

「GMO GPU クラウド」のインフラに Dell PowerEdge XE9680 を採用 国内最速レベルのサービスを提供

多彩な運用管理機能で
大規模 GPU クラスターの
効率的な運用も実現



ビジネス課題

インターネットインフラ事業、並びにインターネット広告・メディア事業を展開する GMO インターネットでは、先進的な ICT サービスを次々と展開しています。その一環として、生成 AI 活用を加速させる新たな GPU クラウドサービスの開発に着手。国内最速レベルのサービスを実現すべく、高い性能と信頼性を備えた GPU サーバーの導入に取り組みました。

導入効果



商用クラウドサービスとして国内最速レベルのGPUクラウドサービスを実現



NVIDIA H200 Tensor コア GPUを搭載したGPUクラウドサービスを国内で最も早く提供することに成功



iDRAC/OpenManage Enterpriseにより大規模GPUクラスターを効率的に統合管理



デル・テクノロジーズのサービスを活用し大量のサーバー群を円滑に導入

国内最速レベルのパフォーマンスを有する GPU クラウドサービスを、他社に先駆けて市場投入することに成功。エンタープライズ分野でのビジネスに大きな弾みを付けることができました。大規模サービス基盤の運用管理についても、iDRAC/OpenManage Enterprise を活用することで、安定稼働と作業負荷軽減を実現しています。

ソリューション

・ Dell PowerEdge XE9680

「GMO GPU クラウド」のサービスを支えるサーバーとして、高性能 GPU「NVIDIA H200 Tensor コア GPU」(以下 NVIDIA H200 GPU) を 8 基搭載した Dell PowerEdge XE9680 を導入。NIC、スイッチ等も NVIDIA 推奨構成とし、国内最速レベルの強力な環境を実現しています。またデル・テクノロジーズの手厚い支援により、高品質なインフラを短期間で構築することにも成功しました。

持ち前のベンチャー精神で 業界をリードする サービス/ソリューションを提供

GMO インターネットグループのインターネットインフラ部門と、インターネット広告・メディア事業を展開する GMO アドパートナーズの統合により、2025 年1月に発足した新生 GMO インターネット。同社では「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下、多彩なサービス/ソリューションを展開している。

「当グループでは、1995 年の創業以来培ってきたマインドを『スピリットベンチャー宣言』として掲げています。新しいことにどんどんチャレンジするベンチャー精神こそが、当社の成長を牽引する原動力。その中でも特にこだわっているのは、『一番』であることです」こう語るのは、GMO インターネット 執行役員 システム本部 本部長 牧田 哲氏だ。事業を行うからには、やはり頂点を目指さなくては意味がない。同社が提供するどのサービス/ソリューションにおいても、その分野でトップであることを追求しているという。

商用クラウドサービスとして 国内最速レベルを達成した 「GMO GPU クラウド」

そうした同社の思いを具現化した新サービスが、2024 年 11 月より提供を開始した GPU クラウドサービス「GMO GPU クラウド」である。本サービスは、スーパーコンピューターの性能ランキング「TOP 500^{※1}」において、世界 37 位、国内第 6 位のランキングを獲得。しかもこの結果は、日本国内向けの商用クラウドサービスとしては第 1 位となる（同社調べ）。また、電力効率を競う「Green500^{※2}」では世界 34 位、国内 1 位も獲得し、高い処理能力とグリーン性能の両面で優れた計算基盤であることが国際的に認められたことを実証した。

GMO インターネット ドメイン・ホスティング事業本部 GPU クラウド事業部 部長 武田 茂氏は、「GMO GPU クラウドは、生成 AI 利活用の急速な拡がりに対応すべく開発したサービスです。最近では大企業を中心に、LLM（大規模言語モデル）などの生成 AI 技術をビジネスに取り入れる機運が高まっています。また、大規模なシミュレーションや解析など、いわゆる HPC 分野への取り組みを行っていた企業でも GPU への関心は強い。当社としても、こうしたお客様のニーズに応えられるサービスが必要と考えました」と語る。

NVIDIA H200 GPU を用いたサービスを 国内で最も早く提供すべく Dell PowerEdge XE9680 を採用

サービス設計にあたってのコンセプトは、もちろん「一番」であることだ。同社では、国内最速レベルのサービスを実現するために、この分野におけるデファクト・スタンダードである NVIDIA 社のリファレンスアーキテクチャを採用。GPU には



“デル・テクノロジーズの支援を活かすことで、NVIDIA H200 GPU を用いた『GMO GPU クラウド』を他社に先駆けてリリースできました。今後はこのサービスを通して、日本企業の生成 AI 活用にしっかりと貢献していきたいと思います。

GMO インターネット株式会社
執行役員 システム本部 本部長
牧田 哲氏

最新の「NVIDIA H200 GPU」を選択し、NIC には「BlueField-3 SuperNIC」、スイッチには「Spectrum-4 Switch」を利用することで、AI ワークロード向けに特化したイーサネットファブリック「Spectrum-X」を実現している。

さらに、パフォーマンス面での「一番」に加えて、今回同社ではもう一つの「一番」に挑んだ。それは、NVIDIA H200 GPU を用いた GPU クラウドサービスを、国内で最も早くリリースするということだ。そして、これを実現するカギとなったのが、NVIDIA H200 GPU を 8 基搭載可能なデル・テクノロジーズの高性能 GPU サーバー「Dell PowerEdge XE9680」（以下、PowerEdge XE9680）である。

デル・テクノロジーズの調達力が サービスの早期リリースに貢献

GM0 GPU クラウドのインフラに PowerEdge XE9680 が



商用クラウドサービスとして
国内最速レベルのGPUクラウドサービスを実現

※1: 2024 年 11 月時点

※2: 2025 年 6 月時点

選ばれたポイントとして、まずは同社サービス基盤における PowerEdge サーバーの稼働実績が挙げられる。「当社では多岐にわたるサービスを提供していますが、その多くで PowerEdge サーバーを採用しています。性能・信頼性が高い上に、コストもリーズナブルであることがその理由です。今回の GMO GPU クラウドには、当社としても力を入れていますので、安心できるベンダーと組みたいと考えました」と牧田氏は語る。

また、PowerEdge XE9680 の実力についても、既に十分な手応えを得ていたとのこと。GMO インターネット 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 嘉昌氏は「当社では、他のサービスで『NVIDIA H100 Tensor コア GPU』（以下 NVIDIA H100 GPU）を搭載した PowerEdge XE9680 の試験導入を行っており、そのパフォーマンスの高さを確認済みです。今回の構成とは GPU が NVIDIA H100 GPU か NVIDIA H200 GPU かだけの違いですから、GMO GPU クラウドのサービス基盤にも最適と判断しました」と語る。

さらに決め手となったのが、グローバルベンダーならではの調達力だ。「NVIDIA H200 GPU 搭載サーバーを提供するベンダーは他にもありますが、実際にモノが手に入らないのでは仕方ありません。その点、デル・テクノロジーズでは、強力な調達力を駆使して当社の希望よりも前倒しで製品を納入してくれました。おかげで無事、我々が目指していた国内最速でのリリースに漕ぎ着けることができました」と牧田氏は語る。

前例のない大規模環境を わずか約6ヶ月でスピード構築

今回の GMO GPU クラウドでは、実に96台に上る PowerEdge XE9680 が導入されている。これまでに例のない大規模環境だけに、インフラ構築にあたっては様々な工夫も凝らされた。「PowerEdge XE9680 は、1台当たり約120kg もの重さがある重量物ですので、通常のサーバー導入のようなわけにはいきません。そこで、デル・テクノロジーズのキッティングサービスを活用し、安全に据え付けを実施。稼働後に不具合が出ないよう、製品出荷時だけでなく納品後にもテストを行ってもらいました」と佐藤氏は振り返る。

また、大量のサーバー群の工事や設定作業についても、綿密な作業スケジュールに基づいて順繰りに実施し、できるだけ時間短縮を図った。事前に懸念されたサーバー排熱の問題についても、CFD シミュレーションを行って効率的に冷却できるエアフローを実現。この結果、実にキックオフからわずか約6ヶ月という、驚異的なスピードでのサービスインを果たしている。

パフォーマンスの高さに加え きめ細かなサポートにも高評価 幅広い分野で利用が拡大

実際にサービスを利用するユーザーからの反響も上々とのこと。武田氏は「高性能 GPU クラスタならではのパフォーマンスはもちろん、お客様からは利便性や使い勝手の良さにも高い

評価を頂いています。その結果、自動車業界における自動運転技術開発や大手 IT 企業での LLM 開発、産業機器分野におけるロボット基盤モデルの開発など、幅広い分野でご利用頂いています」と力強く語る。

こうした成功の理由の一つとして、顧客本位のサービス提供に努めている点が挙げられる。たとえば、GMO GPU クラウドではジョブスケジューラとして「Slurm」を提供しているが、それだけでなく作業手順書や問い合わせ等へのサポートも提供。また無料トライアルなども併せて実施している。「インフラが不得手な AI 開発者の方にも簡単にサービスを使って頂くためには、こうしたきめ細かな配慮が欠かせません」と武田氏。また佐藤氏も「一方、インフラに詳しい AI 開発者の方からは、使い勝手に対する様々なご要望も頂きます。そこでサービスのリリース後も、多くの機能改善を継続的に実施しています」と続ける。

運用管理の効率化も実現 今後もサービスを継続強化し さらなる進化を目指す

PowerEdge XE9680 の運用管理機能も、サービスの安定稼働に大きく貢献。佐藤氏は「たとえばデル・テクノロジーズのサーバー管理コンポーネントである『iDRAC』では、サーバーの電力消費量やプロセッサの温度、ファンの稼働状況などを細かく管理できます。これらを活用することで、最適なサーバー冷却を実現しています」と語る。また、大量のサーバー群の統合管理には、システム管理コンソール「OpenManage Enterprise」が貢献。佐藤氏は「パッチ適用やファームウェア更新などの作業をスケジューリングできますし、推奨アップデートモジュールの検索なども簡単に行えます。大規模環境を効率よく運用して



「GMO GPU クラウドは、優れた性能を誇るだけでなく、インフラが不得手な AI 開発者の方でも簡単に使える仕組みをご用意しています。無料トライアルなども実施していますので、ぜひ一度試して頂きたいですね。」

GMO インターネット株式会社
ドメイン・ホスティング事業本部 GPU クラウド事業部 部長
武田 茂氏



今回採用した Dell PowerEdge XE9680 は、他の PowerEdge サーバーと同じく iDRAC や OpenManage Enterprise による統合管理が可能ですので、大規模 GPU クラスタ環境を効率的に運用できています。

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



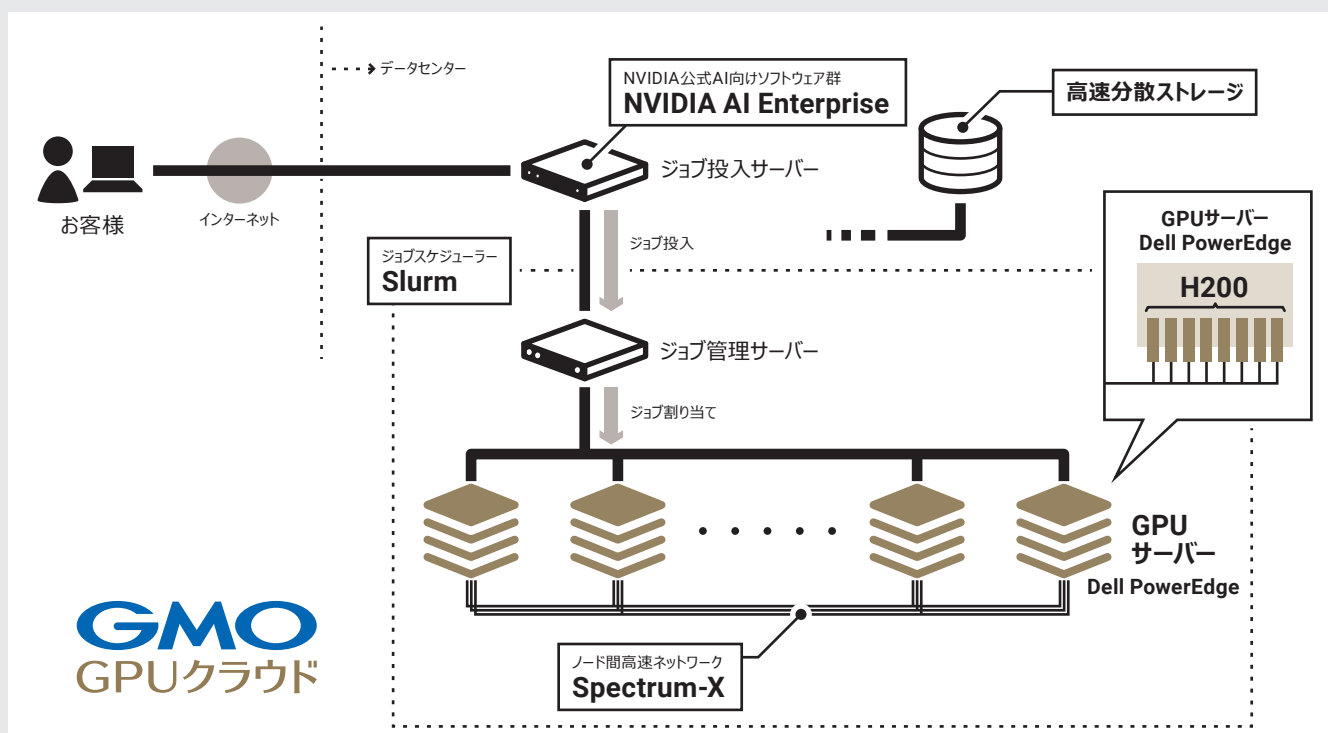
いく上で、こうした仕組みが整っていることは大変ありがたい。デル・テクノロジーズは NVIDIA 社との関係も密接で、NVIDIA 関連のパッチが遅滞なく提供される点も助かっています」と満足感を示す。

このように多くの成果を収めた同社だが、今後も GMO GPU クラウドをより進化させていく考えだ。「まずは、サーバーの台数を現在の 96 台からさらに増やしたい。また、NVIDIA のロードマップでは次世代 GPU も発表されていますので、こちらに向けた検討も並行して進めていきます。ちなみに、現在の

ワークロードは学習がメインですが、今後はモデル最適化や推論の分野にも範囲を広げ、AI クラウドサービスへと発展させていきたい」と武田氏は展望を語る。

「今回、国内最速レベルの GPU クラウドを実現できたのも、当社のサービスを支える大量のサーバー構築や運用、セキュリティのノウハウを連綿と積み上げてきた歴史があるからこそ。このスキル、ノウハウを最大限に活かし、今後お客様にとって最適なサービスをいち早くご提供して参ります」と力強く語る牧田氏。その取り組みを、デル・テクノロジーズも全力で支えていく。

GMO GPUクラウドのシステム構成



デル・テクノロジーズ ソリューションの詳細はこちら

この記事を共有する



DELL Technologies

Copyright © 2025 Dell Inc. or its subsidiaries. All Rights Reserved. Dell Technologies, Dell, EMC の製品およびサービスにかかるその他の商標は米国 Dell Inc. またはその子会社の商標または登録商標です。その他の社名および製品名は各社の商標または登録商標です。本事例は情報提供のみを目的としています。事例内容および事例内に記述された役職名は 2025 年 4 月に行われた取材時のものです。デル・テクノロジーズは、本事例の表現または暗示された内容にいかなる保証もいたしません。

デル・テクノロジーズ株式会社
<https://www.dell.com/ja-jp>