



Japan Cloud Leaders Build NVIDIA AI Infrastructure to Transform Industries for the Age of AI

SoftBank Corp., GMO Internet Group, Highreso, KDDI, Ruteia and SAKURA Internet Adopt NVIDIA Accelerated Computing, Networking and Software to Advance Robotics, Auto, Healthcare, Telecom and Public Sector in Japan

NVIDIA AI Summit Japan—NVIDIA today announced that Japan cloud leaders SoftBank Corp., GMO Internet Group, Highreso, KDDI, Ruteia and SAKURA internet are building AI infrastructure with NVIDIA accelerated computing, networking and software to accelerate transformation across the nation's robotics, automotive, healthcare and telecom industries.

The services from Japan's cloud providers are supported through a Japan Ministry of Economy, Trade and Industry (METI) program to supply critical compute resources across industries. The cloud providers are operating their AI data centers across the country's central, northern and western regions to support development with national and regional NVIDIA accelerated computing infrastructure.

"Japan's companies stand to benefit tremendously from the new industrial revolution powered by AI," said Jensen Huang, founder and CEO of NVIDIA. "Company employees will supercharge their speed and productivity by automating work with AI agents. Industrial companies of tomorrow will operate dual factories — new AI factories to produce software intelligence for the products and machines their factories make today. Working with NVIDIA, Japan's cloud providers are building the AI factories essential to reinvent the country's automotive, robotics, telecommunications and healthcare industries for the age of AI."

Japan Cloud Services Across Nation Support Industries With NVIDIA AI

SoftBank Corp. has adopted the [NVIDIA Blackwell](#) platforms to build Japan's most powerful AI supercomputers, including the world's first [NVIDIA DGX SuperPOD™](#) with [NVIDIA DGX B200](#) systems.

SoftBank Corp. is using its NVIDIA AI infrastructure to accelerate a broad range of industries, including its subsidiary SB Intuitions, which is using NVIDIA AI for the research and development of high-performance [Japanese-native large language models](#) (LLMs).

GMO Internet Group is launching its GMO GPU Cloud, which will be the first local cloud offering in Japan featuring full-stack [NVIDIA H200 Tensor Core GPUs](#), the [NVIDIA Spectrum-X™](#) Ethernet platform for AI, [NVIDIA BlueField™-3 DPUs](#) and the [NVIDIA AI Enterprise](#) software suite. Built on Dell PowerEdge servers, the service is expected to come online this month to provide cloud-based AI computing for businesses in Japan.

GMO GPU Cloud customers will be able to use NVIDIA AI Enterprise software to accelerate their generative AI applications running in production on the platform.

Highreso is accelerating AI development in Japan with the establishment of Highreso Kagawa, a dedicated AI data center powered by NVIDIA accelerated computing. Built with NVIDIA H200 Tensor Core GPUs, the Kagawa data center is expected to begin operations next month and will provide high-performance infrastructure for Highreso's GPUSOROBAN AI Supercomputer Cloud service.

A second Highreso AI data center is scheduled to come online next summer. Together, they will provide researchers and businesses across Japan with access to 1,600 NVIDIA GPUs, helping foster AI development in manufacturing, research and educational institutes.

KDDI is launching AI computing infrastructure built with [NVIDIA HGX™](#) systems to support generative AI and specialized LLM development in collaboration with its ELYZA business group. With NVIDIA software, KDDI customers can use the infrastructure to accelerate AI model training and inference, build digital twins and run simulation workloads for autonomous vehicles, robotic motion control and sensor data processing.

KDDI is also planning a liquid-cooled data center that is expected to feature the [NVIDIA GB200 NVL72](#) platform with NVIDIA GB200 Grace Blackwell Superchips.

Ruteia, an NVIDIA Inception member based in Kyoto, is doubling the NVIDIA Hopper™ computing available in its AI cloud data center. The startup is providing more than 1,000 NVIDIA Hopper GPUs, which will be used for LLM development by customers building foundation models for animation, retail and food services, vision-based editing and data extraction.

Additionally, Ruteia subsidiary AI Fukushima began operations of a new data center in Okuma, Fukushima, in September to support continued industry recovery following the 2011 Tohoku earthquake. AI Fukushima, which operates the infrastructure

implemented by Rutilea, aims to contribute to the revitalization of the regional economy and local development by providing cutting-edge AI computing platforms.

SAKURA internet plans to expand its Koukaryoku cloud services for generative AI from 2,000 to nearly 4,000 NVIDIA Hopper GPUs. It also plans to install [NVIDIA HGX B200](#) infrastructure with Blackwell GPUs at its Ishikari data center, which is expected to be fully powered by renewable energy by 2027. SAKURA internet intends to provide approximately 10,800 GPUs as part of its high-performance AI computing offering.

SAKURA internet customers include the National Institute of Informatics (NII), a public-sector research organization. NII has established the Research and Development Center for Large Language Models and is promoting collaboration among industry, government and academia in Japan through open generative AI research and development activities, including the development of Japanese-centric medical LLMs. Japan-based driving software startup TIER IV is also using NVIDIA-accelerated SAKURA internet computing in its vision to develop intelligent vehicles with its open-source software.

About NVIDIA

[NVIDIA](#) (NASDAQ: NVDA) is the world leader in accelerated computing.

Certain statements in this press release including, but not limited to, statements as to: the benefits, impact, and performance of NVIDIA's products, services, and technologies, including NVIDIA accelerated computing, NVIDIA Blackwell platforms, NVIDIA DGX SuperPOD, NVIDIA DGX B200 systems, NVIDIA H200 Tensor Core GPUs, NVIDIA Spectrum-X Ethernet platform, NVIDIA BlueField-3 DPUs, NVIDIA AI Enterprise software, NVIDIA HGX systems, NVIDIA GB200 NVL72 platform, and Grace Blackwell Superchips; third parties using or adopting our products and technologies, the benefits and impact thereof, and the features, performance and availability of their offerings; Japan's companies standing to benefit tremendously from the new industrial revolution powered by AI; company employees supercharging their speed and productivity by automating work with AI agents; industrial companies of tomorrow operating dual factories — new AI factories to produce software intelligence for the products and machines their factories make today; and working with NVIDIA, Japan's cloud providers building the AI factories essential to reinvent the country's automotive, robotics, telecommunications and healthcare industries for the age of AI 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.

Many of the products and features described herein remain in various stages and will be offered on a when-and-if-available basis. The statements above are not intended to be, and should not be interpreted as a commitment, promise, or legal obligation, and the development, release, and timing of any features or functionalities described for our products is subject to change and remains at the sole discretion of NVIDIA. NVIDIA will have no liability for failure to deliver or delay in the delivery of any of the products, features or functions set forth herein.

© 2024 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, BlueField, NVIDIA DGX SuperPOD, NVIDIA HGX, NVIDIA Hopper and NVIDIA Spectrum-X 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.

Shannon McPhee
NVIDIA Corporation
+1-310-920-9642
smcphee@nvidia.com