



HYPERION RESEARCH

Perspectives on HPC-AI Storage and Interconnects

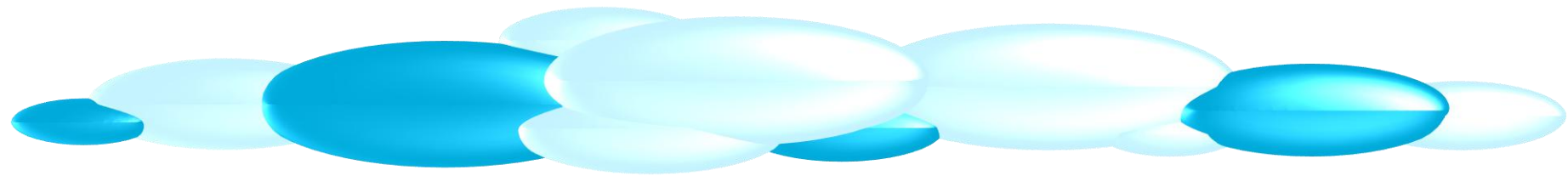
ISC25 Market Update Briefing
June 2025

www.HyperionResearch.com
www.hpcuserforum.com

Mark Nossokoff

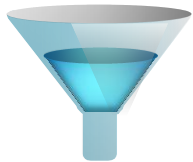
AI Data Pipeline

Diverse range of profiles and requirements



Prep

Checkpoint



Ingest

Train

Inference

AI Data Pipeline Storage Considerations

	Ingest	Prep (ETL)	Train	Checkpoint	Inference
Access Pattern	Sequential	Sequential or Random	Random	Sequential	Sequential
Access Type	Writes	Reads and Writes	Reads	Writes	Reads
Access Frequency	Idle $\leftrightarrow$ Intense	Moderate	Idle $\leftrightarrow$ Intense	Idle $\leftrightarrow$ Intense	Moderate to Intense
Data Size	Small to Large	Small to Large	Mostly Small	Small to Large	Small to Large
Locality	Edge	Edge, Cloud, On-premises	Cloud, On-premises	Cloud, On-premises	Edge, Cloud, On-premises

*ETL – Extract, Transform, Load
Source: Hyperion Research, 2024

- **Training frequency (new foundation, RAG, pre-trained)**
- **Model type and size**
- **Data type (structure, unstructured; file, block, object)**
- **Data mode (text, image, video)**
- **Security**
- **Compliance (what data to save and for how long?)**
- **Parallel file system – is one a requirement?**

What's Happened in Storage and Interconnects Since SC24?

Storage

- **Blackstone \$300M investment in DDN**
- **Vendor announcements**
 - System vendors
 - Dell
 - HPE
 - Data platform vendors
 - DDN
 - Hammerspace
 - VAST
 - Weka

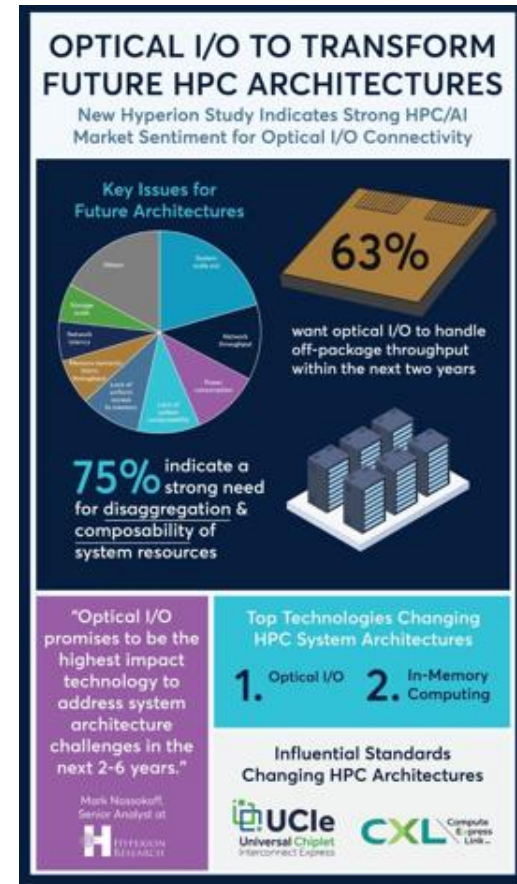
Interconnects

- **GTC items**
 - On-chip optics
 - Heavy emphasis on ethernet; limited mention of InfiniBand
- **Progress in standards**
 - UltraEthernet Consortium (UEC)
 - Contributions to Linux kernel
 - UltraAccelerator Link (UALink) Consortium
 - Released version 1.0 of spec
 - EuroHPC JU NET4EXA
 - Deploy BXIv3 exascale integration
 - Intent to develop BXIv4 roadmap
- **NVLink Fusion for 3rd party integration**

Strong Sentiment Toward Optical I/O

Both users and vendors eagerly anticipating optical I/O

- **Predominant system issues for future architectures**
 - System scale-out
 - Lack of system composability
 - Network throughput
- **Optical I/O was rated as technology that has highest potential to improve HPC architectures in next 2-6 years**
- **75% of respondents felt that there is a strong need for disaggregation of system resources**



Graphic: Ayar Labs, 2023

Prior
Study

Future Research Direction

The more things change, the more they stay the same

- **Continued impact of the AI data pipeline on storage architectures**
- **Evolution of data platform SW stack**
- **AI workflow impacts on interconnect architectures**
- **Evaluating and determining optimized utilization of on-premises and cloud storage resources**
- **Maturity and adoption of optical interconnects**
- **Convergence or differentiation between interconnects (InfiniBand, Ethernet, OmniPath, BXI) as a result of standards activities (UEC, UAL) and changes in vendor strategy (NVIDIA NVLink Fusion)**

Questions?



mnossokoff@hyperionres.com