

報道関係各位

株式会社ジーデップ・アドバンス

2023年8月28日

次世代データセンターワークロードに最適な「NVIDIA L40S」搭載モデル 販売開始のお知らせ

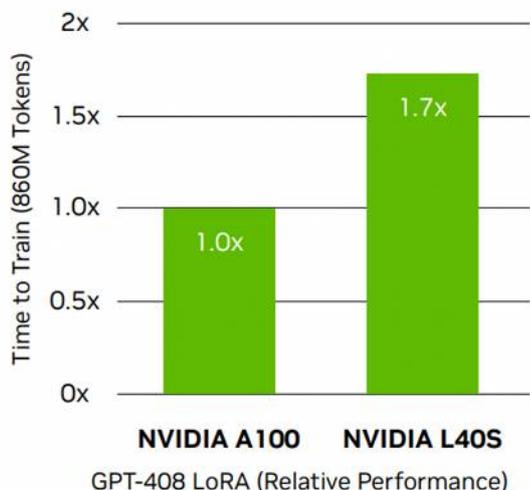


AI とビジュアライゼーションのソリューションカンパニーである株式会社ジーデップ・アドバンス (GDEP Advance, Inc. 代表取締役社長 飯野国道、東証スタンダード : 5885) は、米国 NVIDIA 社より 2023 年 8 月 8 日 (米国太平洋夏時間)、SIGGRAPH (正式名 : Special Interest Group on Computer GRAPHics) 2023 で発表された、強力なユニバーサル データセンター プロセッサ 「NVIDIA L40S」 を搭載したオリジナルのサーバーモデルの受注を開始いたします。

NVIDIA L40S の特徴

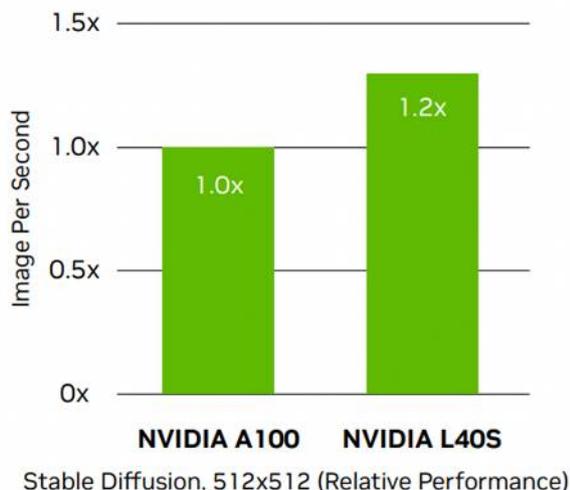
NVIDIA L40S GPU は NVIDIA Ada Lovelace GPU アーキテクチャに基づいて設計され、第 4 世代 Tensor コアと FP8 Transformer Engine が搭載されており、1.45 ペタフロップスを超える Tensor 処理能力を実現します。数十億のパラメータとテキストやビデオなどの複数のデータ モダリティを含む複雑な AI ワークロードの場合、NVIDIA L40S は、NVIDIA A100 Tensor コア GPU と比較して、最大 1.7 倍の生成 AI トレーニングパフォーマンスと最大 1.2 倍の推論パフォーマンスを実現します。

AI Training



Fine-tuning LoRA (GPT-40B): global train batch size: 128 (sequences), seq-length: 256 (tokens). NVIDIA HGX™ A100 8-GPU vs. two systems with 4x L40S GPUs. Performance based on prerelease build; subject to change.

Generative AI



Stable Diffusion v2.1. Relative speedup for 512 x 512 resolution image generation. NVIDIA HGX A100 8-GPU vs. two systems with 4x L40S GPUs. Performance based on prerelease build; subject to change.

GPU リソースの安定確保に向けて

現在、世界的なGPUの急激な需要増加によってNVIDIA H100やNVIDIA A100などのデータセンターGPUの納期が非常に長期化している状況です。ジーデップ・アドバンスではNVIDIA A100を凌駕したAIトレーニングパフォーマンスを誇る新製品 NVIDIA L40S を搭載した製品ラインナップを増強し、多彩なポートフォリオでお客様のご要望にお応えして参ります。

NVIDIA L40S 製品仕様比較

	L40S	A100 80GB SXM
GPU Architecture	NVIDIA Ada Lovelace	NVIDIA Ampere
FP64	N/A	9.7 TFLOPS
FP32	91.6 TFLOPS	19.5 TFLOPS
RT Core	212 TFLOPS	N/A
TF32 TensorCore*	366 TFLOPS	312 TFLOPS
FP16/BF16 Tensor Core*	733 TFLOPS	624 TFLOPS
FP8 Tensor Core*	1466 TFLOPS	N/A
INT8 Tensor Core*	1466 TOPS	1248 TOPS
GPU Memory	48 GB GDDR6	80 GB HBM2e
GPU Memory Bandwidth	864 GB/s	2039 GB/s
L2 Cache	96 MB	40 MB

Power	Up to 350 W	Up to 400 W
Form Factor	2-slot FHFL	8-way HGX
Interconnect	PCIe Gen4 x16: 64 GB/s	PCIe Gen4 x16: 64 GB/s

なお、ジーデップ・アドバンスでは、アカデミックユーザー 様 / 「NVIDIA Inception Program」パートナー企業 様向けに特別価格でご提供しております。

また、お客様のご予算や利用上に合わせてクラウドやレンタルでの提供も可能です。詳しくは、弊社 HP もしくは営業担当までお問い合わせください。

以上

ジーデップ・アドバンスについて

株式会社ジーデップ・アドバンス（東証スタンダード：5885）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 社パートナー認定制度「NPN (NVIDIA Partner Network)」においてエリートパートナー認定を受けて活動しています。

株式会社ジーデップ・アドバンス ウェブサイト：

<http://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。

■製品に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 TEL : 03-6803-0620 ※電話受付(平日)9:00~17:00

■報道関係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：小島 広

E-mail : news@gdep.co.jp TEL : 03-6803-0620 FAX : 03-6803-0450