



国内最速レベルの高性能クラウドサービス 「GMO GPUクラウド」を支える DDN EXAScaler

GMO インターネット株式会社

「すべての人にインターネット」を掲げる総合IT企業。国内最速レベルの高性能クラウドサービス「GMO GPUクラウド H200」を2024年1月より提供開始。2025年12月に、B300(ペアメタル構成)サービスを追加拡張
<https://internet.gmo/>



<https://gpucloud.gmo/>

DDNソリューション

- ・利用用途
高性能クラウドサービス「GMO GPUクラウド」向けストレージ
- ・導入システム

 DDN AI400X2

日本において世界最高水準の高性能クラウドサービスを目指し、GMO インターネットは2024年11月より「GMO GPUクラウド」の提供を開始した。同社はこのサービスに「NVIDIA H200 Tensor コア GPU」と「NVIDIA Spectrum-X™ Ethernet ネットワーク」、DDN社製の超高速ストレージ「EXAScaler」を採用。96台のマルチノード構成で38.06PFLOPSを記録し、スーパーコンピュータの性能を示す世界的ランキング「TOP500」において、世界37位・国内6位にランクインし、商用クラウドとしては1位のパフォーマンスを発揮した*1。この国内最速レベルのAI基盤サービス「GMO GPUクラウド」とそれを支えるDDNストレージについてお話を伺った。

NVIDIAリファレンスアーキテクチャが示した最適解「DDN EXAScaler」

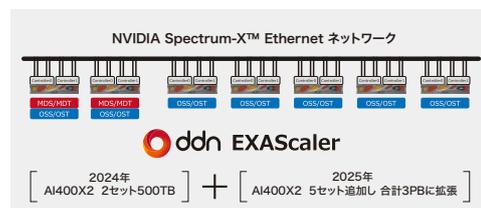
GMO インターネットがGPUクラウド事業への参入を決断したのは、2023年末のことだ。システム本部プロジェクト統括 エグゼクティブリードの佐藤嘉昌氏は「経済産業省のクラウドプログラム公募がきっかけでした。日本においても、世界水準を満たすパソコンに近い性能を持つAIインフラ環境が求められており、ストレージも含めて非常に高速な環境が必要でした」と語る。

2024年4月、経済安全保障推進法に基づく特定重要物資である「クラウドプログラム」の供給確保計画に関する経済産業省の認定を受け*2、GMOインターネットは100億円規模の設備投資の実施と、2024年末までのサービス提供開始を発表した。高性能GPUクラスターにおいて、ストレージは演算性能と同様に重要であったと佐藤氏は選定プロセスを振り返る。「NVIDIAのリファレンスアーキテクチャには必要なI/O性能が定められています。それを満たし、かつコストパフォーマンスに優れたストレージを探しました。複数の製品を比較しましたが、性能、価格に優れ、そしてNVIDIAから認証されていたのは

当時DDNだけでした」

検討の結果、GMO GPUクラウドではDDN EXAScaler(AI400X2) 500TB構成を採用した。システム本部 インフラ技術部 仮想化・共用技術チームの大川将史氏は「GMO GPUクラウドが目指すのは世界トップクラスの性能ですので、それに追従するストレージI/O性能が不可欠でした。DDNのリード170GB/s、ライト140GB/sという性能は初めて見る速度で、率直にすごいと思いました」とDDN EXAScalerの性能の高さに驚きを隠さない。システム本部 インフラ技術部 仮想化・共用技術チームでストレージチームを率いる長谷川泰斗氏も「他のストレージでは性能低下を感じることもありますが、DDNではそれが一切ありません」と評価している。GMO GPUクラウドの引き合いが増え、多くの顧客の需要に対応するなか、2025年夏前、本格利用の増加で容量不足の兆候が現れたため、DDNストレージを3PB構成に拡張する。拡張にあっては性能もリニアにスケールすることが実証された。顧客が利用している環境下においてリード219GB/s、ライト194GB/sを記録している。

また、実際にシステムを使用することも多い大川氏は、ライト性能の速さも評価する。「AI学習におけるチェックポイント書き出しも含め、ローカル設置の高速ストレージと遜色ない時間で完了します。分散ファイルシステムで共有しているにもかかわらず、ボトルネックにならない。困ったことはありません」



GMO GPUクラウドH200/B300 DDNストレージ構成



GMOインターネット株式会社
システム本部 プロジェクト統括
エグゼクティブリード
佐藤 嘉昌 氏



GMOインターネット株式会社
システム本部 インフラ技術部
仮想化・共用技術チーム リーダー
長谷川 泰斗 氏



GMOインターネット株式会社
システム本部 インフラ技術部
仮想化・共用技術チーム
大川 将史 氏

「ベアメタルでもDDNを使いたい」というユーザーからの要望

2025年12月、GMOインターネットは最新のGPUや推論での活用を求める声に応える新サービスとして、NVIDIA HGX B300を搭載したGMO GPUクラウド(ベアメタル構成)の提供を開始した。多様な生成形式への対応といった要求に応えるための新展開である。「お客様から『最新GPUを使いたい』『推論にも使いたい』という要望を受け、AI推論に強みを持つGPUの特性を活かし、大規模言語モデルの高速推論などに対応可能なベアメタル構成での提供をファーストステップとして決定しました」(佐藤氏)

しかしベアメタル構成では、既存GPUクラスターとは異なるテナント分離が必要なことや、推論ではデータの種類や保存期間などの異なる要件がストレージに求められることから、当初はDDNストレージの採用を見送ることも検討していたが、DDNを使いたいという強い要望が上がってきたのである。「『ベアメタルでもDDNがいい』という要望をGMO GPUクラウドのお客様から複数いただきました」と長谷川氏。背景には性能差がある。「DDNの速度は

桁が違う」と長谷川氏は表現し、大川氏も「数十GB/sでも十分速いですが、DDNは200GB/s台。この差は大きいです」と語る。

こういったユーザーの声を反映して、ベアメタルサービスのストレージラインナップにDDNを含めることを検討している。ユーザーの要件に応じて高速性が求められる環境にはDDN EXAScaler、柔軟なテナント分離重視の場合にはネットワークファイルストレージ、データ可搬性・長期保存にはオブジェクトストレージといった形だ。また、ベアメタルでDDNを提供するため、LNET(Lustre Networking)によるネットワーク分離を実装。既存サービスとベアメタルで異なるLNETを構成し、マルチテナント対応を実現した。今後はリードユーザーへの提供を通じてフィードバックを得ながら、その実用性を確認していく予定だ。

「GPUを効率的に使うには、ストレージがボトルネックにならないことが重要です。ライト性能が高ければ、データ持ち込みのリードタイムを大幅に短縮できます」(大川氏)

DDNの技術力と手厚いサポートを高く評価、さらなる進化への支援を期待

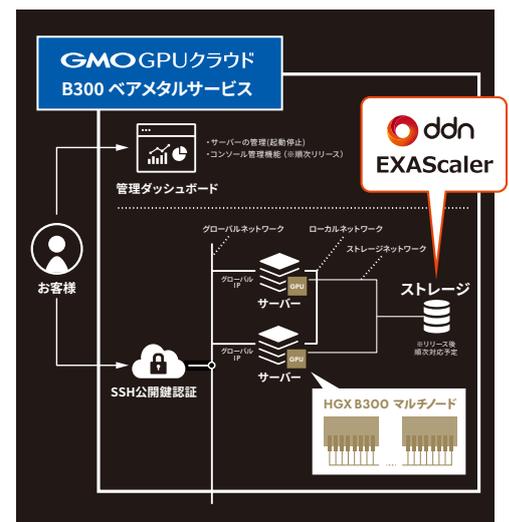
GMO GPUクラウドの次なる展開は、仮想化とコンテナネイティブ化だ。「Kubernetesとの連携を考えています。コンテナ管理でサブシステムを管理しやすくするため、CSIドライバー対応が求められています」(長谷川氏)

また、大川氏はLustreの特性も踏まえた検討を進めている。「Lustreはクラウドネイティブ環境では複雑になる可能性があります。CSI対応は進めたいですが、現在の構成で行くかは検討の余地があります」そこで注目されるのが、DDNがAI向けに新たにリリースしたデータプラットフォームのInfiniaだ。「よりクラウドネイティブに適した新モデルも登場しています。検証したいというより、しなければならぬと考えています」(佐藤氏)

これからもGMOインターネットの挑戦は続くなか、DDNのストレージはその戦略に欠かせないと大川氏は語る。「製品の良さだけでなく、ユーザーを育ててくれるベンダーは貴重です。技術的な使い方も含め、DDNは本当にユーザーを育てられるベンダーだと思います」

長谷川氏もDDNのサポート体制を高く評価している。「当社のストレージチームには若手メンバーも多いですが、DDNのサポートがあるから安心してアサインできます。1on1で迅速に回答いただけるため、大きな信頼感があります」

佐藤氏は最後に、DDNへの信頼と期待を語った。「全体的にとっても安定しています。いくつかトラブルがあっても、お客様に影響を与えることなく解消できており、DDNの技術力を高く評価しています。当社はお客様のデータを預かる立場なので、安全が担保されていることは極めて重要です。今後もDDNと共に、GPUクラウドサービスを進化させていきたいと考えています」



GMO GPUクラウドB300ベアメタル システム構成図

※本事例は2026年1月14日、GMOインターネット株式会社におけるインタビューに基づいて作成しました。

※1 2024年11月時点 <https://group.gmo/news/article/9266/> 参照 ※2 <https://group.gmo/news/article/8933/> 参照