



NVIDIA と世界有数のサーバーメーカーがパートナーを組み、 AI クラウド コンピューティングを推進

Foxconn、Inventec、Quanta、Wistron が NVIDIA HGX リファレンス アーキテクチャーを利用し、ハイパースケール データ センター向け AI システムを構築

台湾、台北 — Computex— (2017 年 5 月 30 日) — NVIDIA は本日、より迅速に AI クラウド コンピューティングの需要に応えるため、世界有数の ODM である Foxconn、Inventec、Quanta、Wistron とのパートナー プログラムを開始しました。

NVIDIA HGX パートナー プログラムを通じ、NVIDIA は、各 ODM に NVIDIA HGX リファレンス アーキテクチャー、NVIDIA GPU コンピューティング テクノロジー、デザイン ガイドラインの早期アクセスを提供します。HGX は、Microsoft の Project Olympus の取り組み、Facebook の Big Basin システム、NVIDIA DGX-1™ AI スーパーコンピューターにおいて利用されているのと同じデータ センター向けに設計されています。

HGX をスターター向けの「レシピ」として利用する ODM パートナーは、NVIDIA と協力し、ハイパースケール データ センター向けの条件を満たす、さまざまな GPU アクセラレーテッド システムをより迅速に設計し、市場化することができます。NVIDIA のエンジニアは、このプログラムを通じて ODM とより緊密に協力し、技術の採用から実動環境への展開までの時間が最短になるよう支援します。

昨年、AI コンピューティングのリソースに対する需要全体が大幅に上昇し、同様に、NVIDIA の GPU コンピューティング プラットフォームも、市場での導入およびパフォーマンスの点で急激な伸びを記録しました。現在、世界上位 10 社のハイパースケール企業のすべてが、それぞれのデータ センターで NVIDIA GPU アクセラレーターを利用しています。

従来型の 3 倍のパフォーマンスを提供する新しい NVIDIA® Volta アーキテクチャー ベースの GPU により、ODM は、市場の需要に対し、利用可能な NVIDIA の最新テクノロジーに基づき、新しい製品を供給することができます。

NVIDIA のアクセラレーテッド コンピューティング担当のゼネラル マネージャーであるイアン・バック (Ian Buck) は、次のように述べています。「わずか 1 年間に Tesla GPU のディープリニング パフォーマ

ンスが 3 倍になったように、アクセラレーテッド コンピューティングは急速に進化しており、このことは、システムの設計方法に重大な影響を与えています。各デバイス メーカーは、当社の HGX パートナー プログラムを通じ、ますます広がるクラウド コンピューティング プロバイダーのコミュニティに、最新の AI テクノロジーを確実に提供できるようになります。」

アップグレードできる柔軟な設計

NVIDIA は、ハイパースケール クラウド環境に特有のハイ パフォーマンス、効率性、大規模な拡張の要件を満たせるよう、HGX リファレンス デザイン を構築しました。ワークロードのニーズに基づき高度な設定変更が可能な HGX は、ハイパフォーマンス コンピューティング、ディープラーニングのトレーニング、ディープラーニングの推論向けに、さまざまな方法で GPU と CPU を容易に組み合わせることができます。

標準の HGX デザイン アーキテクチャーは、SXM2 フォーム ファクターに NVIDIA Tesla® GPU アクセラレーターを 8 基搭載し、NVIDIA NVLink™ 高速インターコネクトと最適化された PCIe トポロジーを利用するキューブ メッシュ内で接続しています。モジュール設計による HGX の筐体は、必要に応じてハイパースケールの CPU ノードを利用し、世界中の既存のデータ センターにあるラックへの展開に適しています。

NVIDIA Tesla P100 と V100 GPU アクセラレーターはいずれも、HGX と互換性があります。これは、V100 GPU が今年中に提供されれば、すべての HGX ベースの製品をすぐにアップグレードできるよう考慮しているためです。

HGX は、新しい NVIDIA GPU Cloud プラットフォームのホスティングを求めるクラウド プロバイダーにとって、理想的なリファレンス アーキテクチャーです。NVIDIA GPU Cloud プラットフォームは、Caffe2、Cognitive Toolkit、MXNet、TensorFlow など、ディープラーニング フレームワーク コンテナを完全に統合し、最適化するカタログを管理します。

Foxconn/Hon Hai Precision Ind Co., Ltd のゼネラル マネージャー兼 Ingrasys Technology Inc.のプレジデントであるタイユー・チョウ (Taiyu Chou) 氏は、次のように述べています。「私たちは、NVIDIA との新しいパートナー プログラムを通じ、お客様の多くは世界最大級のデータ センターの一部を管理していますが、その拡大する需要に、より迅速に応えられるようになります。NVIDIA GPU テクノロジーへの早期アクセスと設計ガイドラインは、お客様の高まる AI コンピューティングのニーズに対し、私たちがより迅速に革新的な製品を発表するのに役立ちます。」

Inventec Corporation の IEC チャイナの運営責任者であるエバン・チェン (Evan Chien) 氏は次のように述べています。「NVIDIA とより緊密に協力すれば、世界中のデータセンターのインフラストラクチャーに新しいレベルのイノベーションを容易に投入できるようになります。緊密な協力を通じ、ハイパースケール クラウド環境を管理する企業の、コンピューティングを駆使する AI のニーズに、より効率的な対処が可能です。」

Quanta Computer Inc. のシニア バイス プレジデント兼 QCT のプレジデントであるマイク・ヤン (Mike Yang) 氏は、このように述べています。「NVIDIA の AI コンピューティングの専門知識を活用すれば、AI の時代の新しいコンピューティング要件を満たす、これまでの仕組みを一変させるようなソリューションをすぐに市場に提供できるようになります。」

Wistron のチーフ テクノロジー オフィサー兼エンタープライズ ビジネス グループのプレジデントであるドナルド・ファン (Donald Hwang) 氏は、次のように述べています。「NVIDIA の長期にわたる協力企業として、当社のハイパースケール データ センターを利用するお客様のますます高まるコンピューティング ニーズに応えられるよう、両社の関係を深めていきたいと考えています。当社のお客様は、さまざまな AI ワークロードに対処するため、より高い GPU のコンピューティング能力を求めており、この新しいパートナーシップを通じ、当社は、新しいソリューションをより迅速に提供できるようになります。」

Microsoft の Azure ハードウェア インフラストラクチャー担当ゼネラル マネージャー兼ディスティングイッシュド エンジニア (Distinguished Engineer) のクシャグラ・ベイド (Kushagra Vaid) 氏は、次のように述べています。「私たちは、Ingrasys および NVIDIA と協力し、新しい AI の時代の高まる需要に応えるため、新しい業界標準の設計を切り開いてきました。HGX-1 AI アクセラレーターは、Microsoft の Project Olympus のコンポーネントとして開発されており、最大で GPU 32 基に対する高帯域幅のインターコネクティビティーのオプションを通じ、きわめて高いパフォーマンスの拡張性を実現します。」

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

NVIDIA が 1999 年に開発した GPU は、PC ゲーム市場の成長に拍車をかけ、現代のコンピューターグラフィックスを再定義し、並列コンピューティングを一変させました。最近では、GPU ディープラーニングが最新の AI、つまりコンピューティングの新時代の火付け役となり、世界を認知して理解できるコンピューター、ロボット、自動運転車の脳の役割を GPU が果たすまでになりました。今日、NVIDIA は

「AI コンピューティングカンパニー」として知名度を上げています。詳しい情報は、
<http://www.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 のパートナーシッププログラムにおける利点・インパクト、NVIDIA Volta アーキテクチャーGPU および、HGX レファレンスデザインおよび、V100 GPU の可用性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、ユーザー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-K の 2017 年 1 月 29 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。

© 2017 NVIDIA Corporation. All rights reserved. NVIDIA、NVIDIA ロゴおよび、NVIDIA DGX、Tesla および、NVLink 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。