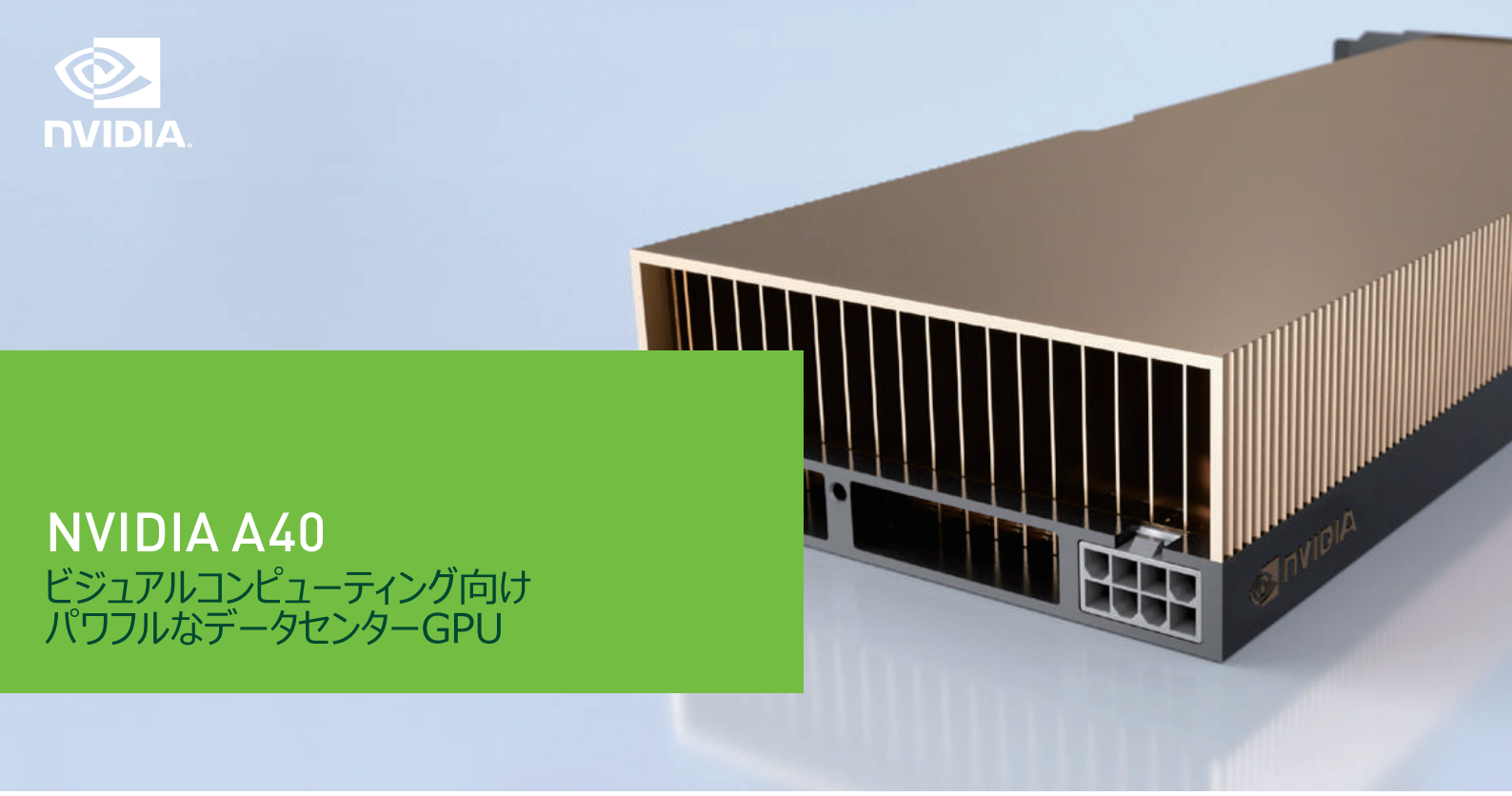
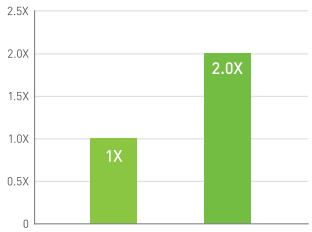


NVIDIA A40

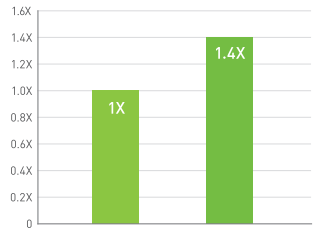
ビジュアルコンピューティング向け パワフルなデータセンターGPU

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て、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。

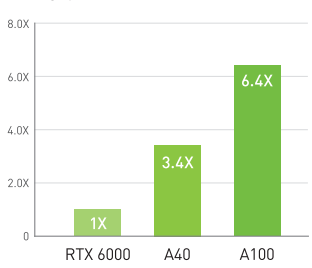
レンダリング性能が最大2倍高速[1]



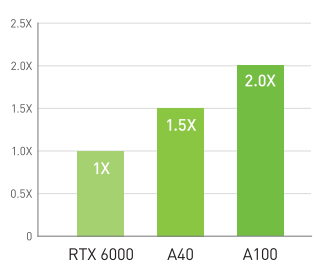
グラフィックス性能が最大40%高速[1]



AI学習が最大3倍高速[2]



単精度[FP32]HPC性能が最大50%高速[2]



仕様

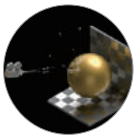
GPU アーキテクチャ	**NVIDIA Ampere アーキテクチャ**
GPU メモリー	**48 GB GDDR6 ECC付き**
メモリー帯域幅	**696 GB/s**
インターコネクト I/F	**NVIDIA® NVLink® 112.5 GB/s (双方向)[3] PCIe Gen4 31.5 GB/s (双方向)**
NVIDIA Ampere アーキテクチャに基づくCUDAコア	**10,752**
NVIDIA 第2世代 RTコア	**84**
NVIDIA 第3世代 Tensor コア	**336**
ピークFP32 TFLOPS (Tensor無し)	**37.4**
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 ピーク INT 4 Tensor TOPS	**299.3 | 598.6*** **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[4] 対応**
最大消費電力	**300 W**
電源コネクター	**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	**仮想GPUライセンスガイドを参照**
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/メジャーブート	**Yes**
NEBS ready	**Level 3**
コンピュートAPIs	**CUDA, DirectCompute, OpenCL™, OpenACC®**
グラフィックス APIs	**DirectX 12.07[5], Shader Model 5.17[5], OpenGL 4.68[6], Vulkan 1.18[6]**
MIG サポート	**No**

* 構造化スパースを適用
** 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

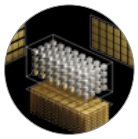
NVIDIA AMPERE アーキテクチャ CUDA コア

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[3]で最大96GBまで拡張可能な超高速GDDR6メモリは、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を提供します。

PCIE 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クセス（DMA）転送も高速化し、GPUとGPUDirect®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。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 Vrtual 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
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

Adobe Premiere Pro
SOLIDWORKS
PLM Software SIEMENS NX
AUTODESK ARNOLD
REDSHIFT
AUTODESK VRED
KeyShot by Luxion
UNREAL ENGINE
blender
octanerender
V-RAY

NVIDIA A40 GPUに関する詳しい情報は, **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、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コンポジット 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、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 2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Windows10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.krhonos.org/conformanceで確認できます。

HPC TECH | NVIDIA

NVIDIA ELITE PARTNER
HPCTECH 株式会社 HPCテック
https://www.hpctech.co.jp
〒103-0006 東京都中央区日本橋富沢町 7-13 洋和ビル 4 F
TEL:03-5643-2681 FAX:03-5643-2682
MAIL:info@hpctech.co.jp