

# NVIDIA Hopper, Ampere GPUs Sweep Benchmarks in AI Training

In industry-standard tests of AI training, NVIDIA H100 Tensor Core GPUs set world records on enterprise workloads; A100 raised the bar in high performance computing.

Author: Dave Salvator

Two months after their debut sweeping MLPerf inference benchmarks, NVIDIA H100 Tensor Core GPUs set world records across enterprise AI workloads in the industry group's latest tests of AI training.

Together, the results show H100 is the best choice for users who demand utmost performance when creating and deploying advanced AI models.

MLPerf is the industry standard for measuring AI performance. It's backed by a broad group that includes Amazon, Arm, Baidu, Google, Harvard University, Intel, Meta, Microsoft, Stanford University and the University of Toronto.

In a related MLPerf benchmark also released today, NVIDIA A100 Tensor Core GPUs raised the bar they set last year in high performance computing (HPC).

H100 GPUs (aka Hopper) raised the bar in per-accelerator performance in MLPerf Training. They delivered up to 6.7x more performance than previous-generation GPUs when they were first submitted on MLPerf training. By the same comparison, today's A100 GPUs pack 2.5x more muscle, thanks to advances in software.

Due in part to its Transformer Engine, Hopper excelled in training the popular BERT model for natural language processing. It's among the largest and most performance-hungry of the MLPerf AI models.

MLPerf gives users the confidence to make informed buying decisions because the benchmarks cover today's most popular AI workloads — computer vision, natural language processing, recommendation systems, reinforcement learning and more. The tests are peer reviewed, so users can rely on their results.

In the separate suite of MLPerf HPC benchmarks, A100 GPUs swept all tests of training AI models in demanding scientific workloads run on supercomputers. The results show the NVIDIA AI platform's ability to scale to the world's toughest technical challenges.

For example, A100 GPUs trained AI models in the CosmoFlow test for astrophysics 9x faster than the best results two years ago in the first round of MLPerf HPC. In that same workload, the A100 also delivered up to a whopping 66x more throughput per chip than an alternative offering.

The HPC benchmarks train models for work in astrophysics, weather forecasting and molecular dynamics. They are among many technical fields, like drug discovery, adopting AI to advance science.

Supercomputer centers in Asia, Europe and the U.S. participated in the latest round of the MLPerf HPC tests. In its debut on the DeepCAM benchmarks, Dell Technologies showed strong results using NVIDIA A100 GPUs.

In the enterprise AI training benchmarks, a total of 11 companies, including the Microsoft Azure cloud service, made submissions using NVIDIA A100, A30 and A40 GPUs. System makers including ASUS, Dell Technologies, Fujitsu, GIGABYTE, Hewlett Packard Enterprise, Lenovo and Supermicro used a total of nine NVIDIA-Certified Systems for their submissions.

In the latest round, at least three companies joined NVIDIA in submitting results on all eight MLPerf training workloads. That versatility is important because real-world applications often require a suite of

diverse AI models.

NVIDIA partners participate in MLPerf because they know it's a valuable tool for customers evaluating AI platforms and vendors.

The NVIDIA AI platform provides a full stack from chips to systems, software and services. That enables continuous performance improvements over time.

For example, submissions in the latest HPC tests applied a suite of software optimizations and techniques described in a technical article . Together they slashed runtime on one benchmark by 5x, to just 22 minutes from 101 minutes.

A second article describes how NVIDIA optimized its platform for the enterprise AI benchmarks. For example, we used NVIDIA DALI to efficiently load and pre-process data for a computer vision benchmark.

All the software used in the tests is available from the MLPerf repository, so anyone can get these world-class results. NVIDIA continuously folds these optimizations into containers available on NGC , a software hub for GPU applications.

Original URL: <https://blogs.nvidia.com/blog/2022/11/09/mlperf-ai-training-hpc-hopper/>