February 21, 2025

To all members of the media

GMO Internet Inc.

High-performance GPU cloud service GMO GPU Cloud powered by NVIDIA technology now supports Multi-Instance GPU (MIG) functionality.

～Optimizing resource usage and improving cost performance～

GMO Internet Group's GMO Internet, Inc. (President and CEO: Masaru Ito, hereinafter referred to as "GMO Internet") has integrated NVIDIA Multi-Instance GPU (MIG) technology into its high-performance GPU cloud service, "GMO GPU Cloud" (https://gpucloud.gmo/), starting today, Friday, February 21, 2025.

With this update, customers using the "Dedicated Plan" of GMO GPU Cloud can now take advantage of MIG functionality at no additional cost. They can choose the optimal instance size according to the scale and nature of their jobs and split a single NVIDIA GPU into up to seven independent GPUs. This enables parallel execution of different workloads, optimizes resource utilization, and improves cost performance.

【Background of NVIDIA MIG Integration】

GMO GPU Cloud, equipped with NVIDIA H200 GPUs, ranks 37th globally and 6th in Japan on the TOP500 list of supercomputer performance. Furthermore, it has been proven to deliver top-tier performance, ranked No.1 among commercial cloud services in Japan(※1).

At the same time, workload requirements have become increasingly diverse. jobs do not always require high computational resources depending on their scale or characteristics, while others demand high throughput through parallel execution of multiple jobs.

To address these evolving needs, GMO Internet has integrated NVIDIA MIG technology to provide customers with a more flexible and efficient use of computing resources, making the most of job schedulers(※2).

(※1)GMO Internet Group, Inc. press release dated November 19, 2024:
"GMO GPU Cloud by GMO Internet Group ranked 37th in the world in the TOP500 supercomputer ranking"
https://www.gmo.jp/news/article/9266/
(※2) A job scheduler is a tool that automates the execution of tasks on a computer system. It can run jobs based on predefined times or conditions to improve system management efficiency.

【About MIG Functionality in GMO GPU Cloud】

MIG (Multi-Instance GPU) allows a single NVIDIA GPU to be split into up to seven instances. Since each server in GMO GPU Cloud is equipped with eight NVIDIA GPUs, up to 56 instances can be created on a single server. Each split instance is allocated its own high-bandwidth memory, cache, and compute cores, and operates in a completely isolated environment.
Customers can flexibly customize the instance configuration according to the scale and nature of their workloads. For example, larger jobs can be assigned one or more full GPUs, while smaller jobs can be assigned individual MIG instances. This enables optimal resource utilization and improves cost efficiency.
MIG technology is available at no additional cost to customers using the "Dedicated Plan" of GMO GPU Cloud.

【About GMO GPU Cloud】(https://gpucloud.gmo/)

Launched on Friday, November 22, 2024, GMO GPU Cloud is one of the fastest GPU cloud services in Japan, powered by NVIDIA technology. It leverages the high-performance NVIDIA H200 GPUs to significantly shorten development and research cycles while reducing costs.
Furthermore, GMO GPU Cloud is the first cloud service provider in Japan to adopt both the NVIDIA H200 GPU and NVIDIA Spectrum-X, designed for AI networking. The combination delivers a high-performance GPU cloud environment optimized for generative AI and machine learning.
Through this service, GMO Internet provides enterprises and research institutions working in the fields of generative AI and high-performance computing (HPC) with a high-performance compute environment that requires no infrastructure tuning, contributing to faster development cycles, cost reductions, and the growth of Japan's AI industry.

■ Features of GMO GPU Cloud

1.Equipped with NVIDIA H200 GPUs

The NVIDIA H200 GPU is optimized with significantly expanded memory capacity and memory bus bandwidth for the development and research of large language models. It offers about 1.7 times the memory capacity and approximately 1.4 times the memory bandwidth of the NVIDIA H100 GPU.

2.First in Japan to adopt NVIDIA Spectrum-X

GMO GPU Cloud is the first cloud provider in Japan to implement NVIDIA Spectrum-X, which dramatically enhances performance and scalability of Ethernet networking for AI workloads.

3.Cloud network acceleration with NVIDIA BlueField-3 DPUs

The NVIDIA BlueField-3 Data Processing Unit (DPU) accelerates GPU access to data, streamlines AI application delivery, and enhances the security of cloud infrastructure.

4.Ultra-high-speed storage by DDN

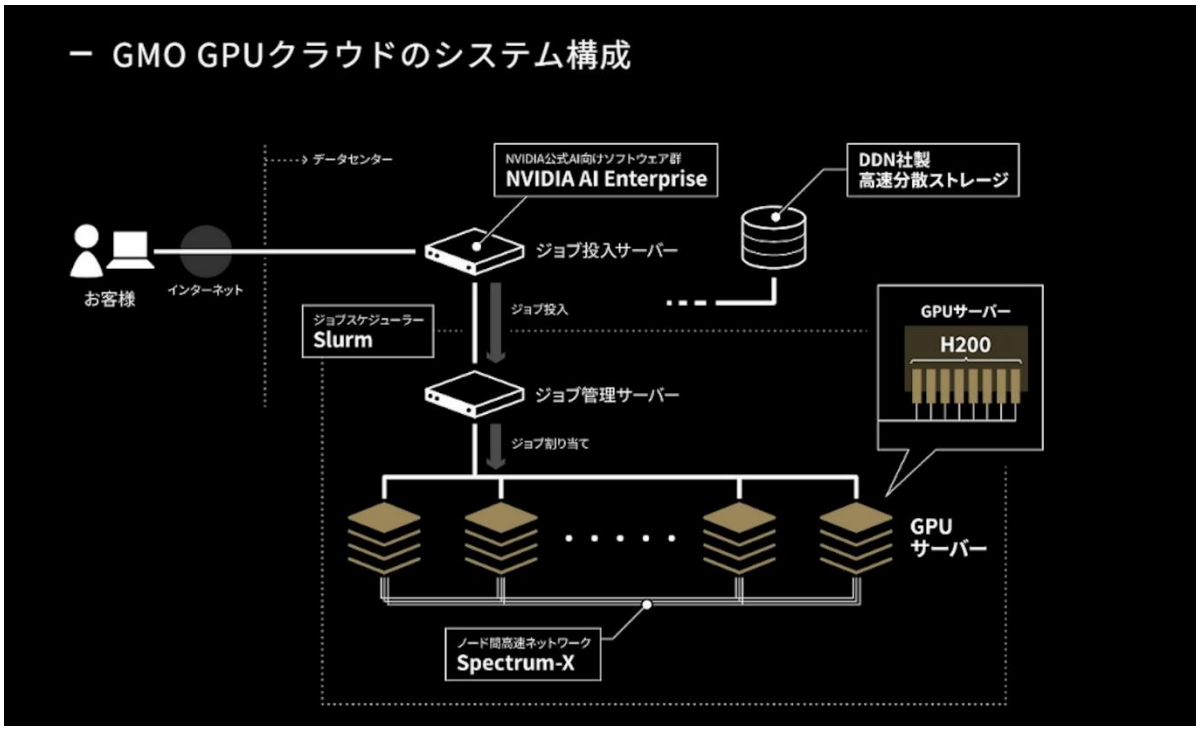
GMO GPU Cloud uses DDN's high-speed storage, optimized for performance with the NVIDIA platform. It delivers a powerful, all-in-one AI development platform.

5.Rapid environment setup and management with NVIDIA AI Enterprise

NVIDIA AI Enterprise is an end-to-end, cloud-native software platform that accelerates data science pipelines and streamlines the development and deployment of production-grade copilots and other generative AI applications.

6.Industry-standard job scheduler Slurm

GMO GPU Cloud adopts Slurm, the industry-standard job scheduler for cluster systems, offering resource allocation, job control, and monitoring functions.



■ Pricing (Excluding Tax)

	Dedicated Plan	Shared Plan
GPU Usage Fee	Quoted individually per client	¥100/min per GPU (※3)
GPU Usage Fee	¥20/min (if used)	
Base Fee	-	50% of the contracted monthly usage time

High-Speed Shared Storage	¥30,000/TiB per month (selectable from 1TiB to 100TiB)	
Local Storage	Free (15TiB per server, temporary job storage)	
Home Directory	Free (100GiB per user)	
Number of Registered Users	50 users per contract	10 users per contract

(※3) Charges apply only for usage exceeding 50% of the contracted monthly time.

■ Use Scenarios

- High-speed training and fine-tuning of large language models
- Training computer vision models using large-scale datasets
- Scientific computing for drug discovery, weather forecasting, and more
- Research and development requiring high-performance computing (HPC)

【Press Inquiry】
GMO Internet, Inc.
Fukui, Public Relations
TEL : +81-90-5313-9226
Contact : https://internet.gmo/contact/press/

GMO Internet Group, Inc.
Koinumaru , PR Team,
Group Corporate Communications Department
TEL: +81-3-5456-2695
Contact : https://www.gmo.jp/contact/press-inquiries/

【Service Inquiry】
GMO Internet, Inc.
GPU Cloud Division
E-mail : contact@gpucloud.gmo

[GMO Internet, Inc.] （URL： https://internet.gmo/）

Company Name	GMO Internet, Inc. (TSE Prime Market Securities Code: 4784)
Location	Cerulean Tower 26-1 Sakuragaoka-cho, Shibuya-ku, Tokyo
Representative	Tadashi Ito, President and CEO
Business	■Internet Infrastructure Domain Registration and Sales (Registrar Business) Cloud and Rental Server (Hosting Business) Internet Connectivity (ISP Business)

	■ Internet Advertising and Media
Capital stock	500 million yen

[GMO Internet Group, Inc.] (URL: https://www.gmo.jp/)

Company Name	GMO Internet Group, Inc. (TSE Prime Market Securities Code: 9449)
Location	Cerulean Tower 26-1 Sakuragaoka-cho, Shibuya-ku, Tokyo
Representative	Masatoshi Kumagai, Founder, Chairman and Group CEO
Business	■Holding Company (Group Management Functions) Internet Infrastructure Internet Security Online Advertising and Media Internet Finance Crypto assets
Capital stock	5 billion yen