Power your next breakthrough with AI-at-scale

Redefine enterprise compute with HPE

Breakthrough HPE Solutions are expanding the possibilities of AI:

- **Built for scale:** Seamlessly scale experiments from tens to thousands of GPUs powered by distributed training, supercomputing, and cluster management on validated infrastructure

- **Collaborative:** Data scientists can easily and safely share resources, experiments, and data, enabling them to build on one another's progress

- **Complete:** Pre-configured, turnkey, and fully installed solution for model development and training at scale with fewer code rewrites or infrastructure changes

- **Flexible:** A foundation for accelerator heterogeneity

- **Trusted:** Established global solution provider with enterprise-level support and services

Supercomputing and AI for the enterprise

The world has entered a new age of growth and competition, where explosive data sets and new technology requirements have prioritized digital transformation. Organizations are working tirelessly to turn this data into actionable insights, using a combination of modeling, simulation, analytics, and artificial intelligence (AI) techniques. This new paradigm has placed rigorous demands on compute to digest troves of data for AI at an extreme scale.

Supercomputing and AI are converging rapidly, driving data processing and computation improvements. A growing number of organizations are adopting supercomputing technologies to run massive workloads and accelerate the path to innovation. In addition, as supercomputing technologies become increasingly accessible, more organizations are shifting to data-centric computing.

Solving the toughest AI/ML problems at scale

AI and machine learning (ML) are transforming entire industries by improving efficiency, reducing costs, and streamlining the development of new products and services. With AI/ML models, enterprises can automate repetitive tasks to operate more efficiently and increase overall performance and revenue. Whether organizations use AI/ML models to hone techniques for real-time fraud detection, run millions of genome sequences to uncover genetic mutations, or better analyze seismic data for productive oil and gas fields, the ability to process large amounts of data can dramatically accelerate results.

Purpose-built solutions that enable AI-at-scale are critical to support diverse AI/ML model training tasks. Top use cases include natural language processing (NLP), generative AI, and computer vision, which improve efficiency and profitability. These applications can be used to analyze financial portfolios, model energy systems, and simulate complex industrial processes—which only scratch the surface of what's possible.

NLP is used in several industries to gain faster, deeper insights from troves of available data. For example, financial institutions apply NLP techniques to read countless publicly accessible stock reports to quickly determine market sentiment for a specific sector, stock, or the entire market. Hence, recommendations that once took teams of analysts weeks to synthesize can be accomplished in a fraction of the time using a trained AI/ML model.

Large language models (LLMs) can perform a variety of NLP tasks, including generating and classifying text, answering questions in a conversational manner, and translating text from one language to another. LLMs use deep learning (DL) transformer models to recognize, summarize, translate, predict, and generate text and other content based on knowledge gained from vast data sets—workloads that require massive processing throughput.

Generative AI can create a wide range of data, including audio, code, images, text, simulations, 3D objects, videos, and more. It does this by learning patterns from existing data and uses the knowledge to generate new and unique outputs. Generative AI is being used to break new ground in the world of product design, art, engineering, and other fields.

Computer vision techniques are critical to automate tasks and improve the accuracy of results. For example, in manufacturing, computer vision AI/ML models perform quality assurance checks on products to rapidly identify and resolve problems. Advanced driver assistance systems use sensors (such as radar and cameras) to perceive the world around the vehicle, then either provide information to the driver or take autonomous action based on what it sees. The same capabilities are used by healthcare organizations to analyze diagnostic images and patient data to provide personalized care and by retailers that offer autonomous shopping experiences with real-time security.

HPE is bringing superior compute performance to more organizations:

The HPE Machine Learning Development System is a new concept allowing enterprises to unlock the true power of supercomputing to tackle diverse, data-heavy workloads. We built this offering on the latest exascale-class compute and storage solutions along with HPE Machine Learning Development Environment to simplify and accelerate the AI-at-scale capabilities for the exascale era.

Unlocking greater opportunities with AI

Hewlett Packard Enterprise is reimagining what infrastructure can do by bringing together leading compute, storage, and software development technologies that are foundational for the use and optimization of AI. Our goal is simple and powerful:

• Reduce complexity and cost of model development and training at scale

• Help eliminate enormous operational difficulties of deploying and managing an AI infrastructure

• Mitigate any other common last-mile challenges of operationalizing AI

The first member of the HPE Cray XD6500 family, the HPE Cray XD670 is optimized for enterprises looking to deploy AI-at-scale. The breakthrough server infrastructure for the enterprise provides high agility and throughput across all system components to enable high performance, flexible expansion, and AI acceleration. With a highly dense architecture, the system delivers the performance needed to run demanding workloads.

The HPE Cray XD670 uses NVIDIA® H100 Tensor Core GPUs and Transformer Engine to dramatically speed up AI performance and bring a trillion-parameter model within reach. This accelerator at scale can carry out complex ML tasks and reduce model training times from weeks to days or even hours.

The HPE Machine Learning Development Environment enables companies to build transformative AI applications at scale by removing the complexity and cost associated with ML model development and making it easier for IT administrators to set up, manage, secure, and share AI compute clusters. The software helps ML practitioners train models faster using state-of-the-art distributed training without changing their model code, automatically finds high-quality models, gets more from GPUs with smart scheduling, and reduces cloud GPU costs by seamlessly using spot instances. It also tracks and reproduces work with experiment tracking.

The HPE Machine Learning Development Environment helps companies build better AI models faster by allowing developers and scientists to focus on innovation and not infrastructure, dramatically increasing productivity for their AI teams and achieving faster time to market for AI applications to uncover new insights from data.

Building on these capabilities, HPE will integrate open-source Pachyderm software with our existing supercomputing and AI solutions to enable faster deployment of large-scale AI/ ML/DL models, delivering a unique architecture that is cost-effective at scale and enables sophisticated data transformations across unstructured and structured data. The software provides auto-scaling and parallelized processing of multistage, language-agnostic pipelines with data versioning and data lineage, enabling faster processing and reproducibility of any outcome.

HPE Machine Learning Development System is a complete solution that allows enterprises to focus on innovation, not infrastructure. Validated and pre-configured hardware, software, and services for model development and training make it faster to deliver compute and technical resources to keep up with the AL/ML demands.

Enterprises can perform ML across accelerated computing clusters without rewriting code or updating infrastructure, enabling teams to iterate and collaborate on inventive models and launch the best ones into production. What's more, with a comprehensive AI infrastructure that is intuitive to use and future-proof, enterprises can improve model quality and train models at scale without delay.

Supporting a broad range of the latest technologies

The new HPE Cray XD670 features the latest advances in hardware and software for the exascale era. The system offers a broad range of technologies—including accelerators, storage, networking, and power and cooling options to create a fully customized environment. Organizations benefit from an extremely dense configuration with optional plug-and-play direct liquid cooling (DLC), so they can get up and running quickly and save data center space without compromising performance for mixed supercomputing and AI workloads.

HPE Slingshot combines the best of high-performance interconnects with Ethernet capability to address the demands for high speed and better congestion control. This capability delivers high connectivity, high bandwidth, and adaptive routing to run even complex applications smoothly, while scaling to thousands of nodes and GPUs in large exascale-class systems, including the world's fastest supercomputer, Frontier.

HPE offers storage solutions to manage the vast data sets collected and processed for these enterprise workloads. High-performance storage is carefully integrated with HPE supercomputing technologies, creating a unified architecture for accelerated computing that can scale rapidly to overcome the biggest data center challenges.

HPE Cray XD670 systems include DMTF-compliant Redfish to make managing a supercomputing environment simple and secure. Built on top of HTTPS, Redfish provides a lot of latitude for implementation choices, including emerging and traditional IT infrastructures. This allows IT staff to focus on business initiatives instead of mastering obscure binary protocols.

Developing system software to manage the complexities of AI-at-scale, along with applications that fully utilize IT resources, can be difficult and time-consuming. HPE offers an extensive software portfolio to help enterprises streamline the development of data-driven applications and launch them into production.

HPE Ezmeral Data Fabric unifies data across ML environments. Data fabric technology breaks down silos that limit the ability to harness vital insights, compiling different data types across edges and clouds. Enterprises have secure and seamless access to analytics-ready data sets and can easily process data in various locations. Multiple applications and users can reuse the same data sets reducing the infrastructure required for analytics. The fabric also supports a variety of open source tools to simplify data management while developing AI/ML models that enterprises can trust.

Enterprises can streamline collaboration with HPE Swarm Learning. As ML tasks shift from centralized data centers to the edge, swarm learning only shares insights without exposing the data. So, learnings can be shared quickly and securely between organizations working on different types of systems for more accurate and less biased AI/ML models—benefiting all.

HPE Cray Programming Environment offers a tightly integrated software suite with compilers and developer tools to enable code portability. Developers can run code when and where they need it, which boosts productivity and optimizes modeling, simulation, analytics, and AI applications.

HPE Cray Operating System is a hardened Linux® operating system that enables reliable supercomputing performance. The software suite is designed to run large, complex applications and scale efficiently to thousands of processor cores.

HPE Performance Cluster Manager is an integrated system management solution for supercomputers and clusters. The software provides complete provisioning, management, and monitoring for a few nodes, all the way to exascale.

To get the most out of your AI environment, HPE Services brings together technology and expertise to accelerate your transformation and prepare for what comes next. Our experts help you plan, troubleshoot, and deploy the right tools and technologies for your requirements.

Accelerating supercomputing and AI convergence

HPE is expanding its efforts in the field of supercomputing to revolutionize enterprise operations. We are pioneering the age of exascale with accelerated computing solutions designed to converge supercomputing and AI. Our goal is to make these capabilities mainstream, enabling more organizations to capture real-time insights and convert data into outcomes.

Integrating our supercomputing and AI solutions, tools, and services has introduced the HPE Cray XD670—which is powerful and flexible for the future. We believe supercomputing technologies for the enterprise will benefit organizations throughout their supercomputing and AI journeys.

HPE has developed multiple generations of solutions to meet the most demanding AI/ML model training requirements. Our continuous research and development efforts have led to our latest AI-at-scale capabilities, which can transform how we work, learn, and live.

Conclusion

AI is a story of progress—the ongoing development of technologies through decades of research and collaboration that transform how enterprises operate and share ideas. The next chapter in this story begins at a crossroads of legacy and evolution. This is where HPE meets you to define the future of enterprise compute and develop AI-at-scale capabilities that will change the face of the modern data center.

HPE has a long-term commitment to technology innovation and delivers the benefits of accelerated computing and AI to all organizations. For that purpose, we are developing unique and future-ready infrastructure crafted for the ever-evolving requirements and applications of AI/ML. We use our combined resources to help enterprises leverage technologies that are tested, proven, and achieving real-world success. The HPE Cray XD670 is the latest addition to our industry-leading portfolio, and we look forward to seeing the future breakthroughs in business, science, and technology it will deliver.

It's time to master the art of the possible with AI-at-scale. Let HPE help you transform. Contact us today to begin your journey.

Learn more at

HPE.com/us/en/compute/ hpc/supercomputing/Cray- exascale-supercomputer.html

HPE.com/us/en/compute/ hpc/supercomputing.html

Visit **HPE GreenLake**

Chat now (sales)

Hewlett Packard Enterprise