

PowerEdge XE8545: パフォーマンス概要

妥協なきコンピュータ能力を提供する、AIインフラストラクチャ

テックノート by

Liz Raymond
シニアパフォーマンスエンジニア

サマリー

この文書は、最新のアクセラレーターサーバー「PowerEdge XE8545」の利用でお客様が得られる、パフォーマンス面の優位性を簡潔に解説した概要です。

本ドキュメントで取り上げるパフォーマンスとテクノロジー特性はすべて、当社がAmericas Data Center (CET) のラボで実施したパフォーマンス測定に基づきます。

測定結果の正確性は2021年3月15日時点のものです。

Ad Ref #G21000042

Dell Technologiesによる最新のアクセラレーション専用モデル

PowerEdge XE8545は、高性能な機械学習を活用するお客様の中でも特に、NVIDIAの最新GPUのイノベーション全てとその演算性能を、既存のデータセンターの冷却方式を大きく変えることなく、ただちに享受されたいというニーズにお応えした製品です。



図1: XE8545とオプションのフロントベゼル

明確に空冷を前提とした製品設計により、スイッチレスで低レイテンシーなNVLinkインターコネクト接続のA100 40GB/400W SXM4 GPUを4基、エネルギー効率の良い35°C環境で稼働します。もう一つの選択肢として、80GB/500W GPUもあります。こちらは少し低い吸気温度の環境（28°C）で、400W稼働のGPUより13~15%高いパフォーマンスを提供します。

500Wサポートの優位性、より高いパフォーマンスを実現

- XE8545は、80GB GPUの駆動を500Wまで解放します。同じGPUモジュールが400Wに制限される他社サーバーと比べ、**13~15%高いパフォーマンスを発揮します。**

他社と違い、Dell Technologiesは製品設計の初期段階でNVIDIAと連携し、XE8545においては高性能なA100 80GB GPUが（空冷のまま）消費電力500Wで稼働することを、確実なものにしてきました。この80GB/500W GPU稼働という選択肢によって、XE8545は各GPUをより高負荷で駆動できるため、より高いパフォーマンスを各GPUから引き出せます。標準的なバッチサイズでの画像分類スピードを測定する、業界で一般的なベンチマーク「ResNet50」のv1.5モデルを使った検証では、500W稼働のGPUが学習に要した時間は67.78分でした。400W稼働のGPUの73.32分と比べ、7.56%速い結果となります。さらにバッチサイズを2倍にした測定では、最大13~15% (!) 高いパフォーマンスが記録されました。学習完了までのスピードを重視するお客様にとって、XE8545は、結果をより速く出せるコンピュータパワーを提供します。

学習：GPU世代間のパフォーマンスの差

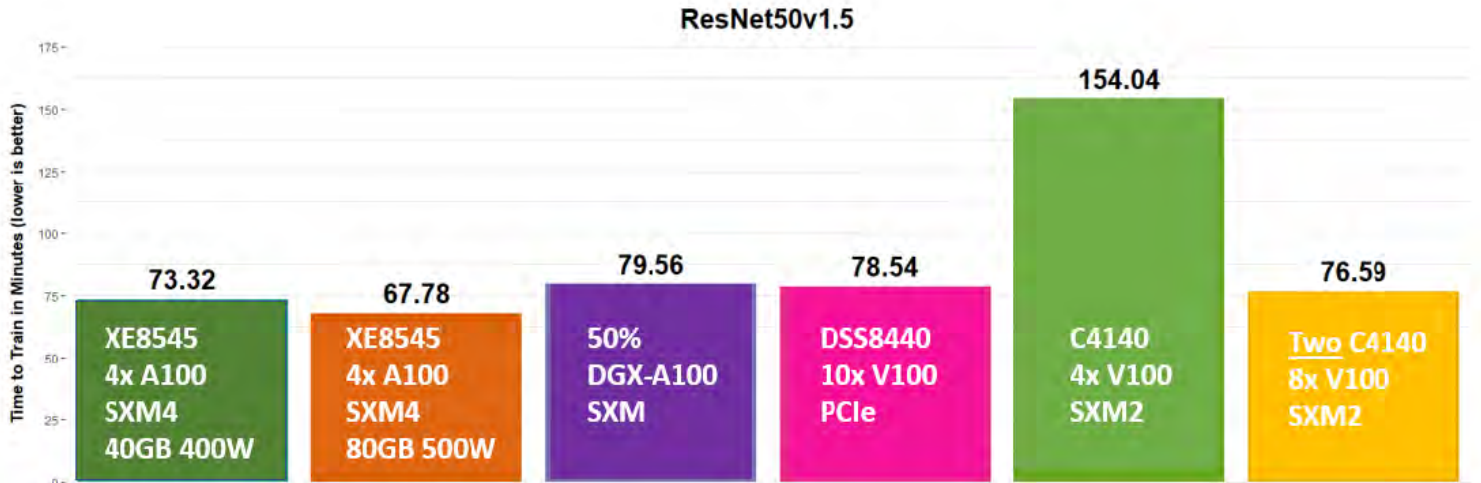


図2: MXNetでの学習パフォーマンスを分単位で表示 (学習の所要時間: 棒グラフが短いほど高いパフォーマンス)

- XE8545は、ResNet50での画像分類の学習を、トップクオリティの精度で、前世代のPowerEdgeサーバーの半分 (1/2) 未満の所要時間で完了。

上のグラフで明らかな通り、40GBメモリ搭載のGPUを稼働するXE8545は、前世代のPowerEdge C4140の**2倍以上のスピード**で画像分類モデルの学習を行います。それどころか、PowerEdge C4140 2台による並列処理よりも高いパフォーマンス (!) を出しています。80GB GPUでは、それをさらに上回る結果が示されています。これは、最新のNVIDIA GPU、最新のAMDプロセッサ、最新世代のPCIeファブリックといった最新テクノロジーの組み合わせが、XE8545に凝縮された結果の表れと言えます。また、A100ではスパース性の活用によって機械学習における行列計算の仕組みが改善されているため、これが効くワークロードでは、さらなるパフォーマンス向上も期待できます。すなわち、従来の手法では、行列に多くのゼロが含まれる場合、計算結果には何の意味もなさない「0の掛け算」が多発して時間がかかっていましたが、その点が改善されています。

PowerEdge XE8545は、あらゆる面で標準的なサーバーの運用となりますが、冷却に関しても、空冷のみで、トップレベルのパフォーマンスを発揮しています。**水冷は必要ではありません。**

画像解析と推論

- XE8545は**1秒間に15万個の画像を解析**できます。これは、A100 SXM4モジュール1基あたり、前世代のPowerEdgeに比べ**毎秒1.46倍多い画像処理数**となります。

GPU同士のピア・トゥ・ピア接続が行われない推論では、パフォーマンスがリニアに拡大する傾向にあります。XE8545は、リニアなパフォーマンス向上能力も持つことが実証されています。したがってXE8545が推論のパフォーマンスにも長けていることに、驚きはありません。学習と同じく、A100 80GB/500W GPUの方がパフォーマンスに秀でており、A100 400Wモデルより10%高速なパフォーマンスを記録しています。(もちろん消費電力も異なります。)

MIG - Multi-Instance GPUs - 推論性能を7倍高速化

A100 GPUと共に発表された革新的なテクノロジー「MIG (Multi-Instance GPU)」を利用すると、XE8545上の各A100 GPUを最大7つの異なる「スライス」にパーティション分割することができます。分割したGPUは完全に隔離され、それぞれに独立した高帯域幅メモリ、キャッシュ領域、コンピュータコアが実装されます。つまりこの技術をフル活用すれば、最大28個のハイパフォーマンスな推論GPUインスタンスを1台のPowerEdge XE8545で稼働できます。各インスタンスのパフォーマンスは、前世代GPUであるV100と同等であることが、NVIDIAにより測定されています。つまり、1枚のA100 GPUは、**前世代GPUの7倍の性能**を持つという考えが成立します。これは、GPU間のピア・トゥ・ピア コミュニケーションが作用しない推論において、特に該当します。

NVIDIA認定システム：機械学習の環境を最短で導入

XE8545は、NVIDIAがデータセンターAIのために推進する包括的な認定プログラム「NVIDIA GPU Cloud (NGC)」を取得済みです。最新のPCI Gen4ネットワーク速度で稼働するシステムとして認定を受けているだけでなく、業界最先端のAI/機会学習およびHPCソフトウェア用に、NVIDIAがチューニング・検証・最適化したフレームワークとコンテナのパッケージ「NGCカタログ」を利用できます。NGC認証済みシステムの採用により、データセンターはより安心して新しい機械学習環境を導入でき、より迅速に成果を出せます。NVIDIA認定システムに関するさらなる詳細は[こちら](#)でご参照いただけます。

新サーバー。新テクノロジー。新たなレベルのパフォーマンスアクセラレーション。

PowerEdge XE8545は、最も要求の高いレベルのAI/機械学習ワークロードと、科学技術系HPCでの演算・解析の、どちらにも対応できる高性能かつアクセラレートされた演算性能を実現する、業界の最新テクノロジーをお届けします。このサーバーは最高レベルの電力あたりパフォーマンスを、標準的な空冷環境で提供するため、オペレーション継続性の観点で、企業データセンターにやさしいという側面も持ち合わせます。



Shop PowerEdge

View All Server Options



Contact Us

For feedback and requests



Follow Us

For PowerEdge news