



Hewlett Packard
Enterprise

Hardware Trends for Large-Scale Scientific Data Analysis (LSSDA)

Larry Kaplan
Senior Distinguished Technologist

October 2025

System Level Trends

Heterogeneity is on the increase

- Driven by desire for efficiency across different computational challenges
- Not only differences in required mathematical precision, also memory BW/capacity, networking, local storage, etc.
- Computations can also occur in multiple places (edge, data center, cloud) with appropriate connectivity

Unified physical memory architectures provide advantages

- Ease on-node data movement requirements and enhance programmability
- Direct access from both CPUs and accelerators
- Tiered memory systems can also help with efficiency but often require software assistance
- On-node storage for some use cases

Need for tightly coupled networks

- Across heterogeneous partitions, systems, and sites
- Performance concerns for network bandwidth and latency
- Provide security and network segmentation (e.g. VLANs)
- Also support for high-bandwidth, low-latency and/or, time-critical data ingest

Computational Component Trends

- 64-bit floating point (FP) hardware not particularly important for LSSDA, getting less industry attention
 - Issues of power, cooling, and reliability in addition to contravening market forces
- AI techniques taking on a significant amount of the analysis workload
 - Not just LLMs, also various forms of NNs
 - Issues: obtaining training data (real world and/or simulated), retraining frequencies, understanding error bounds
- Lower precision FP support getting more attention in the industry
 - GPU industry strongly headed in this direction due to the size of the AI market
 - Special-purpose devices also support this path, and are also being used in places for traditional computation
 - Mixed- and low-precision computation tend to be algorithm and domain specific – numerical analysis required
- Progress in low-precision support expected to move quickly, especially for AI training and inference
 - ISA and other architecture improvements (e.g., TPUs) also helping here
- Several speculative technologies being pursued, including:



Quantum

Big potential gains, but not broadly applicable across the algorithm and application space

Analog

Potentially fast and accurate, avoids discretization issues

Superconducting

Uses different “switching” technique than CMOS, lower energy consumption

Neuromorphic

Uses several different implementations to mimic the brain, potentially most relevant

- All have integration challenges to traditional techniques, including for software
 - Not just programmability, but also communication/conversion, debugging, and tuning



Thank you

