# NVIDIA Grace Hopper Superchips Designed for Accelerated Generative AI Enter Full Production

## GH200-Powered Systems Join 400+ System Configurations From Global Systems-Makers Based on NVIDIA Grace, Hopper, Ada Lovelace Architectures

**COMPUTEX—**NVIDIA today announced that the NVIDIA® GH200 Grace Hopper Superchip is in full production, set to power systems coming online worldwide to run complex AI and HPC workloads.

The GH200-powered systems join more than 400 system configurations powered by different combinations of NVIDIA's latest CPU, GPU and DPU architectures — including NVIDIA Grace™, NVIDIA Hopper™, NVIDIA Ada Lovelace and NVIDIA BlueField® — created to help meet the surging demand for generative AI.

At COMPUTEX, NVIDIA founder and CEO Jensen Huang revealed new systems, partners and additional details surrounding the GH200 Grace Hopper Superchip, which brings together the Arm-based NVIDIA Grace CPU and Hopper GPU architectures using NVIDIA NVLink®-C2C interconnect technology. This delivers up to 900GB/s total bandwidth — 7x higher bandwidth than the standard PCIe Gen5 lanes found in traditional accelerated systems, providing incredible compute capability to address the most demanding generative AI and HPC applications.

"Generative AI is rapidly transforming businesses, unlocking new opportunities and accelerating discovery in healthcare, finance, business services and many more industries," said Ian Buck, vice president of accelerated computing at NVIDIA. "With Grace Hopper Superchips in full production, manufacturers worldwide will soon provide the accelerated infrastructure enterprises need to build and deploy generative AI applications that leverage their unique proprietary data."

Global hyperscalers and supercomputing centers in Europe and the U.S. are among several customers that will have access to GH200-powered systems.

**Hundreds of Accelerated Systems and Cloud Instances**
Taiwan manufacturers are among the many system manufacturers worldwide bringing to market a wide variety of systems powered by different combinations of NVIDIA accelerators and processors. These include AAEON, Advantech, Aetina, ASRock Rack, ASUS, GIGABYTE, Ingrasys, Inventec, Pegatron, QCT, Tyan, Wistron and Wiwynn — all featured in Huang's COMPUTEX keynote address today as key partners.

Additionally, global server manufacturers Cisco, Dell Technologies, Hewlett Packard Enterprise, Lenovo, Supermicro and Eviden, an Atos company, offer a broad array of NVIDIA-accelerated systems.

Cloud partners for NVIDIA H100 include Amazon Web Services (AWS), Cirrascale, CoreWeave, Google Cloud, Lambda, Microsoft Azure, Oracle Cloud Infrastructure, Paperspace and Vultr.

NVIDIA L4 GPUs are generally available on Google Cloud.

**Full-Stack Computing Across Accelerated Systems**
The coming portfolio of systems accelerated by the NVIDIA Grace, Hopper and Ada Lovelace architectures provides broad support for the NVIDIA software stack, which includes NVIDIA AI, the NVIDIA Omniverse™ platform and NVIDIA RTX™ technology.

NVIDIA AI Enterprise, the software layer of the NVIDIA AI platform, offers over 100 frameworks, pretrained models and development tools to streamline development and deployment of production AI, including generative AI, computer vision and speech AI.

The NVIDIA Omniverse development platform for building and operating metaverse applications enables individuals and teams to work across multiple software suites and collaborate in real time in a shared environment. The platform is based on the Universal Scene Description framework, an open, extensible 3D language for virtual worlds.

The NVIDIA RTX platform fuses ray tracing, deep learning and rasterization to fundamentally transform the creative process for content creators and developers with support for industry-leading tools and APIs. Applications built on the RTX platform bring the power of real-time photorealistic rendering and AI-enhanced graphics, video and image processing to enable millions of designers and artists to create their best work.

**Availability**
Systems with GH200 Grace Hopper Superchips are expected to be available beginning later this year.

Learn more about the latest in AI, graphics and NVIDIA-powered systems at COMPUTEX.

**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 industrial 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/.

**For further information, contact:**
Allie Courtney
NVIDIA Corporation
+1-408-706-8995
acourtney@nvidia.com

Certain statements in this press release including, but not limited to, statements as to: the benefits, impact, performance, features and availability of NVIDIA's products, services and technologies, including NVIDIA BlueField, NVIDIA GH200 Grace Hopper Superchips, NVIDIA Grace, NVIDIA Hopper, NVIDIA Ada Lovelace architectures, NVLink-C2C, NVIDIA H100, NVIDIA L4 GPUs, the NVIDIA software stack including NVIDIA AI, Omniverse and RTX technology, NVIDIA AI Enterprise and NVIDIA RTX; our collaborations with system manufacturers including AAEON, Advantech, Aetina, ASRock Rack, ASUS, GIGABYTE, Ingrasys, Inventec, Pegatron, QCT, Tyan, Wistron and Wiwynn, server manufacturers including Cisco, Dell Technologies, Hewlett Packard Enterprise, Lenovo, Supermicro and Eviden, and cloud partners including AWS, Cirrascale, CoreWeave, Google Cloud, Lambda, Microsoft Azure, Oracle Cloud Infrastructure, Paperspace and Vultr, and the benefits, impact, performance and availability thereof; and generative AI rapidly transforming businesses, unlocking new opportunities and accelerating discovery in healthcare, finance, business services and many more industries 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; NVIDIA's 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 NVIDIA's existing product and technologies; market acceptance of NVIDIA's products or its partners' products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of performance of NVIDIA's 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.



Allie Courtney
NVIDIA Corporation
+1-408-706-8995
acourtney@nvidia.com