DeiC HPC TekRef group report on the

Supercomputing 2024 conference

Atlanta, Georgia, USA, November 2024.

Participants:

- Aputsiak Niels Janussen <apja@di.ku.dk>
- Hans Henrik Happe <happe@science.ku.dk>
- Jakob Bruun Hansen <jabh@adm.ku.dk>
- Joachim Sødequist <joachim.sodequist@deic.dk>
- Nina Simone Marstrand Sander Aagaard <nsmsa@its.aau.dk>
- Ole Holm Nielsen <ole.h.nielsen@fysik.dtu.dk>
- Per Ankerstjerne <peraaa@dtu.dk>
- Rainer Bohm <rb@its.aau.dk>
- Martin Rehr <rehr@science.ku.dk>
- Rasmus D. Jensen <rj@its.aau.dk>
- Rune Gamborg Ørum <rune.oerum@deic.dk>
- Tor Skovsgaard <tor.skovsgaard@deic.dk>
- Pietro Bortolozzo <pbor@dtu.dk>
- Eske Christiansen <eske.christiansen@deic.dk>

Preface

The DeiC **HPC TekRef** group, represented by a delegation of *High Performance Computing* (HPC) system administrators from several Danish universities (KU, AAU, and DTU) as well as DeiC participated in the *SC24 International Conference for High Performance Computing, Networking, Storage, and Analysis*[1] annual conference in November 2024 in Atlanta, Georgia, USA.

This report summarizes the fact-finding investigations carried out by the delegation during the conference for the purpose of obtaining and accumulating knowledge about the latest HPC systems, technology, and software for use by DeiC as well as the broader HPC community in Denmark.

The delegation attended a number of conference sessions, visited the conference showfloor with a record number of exhibitors, and in addition held a number of prearranged one-to-one meetings with key technology vendors under *Non-Disclosure Agreements* (NDA). The vendor list was Intel, AMD, NVIDIA, DellEMC, HPE, Lenovo, Supermicro, and Cornelis Networks. The information obtained under NDA terms cannot be disclosed in the present public report where only publicly available information is reported.

Important overall trends

The topics in the following sections represent the delegation's view of the most important trends in HPC, with a particular emphasis on HPC in Denmark. The sections are divided into a number of separate technology topics.

We have noticed some significant changes in technologies as compared to the previous Supercomputing conferences, including:

1. A much stronger focus among many vendors on AI technologies such as GPUs for AI, Machine Learning, and GPU computing.
2. Traditional HPC using CPU computing is advancing rapidly with a number of faster multi-core CPUs on the market now and more products in the vendor roadmaps.
3. Electrical power consumption for both GPU and CPU technologies is skyrocketing along with the performance improvements, leading to much larger electrical power requirements for datacenters. Whereas 10-20 kW of power per rack was commonplace some years ago, presently state-of-the-art GPU racks require up to about 100 kW of power, and some vendors even projected 200-250 kW per rack in the foreseeable future! This will obviously put severe strains on

[1] SC24 https://sc24.supercomputing.org/

datacenter infrastructures in the coming years and lead to significant increases in operational costs.

4. Cooling the power-hungry racks is more important than ever! While it is still possible to air-cool the latest GPUs and CPUs in tall server cabinets by using huge heatsinks and fans, all vendors are now promoting the use of direct liquid cooling of the hottest chips in the system. Liquid cooling infrastructure will become standard in datacenters going forward, however, there are currently no indications of liquid cooling standardization and interoperability among vendors. Many vendors are now offering 75 cm wide and 48U (215 cm) high racks for the Hyperscaler[2] market which require additional datacenter floor space and sufficient vertical space.
5. On the showfloor we did not notice any HPC vendors from China.

Processors for HPC

High-performance processors (CPUs) are crucial for most types of HPC applications. The delegation learned about current products as well as roadmaps for future products from the following vendors.

AMD

AMD launched the 5th Generation AMD EPYC processors codenamed "Turin"[3] in October 2024, with up to 192 "Zen 5" cores per socket and a TDP[4] power of up to 500 Watt. The processors have 12 memory channels with DDR5-6000 RAM memory. The "Zen 5" cores can execute AVX-512 vector instructions with a full 512-bit data path, making it equivalent to Intel's AVX-512 implementation.

The delegation received NDA information on future AMD EPYC "Venice" products with "Zen 6" cores, and also on products to be announced beyond "Venice".

Intel

Intel has announced the Xeon 6 processor family.[5] The high-performance "P-cores" offer the best solution for compute-intensive, vector-based workloads and up to 128 cores per socket. The alternative CPU microarchitecture "E-cores" offer up to 288 cores with optimal performance-per-watt, but the "E-cores" do not support the AVX-512 and AMX vector operations, which are very important for most HPC applications. The Xeon 6

---

[2] Hyperscaler: https://www.google.com/search?q=hyperscaler
[3] AMD Turin: https://www.amd.com/en/products/processors/server/epyc/9005-series.html
[4] TDP power: https://en.wikipedia.org/wiki/Thermal_design_power
[5] Intel Xeon 6: https://www.intel.com/content/www/us/en/products/details/processors/xeon/xeon6-product-brief.html

processors support DDR5 6400 high-speed Memory, and also support the new *Multiplexed Rank DIMM* (MRDIMM) memory at up to 8800 MT/s.

Intel AMX provides up to 16x more multiply accumulate (MAC) operations than Intel AVX-512 for BF16- and FP16-based models to enhance AI performance (a “P-core”-only feature).

The delegation received NDA information about the “Diamond Rapids” processor generation roadmap.

ARM

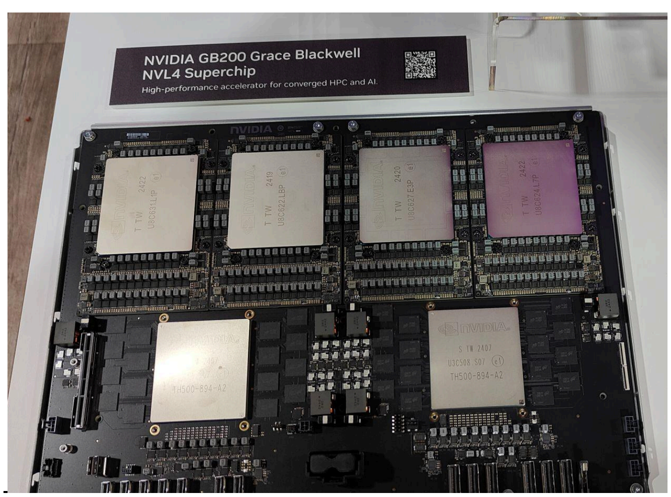
NVIDIA has offered high-performance “Grace” ARM-based CPUs for a while and continues to do so in the latest AI/HPC systems.

In addition, the Fujitsu A64FX ARM-based CPU is powering the “Fugaku” supercomputer which currently holds place no. 6 on the TOP500 list.

Accelerators

NVIDIA

The *Grace Hopper GH200 Superchip*[6] as well as *Blackwell*[7] public information which includes a Technical Brief paper was described in an NDA meeting with NVIDIA. The future *Blackwell Ultra* (B300) and *Rubin* architectures were discussed as well. The launched 8-GPU DGX B200 system was described, and it was available from a number of vendors on the showfloor.



The GB200 NVL72 system[8] implementation was on display from various vendors. One example was the Dell IR7000 rack with PowerEdge XE9712 nodes described below.

On software topics the CUDA-X[9] GPU-accelerated microservices and libraries for AI were discussed.

[6] GH200: https://www.nvidia.com/en-us/data-center/grace-hopper-superchip/
[7] Blackwell: https://www.nvidia.com/en-us/data-center/technologies/blackwell-architecture/
[8] GB200 NVL: https://www.nvidia.com/en-us/data-center/gb200-nvl72/
[9] CUDA-X: https://www.nvidia.com/en-us/technologies/cuda-x/

Features of the new cuPyNumeric[10] Python library were discussed.

NVIDIA is offering the CUDA-Q[11] software platform as the QPU-agnostic platform for accelerated quantum supercomputing. Amongst the CUDA-Q offerings is also a Jupyter based container[12] designed for academic purposes.
NVIDIA has launched the ALCHEMI[13] AI-accelerated chemical and materials discovery workflow software.

AMD

AMD has announced the new Instinct™ MI300[14] series accelerators, following the Instinct MI200 series which has been installed in HPC systems such as LUMI (see footnote 37).

The AMD Instinct™ MI325X GPU accelerator contains 3rd Gen AMD CDNA™ architecture including 256 GB HBM3E memory and 6 TB/s bandwidth. The MI300X accelerator has a slightly smaller 192 GB HBM3E memory. It is still possible to air cool these systems with huge heat sinks and fans (see photo).

The Instinct MI300A accelerated processing units (APUs) contain both 24 "Zen 4" x86 CPU Cores as well as Instinct accelerators. The MI300A is used in the latest TOP500 no. 1 Exascale system "El Capitan"[15].

The delegation also received NDA information on future AMD Instinct accelerators.

[10] cuPyNumeric: https://developer.nvidia.com/cupynumeric
[11] NVIDIA CUDA-Q: https://developer.nvidia.com/cuda-q
[12] CUDA-Q Academic: https://github.com/NVIDIA/cuda-q-academic
[13] ALCHEMI: https://developer.nvidia.com/blog/revolutionizing-ai-driven-material-discovery-using-nvidia-alchemi/
[14] AMD MI300: https://www.amd.com/en/products/accelerators/instinct/mi300.html
[15] El Capitan: https://top500.org/system/180307/

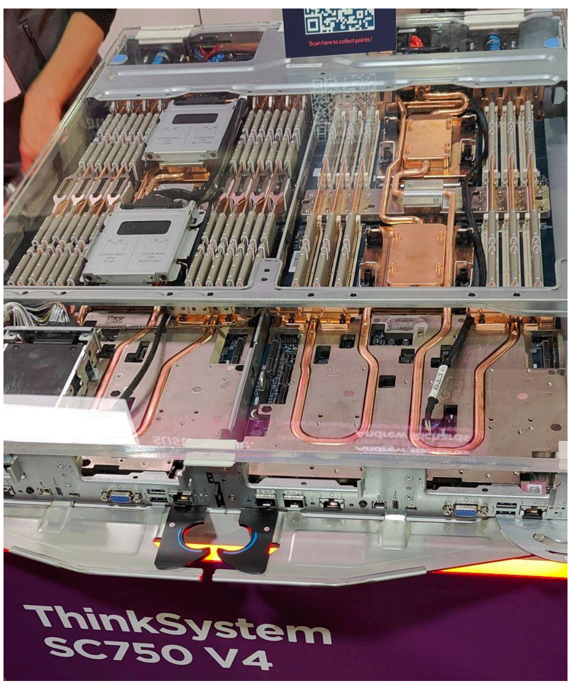
Intel

The Intel *Gaudi 3 AI Accelerator*[16] is built to handle demanding training and inference. The *Gaudi 3* software portfolio[17] includes PyTorch. The delegation received NDA information about the "Shores" family of accelerators. The next-generation *Falcon Shores* GPU is rumored to become available in late 2025. The follow-up *Jaguar Shores*[18] may be launched in 2026.

Server products

The delegation met with several of the major server vendors to learn about the latest products as well as roadmaps for future products.

Lenovo



Lenovo has recently announced the next-generation **N1380[19] water-cooled "Neptune" chassis** which was demonstrated on the SC24 show floor and in an NDA session. The next generations of servers will contain larger CPUs with more memory DIMMs that take up more space on the system board. GPUs are growing in size as well. Lenovo's approach is to mount 8 servers vertically in the N1380 chassis so that up to three 13U chassises can be mounted in an industry standard 42U rack with a 600 mm width.

The rear of the N1380 chassis contains up to four 15 kW *Power Conversion Stations* (PCS) with 3-phase AC power inputs, eliminating the need for traditional *Power Distribution Units*.

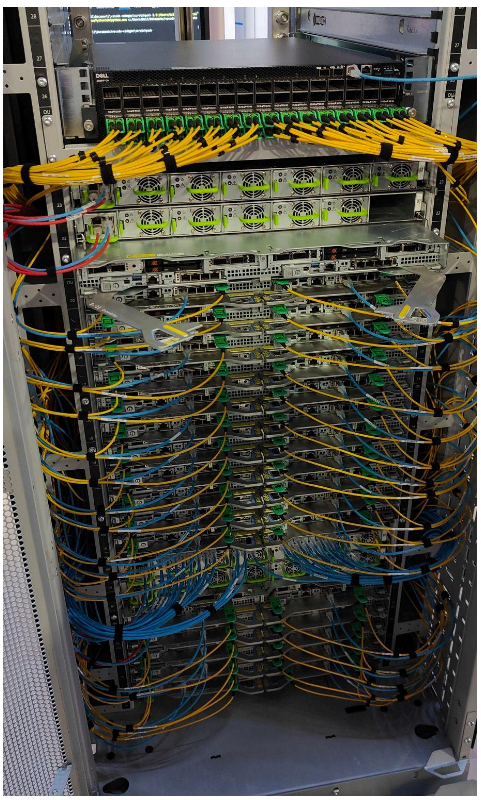
[16] Intel Gaudi 3: https://www.intel.com/content/www/us/en/products/details/processors/ai-accelerators/gaudi.html
[17] Gaudi software: https://www.intel.com/content/www/us/en/developer/platform/gaudi/overview.html
[18] Jaguar Shores: https://www.hpcwire.com/2024/11/19/intel-names-jaguar-shores-as-its-next-generation-ai-chip/
[19] Lenovo N1380: https://lenovopress.lenovo.com/datasheet/ds0187-lenovo-thinksystem-n1380-neptune

Water cooling trays in the rear connect each chassis to a *Cooling Distribution Unit* (CDU).

New HPC/AI servers named **SC750 V4** dual-node CPU-server (see photo), and **SC777 V4** NVIDIA GB200 server with 2 Grace CPUs and 4 Blackwell GPUs are offered initially for the N1380 chassis.

Dell

The **Dell Integrated Rack 7000 (IR7000)**[20] is a 21” wide rails *Open Rack Version 3* (ORV3) based rack infrastructure with dense compute and liquid cooling for High TDP GPUs and CPUs. This *Open Compute Project*[21] (OCP) rack has a 44OU height (about 229 cm) and a 750 mm width, which is larger than current standard racks, since it was designed for large-scale cloud deployments. Please note: A slightly taller rack unit, called an OpenU or “**OU”**, is 48 mm high, whereas the traditional rack unit “U” is 44.5 mm tall.



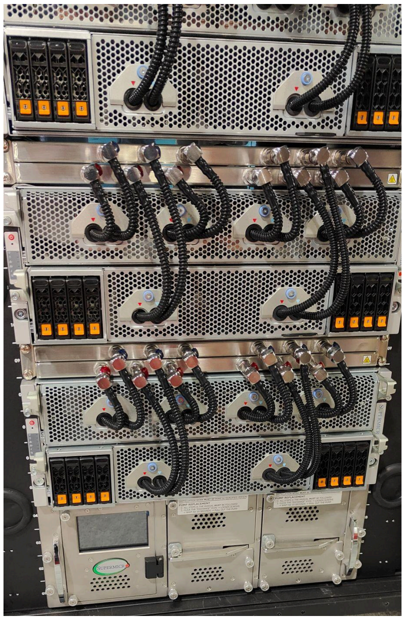
The IR7000 rack contains a 48 V DC low-voltage vertical bus bar in the center rear (see photo). All server nodes and power supply shelves flexibly plug into this bus bar. Currently available AC power shelves support from 33 kW up to 264 kW per rack using compact 3-phase AC power input connectors. Battery-backup shelves for either 48 V DC or 400 V AC power can be plugged in. The rack has liquid cooling manifolds in the left and right rear sides. OCP server nodes, power shelves, and other equipment plug directly into the liquid cooling manifold.

Dell initially delivers a PowerEdge **M7725** supporting two AMD EPYC 5th Generation processors. In addition the **XE9712** expands the *Dell AI Factory* with a NVIDIA GB200 NVL72[22] solution (see photo).

[20] Dell IR7000: https://www.dell.com/en-us/dt/servers/integrated-rack-scalable-systems.htm
[21] OCP: https://en.wikipedia.org/wiki/Open_Compute_Project
[22] NVIDIA GB200 NVL72: https://www.nvidia.com/en-us/data-center/gb200-nvl72/

Supermicro

Supermicro showed a new “OCP-like” rack on the showfloor, albeit with a narrower 600 mm wide 19” rack. New *Supermicro OCP Solutions*[23] for Hyperscale data centers (see footnote 2) are based on 750mm wide 48U high racks offering both air and liquid cooling options.



A number of liquid cooling rack solutions[24] are available in such racks for both HPC and AI systems. An in-rack 4U CDU provides cooling to the rack front side using rear-mounted manifolds and back-to-front liquid hoses in a 1U space between the servers (see photo).

Supermicro also offers an NVIDIA GB200 NVL72 (see footnote 23) rack solution.

Supermicro is setting up distribution channels in Europe, and distributors in Denmark include Danoffice IT and Pedab.

HPE

The HPE ProLiant Compute XD685 server features eight AMD Instinct™ MI325X accelerators and two AMD EPYC™ CPUs.

The HPE Cray XD670 system is a 5U GPU-server containing 8x NVIDIA H100 or H200 SXM 700W TDP GPUs and two Intel Xeon processors.

[23] Supermicro OCP Solutions: https://www.supermicro.com/en/solutions/ocp
[24] Supermicro liquid cooling: https://www.supermicro.com/en/solutions/liquid-cooling

Interconnects for HPC

The delegation met with a number of network interconnect vendors of interest for HPC.

NVIDIA/Mellanox Infiniband

The previous *Mellanox* brand has been rebranded as *NVIDIA Networking*. The NVIDIA® ConnectX®-8 SuperNIC[25] PCIe Gen6 adapter offers both InfiniBand and Ethernet networking at 800 Gbps with 1 or 2 ports, while previous generations offer 400 Gbps and 200 Gbps.

NVIDIA Quantum-X800 InfiniBand switches[26] deliver 800 Gbps bandwidth with 144 ports or 2*36 ports. Both air cooled and liquid cooled solutions are available.

The delegation from DeiC was informed about the roadmap for future Infiniband networking products from NVIDIA/Mellanox under NDA.

Omni-Path

The Omni-Path network fabric is offered by *Cornelis Networks* who recently announced the CN5000 Omni-Path® Product Family[27] featuring one or two ports in a PCIe Gen5 Host Fabric Adapter (HFA). The bandwidth is 400 Gbps with a latency of less than 1 microsecond, and more than 65 billion packets/sec can be transmitted. Edge switches have 48 ports of 400 Gbps, and large Director Switches have up to 576 ports. Both air cooled and liquid cooled solutions are available.

The delegation from DeiC was informed about the roadmap for future Omni-Path networking products under NDA. It is natural that the roadmap includes products delivering 800 Gbps and beyond.

Ultra Ethernet

Several vendors mentioned future *Ultra Ethernet*[28] products promising increased bandwidths of 800 Gbps and up which will be suitable for high I/O workloads from AI applications. However, for HPC computing with MPI message passing, the latency of *Ultra Ethernet* may not be as performant as with *Infiniband* or *Omni-Path*.

---

[25] NVIDIA ConnectX: https://www.nvidia.com/en-us/networking/infiniband-adapters/
[26] NVIDIA Infiniband switches: https://www.nvidia.com/en-us/networking/infiniband-switching/
[27] CN5000: https://www.cornelisnetworks.com/solutions/cornelis-cn5000/
[28] Ultra Ethernet Consortium: https://ultraethernet.org/

HPE/Cray Slingshot

The HPE Cray *Slingshot*[29] network fabric designed for HPC and AI clusters deliver a 200 Gbps bandwidth. The future *HPE Slingshot interconnect 400* will increase the bandwidth to 400 Gbps. Given that *Slingshot* is based on Ethernet silicon, one might surmise that *Ultra Ethernet* could become the standardized technology that could eventually replace *Slingshot*.

Data centers and liquid cooling

As the CPU and GPU *Thermal Design Power* (TDP, see footnote 4) keeps increasing, vendors either have to use ever bigger heatsinks for air cooling, or look for other ways to remove the heat from the servers. This has been a recurring topic in the DeiC HPC TekRef reports, and datacenter planners need to implement flexible and future-proof cooling solutions when designing new server rooms.

Liquid cooling and power infrastructure

Liquid cooling has almost become indispensable for modern servers containing state-of-the-art CPU processors or GPU units. A next-generation CPU will have a TDP of up to 500 W; a single NVIDIA H100 GPU has a TDP of 700 W, and the next-generation B200 GPU will have a TDP of up to 1200 W.

A dual CPU and dual GPU HPC server may therefore use about 1200-2700 W, requiring either an air-cooled chassis with a large 2U-4U height and large, power-hungry fans, whereas a liquid-cooled chassis height of 1U will suffice even for high power densities. A standard 42U high rack may therefore require from 24 kW up to more than 100 kW of power and cooling, depending on server configuration.

Data centers which were designed even a few years ago will be challenged to deliver on the **order of 100 kW power per standard rack**! Electrical transformer stations as well as transmission cables may need to supply several MegaWatt of power for a data center with just a few tens of racks. Obviously, there will be substantial costs of electrical power consumption for such data centers, and data centers may preferentially be placed in regions with competitive costs of electricity.

Servers with liquid cooled CPUs and GPUs (and potentially also RAM memory and other components) require a data center with a robust liquid cooling "Cooling Distribution Unit" (CDU) infrastructure. Unfortunately, there are no multi-vendor industry standards for liquid cooling infrastructure today or in the foreseeable future. The liquid

[29] HPE Cray Slingshot: https://www.hpe.com/psnow/doc/a50002546enw

cooling solutions observed at SC24 include:

- In-rack 4U high CDUs provide a secondary cooling loop within the same rack, with one primary cooling loop per rack connecting to the facility cooling water. Such a CDU may cool up to 200 kW.
- Stand-alone CDUs in a 42U rack form factor connected to the facility cooling water, and with 6 to 12 secondary loop connections for an entire row of racks. A single CDU may cool 600 kW to 1200 kW or more, and multiple CDUs may be combined for redundant N+1 configurations.

There is a lack of industry standards for seemingly trivial items such as 1) liquid pipe connectors, 2) the type of liquid used with either 100% water, or a mixture of 25% propylene glycol and 75% water (denoted as PG25), and 3) the types of additives to the liquid preventing corrosion and bio-growth in the system.

Only Lenovo is offering 100% water based solutions with additives, whereas most other vendors require some form of PG25 liquid which has inherent protection against corrosion and bio-growth and nevertheless requires additives. The PG25 cooling liquid necessitates some safety, handling and environmental considerations.

Quantum computing

The field of *Quantum Computing* (QC) may promise to disrupt the world of high-performance computing and supercomputing centers in the coming years. Multiple vendors with QC related products and research were present at the exhibition floor.

Ass. Prof. Sven Karlsson of DTU led the panel debate on “HPC Meets Quantum Computing: When and How Will Applications Benefit? And Which Ones?”. Here Laura Schulz from LRZ likened the current state of quantum computers to the historical period of traditional computing before the development of the FORTRAN programming language. She thus anticipated a need for the HPC community to work on tools and libraries for QC above the hardware layer. Other panel members suggested that before we might see “quantum supremacy”, it might be possible to achieve “Energy Supremacy” where a simulation on a quantum computer would not necessarily complete in less time than a similar simulation on a HPC system, but instead require less energy to do so.

Quantum computing notes by Pietro Bortolozzo, DTU Computing Center

This section summarizes comments gathered through conversations with various vendors and listening at panel discussions.

Quantum computing holds immense potential, but several often-overlooked issues could make the overall progress harder.

Errors and Quantum Decoherence: Quantum computers are highly impressionable to errors and decoherence when qubits lose their quantum state due to environmental interaction. This can cause "Loss Of Information" and "Inaccurate Computation". Error rates are much higher in quantum systems than in classical systems, making complex calculations challenging. Although researchers are exploring quantum error correction (QEC) to address this, it has challenges, and no quantum computer has yet achieved fault tolerance.

Energy Consumption: Scaling quantum computing systems to more qubits while managing cooling and energy demands is a significant obstacle - error rates, interconnectivity, control signals, and quantum error correction overhead all impact scalabilities. Many quantum computers require extremely low temperatures to function correctly, which is energy-intensive and costly. Researchers are exploring solutions, such as cryogenic optimization, alternative qubit types, hybrid systems, modular quantum computing, and topological quantum computing.

Environmental Impact: Quantum computing has a significant impact on the environment due to its energy consumption, e-waste issue, and data center infrastructure. Reducing this impact requires focusing on energy efficiency, renewable energy, sustainable materials, e-waste management, and green data centers.

Quantum Algorithms: Quantum algorithms must be developed to match hardware advancements. This requires a profound understanding of quantum mechanics and new problem-solving approaches. Not all problems benefit from quantum computation, so it is crucial to identify those where quantum computers offer a significant advantage. The lack of Quantum algorithms restrains practical applications and slows down quantum computing development.

Quantum Application: Writing applications for quantum computers presents various challenges. Traditional programming needs help to contain the complexity of quantum spectacles, and quantum systems are uncontrollable and unpredictable, making debugging and testing more challenging. Hardware limitations, algorithmic complexities and challenges, lack of development tools, and a shortage of skilled developers further complicate quantum application development.

Cybersecurity Risks: Quantum computing poses a significant threat to future cybersecurity infrastructure—the predictions are that it could break used encryption algorithms and jeopardise sensitive data. The "harvest now, decrypt later" threat is also a concern, as competitors can store encrypted data today and decrypt it "tomorrow" with more powerful quantum computers. Reducing these risks requires transitioning to post-quantum cryptography (PQC), building crypto agility, and investing in quantum-resistant security solutions.

Ethical and Societal Issues: The power of quantum computing raises ethical concerns about privacy, security, job displacement, economic disruption, weaponization, and algorithmic issues. Addressing these issues requires developing international cooperation, moral frameworks, embracing diversity, and public engagement.

Quantum computing is the next revolutionary technology, potentially transforming fields from cryptography to drug discovery. However, one of the greatest obstacles facing quantum computing lies within the core physics that make it possible: error rates and quantum decoherence. While classical computers rely on stable binary states (0 or 1) to store and process information, quantum computers use qubits, which can exist in a superposition of both 0 and 1 states. This quantum state is incredibly powerful but also highly fragile, making quantum systems wear towards errors and decoherence. Addressing these challenges is key to making practical, scalable quantum computing a reality.

Storage

HPC obviously has to deal with large to huge amounts of data. Therefore storage systems are integral parts of any HPC data center.

Lustre

Lustre version 2.16 has just been released. In version 2.15 the first work into optimizing the Direct IO path was released. This made 20GB/s single core IO possible, however, Direct IO is hard to program because you need to page-align IO calls. In 2.16 unaligned Direct IO is possible, and a hybrid Buffered/Direct IO solution is planned for 2.17. This will make it transparent to users.

For more details go to the homepage of *Lustre Admin & Dev workshop*[30]. The talk "Lustre 2.17 and Beyond" has a lot of information.

VAST Data

Universities and Research Organizations around the world realize that the days when HPC was the only answer for research is gone. Today HPC and AI infrastructures need to coexist, which demand a new architecture. This is well documented by this video/talk by Dan Stanzione (https://tacc.utexas.edu/about/staff-directory/dan-stanzione/), Vice President for Research at TACC:

https://youtu.be/AxZO034irIs?si=2nmlI2ifJfupVRWg

Already more than 100 universities have implemented a VAST Data platform in their HPC environment.

VAST Data[31] was founded in December 2015 by the former Engineering Director at EMC XtremIO: Renen Hallak and key architects/engineers from his team. Initial investors: NVIDIA, Goldman Sachs and Michael Dell Capital.

Having sold products since 2019 and according to an independent analyst, VAST Data (unofficially) has the largest market share in datacenters globally.

[30] Lustre LAD'24: https://www.eofs.eu/index.php/events/lad-24/
[31] VAST Data: https://www.vastdata.com/

Gartner writes:

"VAST Data is a Leader in this Magic Quadrant. VAST is a unified distributed file system and object storage software platform that has been designed for large-scale multiprotocol deployments targeted at high-performance workloads. VAST achieves higher scalability, reduced latency and improved global efficiency by connecting its front-end stateless protocol nodes and persistent all-flash NVMe QLC media enclosures through NVMe over fabric (NVMe-oF) protocols.

VAST Data is best suited for AI/ML analytics, life sciences and other large-scale, performance-sensitive use cases."

VAST Data or Parallel Filesystems

VAST Data has hired some of the founders of the most used parallel filesystems in the HPC landscape:

- Phil Schwan, CTO at DUG Technology (Australia), wrote Lustre and DUG switched to VAST in 2019.
- Sven Breuner is Field CTO at VAST Data designed BeeGFS and the Fraunhofer parallel file system.
- TACC is a long time Lustre user which moved to VAST in 2023 after extensive testing, and TACC especially appreciates the simplicity and resiliency they have achieved. They are also utilizing the same VAST cluster for their new GPU supercomputer VISTA.
- xAI (Founded by Elon Musk in 2023) chose VAST for their 100.000 GPU supercomputer "Colossus".
- CoreWeave operates a very large GPU Cloud which runs 50% of OpenAI, and the storage for this is S3 on VAST Data.

It seems like with the right underlying storage and namespace architecture, NFS file services (and S3) can scale across a distributed network attached storage array and work just fine for supporting HPC simulation and modeling workloads. With new approaches, this filesystem access protocol can scale to exascale proportions, can meet the needs of new demanding AI applications and (with the right appliance packaging) can dramatically reduce the deployment and administration burden that many supercomputing centers take on.

VAST Data's "Universal Storage" redefines the economics of flash storage, for the first time making flash affordable for all applications, from the highest performance datasets to the largest data archives. The "Universal Storage" concept blends game-changing storage innovations to lower the cost of flash with an exabyte-scale file and object storage architecture breaking decades of storage tradeoffs.

Vast Data platform is a full-stack, end-to-end AI solution that simplifies the creation and expansion of AI deployments.

The End of the Unstructured Data Era

October 1, 2024 - VAST Data Unveils Groundbreaking VAST InsightEngine with NVIDIA to Unlock Insights from All Enterprise Data, Jensen Huang & Renen Hallak | NVIDIA & VAST Accelerate Enterprise AI:

https://youtu.be/awEBdlUlYSM?si=5oM_nmpQATyKcL4H

Collaboration brings industry's first real-time data engine built for enterprise AI – delivering deeper, faster insights from all data and simplifying operations while ensuring scalability, security, and compliance

https://www.vastdata.com/blog/the-end-of-the-unstructured-data-era

Data needs Meta Data (New York University use case)

Attending a session with New York University demonstrated the power of having a Data Platform where Metadata through user defined tags can create an intelligent data structure which can potentially be a foundation for the FAIR principles.

VAST Data can support Unified Data Governance and centrally manage, monitor, and govern diverse data assets by streamlining data governance with a unified framework and the VAST Data Catalog for efficient metadata querying.

https://www.vastdata.com/industry/research-universities

WEKA

In 2013, the WEKA[32] founders set out with a blank sheet of paper, a novel programming language, and a vision to create a fundamentally different approach to managing data that would eradicate the compromises of the past while supercharging the future.

The "WEKA® Data Platform" was designed from the ground up with a clear purpose: Seamlessly and sustainably deliver speed, simplicity, and scale that meets the needs of modern enterprises and research organizations.

The WEKA® Data Platform is a subscription-based software solution that is purpose-built for large-scale, data-intensive environments. Its advanced architecture delivers radical performance, leading ease of use, simple scaling scale, and seamless sharing of your data, so you can take full advantage of your AI and enterprise workloads in virtually any location

[32] WEKA: https://www.weka.io/

Networking, compute, and storage form a sturdy triangle that supports the data center's foundation, enabling data processing, storage, and management. Over the past decade, this triangle has witnessed remarkable advancements and significant improvements in compute and networking which have been pivotal in propelling the AI revolution forward.

Legacy storage systems, however, present significant challenges for modern AI development. These systems struggle to fully utilize the bandwidth of modern networks, limiting the speed at which data can be transferred and processed and struggling to handle large volumes of small files. These challenges can overwhelm storage metadata servers (MDS) and lead to performance bottlenecks, handicapping the tremendous advances in networking speed and compute power. These limitations hinder the scalability and efficiency of AI workflows, impeding tasks such as data preprocessing, model training, and inference.

With WEKA, organizations can deploy a modern data infrastructure that supercharges the storage leg of the triangle and keeps pace with continual advancements in networking and compute. WEKA has a lot of customer use cases, both commercial and institutional, that can be explored here: https://www.weka.io/customers/

DAOS

DAOS (*Distributed Asynchronous Object Storage*)[33] is the open source contender for the likes of VAST and WEKA. This means targeting NVMe and persistent memory using RDMA networks. It is used for the *Aurora* supercomputer at the Argonne National Laboratory which is number 3 on the latest TOP500 list. The capacity is 230 PB and it has a bandwidth of 31TB/s.

Intel started this project to meet the needs of the *Aurora* supercomputer. We learned that Intel is gearing down, so many of the developers have moved to HPE. We also learned that Google has a big stake in DAOS and is a member of the foundation together with HPE, Intel and Argonne.

[33] DAOS: https://docs.daos.io

Middleware and software for HPC

Slurm batch queue resource manager

The *Slurm* resource manager[34] (batch job queue system) was represented at the *SchedMD* company's show floor booth. There was a *Slurm* BOF[35] session presenting the new Slurm 24.11 release, the previous 24.05 release, and other Slurm related tools.

Some of the interesting new features of Slurm 24.11 include:

- New *gpu/nvidia* plugin no longer uses NVIDIA libraries.
- The *slurmd -C* command now detects GPUs.
- Encryption of Slurm daemon network traffic.

TOP500 supercomputers

During SC24 the November 2024 edition of the TOP500 list[36] was released. The new number 1 supercomputer is the "El Capitan" system at the Lawrence Livermore National Laboratory in California, USA which achieved a score of 1742 PFlop/s on the Linpack HPL performance test.

The EuroHPC supercomputer "LUMI"[37] has moved down to the no. 8 position on the current TOP500 list.

As noted by HPCwire[38] China no longer submits new machines to the TOP500 list:

- The Top500 list isn't the global phenomenon it once used to be. It is dominated by eight to ten countries, and the top three systems are installed in labs run by the U.S. Department of Energy. China, previously an enthusiastic participant, has internalized HPC development and is disappearing from the Top500 list.

[34] Slurm: https://slurm.schedmd.com
[35] Slurm BOF: https://slurm.schedmd.com/SC24/Slurm-BoF.pdf
[36] TOP500: https://www.top500.org/
[37] LUMI: https://www.lumi-supercomputer.eu/about-lumi/
[38] HPCwire: https://www.hpcwire.com/2024/11/21/sc24-observations-china-politics-risc-v-and-cxl/

Conclusions

Firstly, the section **Important overall trends** on page 2-3 noted some important changes seen in the market.

The latest generations of high-performance CPUs as well as GPUs continue to make significant impacts on the HPC market. Current CPU and GPU chips may comprise of the order of 100 billion (100.000.000.000) transistors, and even though chip features continue to shrink according to *Moore's Law*, the per-chip heat dissipation continues to grow and imposes limits for computer performance.

A vibrant market exists for products for High Performance Computing, Artificial Intelligence and Machine Learning systems and applications. Very important topics also include storage, networking, cooling and datacenter design. The present report contains snapshots of the current *state-of-the-art* technologies in these areas.

On the horizon looms the future promises of Quantum Computing technology for accelerated computing, provided some fundamental obstacles as discussed above can be overcome.