NEWS Release

報道関係者各位

株式会社ジーデップ・アドバンス
2024 年 7 月 29 日

NVIDIA B200 搭載 GPU アプライアンスサーバー 「NVIDIA® DGX™ B200」受注開始のお知らせ



AI とビジュアライゼーションのソリューションカンパニーである株式会社ジーデップ・アドバンス（GDEP Advance,Inc. 本社：東京都中央区、代表取締役 CEO：飯野匡道、東証スタンダード 証券コード：5885）は、2024 年 7 月 29 日（月）より、NVIDIA B200 GPU 搭載 GPU アプライアンスサーバー「NVIDIA® DGX™ B200」の受注を開始いたします。

NVIDIA B200 GPU 搭載 第 6 世代 ラックマウント型 DGX システム「NVIDIA DGX B200」

NVIDIA DGX B200 は、2024 年 3 月 NVIDIA 主催イベント GTC2024（米国カリフォルニア州サンノゼ開催）で発表された AI モデルのトレーニング、ファインチューニング、推論のための統合 AI スーパーコンピューティングプラットフォームです。

NVIDIA DGX B200 は、Blackwell アーキテクチャを採用した NVIDIA B200 Tensor コア GPU が 8 基搭載され、Blackwell Transformer Engine が 4 ビット浮動小数点（FP4）精度で、最大 144PFLOPS の AI パフォーマンスを実現します。また、GPU メモリは DGX H100 の 2.25 倍にあたる 1440GB へ大幅に増強され、64TB/秒の広帯域なバンド幅により 1 兆パラメータモデルのリアルタイム推論が Hopper 世代と比較して最大 15 倍高速化されます。

さらに、8 基の NVIDIA ConnectX™-7 HCA と 2 基の BlueField®-3 DPU を備え、ネットワーク IO モジュール毎に最大 400 ギガビット/秒のスループットをサポートし、NVIDIA Quantum-2 InfiniBand や、NVIDIA Spectrum™-X Ethernet™ ネットワーキングプラットフォームによって、より高速な相互接続とよりインテリジェントなネットワークを実現します。

NVIDIA DGX B200 は、自然言語処理やディープラーニングによるレコメンデーションモデルなど、大規模で複雑な AI ジョブに対応できるように強化されているだけではなく、開発と展開のための様々なソフトウェアスタックがオプション提供されています。「NVIDIA AI Enterprise」に含まれる事前トレーニング済みの NVIDIA 基盤モデル、フレームワーク、ツールキット、及び NVIDIA NIM マイクロサービスを使用することで推論作業を大幅に加速します。また、NVIDIA DGX プラットフォームを強化する「NVIDIA Base Command」には、AI ジョブスケジューリングとワークロードオーケストレーション、包括的なクラスター管理、コンピューティング、ストレージ、ネットワークインフラストラクチャを高速化するライブラリ、AI ワークロードの実行に最適化されたシステムソフトウエアが含まれます。

究極の AI パフォーマンス

革新的な NVIDIA Blackwell アーキテクチャによる高度なコンピューティングの NVIDIA DGX B200 は、前世代の NVIDIA DGX H100 と比較してトレーニング性能が 3 倍、推論性能が 15 倍の最先端の性能を提供します。NVIDIA Blackwell GPU アーキテクチャを採用した DGX B200 は、大規模言語モデル、レコメンダーシステム、チャットボットなどの多様なワークロードを処理することが可能であり、AI トランスフォーメーションの加速を目指すパワーユーザーに最適です。

NVIDIA DGX B200　製品特徴

◆NVIDIA B200 GPU x8、合計 GPU メモリ 1,440GB の大容量 GPU メモリ
72 PFLOPS の FP8 トレーニングと 144 PFLOPS の FP4 推論性能

◆第 5 世代の NVIDIA NVLink®により、最大 18 本の NVLink 100GB/s の接続で GPU 間の双方向帯域幅の合計は 1.8TB/s（テラバイト毎秒)性能

◆NVIDIA ConnectX-7 NIC 及び、BlueField-3 DPU 400Gb/秒ネットワークインターフェイス x2
接続ごとに最大 400 ギガビット/秒の帯域幅を提供、ピーク時の双方向ネットワーク帯域幅 1.6TB/秒で高速な AI パフォーマンスを実現

◆第 5 世代インテル® Xeon®プロセッサと最大 4.0TB システムメモリ
AI への依存が非常に高い仕事を可能にするパワフルな CPU

◆30TB NVMe ストレージ
最高のパフォーマンスを実現するための高速ストレージ

◆ソフトウエア
DGX OS / Ubuntu ベース – オペレーティングシステム

製品ページ URL：https://www.gdep.co.jp/products/list/v/66016bbc87594/

NVIDIA DGX B200 仕様

GPU	NVIDIA Blackwell GPU（NVIDIA Blackwell アーキテクチャ）
GPU メモリ	合計 1,440GB（HBM3e 180GB x 8）

NVIDIA NVSwitch	NVIDIA SHARP™
NVIDIA NVLink	第 5 世代 NVLink 1,800GB/秒
演算性能	FP4 Tensor Core / 144PFLOPS FP8 / FP6 / INT8 / 72PFLOPS
CPU	インテル第 5 世代 Xeon プロセッサ Platinum 8570 112 コア l, 2.1 GHz（Base), 4 GHz（Max Boost)
システムメモリ	2TB～4TB
ネットワーク	10Gb/s オンボード NIC with RJ45 100Gb/s デュアルポート イーサネット NIC 管理ポート（BMC)with RJ45
高速ネットワーク InfiniBand / Ethernet	ノード：4x OSFP ポート 8x シングルポート NVIDIA ConnectX-7 VPI 最大 400Gb/s ストレージ：2x QSFP 112 デュアルポート NVIDIA BlueField-3 DPU 最大 400Gb/s
ストレージ	システム：2x 1.9TB NVMe M.2（内蔵） データ：前面 8 ベイ x 3.84TB NVMe U.2
冷却方式	空冷
電源	200V 入力 3300W PSU 6 基リダンダント設計
最大消費電力	最大 14.3kW
寸法	10U（ラックマント）H 44.4cm x W 48.2cm x D 89.71cm
動作環境温度	5–30°C（41–86°F）
ソフトウエア	DGX OS

NVIDIA DGX SuperPOD へのスケールアウト

NVIDIA Blackwell GPU を搭載した NVIDIA DGX B200 は、NVIDIA DGX BasePOD 及び NVIDIA DGX SuperPOD といった AI データセンターインフラストラクチャプラットフォームを構成することが可能です。DGX SuperPOD は、生成 AI ワークロードの要求を満たすために、高性能ストレージと統合された完全なデータセンター規模の AI スーパーコンピューターです。メーカーにおいて全てのノードとネットワークスイッチをケーブル接続し、構築及びテストされるためデータセンターでの展開が大幅に高速化されます。

ジーデップ・アドバンスでは、NVIDIA DGX B200 の導入を検討されている開発者ユーザーの為に PoC 環境からのシステム設計・提案、導入・設置、データ運用支援等、当社独自の「GDEP フルスタック」をご用意しております。ファシリティ環境にてシステム導入が困難な場合、データセンターのハウジングサービスをセットにしたプライベートクラウドの形態でもご提供させていただきます。
GDEP フルスタックページ URL：https://www.gdep.co.jp/products/list/v/6672a2bbaa47b/

【ジーデップ・アドバンスについて】

ジーデップ・アドバンスは、「Advance with you」をミッションに、GPGPU をはじめとするアクセラレーターやハイエンドワークステーション、広帯域ネットワークや高速ストレージを用いたクラスターシステム、さらにライブラリやコンパイラ、ジョブスケジューラなどの運用ツールの提供構築から運用支援まで、仕事や研究を前に進めるための手段をオンプレミスやレンタル、クラウドなどあらゆる形態で総合的に提供する AI とビジュアライゼーションのソリューションプロバイダです。NVIDIA 社の「NPN（NVIDIA Partner Network）」においてエリートパートナーとして活動しています。

URL：https://www.gdep.co.jp/

※記載されている会社名、製品名等は各社の登録商標あるいは商標です。

※記載されている仕様は、NVIDIA 社での発表時点での仕様であり、変更される場合があります。

■製品に関するお問い合わせ先

株式会社ジーデップ・アドバンス ソリューション部

E-mail：sales@gdep.co.jp

■報道関係の方からのお問い合わせ先

株式会社ジーデップ・アドバンス 担当：中野

E-mail：news@gdep.co.jp