

ELSA

TESLA 2018.04



NVIDIA[®] TESLA[®]

ビッグデータ解析 / 科学計算 / ディープラーニング
あらゆるGPUコンピューティングを加速

NVIDIA[®] TESLA[®] シリーズ

NVIDIA[®] TESLA[®] シリーズ 総合カタログ

NVIDIAが設計・製造・テストしています

NVIDIA® TESLA® シリーズ 総合カタログ

Accelerated data centers with Tesla GPUs

世界最速の PCIe ベースサーバ向け、 アドバンスドデータセンターアクセラレータ

NV

NVIDIA® TESLA® V100 は、今までにない 210 億トランジスタを積層した世界最大・最速のデータセンター向け
までで最速の性能を実現すると同時に、新たにディープラーニングに特化した「Tensor (テンソル) コア」を搭載す

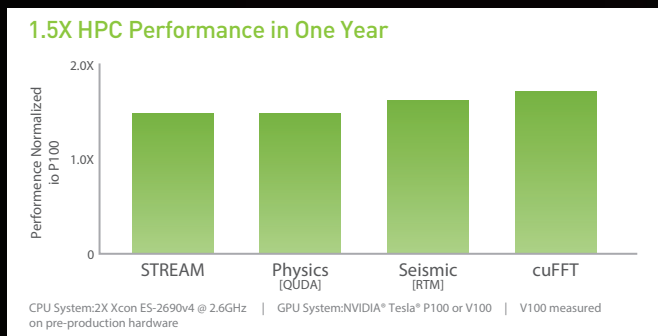
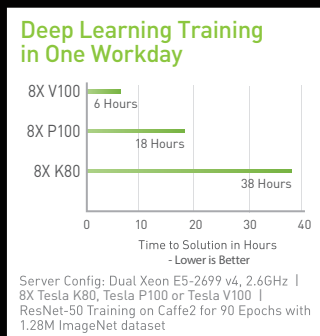
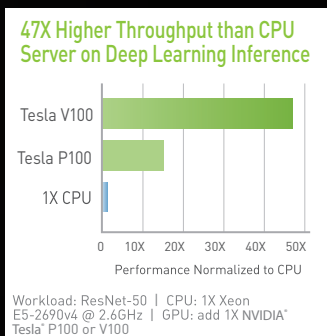


NVIDIA® TESLA® V100

GPU です。前世代 (P100) と比較して 1.43 倍の CUDA コアを搭載し、従来の科学技術計算などの分野でも今
ることで FP16 (半精度浮動小数点演算) 時の性能をトレーニング時で従来比 12 倍へ劇的に向上させています。

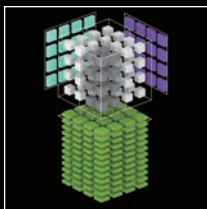
もっとも高性能なデータセンター向け GPU

NVIDIA® TESLA® V100 は世界一高性能なデータセンター用 GPU で、人工知能 (AI)、HPC、科学計算などのワークフローを加速します。最新の GPU アーキテクチャである「NVIDIA Volta」によって、最大で CPU100 個分のパフォーマンスを TESLA® V100 1 台で提供します。これによってデータサイエンティストやリサーチャー、技術者は、かつて不可能と思われていた挑戦に取り組むことが可能になりました。



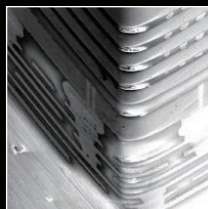
先駆的な技術革新

Tensor コア



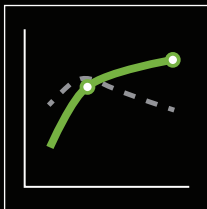
全体で 640 個の Tensor (テンソル) コアを搭載している TESLA® V100 は、ディープラーニングのパフォーマンスにおいて 112TFLOPS を実現しています。これは前世代 (P100) と比較するとディープラーニングのトレーニング時で 12 倍、推論時にも 6 倍の性能を発揮しています。

HBM2



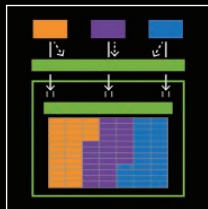
改良された 900GB/秒の RAW 帯域とさらに高効率化されたメモリコントローラーによって、TESLA® V100 は STREAM での測定において前世代 (P100) と比較してメモリ帯域が 1.5 倍に増加しました。

最大効率モード



この新しい最大効率モードによって、データセンターは既存の電力予算内でラックあたり最大 40% の計算処理能力を向上させることが可能になります。このモードにおいて TESLA® V100 は最大限の処理効率を発揮し、従来比で最大 80% の性能を半分の電力消費で提供することが可能になりました。

シンプルなプログラミング



NVIDIA® TESLA® V100 は簡単にプログラミングできるように徹底的に考慮して設計されています。中でも新しい独立スレッドスケジューリングは、極めてスムーズな同期を可能にし、また、細かいジョブに対してリソースをシェアすることで、GPU の利用率を改善します。

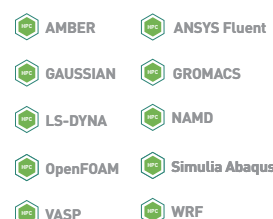
ディープラーニングと HPC における、最も高速で生産性の高い GPU アクセラレータ

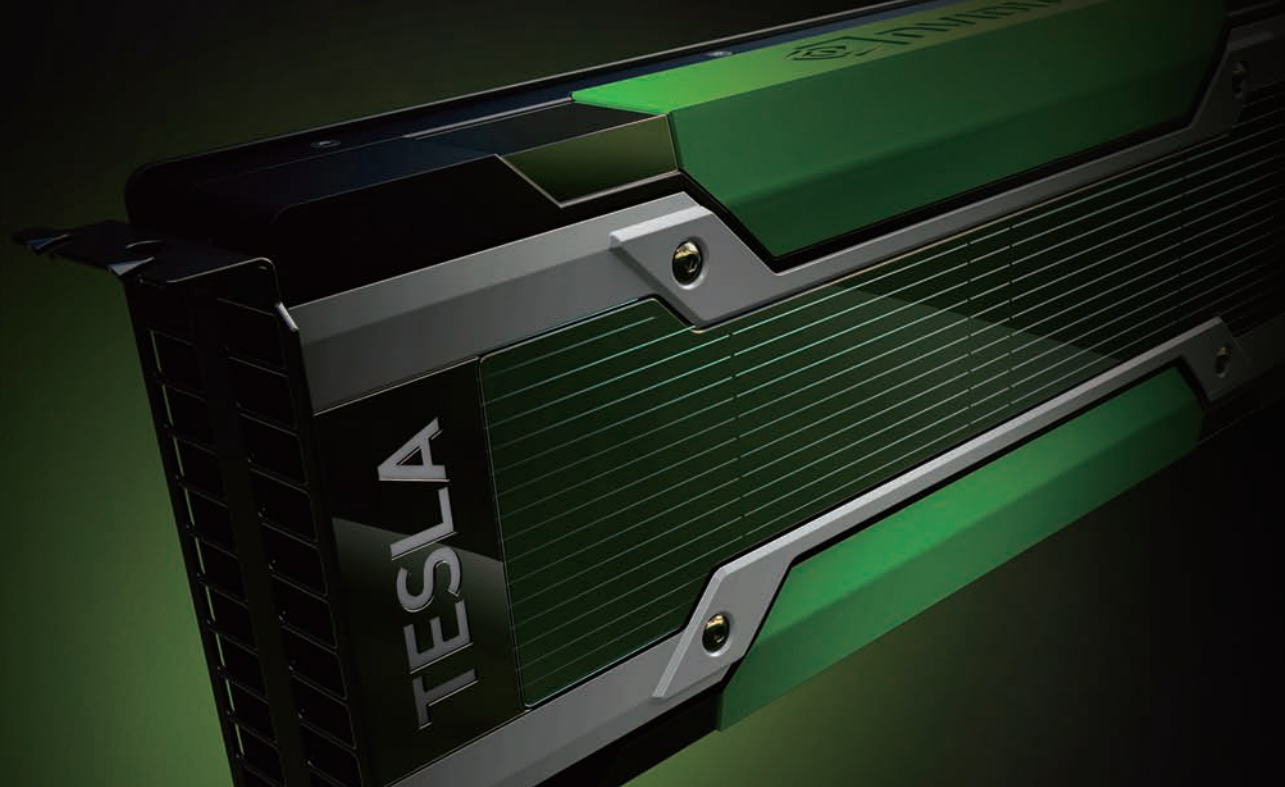
NVIDIA® TESLA® V100 は、ディープラーニング・HPC・科学計算向けデータセンターコンピューティングプラットフォームのフラッグシップ製品です。TESLA プラットフォームは 460 以上の HPC アプリケーションと主要なディープラーニングフレームワークの全てを加速させます。どこからでもデスクトップからサーバー、またはクラウドサービスにアクセスして利用することが可能で、劇的なパフォーマンス向上とコスト削減の両方を同時に実現できます。

EVERY DEEP LEARNING FRAMEWORK



500+ GPU-ACCELERATED APPLICATIONS





「Pascal」アーキテクチャベースの PCIe ベースサーバー向け
アドバンスドデータセンターアクセラレータ

NVIDIA® TESLA® P100

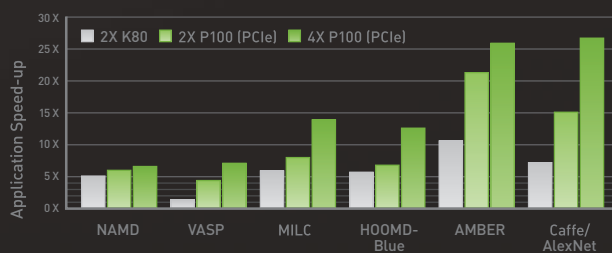
NVIDIA® TESLA® P100 は、「Pascal™」アーキテクチャに基づいた GPU を搭載した PCIe ベースサーバー向けのアドバンスドデータセンターアクセラレータです。従来の TESLA® M シリーズから一新された設計思想によって構成され、各所に革新的な技術を採用してパフォーマンスの更なる飛躍をもたらしました。

パフォーマンスの大きな飛躍

HPC 向けデータセンターは、科学者・研究者の更なる大きな要求を厳しい予算内でサポートする必要があります。多数の汎用的な計算ノードを相互接続して運用する従来のアプローチでは、台数に比例してリニアに性能を向上させることが難しく、実質的にコストの増大を招くとともに巨大なインターコネクトのオーバーヘッドが問題となっていました。

TESLA® P100 では、画期的かつ先進的な NVIDIA® Pascal™ アーキテクチャによってこれを解決。HPC やハイバースケールデータセンターのコストを削減し、より高いスループットを実現します。

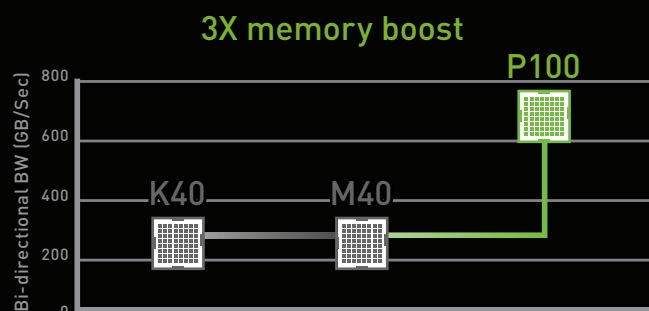
NVIDIA TESLA P100 for PCIe Performance



Dual CPU server, Intel E5-2698 v3 @ 2.3 GHz, 256 GB System Memory, Pre-Production Tesla P100

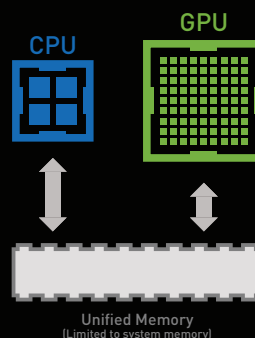
CoWoS HBM2

高速なメモリアクセスを実現する HBM2 (High Bandwidth Memory 2) を Pascal™ アーキテクチャに統合する「CoWoS (Chip-on-Wefer-on-Substrate)」によってメモリ帯域を劇的に向上させることに成功し、従来比 3 倍のメモリアクセス性能を実現しました。



PAGE MIGRATION ENGINE

プログラミングやパフォーマンスチューニングをさらにシンプルにします。これまでは GPU の物理メモリサイズが上限であったユニファイドメモリが、TESLA® P100 では物理メモリサイズに関わらず利用できるようになりました。



24GB の大容量 GDDR5 メモリを搭載 PCIe ベース GPU サーバー向けファンレスモデル

NVIDIA® TESLA® P40

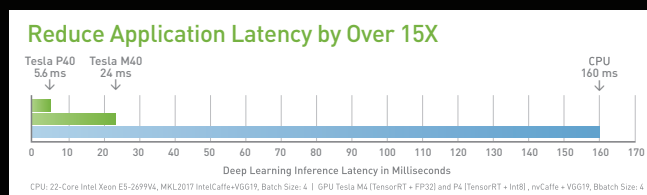
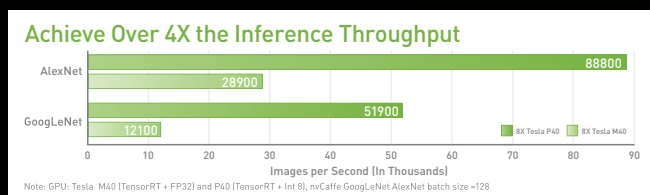
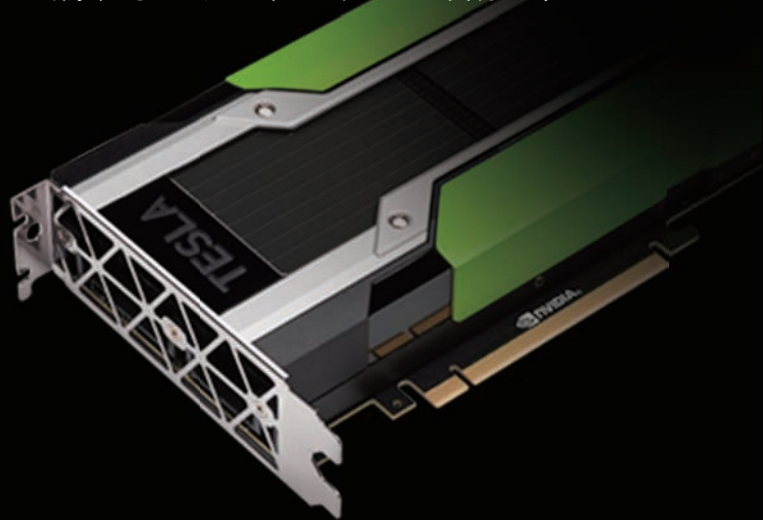
人工知能やインテリジェントマシンの新しい領域であるディープラーニングは、歴史上全く新しいコンピューティングモデルの世界を想像しています。革新的な NVIDIA® Pascal™ アーキテクチャによる GPU パワーは、人工知能の新たな時代を開く計算エンジンを供給し、ディープラーニングをアクセラレートすることによって驚くべきユーザーエクスペリエンスを実現します。

膨大なデータ量に負けない高処理能力

TESLA® P40 は、Pascal™ アーキテクチャに基づいて製造され、ディープラーニング推論において 47TOPS (Tera-Operations-Per-Second) もの性能を有します。例えば8枚の TESLA® P40 を搭載した1台のサーバは、ディープラーニング向けに利用されている140個のCPUが搭載されたサーバと同等の性能を有し、結果として、より低いコストでより高いスループットを実現します。

低レイテンシによるリアルタイムな応答性

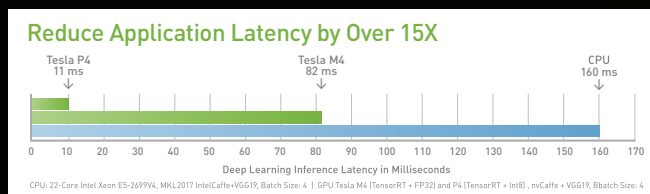
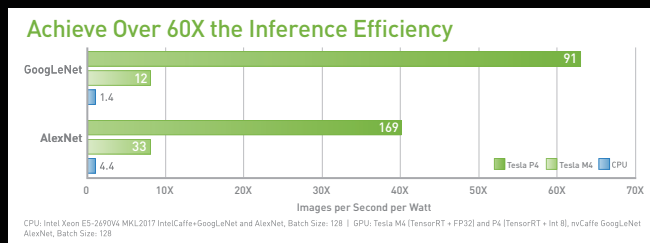
モデルが正確さや複雑さを増す中で、もはや CPU だけではインタラクティブなユーザーエクスペリエンスを実現できません。TESLA® P40 は、最も複雑なモデルであっても、リアルタイムな応答性を CPU よりも 30 倍以上も低いレイテンシで実現します。



8GB の GDDR5 メモリを搭載した小型／薄型サーバー筐体向け TESLA


NVIDIA® TESLA® P4

NVIDIA® TESLA® P4 は、革新的な NVIDIA® Pascal™ アーキテクチャを採用し、ディープラーニングワークロードを実行しているスケールアウトサーバの効率を加速するために設計されたスモールフォームファクターの TESLA® シリーズです。ハイパースケールインフラストラクチャーにおいて CPU と比較して推論レイテンシを 15 分の 1 に削減し、60 倍もの信じられないエネルギー効率を供給します。これにより、レイテンシの制限のために不可能であったこれまでの AI サービスに新しい価値を解放します。



技術仕様

Technical Specifications

製品名	NVIDIA® Tesla® V100 16GB / NVIDIA® Tesla® V100 32GB	NVIDIA® Tesla® P100 16GB / NVIDIA® Tesla® P100 12GB
外観		
CUDAコアプロセッサ数	5120	3584
Tensorコア数	640	-
ベースクロック	16GB 1245 MHz / 32GB 1230 MHz	1189 MHz
ブーストクロック	1380 MHz	1328 MHz
Tensor演算性能	112TFLOPS	-
整数演算性能[Int8]	-	-
半精度浮動小数点演算性能(FP16)	-	18.7 TFlops(最大ブースト)
単精度浮動小数点演算性能(FP32)	14 TFLOPS(最大ブースト)	9.3 TFLOPS(最大ブースト)
倍精度浮動小数点演算性能(FP64)	7 TFLOPS(最大ブースト)	4.7 TFLOPS(最大ブースト)
メモリ	16GB HBM2 / 32GB HBM2(バンド帯域幅 900GB/s)	16GB HBM2(バンド帯域幅 720 GB/s) / 12GB HBM2(バンド帯域幅 540GB/s)
接続バスコネクタ	PCI Express 3.0 × 16	PCI Express 3.0 × 16
最大消費電力	250W	250W
補助電源コネクタ仕様	CPU 8ピン電源コネクタ	CPU 8ピン電源コネクタ
認証規格	WHQL / EU RoHS / JIG / REACH / HF / RCM / BSMI / CE / FCC / ICES / KCC / cUL, UL / VCCI	WHQL / EU RoHS / JIG / REACH / HF / RCM / BSMI / CE / FCC / ICES / KCC / cUL, UL / VCCI
対応OS ^{*1} (※32bit OSはサポート対象外)	Windows® Server 2016 / Windows® Server 2012 R2 / Windows® 10 64bit / Windows® 8.1 64bit / Windows® 7 64bit / Linux 64bit	Windows® Server 2016 / Windows® Server 2012 R2 / Windows® Server 2008 R2 64bit / Windows® 10 64bit / Windows® 8.1 64bit / Windows® 7 64bit / Linux 64bit
外形寸法(プラケッ含まず)	266.7mm(幅)×111.2mm(高さ)×37mm(厚さ) 2スロットサイズ	266.7mm(幅)×111.2mm(高さ)×37mm(厚さ) 2スロットサイズ
冷却機構	パッシブ	パッシブ
型番	ETSV100-16GER / ETSV100-32GER	ETSP100-16GER / ETSP100-12GER
JANコード	4524076070669 / 4524076070713	4524076070478 / 4524076070485
保証期間	3年	3年

製品名	NVIDIA® Tesla® P40	NVIDIA® Tesla® P4 75W
外観		
CUDAコアプロセッサ数	3840	2560
Tensorコア数	-	-
ベースクロック	1303 MHz	885 MHz
ブーストクロック	1531 MHz	1531 MHz (1113 MHz default)
Tensor演算性能	-	-
整数演算性能[Int8]	47 TOPS(最大ブースト)	22 TOPS(最大ブースト)
半精度浮動小数点演算性能(FP16)	-	-
単精度浮動小数点演算性能(FP32)	12 TFLOPS(最大ブースト)	5.5TFLOPS(最大ブースト)
倍精度浮動小数点演算性能(FP64)	-	-
メモリ	24GB GDDR5(バンド帯域幅 347GB/s)	8GB GDDR5 (バンド帯域幅 192GB/s)
接続バスコネクタ	PCI Express 3.0 × 16	PCI Express 3.0 × 16
最大消費電力	250W	75W
補助電源コネクタ仕様	CPU 8ピン電源コネクタ	なし
認証規格	WHQL / EU RoHS / JIG / REACH / HF / RCM / BSMI / CE / FCC / ICES / KCC / cUL, UL / VCCI	WHQL / EU RoHS / JIG / REACH / HF / RCM / BSMI / CE / FCC / ICES / KCC / cUL, UL / VCCI
対応OS ^{*1} (※32bit OSはサポート対象外)	Windows® Server 2016 / Windows® Server 2012 R2 / Windows® Server 2008 R2 64bit / Windows® 10 64bit / Windows® 8.1 64bit / Windows® 7 64bit / Linux 64bit	Windows® Server 2016 / Windows® Server 2012 R2 / Windows® Server 2008 R2 64bit / Windows® 10 64bit / Windows® 8.1 64bit / Windows® 7 64bit / Linux 64bit
外形寸法(プラケッ含まず)	266.7mm(幅)×111.2mm(高さ)×37mm(厚さ) 2スロットサイズ	167.7mm(幅)× 69 mm(高さ)×18mm(厚さ) 1スロットサイズ
冷却機構	パッシブ	パッシブ
型番	ETSP40-24GER	ETSP4-8GER
JANコード	4524076070539	4524076070546
保証期間	3年	3年

*1：オンボードグラフィックスとの共存ができない場合もございますのでNVIDIA Quadro®/NVS® との組み合わせでのご利用を推奨致します。

お問い合わせ先

株式会社エルザ ジャパン

<http://www.elsa-jp.co.jp>

〒105-0014 東京都港区芝 3丁目 42番 10号 三田 UTビル

TEL : 03-5765-7391 FAX : 03-5765-7235

© 2018 NVIDIA Corporation. All rights reserved. NVIDIA, NVIDIAロゴ, Quadro, nView, NVIDIA Pascal, Maxwell, NVIDIA Mosaic, NVIDIA GPU Boost, およびCUDAは、NVIDIA Corporationの米国および/または他国における登録商標または商標です。ELSA (エルザ) は、テクノロジージョイント株式会社の登録商標です。その他の商品名は各社の商標または登録商標です。仕様などは改良のため、予告なしに変更されることがあります。本カタログの掲載内容は2018年1月現在の情報です。