



FlexPod Datacenter for AIで さらに多くのデータを 分析

業務を効率化して収益を伸ばしたいのであれば、人工知能 (AI) と機械学習 (ML) の利用をお勧めいたします。Cisco UCS® C480 ML M5サーバ搭載のFlexPod® Datacenter for AIソリューションを導入すれば、すぐにAI / MLモデルをトレーニングして詳細な分析情報を得られます。

人工知能の可能性

ディープラーニング (DL) や機械学習 (ML) には、今までの手法とは比べものにならないほどすばやく詳細に情報を分析できるポテンシャルがあります。データから学んだ情報を基に迅速に判断する機能を利用すれば、競争が激化する市場で、革新的な製品とサービスを提供する組織として自社を優位に位置付けることができます。調査、パターンの分析、不正行為の検出、顧客関係の強化、サプライチェーンの最適化、プロセスの自動化など、どのような目的でも、DLとMLの手法を用いれば、デジタル情報を活かした優れたビジネスを展開できます。

従来のITインフラでは力不足

AIは、従来のITインフラの限界を超えて進化し続けています (図1)。ますます複雑化するタスクには、今までとはまったく異なるコンピューティング能力と大量のスケラブル ストレージが必要です。データが急増していることから、ディープラーニング モデルのトレーニングには数日から数週間かかるのが普通ですが、大がかりなAI、ML、DLアプリケーションのデータ ボリューム、速さ、変化に対して、コンピューティング ノードやストレージ システム、ネットワークが対応しきれずにいる例がよく見られます。お客様が直面するであろう課題は、これだけではありません。

- **自力での統合は複雑**：既製のハードウェア コンポーネントに既製のソフトウェア コンポーネントを組み合わせる統合方法では、複雑さが増して導入に長い時間がかかります。データサイエンスに投じるべき貴重なリソースをシステムの統合に投じるという無駄を招くばかりか、最適化と管理に深い専門知識が必要な、扱いづらい孤立したITリソースをいくつも生み出す結果となってしまいます。
- **拡張性に優れた予測可能なパフォーマンスの達成が困難**：従来のソリューションによるパフォーマンスの拡張は、ダウンタイムの原因になりかねません。システム停止はデータサイエンティストの生産性を低下させるだけでなく、開発者の生産性低下や抑制の利かない運用コストの増大という連鎖反応も引き起こします。

FlexPod Datacenter for AI

- AI / MLへの取り組みを導入しやすい検証済みソリューションで加速
- 20PB以上に拡張可能なグローバル ネームスペースにより、大容量の学習データセットをONTAPのFlexGroupでサポート
- 重複排除と圧縮の手法で、データストレージの容量要件を最大10分の1に低減
- 開発、テスト、トレーニング、推論の各環境をサポート

Cisco UCS C480 ML M5ラックサーバ

AI / ML環境向けに設計された最新のCisco UCS C480 ML M5ラックサーバをFlexPod環境に追加することで、以下を実現

- 2つの最新のIntel® Xeon® スケーラブル プロセッサによるハイパフォーマンス コンピューティング
- 4ラックユニット (4RU) のフォーム ファクタに8つのNVIDIA Tesla V100-32GB Tensor Core GPUを搭載し、他に例のないGPUアクセラレーションを提供
- 広帯域幅と優れた拡張性を提供するNVIDIA NVLinkテクノロジーを採り入れたマルチGPU構成
- ネットワーク、ストレージ、メモリ、OSの柔軟な選択が可能
 - 最大メモリ容量3TB
 - 最大24本のHDDまたはSSDを搭載可能
 - 最大6本のNVMeを搭載可能
 - 最大4枚のCisco UCS仮想インターフェイス カードを搭載可能

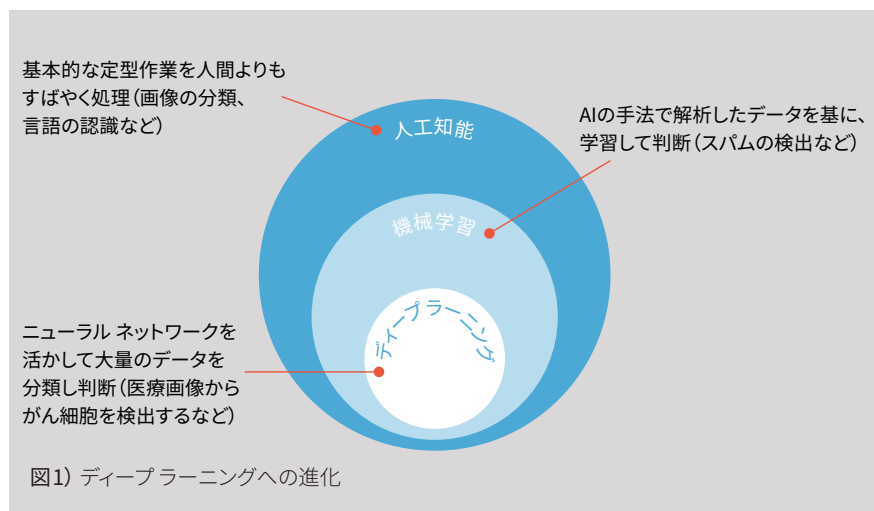
新しいデータの捉え方

FlexPodインフラに格納されているデータには、計り知れない価値があります。既存のAIソリューションやMLソリューションではデータの急増に対応できない、AIやMLソリューションをまだ導入していないというお客様は、Cisco UCS C480 ML M5ラック サーバ搭載のFlexPod Datacenter for AIソリューションを導入すると、情報分析能力の向上という新たな道を開くことができます。特に、データの重力、セキュリティ、規制上の要件により、実際にデータを使用している場所でモデルトレーニングを行う必要があるお客様にお勧めいたします。

FlexPod Datacenter for AIソリューション

FlexPod Datacenter for AIソリューションは、分析ワークロードに最適な統合インフラです。広く普及しているFlexPod Datacenterプラットフォームを基盤に、Cisco UCSブレード サーバとラックサーバ、Cisco Nexus® 9000シリーズ スイッチ、Cisco UCS 6000シリーズ ファブリック インターコネクト、NetApp ONTAP®搭載のNetApp® AFF A800フラッシュ ストレージ アレイで構成された製品です。

- **Cisco UCS C480 ML M5ラック サーバ**：グラフィック処理ユニット (GPU)、高速インターコネクト テクノロジー、高速ネットワークが統合されたこのサーバは、スピーディーなタスク処理を実現する、目的に即した妥協のない製品です。特長は、それぞれ最大28コアを搭載可能な2つのCPUと、インターコネクトにNVIDIA NVLinkを使用した最大で8つのNVIDIA Tesla V100-32GB Tensor Core GPUです。これにより、GPU間での高速通信を実現し、コンピューティングの処理速度を向上させます。NVIDIAによれば、TensorFlowのパフォーマンスはモジュールあたり最大125テラフロップスで、サーバあたり最大1ペタフロップスの処理能力を発揮できます。
- **NetApp ONTAP**：ネットアップ ストレージ システムに搭載されているONTAPソフトウェアにより、分散したデータ ソースからシームレスなデータ レイクを簡単に構築できます。複数のI/Oストリームを同時にサポートする、このデータ レイクを使用すれば、オールフラッシュ アレイからトレーニング環境へデータを高速かつ低レイテンシでストリーミングできます。トレーニングが完了したら、生成された推論モデルをリポジトリにすばやく移して推論テストと仮定検証を実施することで、Cisco UCS C480 ML M5サーバの超高速GPUアクセラレーションにより、すぐに結果が得られます。



あらゆる業界の AI / MLニーズに対応

金融

- 不正の検出
- 暗号通貨
- アルゴリズムトレーディング

ヘルスケア

- 医療画像スクリーニング
- がん細胞の検出
- 創薬
- 医学研究

製造

- 点検
- QA (品質管理)
- オートメーション

メディアとエンターテインメント

- ビデオ字幕
- コンテンツベースの検索
- 仮想現実 (VR) と拡張現実 (AR)

小売

- 購買パターンの予測
- サプライチェーンの最適化
- 自動精算
- 盗難の検出
- ターゲット マーケティング

スマートシティ

- 顔認証、ナンバープレート、不審物の検知
- トラフィックパターンの分析
- 侵入検出
- サイバーセキュリティ対策

導入アーキテクチャ

このソリューションでは、新しいCisco UCS C480 ML M5コンピューティング エンジンにより、FlexPodインフラに格納されたデータの近くで超高速GPUアクセラレーションを使用できます (図2)。Cisco UCS C480 ML M5サーバ同士との接続には、FlexPodに搭載されている別のCisco UCS ブレード サーバとラック サーバと同様、システムのファブリック インターコネクタが使用されます。AIおよびMLのモデルとアプリケーションは、このCisco UCSサーバ上で稼働し、ネットアップ データ ファブリックを使用して、FlexPodインフラのデータ収集ポイントやストレージからコンピューティング エンジンへデータを高速で移動します。これを可能にするのが、データパイプラインを簡素化、高速化して統合するNetApp ONTAPです。一貫したアーキテクチャ、ラーニング モデルワークロードとAIワークロードの高速化というメリットは、この統合アプローチによってもたらされます。

ITとビジネスの両面で優位に

FlexPod Datacenter for AIソリューションには、AIとMLのワークロードとデータベースを強力に支える機能がすべて備わっています。この拡張性にきわめて優れたアーキテクチャを導入すれば、組み込みの先進テクノロジーと統合管理アプローチを活かして、ITとビジネスの両面に多くのメリットをもたらすことができます。ソリューションはKubeflow Pipelinesと統合しているので、プライベートとパブリックの複数のクラウド プラットフォームで連携し、人工知能機能を幅広く利用することが可能です。

管理を簡易化

AIの運用に使用するCisco UCS C480 ML M5サーバは、既存のFlexPodインフラやデータ ソースと一緒に管理できます。FlexPodシステムの管理に使用している使い慣れたツールで管理できるので、余計なコストや管理オーバーヘッドが発生しません。Cisco UCS Managerを使用すれば、ファブリックや論理サーバを管理し、ワークロードに合わせてサーバの属性を変更できる、エラーのない一貫したポリシーベースの調整を実現するモデルを活用できます。

継続的な統合と開発を実現

FlexPodインフラに搭載のFlexClone®機能を利用すると、開発とテストの環境を構築してサンドボックスや構成テストを実施したり、データセットのコピーを作成したりする作業が容易になります。従来の手法でコピーを作成すると数分から数時間かかりますが、FlexCloneテクノロジーを使用すれば大容量ボリュームのクローンも数秒で作成できます。開発者やテスト エンジニアがデータセットにアクセスするために長時間待つ必要がなくなり、生産的な業務に時間をかけられるので、モデルの精度が向上して開発期間を短縮できます。

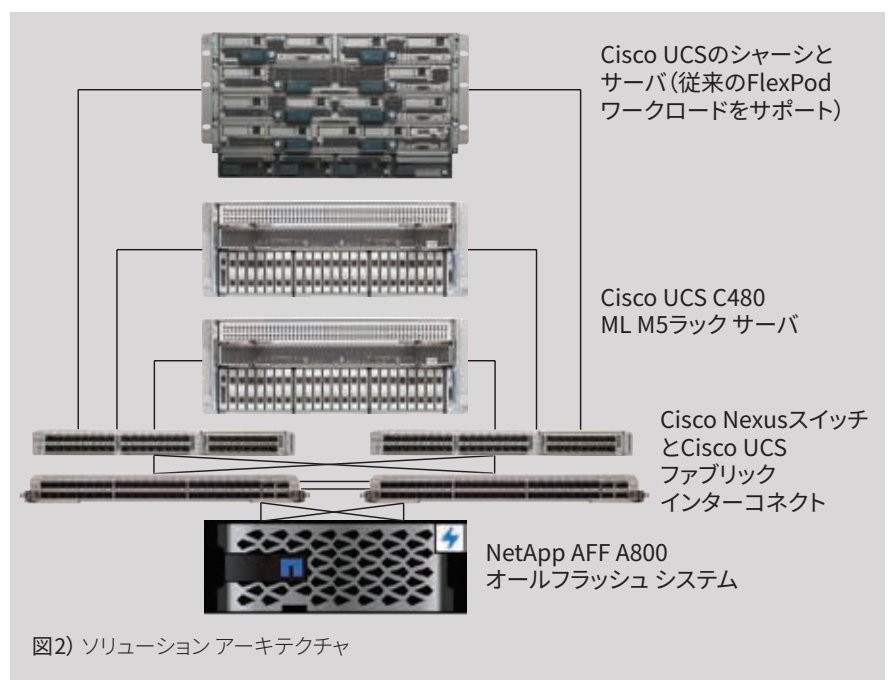
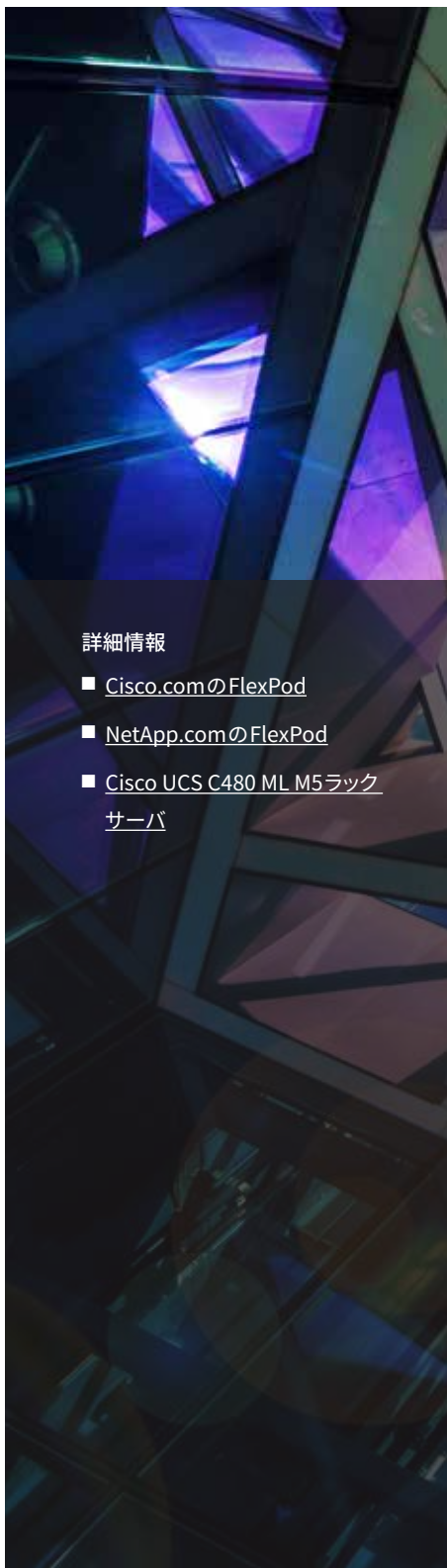


図2) ソリューション アーキテクチャ



詳細情報

- [Cisco.comのFlexPod](#)
- [NetApp.comのFlexPod](#)
- [Cisco UCS C480 ML M5ラックサーバ](#)

コンピューティング、ストレージ、GPUの能力を拡張

AI環境とML環境は、ニーズとデータセットの変化に合わせて拡張することが可能です。アプリケーションに今必要なインフラだけを導入してから、スケールアップやスケールアウトで環境を拡張できます。スケールアップの場合はFlexPodシステムやCisco UCS C480 ML M5サーバにリソースを追加し、スケールアウトの場合はFlexPodインスタンスやコンピューティング エンジンを追加します。Cisco Nexus 9000シリーズ スイッチとONTAPソフトウェアがもたらす優れた拡張性により、グローバル ネームスペースを20PB以上に拡張できる環境を構築して大規模データセットをサポートすれば、データモデルの質が向上します。

ONTAPのFlexGroup機能を使用すれば、複数のストレージ コンポーネントで構成されたスケールアウトNASボリュームを作成して、トラフィックを透過的に自動で共有できます。FlexGroupに自動負荷分散機能を併用すれば、ワークロードが優れた拡張性、高スループット、低レイテンシを必要とする場合に、それをサポートするインフラ リソースを簡単に割り当てることができるので、ストレージ管理の複雑化が避けられます。

既存のデータセットへのアクセスが容易

従来型のインフラ環境では、大容量データを使用したりコピーしたりしようとすると速度が低下し、ストレージシステムが十分に活用されない結果に陥ります。FlexPod Datacenter for AIソリューションにはONTAPソフトウェアが含まれているので、既存のデータセットに高速でアクセスできます。大容量データを、新しく導入したFlexPodプラットフォームのHDDにコピーする必要はありません。

超高速パフォーマンスを実現

短時間で質の高い結果を得るには、インフラのすべてのレイヤにスピードが必要です。ネットアップのオールフラッシュ ストレージ システムが提供する超低レイテンシのパフォーマンスと高速ネットワーク インターコネクトにより、現在の格納場所から処理を行う場所へデータを高速で移動できます。データが移動先に届いたら、Cisco UCS C480 ML M5のマルチGPU構成が提供する、サーバあたり最大1ペタフロップスの処理能力を利用して、すばやく処理を実行できます。

リスクを軽減

テクノロジーの確かな組み合わせが検証済みの、このソリューションを導入すれば、データ ライフサイクルのすべての段階で、より多くの情報を引き出すことが可能です。それだけではありません。ほぼゼロのリカバリ時間 (RT) と目標復旧時点 (RPO) を実現する継続的なデータ保護手法で、AIとMLのデータを常に保護することもできます。AIとMLのデータセットが進化したら、NVMeやNVMe over fabrics (NVMeoF) などの最新の高速フラッシュ テクノロジーやストレージクラス メモリをFlexPodシステムに組み込み、パフォーマンスを継続的に向上させることもできます。大規模アップグレードはもちろんのこと、システムを停止する必要さえありません。

プラスアルファ情報

ビジネスに欠かせない分析情報に高速でアクセスする必要があるお客様は、FlexPod Datacenter ソリューションをぜひご検討ください。エンタープライズ アプリケーションとデータベースに最適なインフラを世界中の数多くの組織に提供し続けているこのアーキテクチャは、AIやML環境向けの拡張にも当然対応しています。分析ワークロード向けに最適化された高度なテクノロジーとインフラにより、AIアプリケーションや最新型アプリケーションを使用して、よりの確かな分析情報を簡単かつ迅速に入手できます。FlexPodを導入済みのお客様なら、Cisco UCS C480 ML M5コンピューティング エンジンを追加するだけで、データセンターに格納している貴重な既存データから、さらに情報を引き出すことが可能です。

シスコシステムズ合同会社

〒107-6227 東京都港区赤坂9-7-1ミッドタウン・タワー
<http://www.cisco.com/jp>
お問い合わせ先：シスココンタクトセンター
0120-092-255 (フリーコール、携帯・PHS含む)
電話受付時間：平日10:00～12:00、13:00～17:00
<http://www.cisco.com/jp/go/contactcenter>

© 2019 Cisco Systems, Inc. All rights reserved.

Cisco、Cisco Systems、およびCisco Systemsロゴは、Cisco Systems, Inc. またはその関連会社の米国およびその他の一定の国における登録商標または商標です。

本書類またはウェブサイトに掲載されているその他の商標はそれぞれの権利者の財産です。

この資料の記載内容は2019年2月現在のものです。

この資料に記載された仕様は予告なく変更する場合があります。

ネットアップ合同会社

TEL:03-6870-7600 Email:ng-sales-inquiry@netapp.com

ネットアップは卓越したコスト削減を実現し、企業競争力を高める

ストレージとデータ管理ソリューションを提供しています。

詳細に関しては、<http://www.netapp.com/jp>をご覧ください。

© 2019 NetApp, Inc. All rights reserved. 本ドキュメントの記載内容は、いかなる箇所においてもNetAppからの書面による事前の許可なく複製、転用することを禁じます。記載事項は、予告なく変更される場合があります。NetApp、NetAppのロゴ、<http://www.netapp.com/jp/legal/netapptmlist.aspx>に記載されているマークは、NetApp, Inc. の商標です。CiscoおよびCiscoのロゴは、米国およびその他の国におけるCiscoまたはその系列会社の商標または登録商標です。Ciscoの商標一覧は、www.cisco.com/go/trademarksでご覧いただけます。本書に記載のサードパーティの商標は、各社に帰属する財産です。「パートナー」または「partner」という用語の使用はCiscoと他社との間のパートナーシップ関係を意味するものではありません。

LE-68702-00-jp 1/19