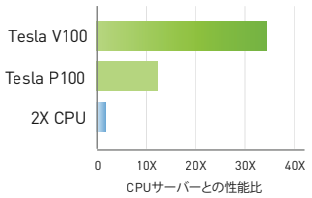
NVIDIA TESLA V100 GPU アクセラレーター

最先端のデータセンター向けGPU

NVIDIA® Tesla® V100 はAI、HPCそしてグラフィクス処理を高速化する史上最高のデータセンター向け GPU です。

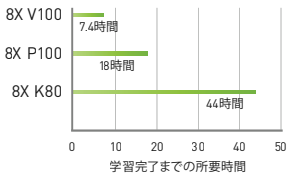
最新のNVIDIA Volta™ アーキテクチャにより、Tesla V100は1基で最大100CPU分のパフォーマンスを発揮し、かつては不可能と考えられていた課題に取り組むデータサイエンティスト、研究者、エンジニアを強力に支援します。

ディープラーニングの推論処理がCPUサーバーの30倍高速



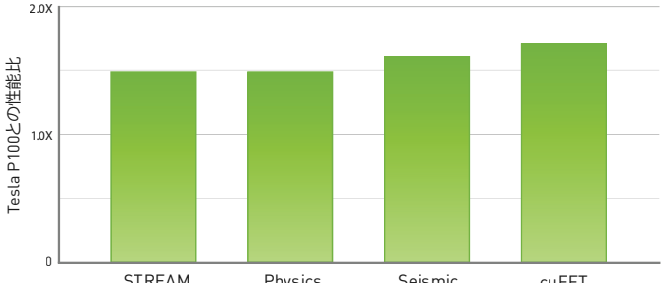
Workload: ResNet-50 | CPU: 2X Xeon E5-2690v4 @ 2.6GHz | GPU: add 1X NVIDIA® Tesla® P100 or V100 at 150W | V100 measured on pre-production hardware.

ディープラーニングの学習を1日で



Server Config: Dual Xeon E5-2699 v4, 2.6GHz | 8x Tesla K80, Tesla P100 or Tesla V100 | V100 performance measured on pre-production hardware. | ResNet-50 Training on Microsoft Cognitive Toolkit for 90 Epochs with 1.28M ImageNet dataset

HPCパフォーマンスが1年で1.5倍に

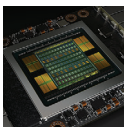


CPU System: 2X Xeon E5-2690v4 @ 2.6GHz | GPU System: NVIDIA® Tesla® P100 or V100 | V100 measured on pre-production hardware

仕様

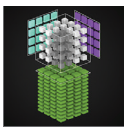
	Tesla V100 PCIe	Tesla V100 SXM2
GPUアーキテクチャ	NVIDIA Volta	
NVIDIA Tensorコア	640	
NVIDIA CUDA® コア	5,120	
倍精度演算性能	7 TFLOPS	**7.5 TFLOPS**
単精度演算性能	14 TFLOPS	**15 TFLOPS**
行列演算性能	112 TFLOPS	**120 TFLOPS**
GPUメモリ	16 GB HBM2	
メモリ帯域幅	900 GB/sec	
ECC	対応	
GPU間接続帯域*	32 GB/sec	**300 GB/sec**
システム接続	PCIe Gen3	**NVIDIA NVLink**
フォームファクタ	PCIe Full Height/Length	**SXM2**
最大消費電力	250 W	**300 W**
冷却方式	パッシブ（冷却ファンなし）	
対応計算API	CUDA, DirectCompute, OpenCL™, OpenACC	

TESLA V100の新機軸



VOLTAアーキテクチャ

CUDAコアとTensorコアの組み合わせにより、Tesla V100搭載サーバーの性能はHPCやディープラーニング用途で100台のコモディティCPUサーバーに匹敵します。



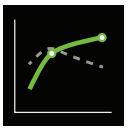
TENSORコア

新たな640個のTensorコアにより、Tesla V100は120 Tensor TeraFLOPSのディープラーニング性能を発揮。これは、NVIDIA Pascal™世代GPUとの比較で学習が12倍、推論は6倍のパフォーマンスです。

新世代のNVLINK

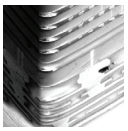
Tesla V100のNVIDIA NVLinkは前世代比で2倍のスループットを提供します。

8基までのTesla V100を最大300GB/sで接続することで、単一サーバー上での最高性能を発揮できます。



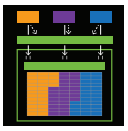
最大効率モード

新たな「最大効率モード」によりデータセンターの電源容量はそのままに、ラックあたり計算性能を最大40%向上させられます。このモードのTesla V100は最大性能の80%を、最大消費電力の半分で実現します。



HBM2

900GB/sに向上した帯域幅と95%向上したDRAM効率により、Tesla V100は STREAM 計測でPascalGPU比1.5倍のメモリ帯域幅を記録しています。



プログラミング効率

Tesla V100 はプログラムをシンプルにするために設計されました。新しい独立したスレッドスケジューリングにより同期の粒度を細かく設定可能で、小さなジョブを複数同時に処理する効率を高められます。

Tesla V100はディープラーニング、HPC、グラフィックスに向けたTesla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450以上のHPCアプリケーションと、全ての主要なディープラーニング フレームワークを高速化します。デスクトップからサーバー、そしてクラウドサービスまであらゆるところで利用でき、劇的なパフォーマンス向上とコスト削減の機会を提供します。

主要なディープラーニングフレームワーク

Caffe2
Microsoft Cognitive Toolkit
mxnet
PYTORCH
TensorFlow
theano

450以上のGPU対応アプリケーション

- AMBER
- ANSYS Fluent
- GAUSSIAN
- GROMACS
- LS-DYNA
- NAMD
- OpenFOAM
- Simulia Abaqus
- VASP
- WRF

Tesla V100の詳細はこちら：www.nvidia.co.jp/v100

© 2017 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, Tesla, NVIDIA GPU Boost, CUDA, and NVIDIA Volta are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. OpenCL is a trademark of Apple Inc. used under license to the Khronos Group Inc. All other trademarks and copyrights are the property of their respective owners. JUL17