



Golden Award

NVIDIA DGX H100





NVIDIA Corporation 香港商輝達香港控股有限公司 台灣分公司

Company Website

Winning Reason

The NVIDIA DGX platform is a fully integrated artificial intelligence supercomputer. What sets it apart from an everyday server is that it features an NVIDIA H100 GPU designed specifically for artificial intelligence all built on top of NVIDIA's trusted DGX platform. The stacked software accelerates the workload of the whole cluster to achieve a seamless expansion. Users have direct access to NVIDIA AI experts to optimize code for faster performance, certified storage and MLOps. NVIDIA DGX was designed to be used in data centers of large organizations. This supercomputer is a reliable, high-performance, plug-and-play system well-suited to helping organizations scale their AI projects.

Product Feature

•8 GPUs in each DGX H100 are interconnected with NVIDIA NVLink. This bypasses PCle bus to provide full bandwidth speed with 640GB of GPU memory •PCle Gen5 doubles data transfer rate from CPU to GPU, improving performance of data intensive AI workloads •Transformer Engine significantly speeds AI performance and capabilities to train large models •DPX accelerates dynamic programming for genomics, proteomics, robot path planning •NVIDIA MIG can partition each GPU into 7 instances isolated with high-bandwidth memory, cache & compute cores •Integrated AI software for orchestration & cluster management, acceleration libraries, storage & network infrastructure & an operating system optimized for AI •DGX H100 scales into NVIDIA DGX SuperPOD infrastructure for off the shelf TOP500-class.