



NVIDIA、世界初のディープ・ラーニング・スーパーコンピュータを発売

NVIDIA DGX-1 はサーバー 250 台分のスループットを実現し、
人工知能以求められる大量コンピューティングに対応

米国カリフォルニア州サンノゼーGPU テクノロジ・カンファレンスー2016 年 4 月 5 日ーNVIDIA（本社：米国カリフォルニア州サンタクララ、社長兼 CEO：ジェンスン・ファン（Jen-Hsun Huang）、Nasdaq：NVDA）は本日、人工知能（AI）で要求される無制限のコンピューティングに対応する世界初のディープ・ラーニング・スーパーコンピュータ、[NVIDIA® DGX-1™](#)を初公開しました。

NVIDIA DGX-1 は、ディープ・ラーニングに特化して設計された初のシステムで、ハードウェア、ディープ・ラーニング・ソフトウェア、および開発ツールを余すところなく統合し、開発を迅速かつ容易にします。このシステムはすぐに稼働可能な状態で提供され、新世代の GPU アクセラレータを搭載し、x86 サーバー250 台分のスループットを実現します¹。

DGX-1 ディープ・ラーニング・システムにより、研究者やデータ科学者は、GPU で加速された計算機能をたやすく活用して、人間のように世界を学び、理解し、知覚する、新たなクラスのインテリジェント・マシンを作ることができます。このシステムにより、かつてないレベルの計算能力が実現して新世代の人工知能の応用が推進され、研究者は、より大規模でより精巧なディープ・ニューラル・ネットワークを発達させるための時間を劇的に削減できます。

NVIDIA では DGX-1 を、科学、ビジネス、そしてますます日常生活のあらゆる面にわたって広がっていく AI 革命を推進するための、新しいコンピューティング・モデルを目指して設計しました。強力なディープ・ニューラル・ネットワークが、大量のデータで作成された新たな種類のソフトウェアを牽引し、それにより、これまで以上にはるかにハイレベルな計算性能が求められます。

NVIDIA の CEO 兼共同創設者であるジェンスン・ファン（Jen-Hsun Huang）は、次のように述べています。「人工知能は、私たちの人生の中で最も遠大な技術進歩です。すべての業界、すべての企業、すべてを変えます。すべての人にメリットをもたらすマーケットが出現するでしょう。データ・サイエンティストと人工知能研究者は現在、自社製の高性能コンピューティング・ソリューションにあまりにも多くの時間を費やしています。DGX-1 は展開が容易で、1 つの目的のために作成されています。人間の能力を超えた力を解き放ち、かつては解決できなかった問題に応用するという目的です」

5つのブレイクスルーの活用

NVIDIA DGX-1 ディープ・ラーニング・システムは、新たな NVIDIA Pascal™ GPU アーキテクチャを基盤とする NVIDIA Tesla® P100 GPU で構築されます。CPU ベースのサーバー250 台分のスループット、ネットワーク接続、ケーブル、およびラックをオールインワンで提供します。

また、DGX-1 には、パフォーマンスを最大化し、使いやすくするためのブレイクスルー・テクノロジーがほかにも 4 つあります。これには、アプリケーションの拡張性を最大化する [NVIDIA NVLink™ 高速インターコネクト](#)、かつてないエネルギー効率を生み出す 16nm FinFET ファブリケーション・テクノロジー、ビッグ・データ・ワークロード用の HBM2 を使用したチップ・オン・ウェハ・オン・サブストレート、ディープ・ラーニングのピーク時性能が 21 テラフロップスを超える新しい半精度浮動小数点命令があります。

さらに、これらの主要な技術的進歩により、Tesla P100 GPU を備えた DGX-1 システムはたった 1 年前の 4-way NVIDIA Maxwell アーキテクチャ・ベースのソリューションの 12 倍以上の速さで学習を行います。

Pascal アーキテクチャには人工知能のエコシステムの強力なサポートがあります。

Facebook の AI 研究担当取締役、ヤン・ルクン（Yann LeCun）氏は、次のように述べています。「NVIDIA の GPU は人工知能の進化を加速しています。ニューラル・ネットワークの規模が拡大するにつれ、より大容量で高速なメモリを備えた高速な GPU が必要となるだけでなく、さらに高速な GPU 間通信に加えて、低精度演算を活用できるハードウェアが必要となります。これはまさに、Pascal そのものです」

Baidu（百度）のチーフ・サイエンティスト、アンドリュー・ウン（Andrew Ng）氏は言います。「人工知能コンピュータは宇宙ロケットのようなもの。大きければ大きいほどいい。Pascal のスループットとインターコネクトは、これまで見たこともないほど大きなロケットとなることだろう」

マイクロソフト・リサーチのチーフ・スピーチ・サイエンティスト、シュエトン・ファン（Xuedong Huang）氏は、次のように述べています。「マイクロソフトは 1,000 階層を超えるスーパー・ディープ・ニューラル・ネットワークを開発しています。NVIDIA Tesla P100 の圧倒的な馬力により、マイクロソフトの CNTK の AI ブレイクスルーを加速できるでしょう」

包括的なディープ・ラーニング・ソフトウェア・スイート

NVIDIA DGX-1 システムには、最適化された [ディープ・ラーニング・ソフトウェア](#)・スイートが揃えられ、研究者やデータ科学者はディープ・ニューラル・ネットワークを短期間で簡単に発達させることができます。

DGX-1 ソフトウェアには、ディープ・ニューラル・ネットワーク（DNN）を設計するためにまとめられたインタラクティブなシステムである、[NVIDIA Deep Learning GPU Training System \(DIGITS™\)](#) が備えられています。また、新しくリリースされた DNN の設計用プリミティブの GPU 加速化ライブラリである、[NVIDIA CUDA® Deep Neural Network library \(cuDNN\) バージョン 5](#) も含まれています。

さらに、広く使用されているディープ・ラーニング・フレームワークである Caffe、Theano、および Torch の最適化バージョンも含まれ、追加でクラウド管理ツール、ソフトウェア・アップデート、およびコンテナ化アプリケーションのリポジトリも提供します。

システム仕様

NVIDIA DGX-1 のシステム仕様は次のとおりです。

- 最大 170 テラフロップス（半精度浮動小数点（FP16）のピーク時性能）
- Tesla P100 GPU アクセラレータ x 8、16GB メモリ/GPU
- NVLink Hybrid Cube Mesh
- 7TB SSD DL キャッシュ
- デュアル 10GbE、クアッド InfiniBand 100Gb のネットワーク接続
- 3U サイズ電源 – 3200W

NVIDIA DGX-1 のオプション・サポート・サービスにより、生産性が向上し、実稼働システムのダウンタイムが削減されます。ハードウェアおよびソフトウェアのサポートには、NVIDIA におけるディープ・ラーニングの専門知識の提供や、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アップデート、およびクリティカルな問題の優先解決が含まれます。詳細については、www.nvidia.com/page/support/dgx1-support を参照してください。

提供開始

NVIDIA DGX-1 ディープ・ラーニング・システムの一般提供開始時期は米国では 6 月、その他の地域では第三四半期始めとなり、NVIDIA と認定されたシステム・インテグレータ から直接入手できます。

参考資料

- [ディープ・ラーニングに関する動画](#)

(1) Caffe/AlexNet が InfiniBand インターコネクタを利用した 2 つのソケット Intel Xeon E5-2697 v3 プロセッサベースシステムの ILSVRC-2012 データセットを学習させる時間と比較。
<https://software.intel.com/en-us/articles/caffe-training-on-multi-node-distributed-memory-systems-based-on-intel-xeon-processor-e5>.

NVIDIA についての最新情報:

・公式ブログ [NVIDIA blog](#)、[Facebook](#)、[Google+](#)、[Twitter](#)、[LinkedIn](#)、[Instagram](#)、NVIDIA に関する動画 [YouTube](#)、画像 [Flickr](#)。

NVIDIA について

1993 年以来、NVIDIA (NASDAQ: NVDA) は、ビジュアル・コンピューティングという芸術的な科学の世界をリードしてきました。ゲーミング、自動車、データセンターおよびプロフェッショナル・ビジュアリゼーションの分野で特化したプラットフォームを提供し続けています。NVIDIA の製品は仮想現実、人工知能、自律走行車の開発においても最新の技術を提供しています。詳しい情報は、

http://www.nvidia.co.jp/object/newsroom_jp.html をご覧ください。

本案件に関するお問い合わせ先:

エヌビディアプラットフォームビジネス本部

広報/マーケティングコミュニケーションズシニアマネージャ中村

Email アドレス : knakamura@nvidia.com

Tel : 03-6743-8712 Fax : 03-6743-8799

本プレスリリースに記載されている、人工知能分野における、NVIDIA DGX-1 ディープ・ラーニングシステムの影響、バーチャル・リアリティおよび、自動運転に関連する開発の利益・影響、は将来予測的なものが含まれており、予測とは大幅に異なる結果が生じるリスクと不確実性を伴っています。かかるリスクと不確実性は、世界的な経済環境、サードパーティーに依存する製品の製造・組立・梱包・試験、技術開発および競合による影響、新しい製品やテクノロジーの開発あるいは既存の製品やテクノロジーの改良、当社製品やパートナー企業の製品の市場への浸透、デザイン・製造あるいはソフトウェアの欠陥、ユーザの嗜好および需要の変化、業界標準やインターフェースの変更、システム統合時に当社製品および技術の予期せぬパフォーマンスにより生じる損失などを含み、その他のリスクの詳細に関しては、Form10-Q の 2015 年 10 月 25 日を末日とする四半期レポートなど、米証券取引委員会 (SEC) に提出されている NVIDIA の報告書に適宜記載されます。SEC への提出書類は写しが NVIDIA のウェブサイトに掲載されており、NVIDIA から無償で入手することができます。これらの将来予測的な記述は発表日時点の見解に基づくものであって将来的な業績を保証するものではなく、法律による定めがある場合を除き、今後発生する事態や環境の変化に応じてこれらの記述を更新する義務を NVIDIA は一切負いません。

© 2016 NVIDIA Corporation. All rights reserved. © 2016 NVIDIA Corporation. All rights reserved. NVIDIA、NVIDIA ロゴ、Tesla、NVIDIA DIGITS、DGX-1、Pascal、Maxwell、CUDA、および、NVIDIA NVLink/またはその他の国における NVIDIA Corporation の商標あるいは登録商標です。その他の企業名および製品名は、それぞれ各社の商標である可能性があります。機能や価格、供給状況、仕様は、予告なく変更される場合があります。