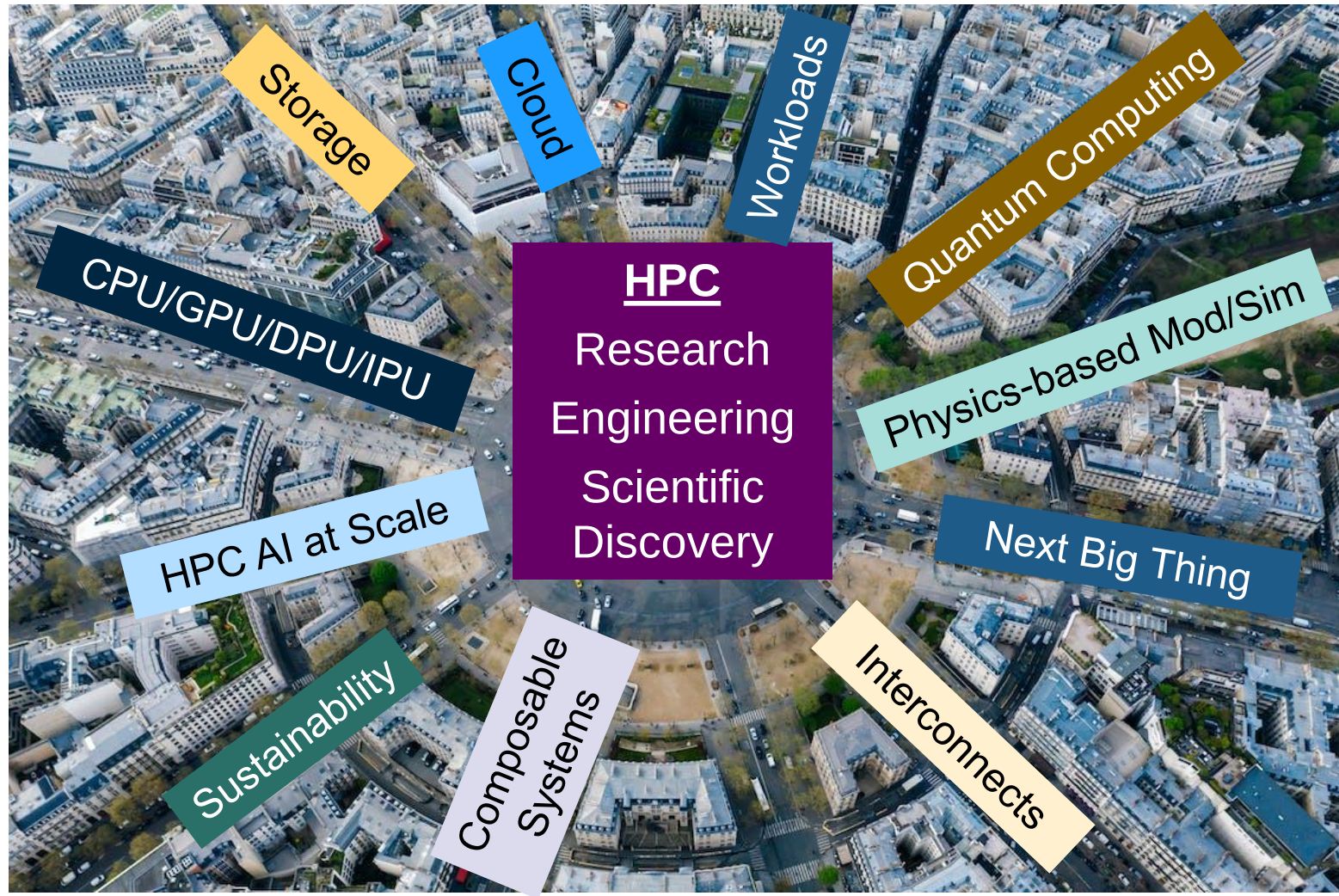
ISC24 Market Update Briefing
May 2024

Mark Nossokoff

www.HyperionResearch.com
www.hpcuserforum.com

Not Your Father's HPC

A Busy Intersection of Complex Challenges

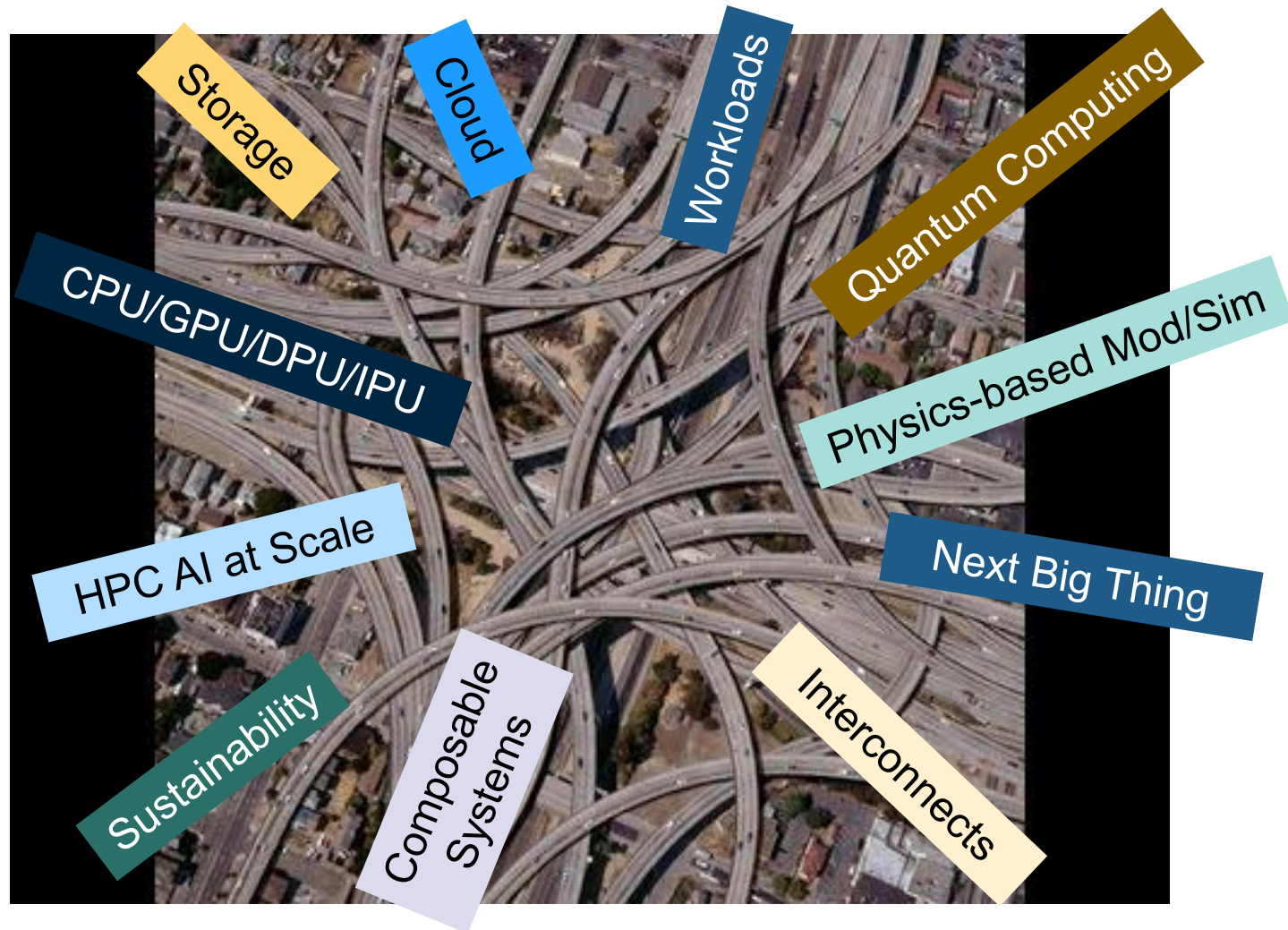


Not Your Father's HPC... ..Now Add AI

An Almost Untenable Increase in Complexity

HPC + AI

Research
Engineering
Scientific
Discovery
Business
Analytics
LLMs
SLMs
Multimodal
GenAI

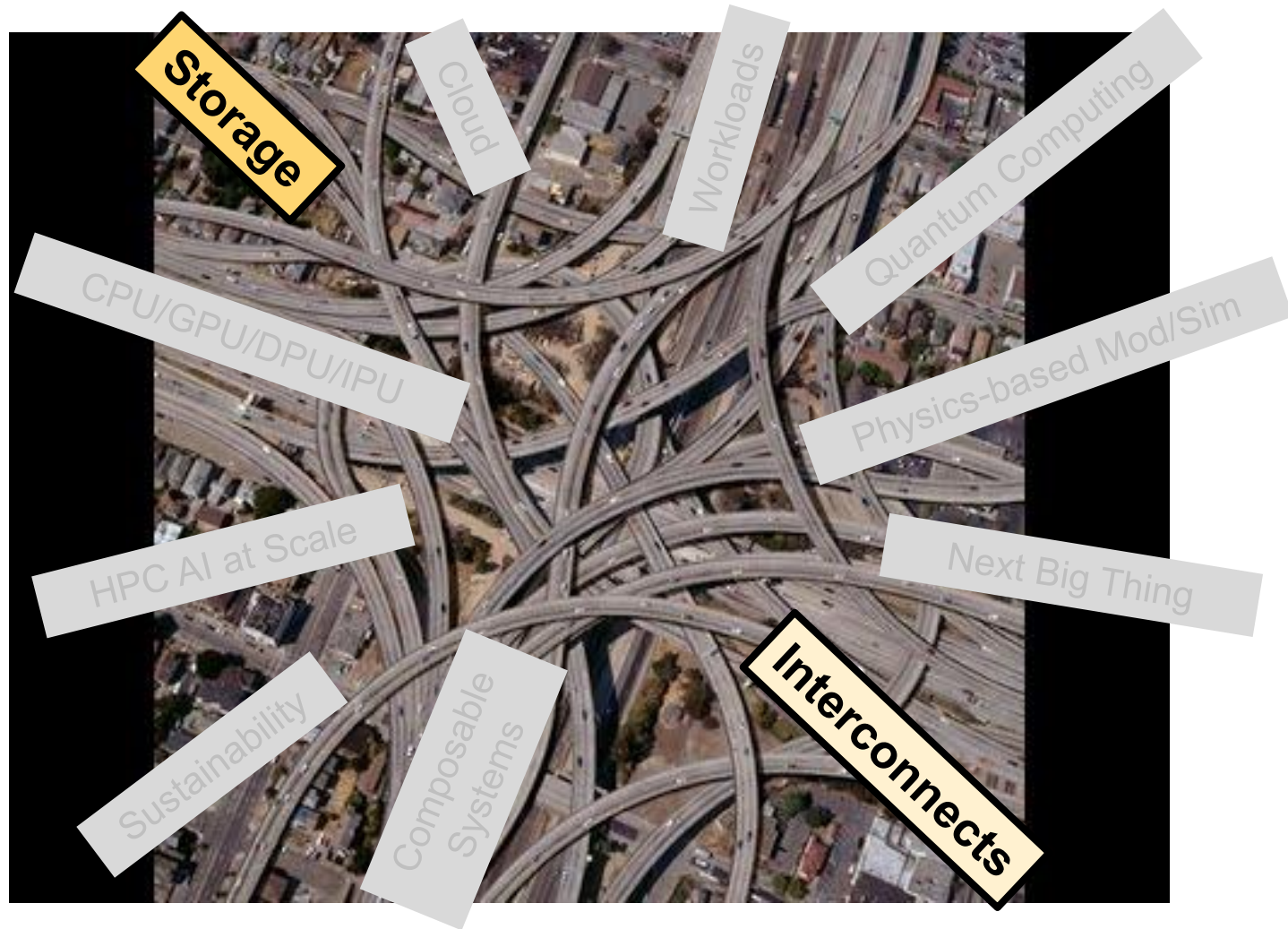


Not Your Father's HPC... ..Now Add AI

An Almost Untenable Increase in Complexity

HPC + AI

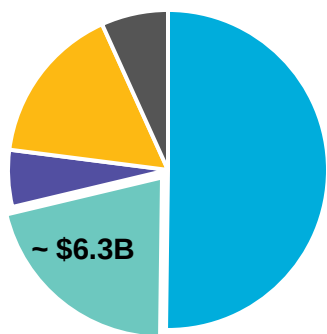
Research
Engineering
Scientific
Discovery
Business
Analytics
LLMs
SLMs
Multimodal
GenAI



HPC/AI Storage Growth Continues

Demand increasing across all sectors and verticals

On-premises HPC/AI Spend - 2023
Total 2023 HPC/AI Spend - \$29.8



■ Server ■ Storage ■ Middleware ■ Applications ■ Service

Source: Hyperion Research, 2024

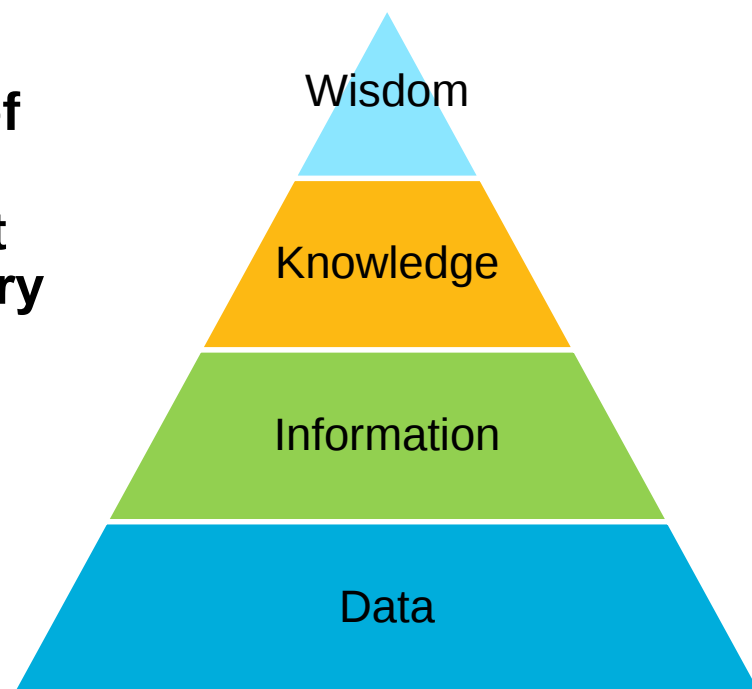
- **Storage continues as highest growth HPC/AI element**
- **Storage represents ~ 21% of on-premises HPC spending and growing**
- **Impact of AI-related requirements and demand still evolving, but anticipated to further increase storage spending**

Area (\$M)	2023	2024	2025	2026	2027	2028	CAGR 23-28
Server	\$14,954	\$16,254	\$18,058	\$19,501	\$21,025	\$22,152	8.2%
Add-on Storage	\$6,282	\$6,972	\$7,862	\$8,552	\$9,288	\$9,786	9.3%
Middleware	\$1,711	\$1,836	\$2,014	\$2,160	\$2,329	\$2,454	7.5%
Applications	\$4,830	\$5,152	\$5,634	\$6,020	\$6,457	\$6,805	7.1%
Service	\$2,014	\$2,050	\$2,167	\$2,186	\$2,334	\$2,458	4.1%
Total Revenue	\$29,791	\$32,265	\$35,735	\$38,420	\$41,433	\$43,656	7.9%

Data Platform Emerges as Key Storage Value Point

While “speeds and feeds” will continue to be important to buyers of storage systems, the primary value point and competitive advantage for data storage solutions will shift to the “Data Platform”

- **Predictable cadence of bandwidth, throughput, and latency performance improvements with each generation of storage system are base table stakes**
- **Speeds and feeds alone do not reflect the value of a storage system’s primary asset – the data**
- **Long term storage system business success will be driven by a data platform:**
 - Providing users their data wherever, whenever, however in a performant, reliable manner
 - Delivering capabilities to users to derive value from the data they feed into their scientific, engineering, and business workloads

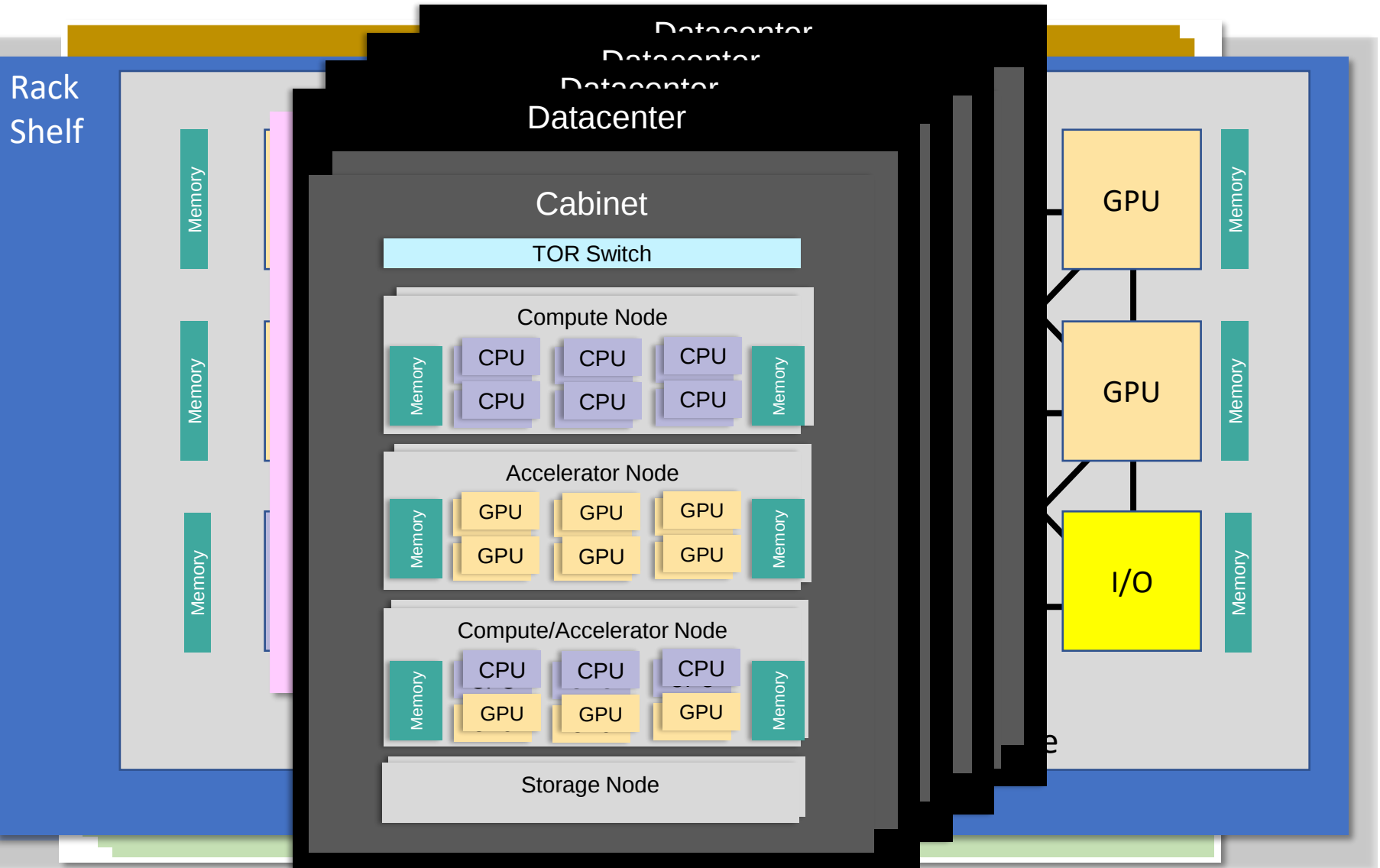


Evolving File System Landscape

AI workflows stressing traditional HPC file systems

- **Incumbent file systems developed for traditional HPC mod/sim being challenged by modern file systems designed with AI in mind from the start**
 - Compute-intensive vs. data-intensive
 - IO profiles (large block sequential vs. small block random)
 - Access methods (file vs. block vs. object)
 - Access frequency (hot vs. archive vs. cold)
 - Locality (centralized datacenter vs. cloud vs. edge)
- **Current file system adoption preferences**
 - Traditional: BeeGFS, Lustre, NFS, OneFS, PanFS, Spectrum Scale
 - Emergent: Hammerspace, VAST, Weka
- **Software business models support broader range of solution choices**
 - Hammerspace, VAST, VDURA (formerly Panasas), Weka

Hierarchy of Networks*



Implications to Interconnect Use Cases

Different interconnect technology options for each use case

- **Ethernet evolving to support current use cases**
- **Protocols critical to satisfying feature and performance requirements**
 - CXL
 - NVMe, NVMeoF
 - Ultra Ethernet
 - Vendor-unique enhancements
 - RDMA, RoCE
 - iSER
 - iWarp
 - SRP
- **Users increasingly want choice in both technology and vendor**

Interconnect Technologies	Use Cases (current and aspirational)					
	Rack to Rack	Node to Node	Node to Storage	xPU to xPU intranode	xPU to Memory	Chiplet to Chiplet
Ethernet	✓	✓	✓	✓		
Infiniband	✓	✓	✓			
Omnipath	✓	✓	✓			
Slingshot	✓	✓	✓			
NVLink		✓		✓		✓
InfinityLink				✓		✓
PCIe			✓	✓		
UCIe				✓		✓
HBM					✓	
OCP BoW					✓	✓
NuLink				✓	✓	✓

What's next for HPC/AI Storage and Interconnects?

Advanced computing architectures will continue to stress existing interconnect capabilities

- **Continued cadence of speeds and feeds**
- **Optimized data platforms to satisfy evolving AI workflow demands**
 - Locality: on-premises, cloud, edge
 - Business model: increasing adoption of SW-only solutions
- **Standards evolution that address new realities will seek industry adoption**
 - Ultra Ethernet
 - UCle
 - Sovereign EuroHPC interconnect

Questions?



**We welcome questions,
comments, and suggestions**

**Please contact us at:
mnosokoff@hyperionres.com**