Hewlett Packard Enterprise

Check if the document is available in the language of your choice.

LAUNCHING A NEW AGE OF INTELLIGENCE WITH ACCELERATED COMPUTING INNOVATION

Unleash revolutionary HPC/AI with the HPE Apollo 6500 Gen10 Plus system

The HPE Apollo 6500 Gen10 Plus system is enhanced for high-performance results by enabling:

- Faster time to insight for competitive advantage, allowing for better use of data
- Superior performance by tightly coupling compute power with industry-leading GPUs for unbeatable job throughput and performance
- A fully tested and configured HPE solution that is A100-ready for demanding HPC, AI, machine learning, and deep learning workloads
- Comprehensive system security and management to help you work and innovate with confidence
- Air-cooled version with nodes of HPE Apollo 6500 Gen10 Plus System in HPE Cray supercomputer architecture—coupled with HPE Slingshot interconnect and supported by HPE Cray OS

ESCALATING HPC AND AI WORKLOADS

In this dynamic global economy, success begins in the data center. Today's organizations rely on the latest technology developments to unlock the value of their data. The ability to innovate is key to unleash exceptional performance, achieve greater intelligence, and improve outcomes that will take businesses to the next level.

Data centers are changing dramatically as the demand for high-performance computing (HPC) and artificial intelligence (AI) skyrocket across a wide range of industries. HPC and AI fuel major improvements in data processing and computation, driving ongoing progress in a variety of scientific, industrial, and societal challenges.

HPC and AI workloads continue to escalate in size and complexity, quickly exhausting the capacity of traditional infrastructure. As a result, many organizations have deployed accelerated computing solutions that provide greater power and memory bandwidth to handle their most data-intensive workloads. Accelerated computing supports HPC, AI, and data analytics at scale by enhancing overall speed and performance. These robust platforms make it possible to manage rising data parameters, execute complex modeling and simulation applications, and run massive training and inference jobs at breakneck speeds.

PREPARING FOR THE EXASCALE ERA

AI is evolving at a pace never seen before, and the exascale revolution is fast approaching. The onset of this era will drive a dramatic shift toward data-centric computing in the enterprise space. Exascale is expected to place rigorous demands on IT infrastructure to digest massive amounts of data for AI at extreme scale. Highly sophisticated workloads will demand maximum durability, greater bandwidth, and high-speed interconnects in order to avoid data bottlenecks. Organizations are racing to prepare for exascale by developing compatible technologies that will ease digital transformation and enable more efficient, cost-effective solutions.

Accelerated computing is the ideal foundation for AI techniques like machine learning and deep learning, which are transforming entire industries with unmatched speed, precision, and insight. A number of vertical markets are reaping the benefits of these game-changing advancements including healthcare and life sciences, energy, manufacturing, government, and financial services. Whether organizations utilize AI to run millions of genome sequences to uncover genetic mutations or monitor smart factories in different geographies, AI vastly exceeds human knowledge and capabilities.

FIGURE 1. HPE Apollo 6500 Gen10 Plus system

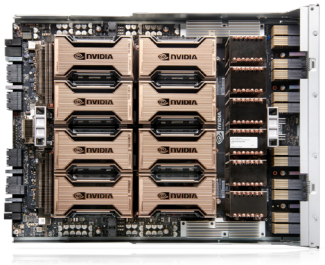

FIGURE 2. NVIDIA A100 Tensor Core GPUs

Yet modern AI workloads are already pushing the boundaries of accelerated computing. Now, savvy organizations are looking for cutting-edge solutions to unleash the full power AI, gain competitive advantage, and solve some of the world's biggest problems.

With so much intelligence at stake, the next phase of accelerated computing is essential to harness the expansion of AI. These breakthrough innovations will not only meet existing standards for I/O, security, and manageability, but they will also provide superior processing at scale, with faster, more reliable data communication to optimize demanding workloads.

A NEW BREED OF ACCELERATED COMPUTING

Hewlett Packard Enterprise is welcoming the coming wave of exascale, with accelerated computing innovations to empower businesses on their AI journeys.

At HPE, our mission is to deploy end-to-end solutions that can tackle any data center workload—from edge to cloud. To achieve this, we are introducing the HPE Apollo 6500 Gen10 Plus system (Figure 1) engineered with performance and density in mind. Backed by the most effective accelerators on the market, HPE systems leverage CPU-GPU heterogeneous compute for a broad range of mission-critical HPC/AI applications. We employ the unparalleled processing capacity of the latest NVIDIA® A100 Tensor Core GPUs (Figure 2) and 2nd Gen AMD EPYC™ Series processors to help eliminate the strain of legacy infrastructure and make way for exascale.

New A100 GPUs ensure low latency at high throughput to enhance these powerful accelerated computing solutions. With third-generation tensor cores, the A100 can efficiently scale up to thousands of GPUs. With NVIDIA Multi-Instance GPU (MIG) technology, it can be divided into seven isolated GPU instances to accelerate diverse workloads. Helping maximize GPU utilization drives revolutionary performance gains of up to 20X from Volta to Ampere architecture. The A100 offers up to 6X out-of-the-box performance for training large models, plus up to 7X performance with MIG for inference to drastically reduce time to insight.

HIGH-DENSITY COMPUTE FOR DATA-DRIVEN INNOVATION

The HPE Apollo 6500 Gen10 Plus system provides vast computational power, leveraging superior memory bandwidth, throughput and data communication. Now, businesses can execute multiple iterations in less time, quickly deploy AI models into production, and expedite time to solution.

The HPE Apollo 6500 Gen10 Plus system is purpose built to deliver unbeatable value:

- Accelerated performance for the most complex HPC and AI applications
- Flexible to meet your diverse workload and data center requirements
- Customized design for reduced costs, improved reliability, and leading serviceability
- Energy-efficient computing with air cooling and liquid cooling system options
- Comprehensive server security and management

HPE enables peak accelerated computing performance to meet the rising complexities of AI. The A100 is a central feature of the HPE Apollo 6500 Gen10 Plus system, with up to 16 GPUs per server to tackle next-level challenges—from deep recommendation engines to conversational AI. These enterprise systems harness the power, frequency, and processing capacity to rapidly capture, analyze, and operationalize intelligence, whatever your workload requires.

NVIDIA NVLink establishes a seamless connection between GPUs, so they can work together as a single robust accelerator. NVLink interconnects provide dedicated communication which enables memory to migrate from GPU to GPU. A single A100 supports up to 12 NVLink connections for a total bandwidth of 600 gigabytes per second.

AMD EPYC processors offer tremendous bandwidth and a high core count to continuously feed information to data-hungry GPUs. High-frequency processors integrated with HDR InfiniBand add up to 200 gigabytes per second of bandwidth for every two GPUs, so even businesses operating at the cluster level can communicate at twice the speed.[1]

In addition, the HPE Apollo 6500 Gen10 Plus system offers extensive storage options, with up to 16 storage drives and choices of SAS, SATA, or NVMe to meet your workload requirements. HPE has plans to roll out new solutions through Q1 2021 that features a staggering 16 NVMe drives for almost 8X greater bandwidth than in our Gen10 servers.

CHOOSING A TRUSTED PARTNER

HPE is uniquely positioned to help organizations meet today's requirements and evolve for tomorrow's challenges. The HPE Apollo 6500 Gen10 Plus system is disrupting the market with accelerated computing capabilities to unleash the future of HPC and AI.

HPE secures your deployments in firmware protection, malware detection, and firmware recovery down to the silicon. The silicon root of trust creates a digital fingerprint in the silicon that helps ensure HPE systems will never boot with compromised firmware. HPE Integrated Lights Out (iLO) server management software enables customers to securely configure, monitor, and update HPE systems seamlessly from any location in the world, so you can operate with confidence.

HPE provides total flexibility that other vendors on the market cannot, equipping customers with end-to-end solutions that are thoroughly tested, secured, and backed by a variety of financial and professional services.

HPE Pointnext Services provide the support and expertise to accelerate innovation and achieve your desired outcomes. HPE professionals collaborate and work with customers to design and implement technology solutions, help optimize processes, smooth skill gaps, and determine the right financial model for your needs.

Key areas of expertise include:

- **Cloud services:** Bring agility and manageability to your technology environment with hybrid cloud
- **Edge services:** Harness the power of data at the edge to achieve better insights and automation
- **IT modernization services:** Modernize IT and data centers with automation and container technologies
- **AI and data-driven services:** Streamline AI adoption, and migrate to unified and secured data platforms with a carefully curated ecosystem of partners

HPE GreenLake is a consumption-based payment model that aligns cash to actual usage. Cloud services from HPE GreenLake deliver business outcomes faster with an as-a-service model. Now, customers can achieve the cloud experience in just a few clicks, without the cost, risk, and time to move data or refactor applications. Cloud services unify your data, centralize operations, boost operational efficiencies, and free up capital with pay-per-use economics—all within the control of your on-premises environment. HPE GreenLake also provides a support team to help customers create a road map from your needs to your ideal solution.

[1] "Introducing 200G HDR InfiniBand Solutions," Mellanox Technologies, 2019

TABLE 1. HPE Apollo 6500 Gen10 Plus system offering a choice of server trays to meet your most demanding AI and HPC workloads

Specifications	HPE Apollo 6500 Gen10 Plus System	
	HPE ProLiant XL645d server	**HPE ProLiant XL675d server**
Density/Scale	6U system supporting a single XL675d or two XL645d	
GPU	NVIDIA HGX A100 4-GPU combines NVIDIA A100 Tensor Core GPUs with new NVIDIA NVLink and NVSwitch AMD MI100 with Infinity Fabric Other leading accelerators as available up to 4 Double Wide PCIe or 8 Single Wide PCIe GPU	NVIDIA HGX A100 8-GPU combines NVIDIA A100 Tensor Core GPUs with new NVIDIA NVLink and NVSwitch AMD MI100 with Infinity Fabric Other leading accelerators as available up to 10 Double Wide PCIe or 16 Single Wide PCIe GPU
Interconnect	Support for up to three high speed fabric interconnects, whether Ethernet, InfiniBand, or HPE Slingshot Additional ports for Ethernet options	Support for up to six high speed fabric interconnects, whether Ethernet, InfiniBand, or HPE Slingshot
Processor	Single AMD EPYC 7002 Series Processor per node, up to 280W for the top frequency and core counts your workloads demand	Dual AMD EPYC 7002 Series Processor per node, up to 280W for the top frequency and core counts your workloads demand
Memory	8 3200MT/s DDR4 SmartMemory	32 3200MT/s DDR4 SmartMemory
Storage	Up to 8 SFF drives total, (optional) M.2	Up to 16 SFF drives, (optional) M.2
System management and System Security	HPE Integrated Lights Out (iLO 5), HPE Performance Cluster Manager (HPCM), HPE Container Platform, HPE OneView iLO 5 Silicon Root of Trust, iLO Advanced (Optional)	
OS Support	HPE Cray OS, Microsoft Windows Server, Red Hat®, Ubuntu, VMware®	
Power	Fully redundant power for all configurations with up to 6 3000W Platinum Hot Plug Power Supplies per chassis Power Capping available at the server and chassis level, and at Rack and Row level with APM	
Cooling	Fifteen 80mm dual rotor hot pluggable chassis fans **NEW:** HPE Apollo 6500 Gen10 Plus System with Direct Liquid Cooling System fully integrated, installed, and supported by HPE	
Storage Controller	Embedded SATA or a selection of HPE Smart Array Controllers	
Warranty (parts, labor, on-site support)	3/3/3	

Resources

hpe.com/us/en/services/pointnext.html

hpe.com/us/en/greenlake.html

Make the right purchase decision. Contact our presales specialists.

Chat **Email** **Call**

CONCLUSION

HPE is bringing the power of accelerated computing to every organization. Across the market, HPE offers one of the deepest set of solutions from data center, to edge, to cloud. Our comprehensive Gen10 portfolio is tailored for HPC/AI on an increasing scale, and we are nearly continuously working to collaborate, build, validate, and deliver leading-edge technologies and services to suit your workloads and economic requirements. Our holistic approach provides best-in-class technologies, an extensive partner ecosystem, management services, and support from experts around the globe to help you succeed. We are committed to be the long-term partner that customers trust to make innovation fast and simple.

Whatever your goals for HPC and AI, HPE can help. To learn more about accelerated computing, visit us online today.

LEARN MORE AT

hpe.com/us/en/compute/hpc/apollo-systems.html

Get updates

a50003152ENW, November 2020

Hewlett Packard Enterprise