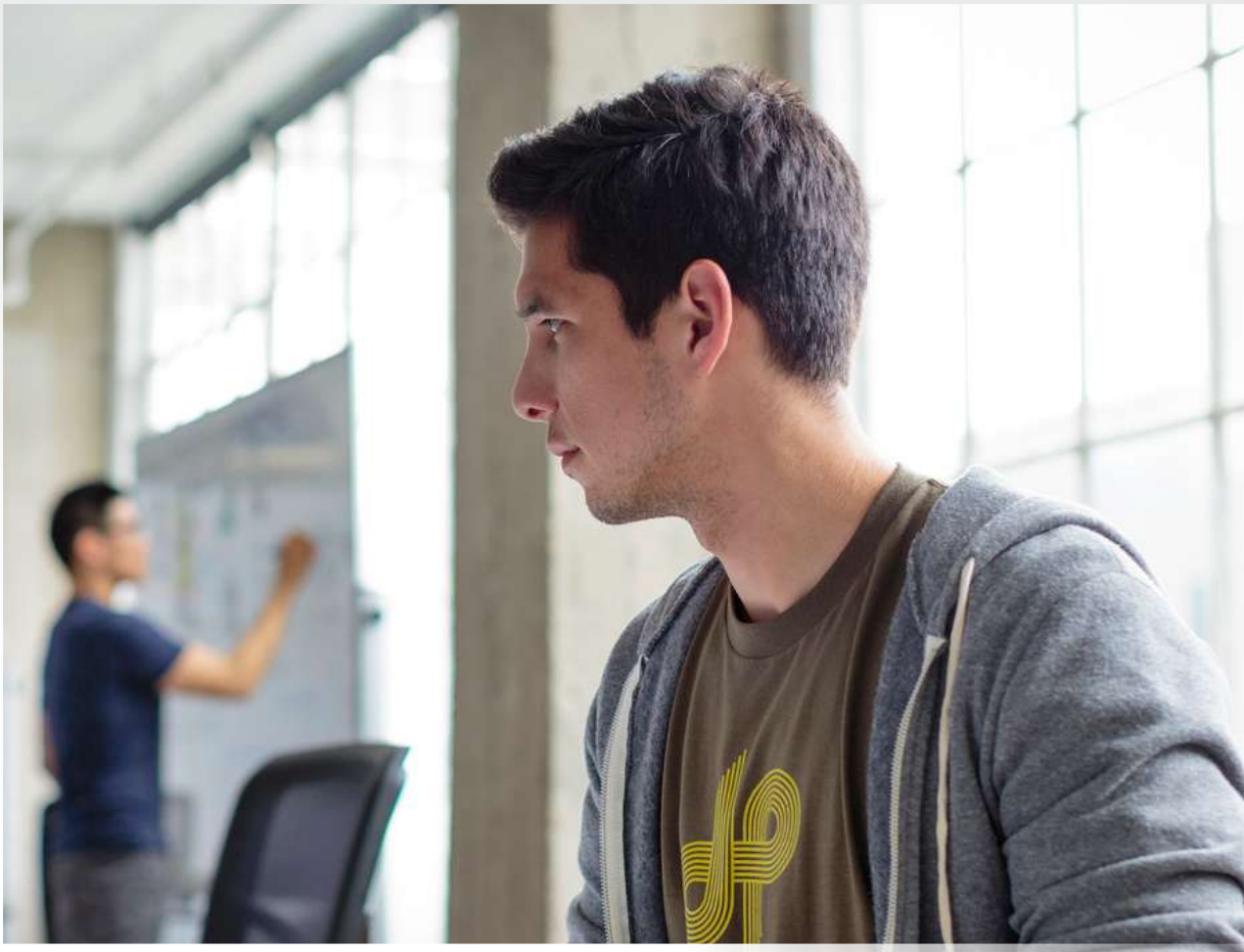
HPE ProLiant Compute

Solution brief



Making its debut in MLPerf Inference benchmarks, the HPE ProLiant DL385 Gen11 immediately secures a no.1 spot

About the MLPerf Inference: Datacenter v5.1 benchmarks

MLPerf Inference: Datacenter v5.1 benchmarks measure the speed, accuracy, and efficiency of data center AI and ML systems in processing inputs and generating results using trained models, a crucial metric for optimizing performance.

Using these metrics, engineers can make reliable assessments of a system's ability to handle cutting-edge AI workloads so they can design reliable, high-performing, and efficient AI products and services that provide valuable insights to users.

Benchmarks for smarter AI infrastructure decisions

As enterprise adoption of artificial intelligence (AI) and machine learning (ML) accelerates, ensuring that underlying hardware delivers optimal performance and efficiency for specific workloads has become both critical and complex. Selecting the right infrastructure requires objective, reproducible data.

Open-source, vendor-neutral benchmarks provide a reliable basis for fair comparisons and informed decisions. MLCommons™, a global consortium of AI leaders, created the MLPerf™ Inference: Datacenter suite as the industry standard for evaluating AI and ML systems. These benchmarks deliver clear insights into system performance, helping organizations maximize AI potential while balancing efficiency and scalability.

No. 1 in Whisper LLM for per-GPU performance

Whisper is a general-purpose speech recognition model. It is trained on a large dataset of diverse audio and is also a multitasking model that can perform multilingual speech recognition, speech translation, and language identification.

Making its first appearance in MLPerf's AI benchmarking, the HPE ProLiant DL385 Gen11 powered by 5th Gen AMD EPYC™ achieved the no.1 performance benchmark for the Whisper speech-to-text model for a PCIe-based system, delivering 3,962.78 samples/second per GPU[1] when equipped with NVIDIA® H200 NVL 141 GB GPUs.

The HPE ProLiant DL385 Gen11 takes the no.1 spot compared to Dell's benchmark result of 3,959.26/Whisper samples/second per GPU.[2]

Secure, scalable, and efficient excellence

The HPE ProLiant DL385 Gen11 offers outstanding flexibility that empowers organizations to meet growing AI workload demands without compromising performance. Businesses can incrementally expand there infrastructure by adding more NVIDIA H200 NVL or NVIDA RTX PRO™ 6000 Blackwell Server Edition GPUs. This method both optimizes budget and enhances resource utilization without overspending on underutilized hardware for a range of enterprise workloads including visual computing and Enterprise AI.

With its 2-GPU configuration, it delivers top-tier performance with industry-leading security and automation to streamline IT operations in a compact, reduced power configuration. The DL385 Gen11 strikes the perfect balance between efficiency, scalability, and affordability for modern AI and enterprise workload deployments requiring advanced acceleration.

[1] **HPE ProLiant DL385 Gen11 delivered** 7,925.55 samples/second with 2 **NVIDIA H200 NVL 141GB GPUs** in the Whisper benchmark. This results in **3,962.78 samples/second per GPU**.

[2] Dell PowerEdge XE7745, 8x H200-NVL-141GB, 2x AMD EPYC 9655 96-Core Processor, **Whisper total samples/sec = 31,674.1, Whisper samples/second per GPU = 3,959.26**

Setting the standard for AI performance and scalability

The HPE ProLiant DL385 Gen11 has proven itself as a leader in AI inference, achieving no.1 per-GPU performance for the Whisper speech recognition model in the MLPerf benchmarks. Its ability to deliver top-tier results in a scalable and economical platform ensures businesses can adapt to growing AI and enterprise workload demands without compromising efficiency or overspending on hardware. As unbiased benchmarks such as MLPerf empower data-driven decision-making, the DL385 Gen11 stands out as a trusted solution for modern AI and ML workloads, balancing cutting-edge performance with real-world affordability.

Visit HPE.com

Learn more at

HPE.com/us/en/HPE-ProLiant-DL385-Gen11.html

Chat now

a00151459ENW

HEWLETT PACKARD ENTERPRISE

hpe.com