



Dirancang untuk Percepatan Generatif AI, NVIDIA Grace Hopper Superchips Masuk Produksi Penuh

Sistem Bertenaga GH200 Bergabung dengan 400+ Konfigurasi Sistem Dari Pembuat Sistem Global Berdasarkan Arsitektur NVIDIA Grace, Hopper, Ada Lovelace

TAIPEI, Taiwan—COMPUTEX—May 30, 2023—NVIDIA hari ini mengumumkan bahwa [NVIDIA® GH200 Grace Hopper Superchip](#) sedang dalam produksi penuh, diatur untuk mendukung sistem yang online di seluruh dunia dalam menjalankan beban kerja AI dan HPC yang kompleks.

Sistem bertenaga GH200 bergabung dengan lebih dari 400 konfigurasi sistem berdasarkan arsitektur CPU dan GPU terbaru NVIDIA — termasuk [NVIDIA Grace™](#), [NVIDIA Hopper™](#) dan [NVIDIA Ada Lovelace](#) — diciptakan untuk membantu memenuhi lonjakan permintaan akan AI generatif.

Di COMPUTEX, pendiri dan CEO NVIDIA Jensen Huang mengungkapkan sistem baru, mitra, dan detail tambahan seputar GH200 Grace Hopper Superchip, yang menyatukan arsitektur CPU NVIDIA Grace dan GPU Hopper berbasis Arm menggunakan teknologi interkoneksi [NVIDIA NVLink®-C2C](#). Hal ini menghasilkan total bandwidth hingga 900GB/dtk — bandwidth 7x lebih tinggi daripada jalur PCIe Gen5 standar yang ditemukan dalam sistem akselerasi tradisional, memberikan kemampuan komputasi yang luar biasa untuk menangani aplikasi AI dan HPC generatif yang paling menuntut.

“Generative AI mengubah bisnis dengan cepat, membuka peluang baru, dan mempercepat penemuan di bidang kesehatan, keuangan, layanan bisnis, dan banyak industri lainnya,” kata Ian Buck, wakil presiden Accelerated Computing di NVIDIA. “Dengan Grace Hopper Superchips dalam produksi penuh, pabrikan di seluruh dunia akan segera menyediakan infrastruktur yang dipercepat yang dibutuhkan perusahaan untuk membangun dan menerapkan aplikasi AI generatif yang memanfaatkan data kepemilikan unik mereka.”

Hyperscalers global dan pusat superkomputer di Eropa dan AS adalah di antara beberapa pelanggan yang akan memiliki akses ke sistem bertenaga GH200.

Ratusan Sistem yang Dipercepat dan Instans Cloud

Pabrikan Taiwan termasuk di antara banyak pabrikan sistem di seluruh dunia yang memasarkan beragam sistem yang ditenagai oleh berbagai kombinasi GPU NVIDIA, DPU, dan sekarang CPU. Mitra sistem NVIDIA Taiwan termasuk AAEON, Advantech, Aetina, ASRock Rack, ASUS, [GIGABYTE](#), Ingrasys, Inventec, [Pegatron](#), QCT, Tyan, Wistron dan Wiwynn — semua ditampilkan dalam [pidato utama COMPUTEX](#) pendiri dan CEO NVIDIA Jensen Huang hari ini sebagai mitra utama.

Selain itu, produsen server global Cisco, Dell Technologies, Hewlett Packard Enterprise, Lenovo, Supermicro, dan Eviden, sebuah perusahaan Atos, menawarkan beragam sistem akselerasi NVIDIA.

Mitra cloud untuk NVIDIA H100 termasuk Amazon Web Services (AWS), Cirrascale, CoreWeave, Google Cloud, Lambda, Microsoft Azure, Oracle Cloud Infrastructure, Paperspace, dan Vultr.

GPU NVIDIA L4 tersedia di Google Cloud.

Komputasi Full-Stack di Seluruh Sistem yang Dipercepat

Portofolio sistem mendatang yang dipercepat oleh arsitektur NVIDIA Grace, Hopper, dan Ada Lovelace memberikan dukungan luas untuk kumpulan perangkat lunak NVIDIA, yang mencakup NVIDIA AI, platform NVIDIA Omniverse™, dan teknologi NVIDIA RTX™.

[NVIDIA AI Enterprise](#), lapisan perangkat lunak platform NVIDIA AI, menawarkan lebih dari 100 kerangka kerja, model yang telah dilatih sebelumnya, dan alat pengembangan untuk merampingkan pengembangan dan penerapan AI produksi, termasuk AI generatif, visi komputer, dan AI ucapan.

Platform pengembangan [NVIDIA Omniverse](#) untuk membangun dan mengoperasikan aplikasi metaverse memungkinkan individu dan tim untuk bekerja di beberapa rangkaian perangkat lunak dan berkolaborasi secara waktu nyata di lingkungan bersama. Platform ini didasarkan pada kerangka [Universal Scene Description](#), bahasa 3D yang terbuka dan dapat diperluas untuk dunia virtual.

[Platform NVIDIA RTX](#) memadukan ray tracing, pembelajaran mendalam, dan rasterisasi untuk secara fundamental mengubah proses kreatif bagi pembuat konten dan pengembang dengan dukungan alat dan API terdepan di industri. Aplikasi yang dibangun pada platform RTX menghadirkan kekuatan rendering fotorealistik waktu nyata dan pemrosesan grafis, video, dan gambar yang disempurnakan AI untuk memungkinkan jutaan desainer dan seniman menciptakan karya terbaik mereka.

Ketersediaan

Sistem dengan GH200 Grace Hopper Superchips diharapkan tersedia mulai akhir tahun ini.

Pelajari lebih lanjut tentang AI terbaru, grafik, dan sistem yang didukung NVIDIA di [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:

Melody Tu
NVIDIA Asia-Pacific
(65) 9355 1454
metu@nvidia.com

Inez Lim
CIZA Concept
(65) 9756 8877
inezlimjie@ciza.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.

© 2023 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, BlueField, NVIDIA Grace, NVIDIA Hopper, NVIDIA Omniverse, NVIDIA RTX and NVLink 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.