



NVIDIA L4

ビデオ、AI、グラフィックスのワークロードを加速

NVIDIA Ada Lovelaceアーキテクチャを搭載したNVIDIA L4 Tensor コア GPUは、ビデオ、AI、ビジュアル コンピューティング、グラフィックス、仮想化などのためのユニバーサルでエネルギー効率の高いアクセラレーションを提供します。

1スロット厚でロープロファイルサイズのフォームファクターにパッケージされたL4は、エッジからデータセンター、クラウドまで、あらゆるサーバーで高いスループットと低遅延を実現する費用対効果、エネルギー効率に優れたソリューションです。

リアルタイムAIビデオ パイプラインの性能を体験

NVIDIA L4のパワーでビデオアプリケーションを変革しましょう。L4を搭載したサーバーは、数百万人の視聴者にライブ配信したり、ユーザーがクリエイティブなストーリーを構築したり、没入感のある拡張現実/仮想現実の体験を提供する場合でも、モバイル ユーザー向けに720p30で最大1040の同時AV1ビデオストリームをホストすることができます。第4世代のTensorコアと1.5倍のGPUメモリを備えたNVIDIA L4 GPUとCV-CUDA™ライブラリの組み合わせは、ビデオコンテンツの理解を新しいレベルに引き上げます。

L4は、エンドツーエンドのパイプライン全体でCPUベースのソリューションよりも120倍高いAIビデオパフォーマンスを実現し、企業がリアルタイムの洞察を得ることで、パーソナライズされたコンテンツの提供、検索の関連性の向上、好ましくないコンテンツの検出、スマートスペース ソリューションの実装を可能にします。

L4 で消費電力と設置スペースを削減

AIや動画が普及するにつれ、効率的でコスト効率の高いコンピューティングへの要求がこれまで以上に高まっています。NVIDIA L4 Tensor コア GPUは、最大120倍のAIビデオ性能を実現し、従来のCPUベースのインフラと比較して、最大99%のエネルギー効率の向上と総所有コストの削減を実現します。

これにより、企業はラックスペースを削減し、二酸化炭素排出量を大幅に削減でき、データセンターをより多くのユーザーに拡張できるようになります。2メガワット(MW) のデータセンターでCPUをNVIDIA L4に切り替えることで節約できるエネルギーは、2,000軒近い家庭の1年間の電力に相当し、10年間で172,000本の木を植えたときのカーボンオフセットに匹敵します。

グラフィックス性能の最適化

第3世代のRTコアとAIを搭載したNVIDIA Deep Learning Super Sampling 3 (DLSS 3)により、NVIDIA L4は AIベースのアバター、NVIDIA Omniverse™の仮想世界、クラウドゲーミング、仮想ワークステーション向けに4倍以上の高い性能を提供します。これらの機能により、クリエイターは、CPUでは不可能な没入感のあるビジュアル体験のために、リアルタイムで映画のような品質のグラフィックスやシーンを構築することができます。

効率的かつ持続的にワークロードを加速

NVIDIA L4は、NVIDIA データセンター プラットフォームにとって不可欠な存在です。ビデオ、AI、NVIDIA RTX™ 仮想ワークステーション (vWS)、グラフィックス、シミュレーション、データサイエンス、データ分析のために構築されたこのプラットフォームは、3,000以上のアプリケーションを加速し、データセンターからエッジ、クラウドまで、あらゆる場所で大規模に利用でき、性能とエネルギー効率を劇的に向上します。

メインストリームでの用途に最適化されたL4は、72Wの低消費電力エンベロープで動作する薄型のフォームファクターを提供し、NVIDIAのパートナーエコシステムのあらゆるサーバーやクラウド インスタンスのための効率的でコスト効率の良いソリューションです。

ジェネレーティブAIの性能を加速

画像やテキストのジェネレーティブAIは、あらゆる業界において、顧客の生活をより便利に、体験をより没入感のあるものにします。NVIDIA L4は、前世代のGPUと比較して最大2.5倍の性能向上を実現することで、計算負荷の高いジェネレーティブAI推論を超高速度化します。また、メモリ容量が50%増加したL4は、前世代のGPUでは不可能だった最大1024×768の大きな画像生成を可能にします。

NVIDIA L4

製品名	NVIDIA L4
製品型番	ETSL4-24GER
JANコード	4524076071543
技術仕様	
GPUアーキテクチャ	NVIDIA Ada Lovelaceアーキテクチャ
GPUメモリ	24GB GDDR6、ECC対応
メモリ帯域幅	300 GB/s
メモリI/F	192 bit
接続バスコネクタ	PCI Express 4.0 x16
CUDAコア数	7,424
Tensorコア数	232(第4世代)
RTコア数	58(第3世代)
FP32 TFLOPS	30.3
TF32 Tensorコア TFLOPS	120 ^{*1}
BFLOAT16 Tensorコア TFLOPS	242 ^{*1}
FP16 Tensorコア TFLOPS	242 ^{*1}
FP8 Tensorコア TFLOPS	485 ^{*1}
ピーク INT8 Tensor TOPS	485 ^{*1}
対応Compute API	CUDA 12.0 / DirectCompute / OpenCL 3.0
認証規格	WHQL / ISO9241 / EU RoHS / JIG / REACH / HF / WEEE / RCM / BSMI / CE / FCC / ICES / KC / cUL, UL / VCCI
ボード外形寸法	167.64mm(長さ)×68.9mm(高さ)、1スロット厚、ロープロファイル、ブラケットを含まず
最大消費電力	72W
補助電源コネクタ	不要
冷却方式	パッシブ
vGPUソフトウェア対応	NVIDIA Virtual PC/Virtual Applications (vPC/vApps)、NVIDIA RTX Virtual Workstation (vWS)
vGPUプロファイル対応	仮想GPUライセンスガイド を参照
NVENC NVDEC JPEGデコーダー	2x 4x 4x(NVENC/NVDECはAV1 エンコード/デコードを含む)
MIGサポート	なし
NVIDIA NVLinkサポート	なし
対応OS ^{*3}	Windows [®] Server 2022 / Windows [®] Server 2019 / Windows [®] 11 / Windows [®] 10 / Linux
付属品	製品保証書、ATXブラケット
保証期間	3年
パッケージ外形寸法 / 重量	295(W)×180(D)×55(H) mm / TBA
必要動作環境	NVIDIA認定システムのみ動作サポート。詳しくは NVIDIA認定システムカタログ を参照。

*1 新しいスパース性機能を利用した場合の実効値。スパース性無しの場合は1/2のスペックになります。*2 vGPU使用時。*3 全て64bit OSのみ対応。ドライバーバージョンにより対応するOSバージョンが異なる場合がございます。詳細は[NVIDIA CUDA Toolkit Release Notes](#)や、該当ドライバーのリリースノート内のSupported Operating Systemsをご参照ください。

お問い合わせ先
株式会社 エルザ ジャパン

<https://www.elsa-jp.co.jp>
〒105-0014 東京都港区芝3-42-10 三田UTビル
TEL:03-5763-7391

コーポレートサイト



オンライン見積



法人様向
貸出しサービス



最新事例一覧

