

2023年8月9日

各位

菱洋エレクトロ株式会社

菱洋エレクトロ、NVIDIA データセンターGPU の新製品、 「NVIDIA L40S」の取り扱いを開始

エレクトロニクス商社の菱洋エレクトロ株式会社（本社：東京都中央区、代表取締役社長執行役員：中村 守孝、以下「菱洋エレクトロ」）は、アクセラレーテッドコンピューティングで世界をリードする NVIDIA が 8 月 8 日（アメリカ時間）に SIGGRAPH^{*1} 2023 で発表した、強力なユニバーサルデータセンタープロセッサである「NVIDIA L40S」の取り扱いを開始いたします。



NVIDIA L40S

生成 AI の爆発的な普及により、あらゆる業界で大規模なコンピューティングリソースを導入する必要性が高まっています。パフォーマンス、効率、ROI^{*2}向上のため、現代のデータセンターには、複雑なワークロード、高速コンピューティング、グラフィックス、ビデオ処理機能を提供する技術が必要です。

NVIDIA L40S は、データセンター向けの最も強力なユニバーサル GPU であり、推論とトレーニング、グラフィックス、およびビデオアプリケーションに対して高性能かつ高速な処理を実現します。マルチモーダル^{*3}生成 AI のプレミアムプラットフォームとして、推論、トレーニング、グラフィックス、ビデオワークフローのエンド to エンドでの高速処理を可能にし、次世代の AI 対応オーディオ、音声、ビデオ、2D および 3D アプリケーションを強化します。

^{*1} Special Interest Group on Computer GRAPHics。アメリカコンピューター学会におけるコンピュータグラフィックスを扱う分科会で、同分科会が主催する国際会議・展覧会の通称。

^{*2} Return On Investment（投資利益率）。事業や施策において、投下した資本に対するの収益性を図る指標のこと。

^{*3} Multi（複数）と Modal（様式）を組み合わせたコンピューター用語。さまざまな種類の情報を利用して高度な判断を行う AI を指す。

【製品の特長】

1. 強力な AI パフォーマンス

第4世代 Tensor コアと Transformer Engine、新しい半精度性能 (FP8) フォーマットは、NVIDIA A100 Tensor Core GPU の推論性能を最大 1.5 倍上回ります。18,176 個の NVIDIA Ada Lovelace GPU アーキテクチャー CUDA® コアを搭載し、A100 の約 5 倍の単精度性能 (FP32) を提供します。

2. 次世代グラフィックス機能

142 基の第3世代 RT コアと 48GB の GDDR6 メモリーにより、NVIDIA Ampere アーキテクチャー世代の最大 2 倍のリアルタイムレイトレーシング性能を実現し、美しく詳細なモデルやシーンの作成や、より忠実なグラフィックスを可能にします。

3. 企業向けデータセンター対応

24 時間 365 日稼働する企業のデータセンター向けに最適化されています。NVIDIA が設計、構築、テスト、サポートすることで、最大限の連続稼働時間とセキュリティを保証します。

※製品仕様は最終ページの添付資料をご覧ください。

※製品情報・出荷開始時期などの最新情報は以下 URL をご覧ください。

<https://ryoyo-gpu.jp/products/data-center/l40s/>

※NVIDIA は、米国またはその他の国における NVIDIA Corporation の商標または登録商標です。

※本リリースに記載されている情報は発表日現在のものです。時間の経過あるいは後発的なさまざまな事象によって内容に変更が生じる可能性があります。あらかじめご了承ください。

菱洋エレクトロ株式会社について <https://www.ryoyo.co.jp/>

半導体/デバイス事業と ICT/ソリューション事業を手掛けるエレクトロニクス商社として、両事業を展開する強みを活かして IoT・AI 導入などを実現するビジネス分野に注力しています。ICT/ソリューション事業では、既存 ICT 製品の販売に加え、コンサルティング、組み込み機器開発、設置、運用保守などのサービスを展開し、あらゆる産業分野における IT 環境の構築をまとめて支援することで、お客様のスムーズな課題解決に取り組んでまいります。

【製品に関するお問い合わせ先】

菱洋エレクトロ株式会社
ソリューション事業本部
ソリューション第5ビジネスユニット
TEL : 03-3546-6211
E-mail : nvidia_ws_info@ryoyo.co.jp

【リリースに関するお問い合わせ先】

菱洋エレクトロ株式会社
企画本部
広報部
TEL : 03-3546-6331
E-mail : pr@ryoyo.co.jp

【製品仕様】※添付資料

アーキテクチャー	NVIDIA Ada Lovelace アーキテクチャー
GPU メモリー	48GB GDDR6
メモリアンターフェース	384bit
メモリー帯域幅	864GB/s
エラー訂正コード(ECC)	対応
NVIDIA Ada Lovelace アーキテクチャーベースの CUDA コア	18,176
NVIDIA 第4世代 Tensor コア	568
NVIDIA 第3世代 RT コア	142
FP32	91.6
FP32 Tensor コア	183/366 ※1
BFLOAT16 Tensor コア	366.5/733 S ※1
FP8 Tensor コア	733/1466 ※1
INT8 Tensor コア	733/1466 ※1
INT4 Tensor コア	733/1466 ※1
システムインタフェース	PCI Express 4.0 x16
消費電力	350W
電源コネクタ	16 ピン
サーマルソリューション	パッシブ
フォームファクター	H4.4 x L10.5 Dual Slot
ディスプレイコネクタ	4x DisplayPort 1.4a
最大同時ディスプレイ表示	4x 4096 x 2160 @ 120Hz 4x 5120 x 2880 @ 60Hz 2x 7680 x 4320 @ 60Hz
エンコード/デコードエンジン	3x エンコード、3x デコード (+AV1 エンコードとデコード)
vGPU ソフトウェア対応	NVIDIA Virtual PC/Virtual Applications (vPC/vApps), NVIDIA RTX Virtual Workstation (vWS)
グラフィックス APIs	DirectX 12, Shader Model 6.6, OpenGL 4.6, Vulkan 1.3
コンピュータ APIs	CUDA 12.0, OpenCL 3.0, DirectCompute
MIG サポート	非対応
NVIDIA NVLink®	非対応
サーバーオプション	1-8 基 GPU 搭載のパートナーおよび NVIDIA Certified Systems™ 対応システム

※1 スパース性を使用。