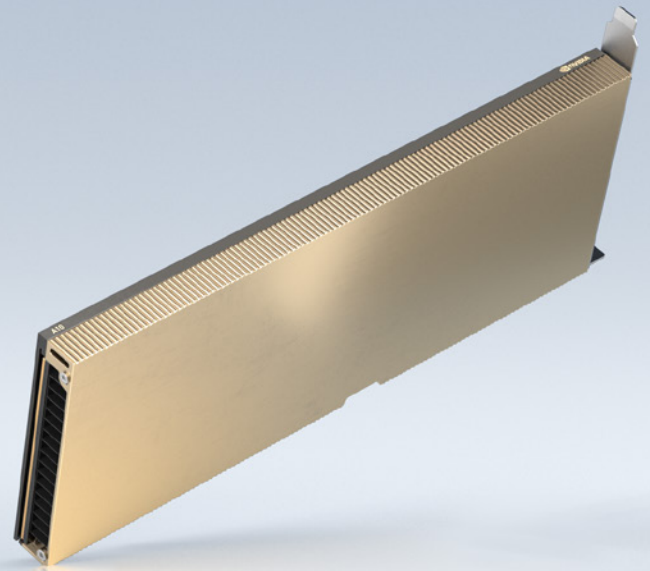




NVIDIA A10

メインストリーム エンタープライズ サーバー
向け AI対応ビデオ、高速グラフィックス



強力なAIでグラフィックスとビデオアプリケーションを充実

NVIDIA A10 Tensorコア GPUは、NVIDIA RTX仮想ワークステーション (vWS)ソフトウェアと組み合わせることで、メインストリームのエンタープライズサーバーでの主力となるグラフィックスとビデオのAIサービスが実現できます。デザイナー、エンジニア、アーティスト、サイエンティストが今日の課題に対応するために必要なソリューションを提供します。

最新のNVIDIA Ampereアーキテクチャに基づいて構築されたA10は、第2世代のRTコア、第3世代のTensorコア、および新しいストリーミングマイクロプロセッサを24ギガバイト(GB)のGDDR6メモリと組み合わせています。多様なグラフィックス、レンダリング、AI、およびコンピューティングパフォーマンス、世界中のどこからでもアクセスできる仮想ワークステーションから、レンダリングノードとしての利用など、さまざまなワークロードをデータセンターで実行できます。A10はシングルワイド、フルハイト/フルレングスのPCIeフォームファクターと150Wの電源で最適なパフォーマンスを提供します。

NVIDIA A10は、NVIDIA-Certified Systems™ (NVIDIA認証システム) の対象で、オンプレミスのデータセンター、クラウド、およびエッジ用途でサポートされます。NVIDIA A10は、NVIDIA NGC™カタログ、CUDA-X™ライブラリの他、230万人を超える開発者、および1,800を超えるGPU最適化アプリケーションから構成されるAIフレームワークの豊富なエコシステムに基づいており、企業がビジネスで最も重要な課題を解決するのに役立ちます。

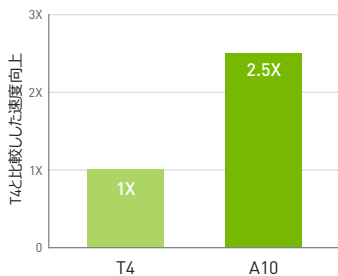
仕様

FP32	31.2 TF
TF32 Tensor コア	62.5 TF 125 TF*
BFLOAT16 Tensor コア	125 TF 250 TF*
FP16 Tensor コア	125 TF 250 TF*
INT8 Tensor コア	250 TOPS 500 TOPS*
INT4 Tensor コア	500 TOPS 1000 TOPS*
RT コア	72
エンコード/デコード	1 エンコーダー 1 デコーダー (+AV1 デコード)
GPU メモリー	24 GB GDDR6
GPU メモリー帯域幅	600 GB/s
インターコネク	PCIe Gen4: 64 GB/s
フォームファクター	1-スロット FHFL
最大消費電力 TDP	150W
vGPU ソフトウェア対応	NVIDIA vPC/vApps, NVIDIA RTX™ vWS, NVIDIA Virtual Compute Server (vCS)
ハードウェアの信頼基点 によるセキュア/メジャー ブート	Yes
NEBS Ready	Level 3
電源コネクタ	PEX 8-pin

*構造化バスを適用

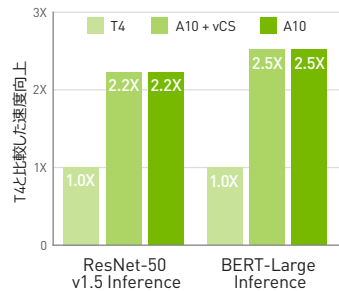
A10は、T4と比較して最大2.5倍
高速な仮想ワークステーションパ
フォーマンスを提供

SPECviewperf 2020



A10は、T4と比較して最大2.5倍の推論
パフォーマンスを提供

NVIDIA vCSは、ヘアメタルと同等のパ
フォーマンスを提供

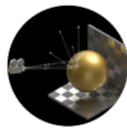


NVIDIA Ampere アーキテクチャの内部



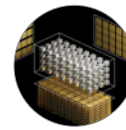
NVIDIA AMPERE アーキテクチャ CUDA コア

単精度浮動小数点 (FP32) 演算の倍速処理と改善された電力効率により、複雑な3Dコンピューター支援設計 (CAD) やコンピューター支援エンジニアリング (CAE) などのグラフィックスおよびコンピューティングワークフローでパフォーマンスが大幅に向上します。



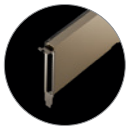
第2世代 RT コア

前世代の最大2倍のスループットと、シェーディングまたはノイズ除去機能のいずれかでレイトレーシングを同時に実行する機能を備えた第2世代のRTコアは、映画コンテンツのフォトリアルなレンダリング、建築デザインのチェック、仮想プロトタイピングなどのワークロードを大幅に高速化します。さらにこのテクノロジーは、レイトレースモーションブラーのレンダリングを高速化し、より高速な結果とより正確なビジュアルを実現します。



第3世代 TENSOR コア

Tensor Float 32 (TF32) の精度は、前世代の最大5倍のトレーニングスループットを提供し、コードを変更することなくAIおよびデータサイエンスモデルのトレーニングを加速します。構造的スパース性に対するハードウェアサポートは、推論のスループットを最大2倍にします。Tensor コアは、ディープラーニングスーパーサンプリング (DLSS)、AIノイズ除去、選択したアプリケーションの強化された編集などの機能を備えたグラフィックスにAIをもたらします。



24 GB GDDR6

超高速GDDR6メモリ。レンダリング、データサイエンス、エンジニアリングシミュレーション、およびその他のGPUメモリを大量に消費するワークロードに600 GB / sの帯域幅を提供します。



PCI EXPRESS GEN 4

PCI Express Gen 4は、PCIe Gen 3の帯域幅を2倍にし、AI、データサイエンス、3Dデザインなどのデータ集約型タスクのCPUメモリからのデータ転送速度を向上させます。より高速なPCIeパフォーマンスは、GPUダイレクトメモリアccess (DMA) 転送も高速化し、GPUとGPU Direct[®] for Videoに対応するデバイス間のビデオデータのより高速な入出力通信を提供し、ライブブロードキャストに強力なソリューションを提供します。A10は、展開の柔軟性のためにPCI Express Gen3と下位互換性があります。



データセンターの効率とセキュリティ

シングルスロット、フルハイト、フルレンジスの電力効率の高い設計を特徴とするNVIDIA A10は、世界中のOEMの幅広いサーバーと互換性があります。NVIDIA A10には、ハードウェアの信頼の基盤に基づいたセキュア、メジャーブートに対応しており、ファームウェアが改ざんされたり破損したりしないようになっています。

NVIDIA A10 Tensorコア GPUは、AIを活用するグラフィックスとビデオに最適です。第2世代RTコアと第3世代Tensorコアは、メインストリームサーバー用として、パワフルなAIと共に強力なグラフィックスとビデオアプリケーションを150Wの消費電力で実現します。

NVIDIA A10は、NVIDIA仮想GPU (vGPU) ソフトウェアと組み合わせて、VDIのグラフィックス性能を強化し、高性能な仮想ワークステーションからAIまで、データセンターのワークロードの高速化にスケラブルに対応します。

全てのディープラーニング フレームワーク

mxnet

PYTORCH

APACHE SPARK

TensorFlow

プロフェッショナル向けRTX対応アプリケーション



AUTODESK REVIT

CATIA

SOLIDWORKS



creo

Rhinoceros

design, model, present, analyze, realize...

SIEMENS

NVIDIA A10 GPUに関する詳しい情報は www.nvidia.com/ja-jp/data-center/products/a10-gpu

1 2x Xeon Gold 6154 3.0GHz (3.7GHz Turbo) 、NVIDIA RTX vWSソフトウェア、VMware ESXi 7 U2、ホスト/ゲストドライバー-461.33を搭載したサーバーでテスト実行 | SPECviewperf 2020サブテスト、およびHD3dsmax-07コンボジット。
2 BERT Large Inference NVIDIA TensorRT7.2、シーケンス長= 128、バッチサイズ= 128、NGCコンテナ : 21.02-py3 | ResNet-50 v1.5 : NVIDIA TensorRT7.2、INT8精度バッチサイズ= 128 NGCコンテナ : 20.12-py3 | vCSソフトウェアを搭載したNVIDIA A10、VMware ESXi 7U2ホスト/ゲストドライバー-461.33

© 2021 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, Certified Systems, CUDA, NGC, RTX, and GPU Direct are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. All other trademarks and copyrights are the property of their respective owners. MAR21

