

NVIDIA H100 and Quantum-2 Systems Announced Worldwide

NVIDIA Accelerates Scientific Discovery on Microsoft Azure Cloud and 50+ Partner Systems; Releases Major cuQuantum, CUDA and BlueField DOCA Acceleration Library Updates; Expands Omniverse Ecosystem Into Scientific Computing

SC22 -- NVIDIA today announced broad adoption of its next-generation <u>H100 Tensor Core GPUs</u> and <u>Quantum-2 InfiniBand</u>, including new offerings on Microsoft Azure cloud and 50+ new partner systems for accelerating scientific discovery.

NVIDIA partners described the new offerings at SC22, where the company released major updates to its cuQuantum, CUDA® and BlueField® DOCA™ acceleration libraries, and announced support for its Omniverse™ simulation platform on NVIDIA A100- and H100-powered systems.

H100, Quantum-2 and the library updates are all part of <u>NVIDIA's HPC platform</u> — a full technology stack with CPUs, GPUs, DPUs, systems, networking and a broad range of AI and HPC software — that provides researchers the ability to efficiently accelerate their work on powerful systems, on premises or in the cloud.

"Al is reinventing the scientific method. Learning from data, Al can predict impossibly complex workings of nature, from the behavior of plasma particles in a nuclear fusion reactor to human impact on regional climate decades in the future," said Jensen Huang, founder and CEO of NVIDIA. "By providing a universal scientific computing platform that accelerates both principled numerical and Al methods, we're giving scientists an instrument to make discoveries that will benefit humankind."

Azure First to Offer NVIDIA Quantum-2 for HPC Workloads

Microsoft Azure adoption of the NVIDIA Quantum-2 InfiniBand networking platform follows news of NVIDIA Quantum-2's general availability, announced at GTC in March.

"The future of transformative enterprise technologies such as AI and HPC is in next-generation cloud platforms like Microsoft Azure, where innovators have the opportunity to deliver a new era of technological breakthroughs," said Nidhi Chappell, general manager of Azure AI Infrastructure at Microsoft. "The NVIDIA Quantum-2 InfiniBand networking platform equips Azure with the throughput capabilities of a world-class supercomputing center, available at cloud scale and on demand, and allows researchers and scientists using Azure to achieve their life's work."

Dozens of New Servers Turbocharged With H100, NVIDIA AI

<u>ASUS</u>, <u>Atos</u>, <u>Dell Technologies</u>, INGRASYS, <u>GIGABYTE</u>, <u>Hewlett Packard Enterprise</u>, <u>Lenovo</u>, <u>Penguin Solutions</u>, QCT and <u>Supermicro</u> are among NVIDIA's many partners that are announcing H100-powered servers in a wide variety of configurations.

A five-year license for NVIDIA AI Enterprise, a cloud-native software suite that streamlines the development and deployment of AI, is included with every H100 PCIe GPU. This ensures organizations have access to the AI frameworks and tools they need to build H100-accelerated AI solutions, from medical imaging to weather models to safety alert systems and more.

Among the wave of new systems is the Dell PowerEdge XE9680, which tackles the most demanding AI and high-performance workloads. This is Dell's first eight-way system based on the NVIDIA HGX™ platform, which is purpose-built for the convergence of simulation, data analytics and AI.

"Al is propelling innovation unlike any technology before it," said Rajesh Pohani, vice president of portfolio and product management for PowerEdge, HPC and Core Compute at Dell Technologies. "Dell PowerEdge servers with NVIDIA Hopper GPUs support customers to push the boundaries and make possible new discoveries across industries and institutions."

Major Updates to Acceleration Libraries

To help boost scientific discovery, NVIDIA has released major updates to its CUDA, cuQuantum and DOCA acceleration libraries:

- NVIDIA CUDA libraries now include a multi-node, multi-GPU Eigensolver enabling unprecedented scale and performance for leading HPC applications like VASP, a package for first-principles quantum mechanical calculations.
- The NVIDIA cuQuantum software development kit for accelerating quantum computing workflows now supports
 approximate tensor network methods. This allows researchers to simulate tens of thousands of qubits, as well as
 automatically enables multi-node, multi-GPU support for quantum simulation with unparalleled performance using the
 cuQuantum Appliance.
- NVIDIA DOCA, the open cloud SDK and acceleration framework for NVIDIA BlueField DPUs, includes advanced programmability, security and functionality to support new storage use cases.

These libraries enable researchers to scale across multiple servers and equip them with massive performance boosts to drive scientific discovery. The NVIDIA HPC acceleration libraries are available on leading cloud platforms AWS, Google Cloud, Microsoft Azure and Oracle Cloud Infrastructure.

Omniverse Ecosystem Expands to Scientific Computing

Also at SC22, NVIDIA announced that the scientific computing community can use NVIDIA Omniverse on NVIDIA A100 systems and forthcoming NVIDIA H100 systems for a variety of workloads, described in today's press release "NVIDIA Omniverse Opens Portals for Scientists to Explore Our Universe."

For more about NVIDIA's high performance computing platform, watch the SC22 special address.

About NVIDIA

Since its founding in 1993, NVIDIA (NASDAQ: NVDA) has been a pioneer in accelerated computing. The company's invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined computer graphics, ignited the era of modern AI and is fueling the creation of the metaverse. NVIDIA is now a full-stack computing company with data-center-scale offerings that are reshaping industry. More information at https://nvidianews.nvidia.com/.

Certain statements in this press release including, but not limited to, statements as to: the benefits, impact, performance and availability of our products and technologies, including NVIDIA H100 GPUs, Quantum-2 InfiniBand, cuQuantum, CUDA and BlueField DOCA acceleration libraries, NVIDIA's HPC platform, NVIDIA AI Enterprise, and the NVIDIA Hopper Superchip; the adoption of H100; the scientific method rapidly transforming into an approach driven by data analytics, Al and physicsbased simulation; giving scientists an instrument to make discoveries that will benefit humankind; partners announcing H100-powered servers; Al propelling innovation unlike any technology before it; and the future of transformative enterprise technologies such as AI and HPC are forward-looking statements that are subject to risks and uncertainties that could cause results to be materially different than expectations. Important factors that could cause actual results to differ materially include: global economic conditions; our reliance on third parties to manufacture, assemble, package and test our products; the impact of technological development and competition; development of new products and technologies or enhancements to our existing product and technologies; market acceptance of our products or our partners' products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of performance of our products or technologies when integrated into systems; as well as other factors detailed from time to time in the most recent reports NVIDIA files with the Securities and Exchange Commission, or SEC, including, but not limited to, its annual report on Form 10-K and quarterly reports on Form 10-Q. Copies of reports filed with the SEC are posted on the company's website and are available from NVIDIA without charge. These forward-looking statements are not guarantees of future performance and speak only as of the date hereof, and, except as required by law, NVIDIA disclaims any obligation to update these forward-looking statements to reflect future events or circumstances.

© 2022 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, CUDA, BlueField, DOCA, NVIDIA HGX, NVIDIA Hopper and NVIDIA Omniverse are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated. Features, pricing, availability and specifications are subject to change without notice.

Alex Shapiro Enterprise Networking 1-415-608-5044 ashapiro@nvidia.com