

NVIDIA DGX-1

DEEPラーニングシステム



新しいコンピューティングモデルのための新しいコンピュータ

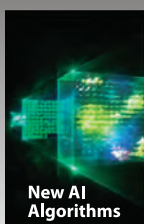
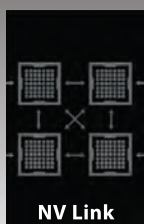
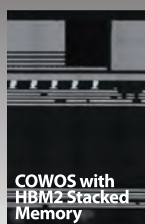
NVIDIA DGX-1™は、GPUの最新アーキテクチャPascal™採用の新世代GPU「Tesla®P100」を8枚搭載し、NVIDIAの次世代バス規格「NVLink™」を採用。CPUには「Dual 20-core Intel Xeon® E5-2698 v4」、512GBメモリ、1.92TB SSDを4基、ネットワークには4つの「InfiniBand EDR」と2つの10Gイーサネットを採用しています。NVIDIA DGX-1は、170TFLOPS（半精度浮動小数点演算）という、従来のCPUサーバの約250台分相当の性能を持ち、それを3Uの大きさのサーバで実現しています。



Pascalアーキテクチャ

ソフトウェアスタックとして、OS（Linux）やディープラーニング用の開発フレームワーク、各種ドライバ、管理ツールなど、ディープラーニングに必要な全てが搭載されて提供されるため、ユーザは、サーバの選定やボードの接続、開発環境の入手などの手間もなく、短期間で最適な構成が入手できます。

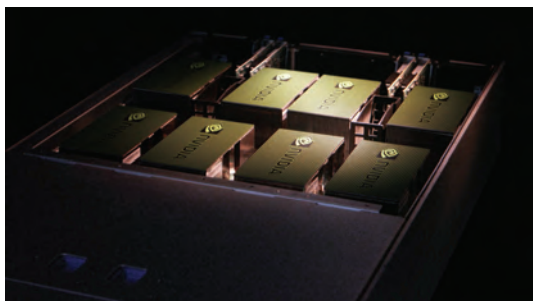
NVIDIA DGX-1 の特徴



NVIDIA PASCAL アーキテクチャ

NVIDIA DGX-1は、画期的な NVIDIA™ Pascal搭載のTesla P100 アクセラレータで構築され、NVIDIA NVLink™による相互接続を実装した最初のシステムです。Pascalは、無限の計算需要に満ちたこの世界を認識し、理解しシミュレートするコンピュータエンジンとして開発されました。

NVIDIA NVLinkは、TSMCの16nm FinFETプロセスで製造されるディープラーニングに適したGPUであり、メモリには各GPUに16GBのHBM2を採用しています。



投資効果を最大化

NVIDIA DGX-1サポートは、お客様の生産性を向上させ、システムのダウンタイムを削減します。ハードウェアとソフトウェアのサポートでは、NVIDIAのディープラーニングに関する専門知識へのアクセスに加え、クラウド管理サービス、ソフトウェアのアップグレードと更新、重大な問題の優先的解決を提供します。NVIDIA DGX-1サポートを利用することで、総合的なシステムサポートを受けられるだけでなく、NVIDIAのクラウド管理ポータルにもアクセスできるため、NVIDIA DGX-1システムに対する包括的サービスを最大限に活用できます。

NVIDIAクラウド管理では、コンテナ化されたアプリケーション管理を利用したディープラーニングの合理化、ジョブの開始、ステータスの監視、ソフトウェア更新のダウンロードを行うことができます。

圧倒的なパフォーマンス

高性能なトレーニングにより、生産性が向上し、洞察を得るまでの時間と市場投入までの時間が短縮されます。

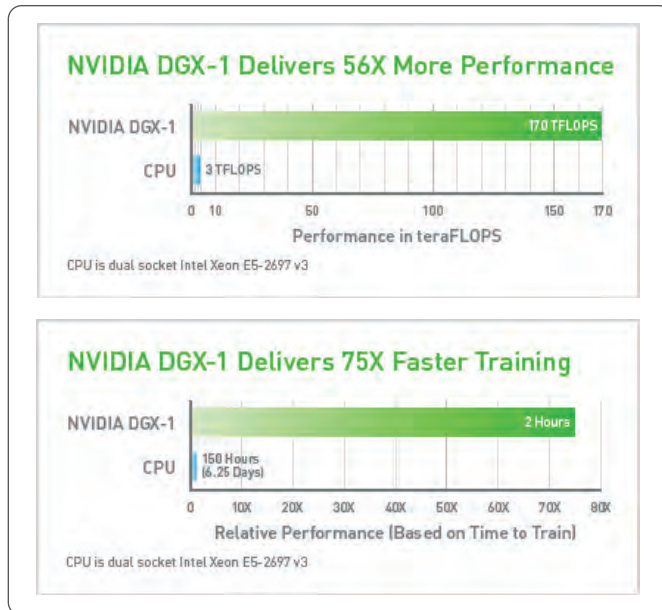
例えば、Xeon E5-2697v3プロセッサを2基搭載したノードと比較すると、DGX-1は37倍の性能差となっています。DualXeon E5-2697v3プロセッサで150時間かかる学習がDGX-1では僅か2時間という驚異的なパフォーマンスを実現します。

また、従来の平均的なスペックを持つCPUサーバの約250台分相当の性能を持ち、それを3Uの大きさのサーバで実現していますので、コスト面に加え、消費電力及びトータルの運用管理費用も大幅に削減可能と想定されます。

ディープラーニングSDKのパフォーマンスを強化

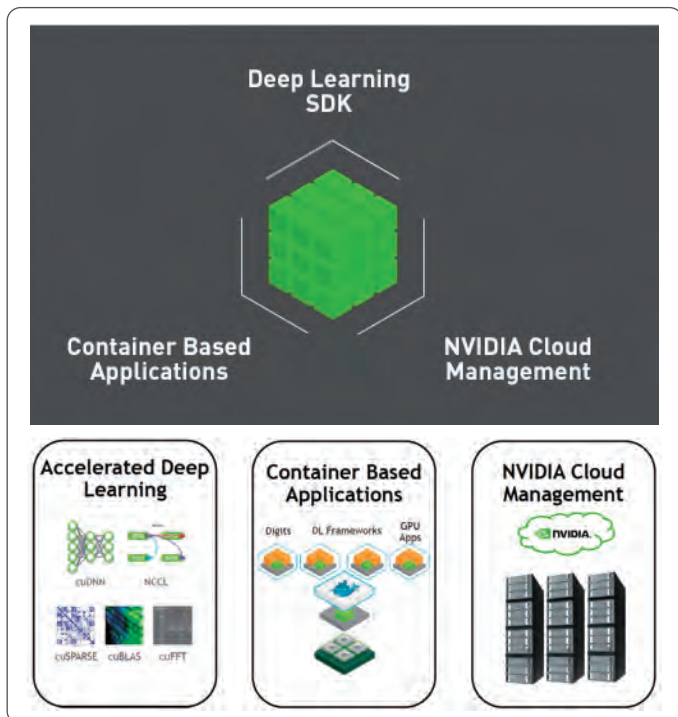
■ディープラーニングのパフォーマンスを大幅に向上

NVIDIA DGX-1のソフトウェアスタックには、ディープニューラルネットワーク(DNN)の迅速な設計を目的として、主要なディープラーニングフレームワーク、NVIDIAディープラーニングSDK、DIGITS™ GPUトレーニングシステム、ドライバ、およびCUDA®が含まれています。この強力なシステムからクラウド管理サービスにアクセスし、コンテナの作成と展開、システムアップデート、アプリケーションリポジトリを実行することができます。PascalベースのTesla GPUでこれらのソフトウェア機能を連動させると、アプリケーション実行速度は以前のGPUアクセラレーションによるソリューションの12倍に達します。また、それらのアップデートもNVIDIA CLOUDを利用することにより簡単に行うことができます。



■NVIDIA DGX-1の仕様

GPUs	8 x Tesla P100
TFLOPS (GPU FP16/CPU FP32)	170
GPU Memory	16GB per GPU
CPU	Dual 20-core Intel Xeon E5-2698 v4 2.2GHz
NVIDIA CUDA Cores	28672
System Memory	512GB 2133MHz DDR4 LRDIMM
Storage	SSD RAID Array 4 x 1.92TB SSD RAID 0 SSD OS Drive 1 x 480GB
Network	Dual 10 GbE, 4 InfiniBand EDR
Software	Ubuntu Server Linux OS DGX-1 Recommended GPU Driver
重量	lbs(60.7Kg)
大きさ	866 D x 444 W x 131 H (3U) [mm]
最大電気容量	3200W
動作温度	10-35 °C



Copyright © 2016 GDEP Solutions, Inc. All Rights Reserved. 社名・製品名は一般に各社の商標または登録商標です。本内容は予告なく変更される場合があります。ご了承ください。

●お問合せは