NVIDIA、革新的な Volta GPU プラットフォームを発表、
新時代の AI とハイパフォーマンスコンピューティングを加速

Volta ベースの Tesla V100 データセンター GPU、
120 テラフロップスのディープラーニングの壁を打破

米国カリフォルニア州サンノゼ--GPU テクノロジカンファレンス -（2017 年 5 月 10 日）-
NVIDIA は本日、人工知能 (AI) とハイパフォーマンスコンピューティングにおける進化の次の波を推進するために開発された、世界で最もパワフルなコンピューティングアーキテクチャー、Volta™を発表しました。

また、初の Volta ベースのプロセッサーNVIDIA® Tesla® V100 データセンター GPU も発表しました。このプロセッサーは、AI の推論とトレーニング向けに、また、HPC とグラフィックスのワークロードを加速するため、並外れたスピードとスケーラビリティーを実現します。

GTC の基調講演において Volta を発表した、NVIDIA の創設者兼 CEO ジェンスン・フアン (Jensen Huang) は、次のように述べています。「AI は、人類の歴史において最も優れたテクノロジーの進化を推進しています。インテリジェンスを自動化し、産業革命以降、類のない社会の進化の波に拍車をかけるでしょう。」

「ディープラーニングは、学習するコンピューターソフトウェアを創造する革新的な AI アプローチであり、処理能力に対して飽くなき要求があります。何千人もの NVIDIA のエンジニアは、Volta の構築に 3 年を費やしてこのニーズに応え、業界は、人生を変えるような AI の可能性を実感できるようになりました」とフアンは語りました。

NVIDIA の第 7 世代の GPU アーキテクチャーである Volta は、210 億のトランジスタから構成され、ディープラーニング向けに CPU 100 基に相当するパフォーマンスを実現します。

Volta は、ピーク性能で、NVIDIA GPU アーキテクチャーの現行世代である Pascal®の 5 倍、2 年前に発表された Maxwell®アーキテクチャーの 15 倍の処理能力を提供します。このパフォーマンスの向上は、ムーアの法則で予測されていた値の 4 倍を超えるものです。

AI の加速に対する要求は、かつてないほど大きくなっています。がんとの闘い、自動運転車によって交通をより安全にする取り組み、インテリジェントで新しい顧客体験の提供など、次の進化を推進するため、開発者、データサイエンティスト、研究者のニューラルネットワークに対する依存度はますます高くなっています。

ネットワークが複雑になるにつれて、データセンターの処理能力を急激に増大させる必要があります。また、効率的に規模を拡大し、自然言語のバーチャルアシスタントやパーソナライズされた検索と推奨システムなど、きわめて正確な AI ベースのサービスが急速に導入されるのをサポートする必要があります。

Volta は、ハイパフォーマンスコンピューティングの新しい標準になります。Volta は、インサイトを発見するという点では、コンピューティング科学とデータ科学の両面で他に勝る HPC システム向けプラットフォームを提供します。統合されたアーキテクチャー内で CUDA®コアと新しい Volta Tensor Core を組み合わせると、Tesla V100 GPU を備えた 1 台のサーバーは、従来の HPC 向けの数百基におよぶコモディティー CPU の代わりになります。

画期的なテクノロジー

Tesla V100 GPU は、100 テラフロップスのディープラーニングパフォーマンスの壁を打破できる画期的なテクノロジーにより、前世代の NVIDIA GPU から一気に進化しました。Tesla V100 GPU は、以下の要素から構成されています。

- **Tensor Core** は、AI ワークロードを加速できるよう設計されています。V100 は、Tensor Core640 基を備え、120 テラフロップスのディープラーニングパフォーマンスを実現します。これは、100 基の CPU に相当するパフォーマンスです。
- **新しい GPU アーキテクチャー**は、210 億を超えるトランジスタを備えています。統合されたアーキテクチャー内で CUDA コアと Tensor Core を組み合わせ、単一の GPU において AI スーパーコンピューターのパフォーマンスを提供します。
- **NVLink™**は、次世代の高速インターコネクトを提供し、GPU 同士、または GPU と CPU をリンクします。スループットは、前世代の NVLink に対し最大 2 倍です。
- **900 GB/秒の HBM2 DRAM** は、Samsung とのコラボレーションにより開発され、前世代の GPU よりメモリ帯域幅を 50% 増強しています。Volta の並外れたコンピューティングスループットをサポートするために不可欠です。
- **Volta 向けに最適化されたソフトウェア** (CUDA、cuDNN、TensorRT™ソフトウェアなど) は、AI と研究を加速するため主要フレームワークとアプリケーションで活用できます。

Volta に対するエコシステムサポート

Volta は、世界中の主要な企業や組織から、以下のように幅広い業界サポートを受けてきました。

「NVIDIA と AWS は、お客様がコンピューティングを駆使した AI ワークロードをクラウドで実行できるように、長い間協力してきました。私たちは、2010 年に GPU 向けに最適化された初のクラウドインスタンスを発表し、昨年は、クラウドにおいて利用可能な最もパワフルな GPU インスタンスを発表しました。AWS には、今日の最も革新的かつ創造的な AI アプリケーションの一部が存在します。Volta が今年中に提供され、

お客様が次世代の汎用 GPU インスタンス群を利用して、新しい優れたアプリケーションを構築するのを支援できることを楽しみにしています。」
― Amazon Web Services、コンピュートサービス担当バイスプレジデント、マット・ガーマン (Matt Garman) 氏

「NVIDIA が最近 Volta を発表したことに対し、お祝い申し上げます。Baidu Cloud から Intelligent Driving まで、Baidu は、オープン AI プラットフォームの構築における取り組みを強化しています。当社は NVIDIA とともに、グローバルな AI テクノロジーの開発と応用を加速し、社会全体のため、より大きな機会を創造できると確信しています。」
― Baidu、プレジデント、張亜勤 (Yaqin Zhang) 氏

「NVIDIA と Facebook はすばらしいパートナーです。私たちは、Facebook の Caffe2 および PyTorch に対する NVIDIA の貢献をうれしく思います。NVIDIA の新しく、ハイパフォーマンスな Volta のグラフィックスアーキテクチャーが実現する、AI の進化を楽しみにしています。」
―Facebook、チーフテクノロジーオフィサー、マイク・シュローファー (Mike Schroepfer) 氏

「NVIDIA の GPU は、Google Cloud Platform のお客様に大幅なパフォーマンスの向上をもたらします。GPU は、当社のインフラストラクチャーの重要な部分であり、Google や当社の企業のお客様に、機械学習やハイパフォーマンスコンピューティング、データ分析向けの特別なコンピューティング能力をもたらします。Volta のパフォーマンス向上により、GPU は一層パワフルになるでしょう。私たちは、Volta GPU を GCP 上で提供しようと計画しています。」
― Google、Google Cloud Platform のエンジニアリング担当バイスプレジデント、ブラッド・カルダー (Brad Calder) 氏

「Microsoft と NVIDIA は、Microsoft Azure N シリーズ、Project Olympus、Cognitive Toolkit など、AI テクノロジーに関して長年パートナーシップを組んできました。新しい Volta のアーキテクチャーは、Microsoft のお客様に並外れた新しい能力を解放するでしょう。」
― Microsoft、Microsoft AI およびリサーチグループ担当エグゼクティブバイスプレジデント、ハリー・シャム (Harry Shum) 氏

「オークリッジ国立研究所は、今夏、先導的な次世代コンピューティングシステム Summit の構築を開始します。Summit は、Volta GPU を備え、2018 年に構築が完了すれば、科学の発見に役立つ米国トップのスーパーコンピューターになるでしょう。Summit により、米国は科学研究の最先端の地位を維持し、エネルギー省は、コンピューティング科学と AI 支援による発見に伴う複雑な課題に対処できるでしょう。」
―オークリッジ国立研究所、コンピューティングおよび計算科学のアソシエートラボラトリーディレクター、ジェフ・ニコルス (Jeff Nichols) 氏

「WeChat の音声テクノロジー、QQ と Qzone のフォトとビデオのテクノロジー、Tencent Cloud に基づくディープラーニングプラットフォームなど、当社の幅広い製品はすでに AI に依存しています。私たちは、Volta が当社の AI 開発者向けにこれまでにないコンピューティングパワーを提供すると確信しており、Tencent Cloud から、このような能力がさらに多くのお客様にまもなく解放されることをたいへんうれしく思います。」
—Tencent、シニアエグゼクティブバイスプレジデント、湯道生 (Dowson Tong) 氏

NVIDIA の最新情報は、以下の方法で入手できます。
公式ブログ NVIDIA blog、Facebook、Google+、Twitter、LinkedIn および 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、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、[挿入ください]の影響と利点、有用性と価格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。