

2025 年 12 月 16 日

報道関係各位

GMO インターネット株式会社

GMO GPU クラウド「NVIDIA HGX B300」の クラウドサービスを国内最速クラスで提供開始

～「NVIDIA Blackwell Ultra GPU」搭載、AI 推論処理に最適な環境を提供～

GMO インターネットグループの、GMO インターネット株式会社（代表取締役 社長執行役員：伊藤 正 以下、GMO インターネット）は、NVIDIA テクノロジーで構築した高性能 GPU クラウドサービス「GMO GPU クラウド」において、「NVIDIA Blackwell Ultra GPU」（URL: <https://www.nvidia.com/ja-jp/data-center/technologies/blackwell-architecture/>）を搭載した「NVIDIA HGX B300 AI インフラストラクチャ（以下、HGX B300）」を導入したクラウドサービスを、2025 年 12 月 16 日（火）より国内最速クラスで提供開始いたしました。

サーバー1 台を専有するベアメタル構成により、AI 推論モデルに最適化された HGX B300 AI インフラストラクチャの性能を引き出し、大規模言語モデルの高速推論やエージェント型 AI 開発など、最先端 AI ワークロードに最適な高性能環境を国内から提供します。GMO インターネットでは、AI・ロボティクス時代の研究開発のみならず、その利活用を支える新たな計算基盤として展開してまいります。



【サービス提供の背景】

近年、企業による独自 LLM（大規模言語モデル）の開発や、産業分野における AI・ロボティクスなどの急速な進化に伴い、高度な演算処理を可能にする大規模な計算資源の需要が高まっています。AI モデルの巨大化、推論のリアルタイム化、多様な生成形式への対応といった要求に応えるため、クラウドインフラも日々進化し続ける「動的な技術基盤」として、アップデートすることが技術革新の鍵となります。

GMO インターネットは、2024 年 4 月 15 日に経済安全保障推進法に基づく特定重要物資「クラウドプログラム」^(※1) の供給確保計画に関する経済産業省の認定を受け、同年 11 月 22 日から国内最速クラスで NVIDIA Hopper GPU (<https://www.nvidia.com/ja-jp/data-center/h200/>) を提供^(※2) しています。こ

うした市場環境や技術的要請を踏まえ、当社は 2025 年 5 月 14 日公表の Blackwell アーキテクチャの最新世代「HGX B300」への切り替えを決定いたしました。

これにより、「GMO GPU クラウド」は、今後の AI 開発・運用に求められる高い計算性能・拡張性・低レイテンシ通信に対応し、さらに進化したインフラ基盤として展開してまいります。

(※1) 経済安全保障推進法に基づく特定重要物資の安定的な供給の確保

https://www.meti.go.jp/policy/economy/economic_security/cloud/index.html

(※2) プレスリリース「GMO インターネットグループ、スパコンランキング TOP500 ランクインの

「GMO GPU クラウド」を提供開始」 <https://internet.gmo/news/article/27/>

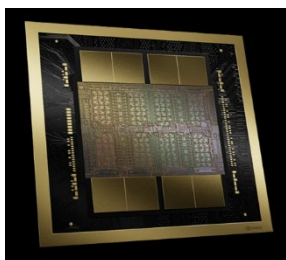
GMO GPUクラウド

【「GMO GPU クラウド (NVIDIA B300)」について】

1. GMO GPU クラウド概要

サービス名	GMO GPU クラウド ベアメタルプラン
提供開始日	2025 年 12 月 16 日
GPU 構成	NVIDIA Blackwell Ultra GPU
GPU メモリ	2.3 TB
CPU 構成	Intel Xeon 6767P x2
メインメモリ	3 TB
ローカルストレージ	3.84TB NVMe Gen5 SSD x8
ネットワークストレージ	NFS ファイルストレージ・オブジェクトストレージ（近日提供予定）
グローバル・ローカル・ストレージ回線（共用）	NVIDIA BlueField-3 DPU 200Gbps x2
インターコネクト	NVIDIA Connect X-8 SuperNIC x8 合計 6,400Gbps
ユースケース	- 大規模言語モデルの高速学習とファインチューニング - AI のリーズニングと推論処理の高速化 - コンピュータビジョンモデルの大規模データセットを用いた学習など
価格	各社ごとにお見積り

2. NVIDIA Blackwell Ultra GPU (GPU メモリ) の特徴



スペック (NVIDIA H200 比)

- メモリ容量は 288GB (H200 比 204%)
- 新たに FP4 演算精度に対応
- FP8 Tensor Core 性能も最大 2.25 倍に向上

3. NVIDIA B300 の特徴



大規模学習の効率化

- 学習性能は最大 4 倍（前世代比：H200）へ向上
- GPU メモリ容量や帯域幅など、主要性能が大幅に強化
- ワットあたりの性能も向上

推論がより強化

- 推論性能 11 倍（前世代比：H200）へ向上
- FP4 対応により、VRAM 使用量を削減しつつスループットを改善
- GPU メモリ容量が過去最大

これにより、大規模モデルの学習をより高速かつ省電力で実行可能にします。

4. ネットワーク構成

HGX B300 AI インフラストラクチャは、8 基の NVIDIA ConnectX-8 SuperNICs、合計 6,400Gbps (<https://www.nvidia.com/ja-jp/networking/products/ethernet/supernic/>) と複数の GPU を接続および拡張させる NVIDIA Spectrum-X Ethernet スイッチ(<https://www.nvidia.com/ja-jp/networking/spectrumx/>)により超高速通信を実現します。低レイテンシ設計により、大規模な AI 学習時のデータ転送を高速化し、複数サーバーを組み合わせた分散学習環境でも安定した高性能を発揮します。



NVIDIAConnectX-8
SuperNIC



NVIDIA BlueField-3 DPU



NVIDIA Spectrum-4

5. NVIDIA B300 と NVIDIA H200 の比較

	NVIDIA HGX B300	NVIDIA HGX H200
GPU メモリ	計 2.3TB (288GB HBM3e x8)	計 1.1 TB (141GB HBM3e x8)
FP4 Tensor コア	144PFLOPS/108PFLOPS	-
FP8 Tensor コア	72PFLOPS	32PFLOPS
FP16 Tensor コア	36PFLOPS	16PFLOPS
FP32	600TFLOPS	540TFLOPS
最大インターコネクト帯域	6,400Gbps	3,200Gbps

NVLINK (GPU 間帯域/GPU1 基ごと)	第 5 世代 (1,800GB/s)	第 4 世代 (900GB/s)
------------------------------	-----------------------	---------------------

【ベアメタルプランについて】

今回、国内最速で提供を開始する「GMO GPU クラウド ベアメタルプラン」では、仮想化レイヤーを介さず GPU に直接アクセスすることで、演算性能を最大限に引き出す完全専有環境を実現します。AI 推論処理に特化した演算能力を活かし、大規模言語モデルの高速推論や、エージェント AI 開発など、多様なワークロードにも適しています。

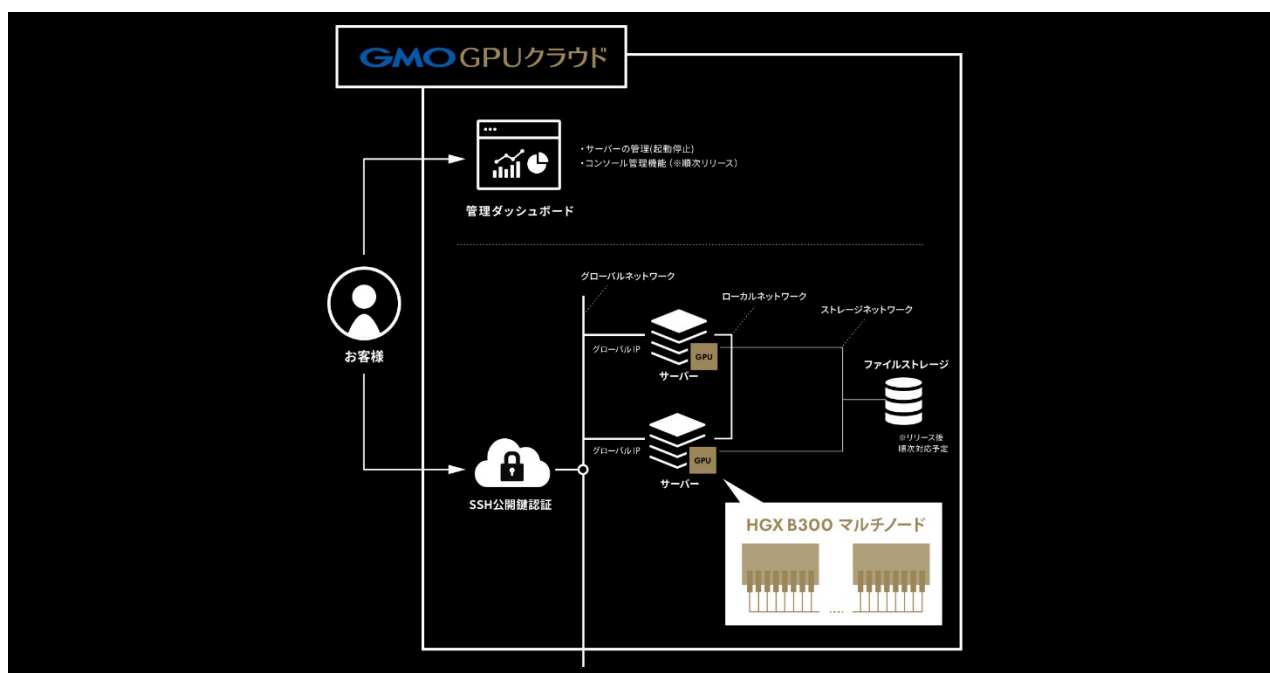
また、ベアメタルならではの自由度を保ちつつ、初期設定やネットワーク構築の手間を軽減する仕組みを備えています。

■主な特徴：

- Ubuntu OS 初期セットアップ済みで提供
- インターコネクトネットワーク設定済み（ローカル NW も利用可能）
- 管理ダッシュボードによるサーバー管理機能
- 追加ストレージの利用が可能

これにより、ユーザーは構築工数を抑えながら、高いパフォーマンスと柔軟なチューニング性を両立した GPU 環境を利用できます。

■NVIDIA B300 におけるサーバー構成イメージ



<代表取締役 社長執行役員 伊藤 正のコメント>

「GMO GPU クラウド」は、2024 年 11 月のサービス開始以来、先端の AI 開発のインフラを支えてまいりました。最先端で高性能な技術への取り組みと、安定した運用環境、きめ細かな技術サポートに一貫してこだわり続け、お客様の AI 開発を支援することに挑み続けてまいりました。NVIDIA B300 を国内最速で提供することは、日本発の AI イノベーションを次のステージへ加速させる取り組みにつながると確信しています。今後も NVIDIA 様との技術連携を深め、エンジニアの挑戦を支える AI 開発基盤を磨き上げ、世界に通用する次世代 AI の創出を力強く後押ししてまいります。



<エヌビディア 日本代表 兼 米国本社副社長 大崎 真孝 氏のコメント>

日々、AI 開発に挑む開発者・研究者にとって、莫大な計算資源にアクセスできる高性能かつ拡張性の高いコンピューティング環境は不可欠です。高効率な学習と高速推論を可能にする NVIDIA HGX B300 を搭載した「GMO GPU クラウド」を国内で提供されることは、日本の AI インフラストラクチャをリードする大きな一歩と言えます。



【今後の展開】

GMO インターネットは、「GMO GPU クラウド」を中核とした AI インフラ戦略により、急速に進化する AI・ロボティクス分野の技術革新に貢献してまいります。今後も最新の AI 計算基盤の提供と、お客様のニーズに応じた柔軟なクラウド環境の構築により、日本の AI 産業に不可欠な国産 AI 基盤として、社会と産業の AI イノベーション創出に貢献してまいります。

【「GMO GPU クラウド」について】

(URL : <https://gpucloud.gmo/>)

NVIDIA テクノロジ上で構築された国内最速クラスの GPU クラウドサービス「GMO GPU クラウド」は、2024 年 11 月に高性能な「H200 GPU」を採用したサービスを開始しました。今回、NVIDIA の最新アーキテクチャ「Blackwell Ultra」を導入した GMO GPU クラウドの「NVIDIA B300」インスタンスは、「NVFP4」^(※3)に対応した第 5 世代 Tensor Core、2.3TB に達する HBM3e メモリ^(※4)、そして第 5 世代の NVIDIA NVLink と NVLink Switch による高速通信を備えた超大規模 AI ファクトリー構築向け GPU です。これにより従来の H200 より大規模言語モデルの学習時間を大幅に短縮することで、さらに AI 開発の効率を大幅に向上させます。

GMO インターネットは、本サービスを通じて、生成 AI 分野に取り組む企業や研究機関に対し、最適化されたインフラ基盤と、お客様のワークロードに応じた柔軟でカスタマイズ可能な計算環境を提供し、開発期間の短縮とコスト低減に貢献、国内 AI 産業の発展を促進します。

(※3) 「NVFP4」は NVIDIA が設計した 4 ビット浮動小数点フォーマットで、他の FP4 フォーマットよりも高い精度を実現。

(※4) 「HBM3e」高帯域メモリの最新規格。従来比で帯域幅が向上し、大規模 AI モデルの学習・推論処理を高速化。

■過去参考リリース

2024 年 4 月 19 日	NVIDIA H200 Tensor コア GPU を採用した生成 AI 向けの GPU クラウドサービスを国内最速提供へ
2024 年 6 月 11 日	生成 AI 向け GPU クラウドサービスに NVIDIA Spectrum-X を国内クラウド事業者として初採用
2024 年 8 月 29 日	「GPU クラウド利用実態調査」～国内利用率わずか 5.4%、約 9 割が海外サービスを利用～
2024 年 9 月 26 日	「NVIDIA H200 GPU」搭載環境の性能を実証
2024 年 11 月 13 日	「NVIDIA AI Summit」で AI・ロボティクス時代のインフラ基盤とセキュリティを解説
2024 年 11 月 19 日	「GMO GPU クラウド」、世界のスーパーコンピュータランキング TOP500 に初ランクイン
2024 年 11 月 22 日	スパコンランキング TOP500 ランクインの「GMO GPU クラウド」を提供開始
2025 年 2 月 21 日	NVIDIA テクノロジーを搭載した高性能 GPU クラウドサービス「GMO GPU クラウド」に「マルチインスタンス GPU (MIG) 機能」を追加
2025 年 5 月 7 日	AI ロボット協会 (AIRoA) の次世代ロボット開発基盤として「GMO GPU クラウド」の正式採用が決定
2025 年 5 月 12 日	「GMO GPU クラウド」がチューリングの自動運転向けマルチモーダル生成 AI 開発基盤に採用
2025 年 5 月 14 日	「GMO GPU クラウド」の追加投資決定
2025 年 6 月 11 日	「GMO GPU クラウド」電力効率を競う世界ランキング「Green500」で世界 34 位、国内 1 位を獲得
2025 年 7 月 1 日	GMO インターネットとマクニカ、NVIDIA で高速化された「GMO GPU クラウド」における生成 AI 開発と活用支援にて協業開始
2025 年 8 月 4 日	GMO GPU クラウド「NVIDIA Blackwell Ultra GPU」を採用
2025 年 10 月 2 日	『GMO GPU クラウド』と低遅延回線『IOWN APN』を活用した次世代分散型 AI インフラの技術実証を開始
2025 年 11 月 7 日	Grafana を活用したモニタリングダッシュボード機能を追加
2025 年 11 月 10 日	プライベートコンテナレジストリ機能を提供開始
2025 年 11 月 10 日	GMO インターネットと CTC、GPU クラウド事業における戦略的販売パートナー契約を締結
2025 年 12 月 12 日	GPU クラウドサービス「GMO GPU クラウド」 Open OnDemand による Web ポータル機能を追加

【GMO インターネット株式会社について】

GMO インターネット株式会社は、GMO インターネットグループのインターネットインフラ事業と広告・メディア事業の強みを融合すべく、2025 年 1 月 1 日に新体制で始動しました。

インターネットインフラ事業の強固な収益基盤と、インターネット広告・メディア事業のそれぞれの強みを最大限に活かし、「すべての人にインターネット」というコーポレートキャッチのもと、関わるすべての方に「笑顔」と「感動」をお届けし、AI で新たな未来を創る価値創造に挑戦してまいります。

以上

【報道関係お問い合わせ先】

●GMO インターネット株式会社

広報担当 福井

TEL : 03-5728-7900

お問い合わせ :

<https://internet.gmo/contact/press/>**【サービスに関するお問い合わせ先】**

●GMO インターネット株式会社

ドメイン・クラウド事業本部 GPU クラウド事業部

お問い合わせ :

<https://gpucloud.gmo/form/>

●GMO インターネットグループ株式会社

グループ広報部 PR チーム 小犬丸

TEL : 03-5456-2695

お問い合わせ :

<https://www.group.gmo/contact/press-inquiries/>**【GMO インターネット株式会社】(URL : <https://internet.gmo/>)**

会 社 名	GMO インターネット株式会社 (東証プライム市場 証券コード : 4784)
所 在 地	東京都渋谷区桜丘町 26 番 1 号 セルリアンタワー
代 表 者	代表取締役 社長執行役員 伊藤 正
事 業 内 容	■ インターネットインフラ事業 ドメイン登録・販売 (レジストラ) 事業 クラウド・レンタルサーバー (ホスティング) 事業 インターネット接続 (プロバイダー) 事業 ■ インターネット広告・メディア事業
資 本 金	5 億円

【GMO インターネットグループ株式会社】(URL : <https://www.group.gmo/>)

会 社 名	GMO インターネットグループ株式会社 (東証プライム市場 証券コード : 9449)
所 在 地	東京都渋谷区桜丘町 26 番 1 号 セルリアンタワー
代 表 者	代表取締役グループ代表 熊谷 正寿
事 業 内 容	持株会社 (グループ経営機能) ■ グループの事業内容 インターネットインフラ事業 インターネットセキュリティ事業 インターネット広告・メディア事業 インターネット金融事業 暗号資産 (仮想通貨) 事業
資 本 金	50 億円

Copyright (C) 2025 GMO Internet, Inc. All Rights Reserved.