

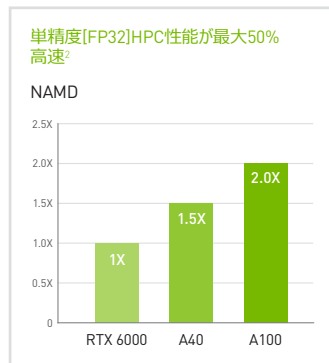
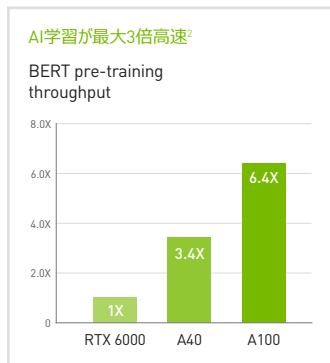
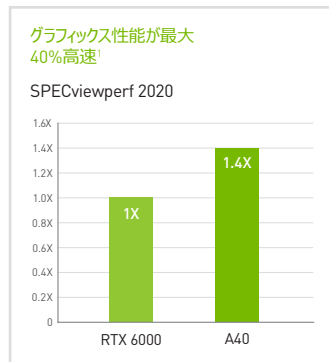
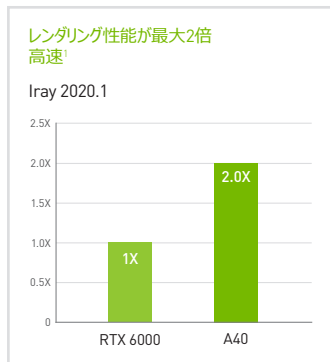
# NVIDIA A40

## ビジュアルコンピューティング向け パワフルなデータセンターGPU

NVIDIA A40は、最新のNVIDIA AmpereアーキテクチャのRTコア、Tensorコア、およびCUDA®コアと48 GBのグラフィックメモリを組み合わせ、最も要求の厳しいビジュアルコンピューティングワークロードのデータセンターでの処理を高速化します。NVIDIA A40は、どこからでもアクセスできる強力な仮想ワークステーションから専用のレンダリングノードまで、次世代のNVIDIA RTX™テクノロジーをデータセンターにもたらし、最先端のプロフェッショナル向けビジュアルワークロードを実現します。

### 仕様

GPU アーキテクチャ	NVIDIA Ampere アーキテクチャ
GPU メモリー	48 GB GDDR6 ECC付き
メモリー帯域幅	696 GB/s
インターコネクティブ	NVIDIA® NVLink® 112.5 GB/s (双方向)³ PCIe Gen4 31.5 GB/s (双方向)
NVIDIA Ampere アーキテクチャに基づく CUDAコア	10,752
NVIDIA 第2世代 RTコア	84
NVIDIA 第3世代 Tensor コア	336
ピークFP32 TFLOPS (Tensor無し)	37.4
ピークFP16 Tensor TFLOPS FP16 アキュムレート有り	149.7   299.4*
ピークTF32 Tensor TFLOPS	74.8   149.6*
RT コア 性能 TFLOPS	73.1
ピークBF16 Tensor TFLOPS FP32 アキュムレート有り	149.7   299.4*
ピークINT8 Tensor TOPS	299.3   598.6*
ピークINT4 Tensor TOPS	598.7   1,197.4*
フォームファクター	4.4" (H) x 10.5" (L) デュアルスロット
ディスプレイポート	3x DisplayPort 1.4**、NVIDIA Mosaic と Quadro® Sync <sup>+</sup> 対応
最大消費電力	300 W
電源コネクタ	8-pin CPU
サーマルソリューション	パッシブ
仮想 GPU (vGPU) ソフトウェアサポート	NVIDIA vPC/vApps, NVIDIA RTX 仮想ワークステーション, NVIDIA Virtual Compute Server
vGPU プロファイルサポート	<a href="#">仮想GPUライセンスガイドを参照</a>
NVENC   NVDEC	1x   2x (AV1 デコードを含む)
ハードウェアの信頼基点によるセキュア/メジャーアップデート	Yes
NEBS ready	Level 3
コンピュータAPIs	CUDA, DirectCompute, OpenCL™, OpenACC®
グラフィックス APIs	DirectX 12.07 <sup>5</sup> , Shader Model 5.17 <sup>5</sup> , OpenGL 4.68 <sup>6</sup> , Vulkan 1.18 <sup>6</sup>
MIG サポート	No



\* 構造化スパスを適用

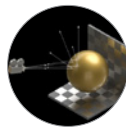
\*\* A40は、デフォルトで仮想化用に構成されており、物理ディスプレイコネクタは無効になっています。ディスプレイ出力は管理ソフトウェアツールを介して有効にできます。

# NVIDIA Ampere アーキテクチャの内部



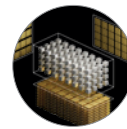
## NVIDIA AMPERE アーキテクチャ CUDA コア

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。



## NVLINK対応 48 GB GDDR6 メモリー

NVLINK<sup>3</sup>で最大96GBまで拡張可能な超高速GDDR6メモリーは、データサイエンティスト、エンジニア、クリエイティブ プロフェッショナルに、データサイエンスやシミュレーションなどの大規模なデータセットやワークロードを処理するために必要な大容量メモリーを提供します。



## 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メモリー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<sup>®</sup> for Video<sup>™</sup>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、ライブロードキャストに強力なソリューションを提供します。A40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

デュアルスロットの電力効率の高い設計を特徴とするNVIDIA A40は、前世代の最大2倍の電力効率であり、世界中のOEMの幅広いサーバーと互換性があります。NVIDIA A40には、ハードウェアの信頼の基点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 A40 GPUは、リアルタイムのレイトレーシング、AIアクセラレーション、ディープラーニング、データサイエンス、コンピューティングベースのワークロードを加速するマルチワークロードの柔軟性など、最先端のビジュアル コンピューティング機能を提供します。NVIDIA A40、NVIDIA RTX仮想ワークステーション (vWS)、およびNVIDIA Virtual Compute Serverソフトウェアを搭載した仮想ワークステーションは、最適なパフォーマンスと安定性を実現するために、幅広い業界アプリケーションとプロフェッショナル ソフトウェアにわたる広範な検証テストを実施しています。

### 全てのディープラーニング フレームワーク

mxnet

PYTORCH

SPARK

TensorFlow

### プロフェッショナル向けRTX対応アプリケーション

Adobe Premiere Pro

SOLIDWORKS

SIEMENS NX

AUTODESK ARNOLD



REDSHIFT

AUTODESK VRED

KeyShot

UNREAL ENGINE

blender

octane render

v-ray

NVIDIA A40 GPUに関する詳しい情報は、[www.nvidia.com/ja-jp/data-center/a40/](http://www.nvidia.com/ja-jp/data-center/a40/)

1 レンダリングおよびグラフィックステストは、2x Xeon Gold 6126 2.6GHz (3.7GHz Turbo)、256GBのシステムメモリー、NVIDIAドライバー461.09で実施。レンダリングテスト：Iray 2020.1、NVIDIA Endeavorシーンのレンダリング時間。グラフィックステスト：SPECviewperf 2020サブテスト、4K医療-03コンボットでの結果です。2 AIおよびHPCテストはAMD EPYC 7742@2.25GHz (3.4GHz Turbo)、512GBのシステムメモリー、NVIDIAドライバー460.14で実施、AIトレーニング：BERTの事前トレーニングスループット、PyTorch (2/3) フェーズ1および (1/3) フェーズ2。RTX6000用の高精度FP32およびA40およびA100用のTF32。フェーズ1のシーケンス長= 128。フェーズ2 = 512。単精度HPC：NAMDバージョン3.0a7、stmv\_nve\_cuda; Precision = FP32; ns / day、CUDAバージョン= 11.1.74の結果です。3 3つのNVIDIA A40カードをNVLINKに接続して、パフォーマンスとメモリー容量を96 GBに拡張できるのは、アプリケーションがNVLINKテクノロジーをサポートしている場合のみです。アプリケーションプロバイダーに連絡して、NVLINKのサポートを確認してください。4 Quadro SyncIIカードは別売りです。モザイクはWindows 10とLinuxでサポートされています。5 GPUはDX12.0 API、ハードウェア機能レベル12 +1をサポートします。6 製品は、公開されているKhronos仕様に基づいており、入手可能な場合はKhronos適合性テストプロセスに合格することが期待されています。現在の適合状況はwww.khronos.org/conformanceで確認できます。

© 2021 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, CUDA, GRID, GPUDirect, NVLink, OpenACC, Quadro, and RTX are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated. All other trademarks are property of their respective owners. FEB21

