Solution brief

High-performance, efficient compute for AI and ML from HPE and AMD

Driving greater business insights and innovation

With 43 world records,[1] HPE ProLiant Gen11 servers with 4th Generation AMD EPYC™ Processors are optimized to power advanced AI and ML workloads

Reshaping the world with new data-driven insights

From improving healthcare outcomes to elevating the retail shopping experience, outsmarting financial services fraud, and optimizing production lines, artificial intelligence (AI) and machine learning (ML) are reshaping industries and lifestyles around the world.

Imagine integrating real-time healthcare data from sensors with critical-care applications that enable physicians to pinpoint diagnoses and create a personalized treatment plan for each patient. Or using video analytics to understand shopper flow and density patterns in retail spaces to ensure customer safety or improve customer experiences. What if you could analyze patterns in network activity to detect and stop bad actors from accessing private financial data before they can do any harm? Or use vibration and temperature analysis to predict when a piece of machinery requires maintenance and avoid downtime?

All of these scenarios and many more are possible by applying AI and ML to data, producing valuable intelligence in a fraction of the time that humans need—if they were even capable of returning the same results. However, such advanced solutions as AI and ML require a high-performing infrastructure that can process vast amounts of data, as well as scale efficiently as data volumes continue to grow. Hewlett Packard Enterprise and AMD answer this need with the latest line of HPE ProLiant Gen11 servers built on the latest generation of AMD EPYC processors.

Compute infrastructure optimized for AI and ML workloads

The new HPE ProLiant DL385 Gen11 server delivers a high level of performance and scalability demanded by AI/ML workloads while maintaining a space- and energy-efficient footprint. This accelerator-optimized 2U and 2P design produces exceptional compute performance for AI/ML, featuring 33% more GPU capacity than previous server generations—supporting up to 8 single-wide GPUs or 4 double-wide GPUs per server.

[1] AMD EPYC Leaderboard: amd.com/en/processors/epyc-world-records

AI and ML are rapidly becoming major driving forces for business innovation. HPE and AMD are leading the way with:

Increased application performance
- New CPU architectures that deliver more cores and higher frequencies to enable more transactions and iterations per cycle
- Large/fast memory footprints for improved analytics and data processing with fewer corrections or loss
- Flexible storage configurations and performance potential to meet increased workload demands
- Industry-leading GPU density—up to 8 single-wide GPUs or 4 double-wide GPUs per server—to do more in less space
- Latest instruction set extensions increase throughput and reduce latency for certain deep learning data types

Enhanced business intelligence
- Tight application integration and automation to create efficient, integrated data pipelines
- Data ingestion from the edge to the data center to the cloud in a seamless and secure design to protect data and IP end to end
- Intelligent video applications; safety, regulatory intelligence, and process control; Natural Language Processing; HPE Ezmeral ML Ops
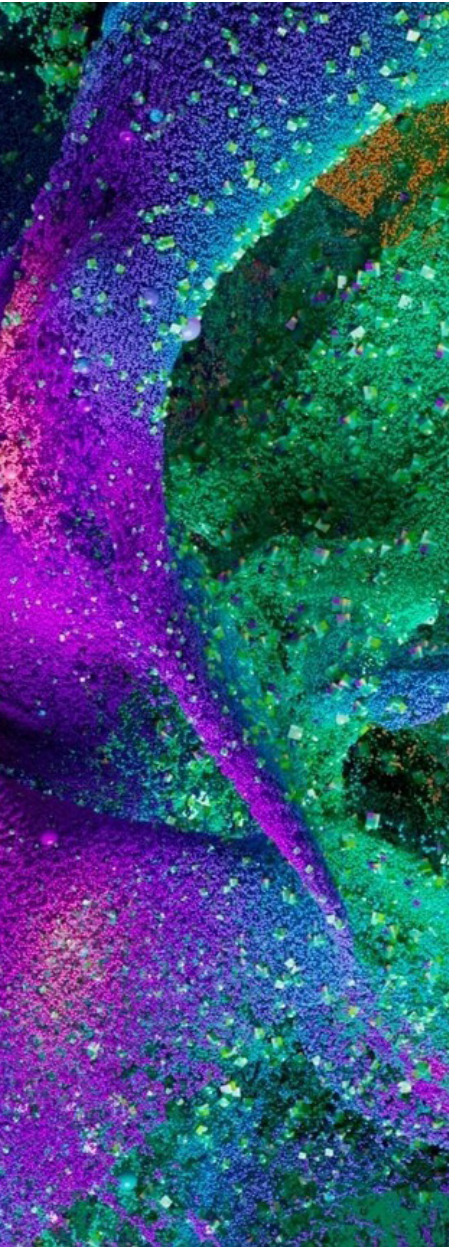
And with GPU clusters you get predictable performance at scale to accelerate business outcomes while lowering the TCO. The HPE ProLiant DL385 Gen11 server is also optimized for efficiency, delivering up to 99% higher performance[2] and 43% more energy efficiency[3] when compared to the previous generation.

HPE ProLiant DL385 Gen11 servers are built on 4th Generation AMD EPYC Processors, featuring next-generation 5 nm technology and support for up to 128 cores per socket, providing a performance-driven foundation for AI/ML workloads. Each server is configured with a next-level energy-efficient DDR5 memory bandwidth of up to 6 TB, further boosting performance. Moreover, the servers are equipped with the PCI Express 5.0 bus, doubling data transfer rates and bandwidth for improved performance. In fact, the EPYC CPUs provide an astounding 160 PCIe 5 lanes on the two-socket HPE ProLiant DL385 Gen11 server. This is critical to improving performance for the demands of today's AI and ML workloads using greater numbers of accelerators.

Secure, efficient management from edge to cloud

Deploying HPE ProLiant Gen11 servers through the HPE GreenLake edge-to-cloud platform enables you to deliver AI/ML workloads wherever your data scientists are located and help them drive more innovation. Additionally, advanced monitoring capabilities designed into the HPE ProLiant Gen11 servers, combined with HPE GreenLake for Compute Ops Management, keep the management of your AI/ML infrastructure streamlined and efficient.

For example, HPE iLO 6, integrated into every HPE ProLiant Gen11 server, allows you to securely configure, monitor, and update the servers from anywhere. Additionally, HPE iLO 6 provides automation and streamlines IT operations with HPE OneView management. HPE GreenLake for Compute Ops Management adds real-time views of your entire server estate and can perform rapid, bulk actions to quickly provision additional AI/ML workloads. You can also count on automated firmware updates with an as-a-service model that provides a consistent experience that is up-to-date with new features.

All this is secured with the zero trust security posture from HPE that spans HPE ProLiant Gen11 servers at every level. For example, the silicon root of trust technology from HPE is embedded in every HPE ProLiant Gen11 server, anchoring an exclusive ASIC to the AMD Secure Processor to help ensure the server is authenticated at boot-up. And the HPE secure supply chain means that server security starts with corruption-free manufacturing, including auditing the integrity of every component.

Enabling the next level of AI/ML innovation

AI and ML workloads hold the potential to uncover vast new insights and opportunities for breakthroughs that improve vital aspects of our world—from healthcare to transportation, entertainment, and shopping. HPE and AMD offer the compute infrastructure needed to deliver the next level of AI/ML innovation with high performance, efficiency, and security. Find out where we can take your business next.

Connection
we solve IT®

Contact your Connection Account Team for more information.

Business Solutions
1.800.800.0014

Enterprise Solutions
1.800.369.1047

Public Sector Solutions
1.800.800.0019

www.connection.com/AMD/server

[2] Compared to previous generation. SPEC and the names SPEC CPU, SPECrate, SPECspeed are registered trademarks of the Standard Performance Evaluation Corporation (SPEC). The stated results are published as of 11-10-22; see spec.org. All rights reserved.

[3] Comparing HPE ProLiant DL385 Gen11 to HPE ProLiant DL385 Gen10 Plus v2. SPEC and the name SPECpower_ssj are registered trademarks of the Standard Performance Evaluation Corporation (SPEC). The stated results are published as of 11-10-22; see spec.org. All rights reserved.

Hewlett Packard
Enterprise