



NVIDIA MGX Memberikan Kreator Sistem Arsitektur Modular untuk Penuhi Beragam Kebutuhan Komputasi Akselerasi dari Pusat Data Dunia

QCT dan Supermicro di Antara yang Pertama Menggunakan Spesifikasi Server Mengaktifkan 100+ Konfigurasi Sistem untuk Mempercepat AI, HPC, Beban Kerja Omniverse

TAIPEI, Taiwan—COMPUTEX—May 30, 2023—Untuk memenuhi beragam kebutuhan komputasi terakselerasi dari pusat data dunia, NVIDIA hari ini meluncurkan spesifikasi server [NVIDIA MGX™](#), yang memberikan produsen sistem arsitektur referensi modular untuk membangun lebih dari 100 variasi server dengan cepat dan hemat biaya agar sesuai dengan berbagai AI, komputasi performa tinggi, dan aplikasi Omniverse.

ASRock Rack, ASUS, GIGABYTE, Pegatron, QCT dan Supermicro akan mengadopsi MGX, yang dapat memangkas biaya pengembangan hingga tiga perempat dan mengurangi waktu pengembangan hingga dua pertiga menjadi hanya enam bulan.

“Perusahaan mencari opsi komputasi yang lebih cepat saat merancang pusat data yang memenuhi kebutuhan bisnis dan aplikasi khusus mereka,” kata Kaustubh Sanghani, wakil presiden produk GPU di NVIDIA. “Kami menciptakan MGX untuk membantu organisasi mem-bootstrap AI perusahaan, sekaligus menghemat banyak waktu dan uang.”

Dengan MGX, pabrikan memulai dengan arsitektur sistem dasar yang dioptimalkan untuk komputasi yang dipercepat untuk sasis server, lalu memilih GPU, DPU, dan CPU mereka. Variasi desain dapat mengatasi beban kerja yang unik, seperti HPC, ilmu data, model bahasa besar, komputasi edge, grafik dan video, AI perusahaan, serta desain dan simulasi. Beberapa tugas seperti pelatihan AI dan 5G dapat ditangani pada satu mesin, sementara peningkatan ke generasi perangkat keras masa depan dapat berjalan mulus. MGX juga dapat dengan mudah diintegrasikan ke dalam cloud dan pusat data perusahaan.

Kolaborasi dengan Pemimpin Industri

QCT dan Supermicro akan menjadi yang pertama tersedia di pasar, dengan desain MGX muncul pada bulan Agustus. Sistem ARS-221GL-NR Supermicro, yang diumumkan hari ini, akan menyertakan Superchip CPU NVIDIA Grace™, sementara sistem S74G-2U QCT, yang juga diumumkan hari ini, akan menggunakan [Superchip Grace Hopper NVIDIA GH200](#).

Selain itu, SoftBank Corp. berencana untuk meluncurkan beberapa pusat data hyperscale di seluruh Jepang dan menggunakan MGX untuk mengalokasikan sumber daya GPU secara dinamis antara AI generatif dan aplikasi 5G.

“Ketika AI generatif meresap ke seluruh bisnis dan gaya hidup konsumen, membangun infrastruktur yang tepat dengan biaya yang tepat adalah salah satu tantangan terbesar operator jaringan,” kata Junichi Miyakawa, presiden dan CEO di SoftBank Corp. “Kami berharap NVIDIA MGX dapat mengatasi tantangan tersebut. dan memungkinkan AI multi guna, 5G, dan lainnya tergantung pada persyaratan beban kerja waktu nyata.”

Desain Berbeda untuk Kebutuhan Berbeda

Pusat data semakin harus memenuhi persyaratan untuk meningkatkan kemampuan komputasi dan mengurangi emisi karbon demi memerangi perubahan iklim, sekaligus menekan biaya.

Server komputasi yang dipercepat dari NVIDIA telah lama memberikan kinerja komputasi dan efisiensi energi yang luar biasa. Sekarang, desain modular MGX memberi pabrikan sistem kemampuan untuk secara lebih efektif memenuhi anggaran khusus, pengiriman daya, desain termal, dan persyaratan mekanis setiap pelanggan.

Berbagai Form Factor Menawarkan Fleksibilitas Maksimum

MGX bekerja dengan berbagai faktor bentuk dan kompatibel dengan perangkat keras NVIDIA generasi sekarang dan mendatang, termasuk:

- Sasis: 1U, 2U, 4U (pendingin udara atau cairan)
- GPU: Portofolio GPU NVIDIA lengkap termasuk H100, L40, L4 terbaru
- CPU: NVIDIA Grace CPU Superchip, GH200 Grace Hopper Superchip, x86 CPU
- Jaringan: NVIDIA BlueField®-3 DPU, adapter jaringan ConnectX®-7

MGX berbeda dari NVIDIA HGX™ karena menawarkan kompatibilitas multi-generasi yang fleksibel dengan produk NVIDIA untuk memastikan bahwa pembuat sistem dapat menggunakan kembali desain yang ada dan dengan mudah mengadopsi produk generasi berikutnya tanpa desain ulang yang mahal. Sebaliknya, HGX didasarkan pada baseboard multi-GPU yang terhubung dengan NVLink® yang disesuaikan dengan skala untuk menciptakan sistem AI dan HPC terbaik.

Perangkat Lunak untuk Mendorong Akselerasi Lebih Jauh

Selain perangkat keras, MGX didukung oleh kumpulan perangkat lunak lengkap NVIDIA, yang memungkinkan pengembang dan perusahaan membangun dan mempercepat AI, HPC, dan aplikasi lainnya. Ini termasuk [NVIDIA AI Enterprise](#), lapisan perangkat lunak platform NVIDIA AI, yang menampilkan lebih dari 100 kerangka kerja, model yang telah dilatih sebelumnya, dan alat pengembangan untuk mempercepat AI dan ilmu data untuk pengembangan dan penyebaran AI perusahaan yang didukung penuh.

MGX kompatibel dengan rak server Open Compute Project dan Electronic Industries Alliance, untuk integrasi cepat ke pusat data perusahaan dan cloud.

Tonton pendiri dan CEO NVIDIA Jensen Huang mendiskusikan spesifikasi server MGX dalam pidato utamanya di [COMPUTEX 2023](#).

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 our products, collaborations, services and technologies, including the NVIDIA MGX server specification, Omniverse, NVIDIA GPUs including H100, L40 and L4, NVIDIA DPUs including BlueField-3, NVIDIA CPUs including x86 CPUs, NVIDIA HPCs, large language models, and edge computing, Grace CPU Superchip, GH200 Grace Hopper Superchip, chassis, ConnectX-7 network adapters, NVIDIA HGX, NVLink, NVIDIA AI Enterprise, and the NVIDIA AI platform; our collaborations with QCT, Supermicro, ASRock Rack, ASUS, GIGABYTE, Pegatron and SoftBank Corp., and the benefits, impact, performance and availability thereof; enterprises seeking more accelerated computing options when architecting data centers to meet their specific business and application needs; and data centers increasingly needing to meet requirements for growing compute capabilities and decreasing carbon emissions, while also keeping costs down 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.

© 2023 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, BlueField, ConnectX, NVIDIA Grace, NVIDIA Grace Hopper, NVIDIA HGX, NVIDIA MGX 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.