

2019年3月28日

報道関係各位

GMO インターネット株式会社

機械学習・ディープラーニング向け GPU サーバー

「GPU クラウド byGMO」提供開始

～GMO インターネットグループ内の GPU 環境整備と研究開発・サービス活用も加速～

GMO インターネット株式会社（以下、GMO インターネット）は、機械学習やディープラーニング/AI（人工知能）分野などに求められる高速演算処理を行える、NVIDIA 社の最新 GPU「Tesla® V100」を搭載したクラウドサービス「GPU クラウド byGMO (<https://gpu.cloud.gmo.jp>)」を、本日 2019 年 3 月 28 日（木）より提供開始いたします。

また GMO インターネットでは、AI や機械学習を用いた先端技術への取り組みの一環として、GPU を搭載したインフラ基盤を順次整備し、GMO インターネットグループにおけるネット証券、EC、ネット広告をはじめとする AI や機械学習と親和性の高い関連事業分野での技術研究やサービス活用を進めてまいります。

これに合わせ、GPU 環境をまとめて保有することによるリソースの有効活用として、GMO インターネットグループのエンジニア職は、「社内エンジニア制度」を通じていつでも GPU リソースを利用可能といたします。エンジニアが使いたいときに使える環境を社内に整えることで、研究やサービスへの活用スピードが早まることが期待できるうえ、AI や機械学習などだけでなく、GPU 環境を活用した新しいアイデアを生み出す土壌を作りたいと考えています。



【「GPU クラウド byGMO」とは】

GPU (Graphics Processing Unit) は、グラフィックス プロセッシング ユニット 高度な並列演算処理が求められる 3D グラフィックスや映像などの画像処理（レンダリング）のために開発され、昨今では機械学習やディープラーニング、AI など高度な計算処理が求められる場面でニーズが高まっているプロセッサ（演算装置）です。

「GPU クラウド byGMO」は、NVIDIA 社の最新 GPU「Tesla® V100」を採用し、加えて高速インターコネクト「InfiniBand EDR」と分散ファイルシステム「Lustre」を組み合わせた高速演算リソースをベアメタルサーバー^{※1}としてクラウド経由で提供するサービスです。お客様は、自社で設備投資する必要が無く、必要なときに使いたい分だけ計算（GPU）リソースの調達が可能となります。

また、「GPU クラウド byGMO」で採用している「Tesla® V100」は、1 基で最大 100CPU 分のパフォーマンスを発揮し、これまで難しいとされていた大量のデータ処理が求められるような課題に取り組むエンジニアや研究者を強力に支援するものです。

※1：物理サーバーのこと。

■ 「GPU クラウド byGMO」の主な特長

(1) 必要なときに使いたい分だけ使える GPU 環境

クラウドサービスなので導入に時間がかからず、お申し込みから最短当日に利用開始できます。また、追加で計算リソースを必要とする場合や、自社で保有している計算リソースの不足分を補うスポット利用など、柔軟にリソース調達が可能です。なお、サービス開始時は月額利用ですが、今後、時間課金のプランも提供開始する予定です。

(2) 開発環境は予めセットアップ。すぐ使える GPU 環境

GPU 向けのプログラミング環境として、「OpenCL」や NVIDIA 社の提供する「CUDA®」などのソフトウェア開発環境をセットアップした状態で提供します。また、コンテナエンジンに「Singularity」を採用することで、マルチノードによる大規模な学習に対応し、Docker イメージ^{※2}をそのままインポートすることも可能で、環境に合わせたセットアップを最小限に抑え、簡単に実行できます。

※2：Docker とは、1 つの OS 環境でありながらファイルを独立させることができるコンテナ型の仮想化環境を提供するオープンソースソフトウェアで、Docker イメージはそのコンテナを作るために必要となるファイルシステム。

(3) 高速インターコネクト「InfiniBand EDR」で広帯域/低レイテンシを実現

GPU の特徴を最大限に生かすためには、ファイルストレージおよび GPU サーバー間の通信において、広帯域かつ低レイテンシ^{※3}の高速通信を実現する必要があります。そこで、科学技術分野における並列計算機や HPC（ハイパフォーマンスコンピューティング）で利用される高速インターコネクト「InfiniBand EDR」を採用し、GPU サーバー間および GPU サーバー/ファイルストレージ間の通信において 100Gb/s の超高速転送を可能にしています。

※3：レイテンシとは、通信の遅延のこと。

(4) 分散ファイルシステム「Lustre」を採用

高い計算処理性能を発揮させるには、膨大な元データを格納しているファイルサーバーの能力も求められます。HPC や、ゲノム分野など、大規模な解析、シミュレーション用システムにおいては、こうした課題を解決するために、分散ファイルシステム^{※4}を採用しており、中でも「Lustre」は多くの利用実績を持ちます。「Lustre」を導入することにより、ファイルサーバーのデータへの GPU サーバーの高い読み込みと書き込みを可能にするとともに、数台~数万台の GPU サーバー間でのデータ共有を実現します。これにより、GPU サーバーでの大量のデータ処理を可能にします。

※4：分散ファイルシステムとは、データを複数のコンピューターにまたがって格納しながらも、仮想的に 1 つのように扱うことを可能にするファイルシステム。

(5) GPU 間の NVLINK 接続

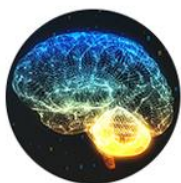
超高速ノード間接続を可能にする NVIDIA® NVLink™ テクノロジーにより PCIe^{※5}の帯域幅を拡張し、1 台のサーバーにつき最大 4 基の「Tesla®V100」を相互接続したシステム構成のプランをご用意しています。これは、通常の PCIe のみで相互接続した際の帯域幅と比べ約 5 倍の速度を実現しており、高いパフォーマンスを発揮します。

※5：PCI Express の略で、コンピューターのパフォーマンスを向上させるための拡張スロット標準仕様の一つ。

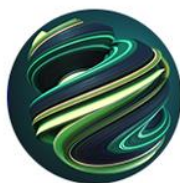
GMO インターネットは、「GPU クラウド byGMO」の提供を通じて、膨大な計算を必要とする機械学習やディープラーニング/AI 分野における重要課題の解決を支援します。

<主なユースケース>

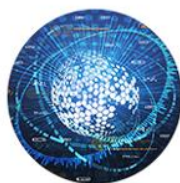
「GPU クラウド byGMO」は、高度な計算リソースの需要にお応えします。具体的には、ディープラーニングや機械学習など AI 技術分野をはじめ、創薬分野の分子動力学シミュレーションや応用シミュレーションなど、各種シミュレーション、科学技術計算などの利用に最適です。特に AI 技術分野では、業務の効率化や異常検知、マッチング、予測など、様々な産業において活用が期待されていることから、AI の需要の高まりとともに GPU サーバーのニーズが高まっています。また、医療用の 3D 解析ビューアーのリアルタイムでの高画質ボリュームレンダリングといった、高負荷な画像処理も可能です。



AI/ディープ
ラーニング



レンダリング



機械学習



FinTech



科学技術計算

■ 「GPU クラウド byGMO」の主な仕様とプラン

全 3 プランご用意しており、GPU の搭載数に応じてお選びいただけます。

プラン名	1GPU プラン	2GPU プラン	4GPU プラン	
初期費用	882,000 円	1,560,000 円	2,880,000 円	
(初期分割)	(73,500 円)	(130,000 円)	(240,000 円)	
月額費用	97,000 円	142,000 円	230,000 円	
GPU 型式	NVIDIA® Tesla® V100 SXM2			
GPU メモリ	16GB	32GB	64GB	
NVIDIA Tensor コア数	640	1,280	2,560	
NVIDIA CUDA®コア数	5,120	10,240	20,480	
CPU	Intel® Xeon® Gold 5122 4 コア 3.60GHz ×2			
OS	Ubuntu 16.04			
メモリ	192GB			
ストレージ	SSD480GB+Lustre ファイルストレージ			
設置場所	国内			
インターコネクト	InfiniBand EDR 100Gbps			
演算性能	単精度浮動小数点 数演算	約 15.7TFLOPS	約 31.4TFLOPS	約 62.8TFLOPS
	倍精度浮動小数点 数演算	約 7.8TFLOPS	約 15.6TFLOPS	約 31.2TFLOPS

※上記記載の内容は変更になる可能性があります。

【GPU クラウド byGMO】提供の背景と活用に向けた取り組み

昨今、GPU を搭載した GPU サーバーは、ディープラーニングや機械学習を含む AI の分野をはじめ、様々な産業分野から注目を集めています。

一方 GMO インターネットは、インターネット産業の黎明期から、多くの産業でのインターネットの普及と活用を推進してまいりました。現在では、インターネットインフラサービスにおいては国内シェア No.1、ご契約数も延べ 1,000 万件に達し、日本のインターネット産業の発展を支える基盤となっています。

こうした中、GMO インターネットでは、今後 GPU を用いたサービスが、AI 技術分野はもちろんのこと、ロボティクスや医療、自動運転、金融、レイトレーシングなど、各産業分野の発展に不可欠なテクノロジーを支える基盤になると考えています。

そこで GMO インターネットでは、膨大な計算を必要とするこれらの分野における課題を解決し、日本の産業を支える基盤となるべく、「GPU クラウド byGMO」を提供することといたしました。

GMO インターネットグループ内でも、すでに GMO クリックグローバルマーケティング株式会社が GPU クラウドを利用した金融向け研究に取り組んでいます。GMO インターネットでは、社内での技術研究やサービスへの活用を進めることで得たノウハウや知見を「GPU クラウド byGMO」に投入し、よりお客様のサービスや研究に GPU を利活用しやすい環境を提供してまいります。

【GMO インターネットグループにおける「GPU インフラ基盤」について】

GMO インターネットグループでは、アドテクノロジーを活かした広告配信サービスや、世界 No.1 の取引高を誇る FX 取引サービス、有料ショッピングサイト数も国内最大規模の EC サイト構築支援サービスを展開しています。また、機械学習の研究にも力を入れており、これらの領域において、GPU クラウドを導入し積極的に活用してまいります。

また、GMO インターネットグループのエンジニア向けにも自由に使える環境を用意することで、積極的に活用を促し、データサイエンティストの育成や、新技術に触れることによる新しいアイデアやビジネスの創出につなげていきたいと考えています。

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
グループ広報・IR 部 石井
TEL : 03-5456-2695 E-mail : pr@gmo.jp

【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
クラウド事業部 高田
E-mail : info@gpu.cloud.gmo.jp

【GMO インターネット株式会社】 (URL : <https://www.gmo.jp/>)

会社名	GMO インターネット株式会社 (東証第一部 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	代表取締役会長兼社長・グループ代表 熊谷 正寿		
事業内容	■インターネットインフラ事業	■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	■仮想通貨事業	
資本金	50 億円		