



新しい NVIDIA Hyperscale アクセラレータ、 ウェブ・データセンターにおける機械学習のスループットを増強

人工知能の活用によるスマートなアプリケーションの構築が
ウェブサービス大手に広がり、機械学習の作業負荷が爆発的に増大中

2015年11月10日 – カリフォルニア州サンタクララ – 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・フアン(Jen-Hsun Huang)、Nasdaq：NVDA）は本日、エンドツーエンドのハイパースケール・データセンター・プラットフォームを発表しました。ウェブサービス企業が直面する膨大な[機械学習](#)の作業負荷を高速化することができます。

NVIDIA ハイパースケール・アクセラレータは、現在、ふたつあります。人工知能（AI）を活用するアプリケーションが増えていますが、ひとつは、そのディープ・ニューラル・ネットワークを研究者がすばやく開発・設計するためのものです。もうひとつは、低消費電力のアクセラレータで、開発されたネットワークをデータセンターにデプロイするためのものです。GPU アクセラレーテッド・ライブラリのスイートも用意されています。

これらを組み合わせると、パワフルな [Tesla アクセラレーテッド・コンピューティング・プラットフォーム](#) を活用し、ハイパースケール・データセンターで機械学習を行い、AI ベースの新たなアプリケーションを生み出すことが可能になります。

NVIDIA の共同創業者兼 CEO、ジェンスン・フアン（Jen-Hsun Huang）は、次のように述べています。「人工知能は、いま、開発競争が激化しています。機械学習が、いま、コンピューティングの世界で注目の的になっていることはまちがひありません。PC のレベルでも、インターネットでも、また、クラウド・コンピューティン

グの世界でも、です。その結果、一般消費者向けのクラウド・サービスから自動車、ヘルスケアにいたるまで、幅広い産業で変革が次々に起きています。」

「機械学習は、いまの世代にとって一番とも言えるコンピューティング課題です。だから、NVIDIA では、機械学習を 10 倍に高速化できる Tesla ハイパースケール・アクセラレータ製品ラインを開発しました。データセンターで消費される時間とコストを大きく削減できるはずですよ。」

いま、山のようなウェブ・アプリケーションが AI 機能の搭載にしのぎを削っていますが、開発されたハードウェアとソフトウェアは、そのようなアプリケーションを高速化するためのものです。機械学習が大きく進化した結果、AI 手法を活用してスマートなアプリケーションやサービスを生みだせるようになったのです。

機械学習は、音声認識の精度向上にも使われています。動画や写真に写っている物やシーンを自動的に認識してタグをつけ、後々検索しやすいようにする処理にも使われています。顔の一部が不鮮明であっても、動画や写真から顔認識をすることもできます。個人の趣味や興味に対応できるアプリケーションとし、スケジュールを管理したりニュースを収集したり、さらには、普通の会話という感じで音声コマンドに精度よく対応したりすることもできます。

いずれも魔法のように思えますが、機械学習を使えば実現可能なのです。もちろん、問題もあります。次々に生みだされるディープ・ニューラル・ネットワークの進化と訓練には膨大なスーパーコンピューティング能力が必要ですし、そして生まれたサービスに対し、消費者から寄せられる何十億件ものクエリを瞬間的に処理するにも膨大なスーパーコンピューティング能力が必要となります。それほどの能力をいかにして得るのか——この大きな問題に対する答えが、NVIDIA ハイパースケール・アクセラレータです。この製品ラインを活用すれば、このような作業負荷をすばやく処理することも、データセンターのスループットを劇的に高めることもできます。

今回、以下の新製品が NVIDIA Tesla プラットフォームに加わりました。

- [NVIDIA® Tesla® M40 GPU](#) – ディープ・ニューラル・ネットワークのトレーニング用として開発された最もパワフルなアクセラレータです。
- [NVIDIA Tesla M4 GPU](#) – 機械学習の推論や、映像のストリーミング、動画の処理などに最適な小型・低消費電力のアクセラレータです。
- NVIDIA Hyperscale Suite – 機械学習と動画処理に最適化されたソフトウェアが豊富に用意されたスイートです。

NVIDIA Tesla M40 GPU アクセラレータ

総合的な精度を高めようと膨大な量のデータでディープ・ニューラル・ネットワークのトレーニングを行う場合、NVIDIA Tesla M40 GPU アクセラレータを使えば、何日も、あるいは何週間も所要期間を短縮することができます。主な特長は、以下のとおりです。

- **機械学習に最適化** – トレーニングに要する期間が CPU の 1/8 程度まで短縮されます（典型的な AlexNet トレーニングが 10 日から 1.2 日まで短縮される）。
- **信頼性が高く連続稼働が可能** – データセンター環境で高い信頼性が発揮できるように開発・検証されています。
- **スケールアウトのパフォーマンス** – NVIDIA GPUDirect のサポートにより、マルチノード・ニューラル・ネットワークをすばやくトレーニングすることが可能です。

NVIDIA Tesla M4 GPU アクセラレータ

NVIDIA Tesla M4 アクセラレータは、ハイパースケール環境専用の低消費電力 GPU で、動画のトランスコーディング、静止画や動画の処理、機械学習の推論など、要求が厳しく、かつ、急速に成長しているウェブサービス・アプリケーションに最適です。主な特長は、以下のとおりです。

- **高スループット** – トランスコーディングやエンハンス、分析などの処理を同時に行えるビデオ・ストリームの本数が CPU の最大で 5 倍となります。
- **低消費電力** – Tesla M4 の消費電力は 50~75 ワットで（パワー・プロファイルはユーザが選択可能）、動画や機械学習アルゴリズムの処理におけるエネルギー効率が CPU の最大で 10 倍に達します。
- **小型** – ロープロファイル PCIe で、ハイパースケール・データセンターで使われるエンクロージャに搭載することができます。

NVIDIA Hyperscale Suite

新しい NVIDIA Hyperscale Suite には、デベロッパ向けのツールとデータセンタ・マネジャー向けのツールが用意されています。ウェブサービスのデプロイメントを念頭に開発されたツールです。主なツールは以下のとおりです。

- **cuDNN** – AI アプリケーションに使われるディープ・ニューラル・ネットワークの一種、畳み込みニューラル・ネットワーク用として業界で大人気となっているアルゴリズム・ソフトウェアです。
- **GPU アクセラレーテッド FFmpeg マルチメディア・ソフトウェア** – 人気の FFmpeg ソフトウェアを使い、動画のトランスコーディングや動画処理を高速化することができます。
- **NVIDIA GPU REST エンジン** – 画像のダイナミックなリサイズ、検索の高速化、画像分類といった幅広いタスクを処理する高スループット、ローレイテンシのアクセラレーテッド・ウェブサービスを簡単に構築し、デプロイすることができます。
- **NVIDIA Image Compute Engine** – REST API 対応の GPU アクセラレーテッド・サービスで、画像のリサイズを CPU に対して 5 倍の高速で行えます。

Mesosphere がサポート

Tesla アクセラレーテッド・コンピューティング・プラットフォームは業界で幅広い支持を受けていますが、今回、新たに、Mesosphere からも、NVIDIA と協力して [Apache Mesos](#) と Mesosphere DCOS (Datacenter Operating System) で GPU テクノロジーをサポートするとの発表がありました。今後、ウェブサービス各社にとって、次世代アプリケーションが動かせるアクセラレーテッド・データセンタを構築し、デプロイすることが容易になるものと思われます。

供給状況

Tesla M40 GPU アクセラレータと Hyperscale Suite ソフトウェアは、年内に発売する予定です。Tesla M4 GPU は 2016 年第 1 四半期の発売を予定しています。詳しくは [NVIDIA Tesla ウェブサイト](#) をご覧ください。

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。ゲームや研究者、消費者からエンタープライズ各社にいたるまで、あらゆる人々にとって、ディスプレイの世界はインタラクティブな発見の世界へと変化しつつありますが、その源となっているのが NVIDIA のテクノロジーです。詳しい情報は、<http://nvidianews.nvidia.co.jp/>や <https://blogs.nvidia.co.jp/> をご覧ください。

#

本プレスリリースに記載されている、NVIDIA Tesla M40 GPU アクセラレータ、NVIDIA Tesla M4 GPU アクセラレータ、および NVIDIA ハイパースケールを含む NVIDIA ハイパースケールアクセラレータおよび機会学習とそれ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、Form 10-Q の 2015 年 7 月 26 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。

© 2015 NVIDIA Corporation. All rights reserved. NVIDIA、NVIDIA ロゴ、PGI アクセラレータおよび、OpenACC は、米国および/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。

【添付画像】

NVIDIA® Tesla® M40 GPU

NVIDIA® Tesla® M4 GPU