DELL Technologies

GPUで高速になったAIを活用してイノベーションを加速する

デル・テクノロジーズとNVIDIAのソリューションで、人工知能のポテンシャルを最大限引き出す

開始する

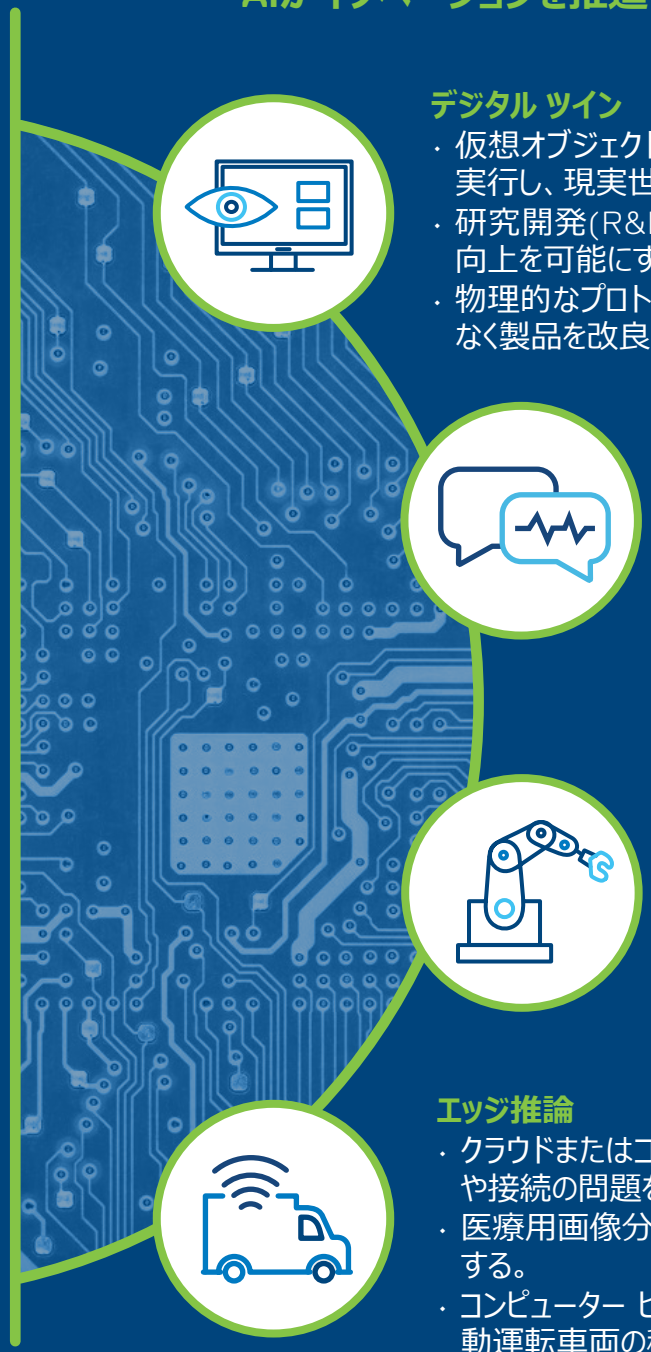
NVIDIA

Table of Contents

日々変化する世界にAIが力を与える

AIの時代には、データへのアクセス増加と新しいデータ管理技術により、あらゆるタイプと規模の組織にAI主導のインサイトを生み出す機会がもたらされます。処理能力の向上からエンタープライズ向けマルチクラウドの台頭まで、AIは至る所で採用されています。企業はオンプレミス、プライベートクラウド、パブリッククラウド、エッジでAIを活用して、さまざまな新規のワークロードに対応できるようになります。

AIがイノベーションを推進する主な領域



デジタル ツイン

- 仮想オブジェクト、システム、プロセスでシミュレーションを実行し、現実世界での動作を予測する。
- 研究開発(R&D)サイクルの改善、高速化、コスト効率向上を可能にする。
- 物理的なプロトタイプ作成にコストや時間を浪費することなく製品を改良する。

生成AI/自然言語処理(NLP)/大規模言語モデル(LLM)

- 機械が人間の言語を理解できるようにする。
- コンピューター システムと人間との個人的な対話を支援する。
- GPT/Transformerモデル、チャットボット、デジタル アシスタント、感情分析、不正行為検出などに使用する。

コンピューターの支援によるデザイン、製造、エンジニアリング(CAD/CAM/CAE)

- 製品の設計や製造の革新的な手法に関するインサイトを得る。
- 製品の革新性と品質を向上させながら、市場投入までの時間を短縮する。
- デザインやエンジニアリングを変革し、未来の製造現場に役立てる。

エッジ推論

- クラウドまたはコアからのデータ転送に関するレイテンシーや接続の問題を克服する。
- 医療用画像分析に使用して、救急医療対応をサポートする。
- コンピューター ビジョンを活用し、現場の設備の分析や自動運転車両の稼働を行う。

AI導入の障壁を取り除く

Aiに対応するデータサイエンスの専門技術とITインフラストラクチャのスキルが不足。
スキル不足は、AIの採用または利用拡大に対する最大の障害の1つです。

データ処理作業の量と複雑さの増大。
データの量と複雑さが増しているため、従来の分析手法では対応できなくなっています。

結果を得るまでのスピードが遅い。
処理能力とスキルが不十分であるため、データの価値の認識に時間がかかります。

AIにおける主導的立場を取る

62%
人材、プロセス、テクノロジーなど、AIの支出を増やす計画をしている組織の割合。[1]

2倍
適切なスキルがある場合に、製造にAIが導入される可能性。[2]

86%
AIでの成功を目指すにあたって技術的な障害が1つ以上あると認識している組織の割合。[1]

69.3%
AIワークロードを実行しているAI分野のリーダーの増加率（平均で8基のGPUを搭載したサーバーを使用）。[3]

「AIを利用可能」から「AIとの互換性あり」へ

デル・テクノロジーズは、AI導入のあらゆる過程で企業のニーズに対応します。AI活用を始めたばかりの企業から、ディープ ラーニング(DL)クラスター導入の準備が整った企業まで、デル・テクノロジーズの包括的なソリューションのポートフォリオが、未開拓の市場機会の発見と活用をお手伝いします。

Dell PowerEdgeサーバーはAIソリューションの基礎となる構成要素です。AIの利用を開始し、必要に応じて拡大するために必要なパフォーマンス、密度、効率性が得られます。さらに、最大12基のNVIDIA®グラフィックス処理ユニット(GPU)もサポートされるため、AIワークロードが加速し、より迅速に結果が得られます。

先進のAIソフトウェア企業と提携しているため、データやAIポートフォリオに関するサポートが必要な場合に、どのようなニーズにも適切なソリューションで対応できます。ワークステーションからデータセンター、エッジ、クラウドに至るまで、技術革新の統合エコシステムを活用し、AIに対する総合的なアプローチで成功へと導きます。

インテリジェント ビジネス向けソリューションでタイムトゥバリューを短縮

AIを迅速かつシンプルに導入できるように、DellではValidated Designs for AIのポートフォリオを用意しています。このソリューションは以下のような特長を備えています。

シンプルなAI

Dell Validated Designs for AIは、NVIDIA GPU、NVIDIA AI Enterpriseスイート、その他のNVIDIAテクノロジーと連携した技術を導入し、検証を受けています。AI導入を推進するように最適化されたソリューション スタックを迅速かつ簡単に導入できるようにします。

より迅速なAIによるインサイト

NVIDIAのGPUによる高速化構成は、最適化されたインフラストラクチャのAIツールおよびフレームワークとともに提供されるため、開発チームとITチームの実稼働までの時間を短縮できます。

実績あるAIの専門知識

世界レベルのデル・テクノロジーズのサービスとサポートに裏打ちされた、エンジニアリングテスト済みのAIソリューションであるため、自信を持って導入することができます。ソフトウェアおよびハードウェア サポートの一元化された窓口として、ProSupport Plusをお選びください。

1日目
AIモデルを活用するための準備状況[4]

10倍
モデル生成の速度[4]

60%減少
AIインフラストラクチャ管理に要する時間[5]

20%高速
カスタマイズされたシステムを使用したAIプロジェクトのタイムトゥバリュー[5]

50%高速
AI開発にかかる時間[6]

AIインサイトを迅速に取得できるように構築

Dell PowerEdgeサーバーでAIのメリットを最大限に活用

Dellは、場所と方法を問わず、AIを業務に導入し、イノベーションを加速するお手伝いをします。NVIDIA GPUで高速化された新しいPowerEdgeサーバーが、迅速に取得できるインサイトでAIワークロードを強化します。Dell PowerEdgeは、コンピューティングの高速化を支援し、強化されたAIワークロードの成果とより優れたインサイト、推論、可視化を推進します。

あらゆる場所でトランスフォーメーションを加速するDell PowerEdgeサーバー

AIを活用してイノベーションを加速	ゼロ トラストの導入の促進	オートメーションの促進
エッジからクラウドまでモダナイズ	セキュリティの強化	運用効率の向上

サステナビリティ

Dell PowerEdge XEサーバー

高速化のために最適化された特定用途向け設計。複雑なコンピューティング、AI/ML/DL、ハイパフォーマンス コンピューティング(HPC)の高負荷ワークロードに対応

	PowerEdge XE9680 妥協のない高速なAIを実現する、パワフルで柔軟な機能	PowerEdge XE9640* リアルタイムのAIインサイトを提供する高密度のスマート冷却サーバー	PowerEdge XE8640* GPU最適化設計の卓越したパフォーマンス	PowerEdge XE8545 AI、機械学習(ML)、HPC向けオールインワン サーバー
アプリケーションとユース ケース	・AI/ML/DLトレーニング、HPC、CRISP ・生成AI ・医療、クラウド サービス プロバイダー(CSP)、金融、学術研究	・AI/ML/DLトレーニング、HPCモデリングおよびシミュレーション	・中規模データ セット言語モデル、NLP、モデリング、シミュレーション ・AI/ML/DLのトレーニングと推論、画像認識	・AI/MLのトレーニングと推論、小規模および中規模データ セット言語モデル
プロセッサー	・2 x 第4世代インテル® Xeon® スケーラブル プロセッサー	・2 x 第4世代インテル Xeon スケーラブル プロセッサー	・2 x 第4世代インテル Xeon スケーラブル プロセッサー	・2 x 第3世代AMD® EPYC™プロセッサー
GPUサポート	・最大8 x NVIDIA H100 SXM5またはNVIDIA A100 SXM4 GPU、NVLink™完全接続付き	・最大4 x インテルGPU	・最大4 x NVIDIA H100 SXM5 GPU、NVLink完全接続付き	・最大4 x NVIDIA A100 SXM4 GPU、NVLink付き
機能	・6Uラック高 ・空冷、35°Cまで ・32 x DDR5 DIMM ・最大10 x 16 PCIe Gen5スロット	・2Uラック高 ・CPUおよびGPUは液体冷却で稼働 ・32 x DDR5 DIMM ・最大2 x PCIe Gen5スロット	・4Uラック高 ・空冷、35°Cまで ・32 x DDR5 DIMM ・最大4 x PCIe Gen5スロット	・4Uラック高 ・空冷、35°Cまで ・32 x DDR4 DIMM ・最大4 x 16 PCIe Gen4スロット

* 2023年上半期に提供

Dell PowerEdgeラックサーバー

柔軟性の高い、メインストリーム コンピューティングの基礎となるサーバー。広範囲にわたるアプリケーション、ユース ケース、ワークロードに対応

	PowerEdge R760xa* GPUベースのワークロード向けフラッグシップ サーバー	PowerEdge R750xa 特定用途向けに設計された柔軟性	PowerEdge R750/7525/7515 R650/6525/6515 メインストリームのパフォーマンス	PowerEdge XR12 エッジのパフォーマンス
アプリケーションとユース ケース	• AI/ML/DLのトレーニングと推論、分析、HPC • 生成AI、密度推論 • VDI、高性能グラフィックス	• AI/ML/DLのトレーニングと推論、分析、HPC • VDI、高性能グラフィックス	• 低負荷のAI/ML/DLのトレーニングと推論 • VDI、高性能グラフィックス • エッジ	• エッジAIのトレーニングと推論 • 電気通信 • レンダリング/モデリング
プロセッサー	• 2 x 第4世代インテル Xeon スケーラブル プロセッサー	• 2 x 第3世代インテル Xeon スケーラブル プロセッサー	• 最大2 x 第3世代インテル Xeon スケーラブルまたは第3世代AMD EPYCプロセッサー	• 1 x 第3世代インテル Xeon スケーラブル プロセッサー
GPUサポート	• 最大4 x ダブルワイドまたは12 x シングル ワイドのNVIDIA PCIe GPU	• 最大4 x ダブルワイドまたは6 x シングル ワイドのNVIDIA PCIe GPU	• 最大3 x ダブルワイドまたは6 x シングル ワイドのNVIDIA PCIe GPU	• 最大2 x ダブルワイドまたはシングル ワイドのNVIDIA PCIe GPU
機能	• 2Uラック高 • 空冷、35°Cまで • 32 x DDR5 DIMM • 最大4 x PCIe Gen5スロット	• 2Uラック高 • 空冷、35°Cまで • 32 x DDR5 DIMM • 最大4 x PCIe Gen4スロット	• 1Uまたは2Uラック高 • 空冷、35°Cまで • 32 x DDR4 DIMM • 最大8 x PCIe Gen4スロット	• 2Uラック高 • 運用上の許容範囲–5°C～55°C • 最大4 x PCIe 4 Gen4スロット

* 2023年上半期に提供

ベア メタルに匹敵するパフォーマンスを達成

97.5%
VMwareを使用したベア メタルのパフォーマンスとの比較[7]

66%
ワットあたりのパフォーマンス向上率[8]

66%
High-Performance Linpack (HPL) パフォーマンスの向上率[9]

NVIDIA GPUでAIの能力を解放する

デル・テクノロジーズはNVIDIAと提携しています。NVIDIAは、HopperおよびAmpere GPUを備え、エントリーレベルからメインストリーム、最高レベルのパフォーマンスにまで対応する、包括的なポートフォリオを提供している唯一のベンダーです。エッジ、クラウド、オンプレミスを問わず、幅広いAIアプリケーションを加速させる汎用性を備えています。

H100 SXM
最高レベルのパフォーマンス重視のAI、MLのトレーニング、エクサスケールのHPC

- 3,958 TFLOPS FP8 Tensorコア*
- NVLink：900GB/秒のPCIe Gen5
- 最大7 x MIG、各10GB

H100 PCIe
最高レベルのパフォーマンス重視のAI、MLのトレーニング、エクサスケールのHPC

- 3,026 TFLOPS FP8 Tensorコア*
- NVLink：600GB/秒のPCIe Gen5
- 最大7 x MIG、各10GB
- NVIDIA AI Enterpriseソフトウェア付属
- NVIDIA vGPU
- ソフトウェア サポート

A100
パフォーマンス重視のAI、MLのトレーニングと推論

- 312 TFLOPS FP16 Tensorコア*
- 最大2基のGPUに対応するNVLink Bridge：600 GB/秒
- 最大7 x MIG、各10GB
- NVIDIA AI Enterpriseソフトウェア付属
- NVIDIA vGPUソフトウェア サポート

A30
メインストリームのグラフィックスおよびAI推論

- 165 TFLOPS TF32 Tensorコア*
- 最大2基のGPUに対応するNVLink Bridge：200 GB/秒
- 最大4 x GPUインスタンス、各6GB
- NVIDIA AI Enterpriseソフトウェア付属
- NVIDIA vGPUソフトウェア サポート

A10
AIを使用した、メインストリームのエンタープライズ サーバー向けの高速グラフィックスとビデオ

- 250 TFLOPS FP16*
- 16 x PCIe Gen4
- NVIDIA AI Enterpriseソフトウェア付属
- NVIDIA vGPUソフトウェア サポート

桁違いの飛躍：NVIDIA H100 TensorコアGPU

H100 GPUをデータセンターの規模で導入すると極めて高いパフォーマンスを確保でき、次世代のエクサスケールのHPCや、兆単位のAIパラメーターに手が届くようになります。

*構造的スパース性を有効にした場合。

9倍
最大モデルでのAIトレーニングの速度[10]

30倍
最大モデルでのAI推論パフォーマンスの速度[11]

3,958
TensorコアのFP8 TFLOPS[12]

L40
最高レベルのパフォーマンス重視のグラフィックスとレンダリング

- 90.5 FP32 TFLOPS（非Tensor）
- 724.1 FP8 Tensor TFLOPS、FP32 accumulate有効*
- NVIDIA AI Enterpriseソフトウェア付属
- NVIDIA vGPUソフトウェア サポート
- NVIDIA OmniverseのOVXサポート

A40
パフォーマンス重視のグラフィックスとレンダリング

- 299.4 BF16 Tensor TFLOPS、FP32 accumulate有効*
- NVLink 112.5 GB/秒（双方向）
- NVIDIA AI Enterpriseソフトウェア付属
- NVIDIA vGPUソフトウェア サポート

A16
マルチメディア対応のVDIで、CAD/CAM/CAEを含むリモートワークが可能

- 4 x 35.9 TFLOPS FP16*
- 16 x PCI Express Gen 4
- NVIDIA AI Enterpriseソフトウェア付属
- NVIDIA vGPUソフトウェア サポート

L4
効率的なビデオ、グラフィックス、AIを実現する、画期的な汎用アクセラレーター

- 485 TFLOPS FP8*
- 16 x PCIe Gen4
- NVIDIA AI Enterpriseソフトウェア付属
- NVIDIA vGPUソフトウェア サポート

A2
エッジでのAI推論向けのエントリーレベルGPU

- 36 TFLOPS FP16 Tensorコア*
- 8 x PCIe Gen4
- NVIDIA AI Enterpriseソフトウェア付属
- NVIDIA vGPUソフトウェア サポート

Dell PowerEdgeサーバーのNVIDIA GPUサポート状況については、
GPUマトリックスを参照してください。

*構造的スパース性を有効にした場合。

NVIDIAテクノロジーを内蔵

ソリューションの中核をなすDell PowerEdgeサーバーは、NVIDIAテクノロジーを統合してAIワークロードを加速し、成果を得るまでの時間を短縮します。

NVIDIA仮想GPU (vGPU)

NVIDIA vGPUソフトウェアを使用すると、複数のVMでGPUリソースを共有し、どこからでもあらゆるデバイスにアクセス可能になります。

NVIDIAマルチ インスタンスGPU (MIG)

NVIDIA MIGは、GPUを7つのインスタンスに分割してそのパフォーマンスと価値を拡張し、すべてのワークロードをサポートします。また、加速されたリソースをより多くのユーザーが使用できるようにします。

NVIDIA H100 GPU

NVIDIA H100 TensorコアGPUは、すべてのデータセンターにこれまでにないパフォーマンス、拡張性、安全性をもたらします。NVIDIA H100 PCIe GPUにはNVIDIA AI Enterpriseソフトウェア スイートが付属しているため、AI開発および導入を合理化することができます。AIトレーニングが9倍[10]、最大モデルでのAI推論パフォーマンスが30倍速くなります。[11]

NVIDIA A100 GPU

AIワークロードを加速して、前世代と比較してパフォーマンスを最大20倍向上させます。A100は、世界初の高速GPUインターコネクト、NVLink Bridgeをサポートし、従来のPCIeベースのソリューションより大幅に高速なマルチGPUシステムを提供します。[14]

VMware vSphere上のNVIDIA AI Enterprise

NVIDIA AI Enterpriseは、エンドツーエンドのクラウド ネイティブ スイートです。AIの専門知識は不要で、コンテナ、フレームワーク、ワークフローを使用してAI導入の開始をお手伝いします。このスイートは、デル・テクノロジーズよりNVIDIA Certified Systems™での動作認定を受けています。またAI開発および導入ツール、インフラストラクチャ最適化ソフトウェア、グローバル エンタープライズ サポートが付属しており、AIプロジェクトを予定通り進めることができます。このスイートを使用すると、インフラストラクチャの導入ではなく、AIのビジネス バリューの活用に注力できるようになります。

NVIDIA-Certified Systems

NVIDIA Certified Systems™であるDell VxRail HCIとDell PowerEdgeは、NVIDIA GPU、NVIDIA ConnectX®スマート ネットワーク インターフェイス カード(SmartNIC)、NVIDIA BlueField® 最適化された構成でDPUを統合します。これらはパフォーマンス、管理性、安全性、拡張性について認定済みで、NVIDIAとデル・テクノロジーズからのエンタープライズ グレードのサポートも用意されています。

NVIDIA LaunchPad

厳選された無料のラボ体験です。必要なハードウェア/ソフトウェア スタックにすぐに短期間アクセスして、AI、データサイエンス、3Dデザイン コラボレーションおよびシミュレーションなど、エンドツーエンド ソリューションのワークフローを体験できます。NVIDIA LaunchPadは、Dell PowerEdgeサーバーで構築されています。詳細については、nvidia.com/dell-launchpadを参照してください。

NVIDIA BlueFieldデータ プロセシング ユニット(DPU)

BlueField DPUは、広範囲にわたる高度なネットワーキング、ストレージ、セキュリティサービスの負荷をオフロードし、高速化と分離を行って、クラウドからデータセンター、エッジに至るあらゆる環境のすべてのワークロードに、安全かつ高速なインフラストラクチャを提供します。

推奨構成

ワークロード	ユース ケース	推奨構成	
HPC/AI/ML/DLトレーニング 生成AI	• 自然言語処理(NLP) • 大規模言語モデル(LLM) • 大規模推奨エンジンのトレーニング • HPC、モデリング、シミュレーション	• PowerEdge XE9680	• H100 SXM GPU
HPC/AI/データベース/分析	• HPC • AI/ML/DLトレーニングと推論 • 中規模データ セット言語モデル • NLP • 画像認識 • モデリングとシミュレーション • 分子動力学 • ゲノム配列解析	• PowerEdge XE9680 • PowerEdge XE8640	• H100 SXM GPU • A100 SXM GPU
高性能グラフィックス/VDI/モデリング	• デジタル ツインと3Dワールド/メタバース • 高性能グラフィックス • CAD/CAM/CAE • 仮想化 • HPC	• PowerEdge R760xa • PowerEdge R750xa • PowerEdge R750 • PowerEdge R7525	• L40 GPU • A40 GPU
メインストリームAI	• HPC • 分析 • GPUによるデータベースの高速化 • AI/MLのトレーニングと推論 • 低負荷のAIトレーニング • A/MLトレーニングと推論	• PowerEdge R960 • PowerEdge R760xa/R750xa • PowerEdge R760/R750 • PowerEdge R7625/7525 • その他のラックサーバー	• A2、A10、A30、A100 GPU • L4 GPU
VDIと仮想化	• パワー ユーザー向けの高機能コラボレーション • モバイルワーカー向けVDI	• PowerEdge R760xa/R750xa • PowerEdge R760/R750 • PowerEdge R7625/R7525 • PowerEdge R960	• A10またはA16 GPU • L4 GPU
メインストリームのグラフィックスおよびVDI	• グラフィックスのレンダリング	• PowerEdge R760/R750 • PowerEdge R7625/R7525 • その他のラックサーバー	• A10 GPU • L4 GPU
推論/エッジ/VDI	• エッジ推論	• PowerEdge XR12 • PowerEdge R760xa/R750xa • PowerEdge R760/R750 • PowerEdge R7626/R7525 • その他のラックサーバー	• A2 GPU • L4 GPU

お客様の成功事例

1 大規模なコンテンツ推奨システムを提供しているTaboola

Taboola®は、驚異的なコンピューティング性能とシンプルな管理を通じて最大限のパフォーマンス、拡張性、オートメーションを確保し、毎日数十億に及ぶ関連推奨コンテンツを提供するAIモデルのトレーニングと実行を可能にしました。

150,000
毎秒処理されるAI駆動型リクエスト数

6倍
AIベースの推論が改善

50ミリ秒
リアルタイムの推奨コンテンツを提供するまでの時間

「今では、AIベースの推論でパフォーマンスが最大で6倍になっています…このおかげでコストを削減できました。」

— Taboola、ITおよびサイバーセキュリティ部門VP、Ariel Pisetzky氏

事例を読む。

2 列車を全速力で動かしているDuos Technologies

Duos Technologies®は、NVIDIA GPUで高速化されたDell PowerEdgeサーバーとエッジのAIを連携させています。データの処理と分析をリアルタイムで実行し、列車が点検のために停車する必要がないよう、迅速に実用的なインサイトを取得しています。

120分の1
点検時間の短縮

1.3TB
1日に処理および分析するサイトあたりのデータ量

$3,000 USD
サーバーのリカバリーで節約されたインスタンスあたりの金額

「PowerEdgeサーバーには本当に助かっています。AIモデルを使用して、カメラやセンサーから常時送られてくる画像などのデータの処理と分析を行ってくれます。」

— Duos Technologies、最高技術責任者、David Ponevac氏

事例を読む。ビデオを見る。

3 科学的発見を加速させたケンブリッジ大学

デル・テクノロジーズは、非常に高い性能が要求される、現代のデータ主導型シミュレーションとAIの課題解決に役立てるため、ケンブリッジ大学によるHPCとデータストレージ システムの構築を支援しました。

3.8
ペタフロップス

74,000
コア

500
ギガバイト/秒

「研究者からのコンピューティング能力の要望は尽きません。提供するとすぐ使い果たしてしまうでしょう。ケンブリッジのスーパーコンピューターは、AIの作業に必要なスーパーコンピューティング能力を、高速かつ手頃な価格で研究者に提供することができます。」

ーケンブリッジ大学、リサーチ コンピューティング サービス担当ディレクター、Paul Calleja博士

事例を読む。

4 AIの能力を拡大したピサ大学

デル・テクノロジーズ、VMware、NVIDIAのソリューションのおかげで、ピサ大学は従来のワークロードとAIワークロードを同じシステムで実行し、ニーズに柔軟に対応しつつ、ITリソースを活用できるようになりました。

ゼロ
AI固有のシステムのサイロの数

1つ
仮想デスクトップとアプリを導入したプラットフォーム

複数
同じインフラストラクチャでサポートされているワークロード

「仮想GPUの最大のメリットは柔軟性です。エンタープライズ インフラストラクチャをデザインし、AIワークロードに適合させることができるのです。」

ー ピサ大学、最高技術責任者、Maurizio Davini氏

事例を読む。

デル・テクノロジーズを選ぶ理由

世界各地のCustomer Solution Centerとコラボレーション

世界各地にあるCustomer Solution Centerのいずれかのデル・テクノロジーズ エンジニアリング チームと協力し、HPCとAIのCenters of Excellenceのリソースを活用したり、HPCおよびAIイノベーション ラボで、現実世界のシステムのテストやチューニングを実行したりできます。

Dell APEXで、AIアズ ア サービスを存分に使用

Dell APEXでは、アズ ア サービス(aaS)で提供されるシンプルかつ一貫性のあるクラウド体験で、どのような場所でも、インテリジェントな成果を迅速に得るために必要な、AIに最適化されたソリューションを入手できます。Dell APEXは、オンプレミス、オフプレミス、エッジのAIのクラウド運用モデルを提供するため、どのような規模のデータからでも測定可能な価値を作り出すことができます。

サービスで成功までの時間を短縮

Dell Technologies Servicesには、コンサルティング、導入、サポート、教育などが含まれます。初期セットアップや人材のスキル向上から継続的サポートまで、さまざまな形でAI環境の迅速な導入と最適化を支援します。マネージド サービスとレジデンシー サービスは、IT管理のコスト、複雑さ、リスクの軽減に役立ちます。そのため、デジタル イノベーションとトランスフォーメーションにリソースを集中して投入できます。

35,000人以上

AIによる成功のロードマップ作成に関わるサービスおよびサポートのメンバー[15]

0ドル

デル・テクノロジーズのAI専門家とのコラボレーションの費用[16]

10か所

世界各地にあるDell HPCとAIのCenter of Excellenceの数[17]

インテリジェントな成果を推進

デル・テクノロジーズは、あらゆるタイプと規模の組織が機会を見つけ出し、所有するデータの持つポテンシャルを最大限に引き出すためのお手伝いをします。デル・テクノロジーズでは、35以上のデータサイエンス チームが450以上のAIプロジェクトを進めており、1,800名以上のチーム メンバーがデータからインサイトを抽出するため奮闘しています。デル・テクノロジーズは、実証済みのAIの専門知識で、ITを効率化し、リスクを低減して、お客様のインサイトとエクスペリエンスを改善します。そして、オンプレミス、オフプレミス、エッジにわたるハイブリッドクラウド全体において、一貫性のある方法でこれを達成します。

デル・テクノロジーズは、AI時代における成功を支援します。

詳細

Dell.com/PowerEdge

デル・テクノロジーズとNVIDIA

AIワークロードの実現と高速化

デル・テクノロジーズとNVIDIAは連携して、AI、ML、DLワークロードを高速化するため、技術的に検証済みのハードウェアとソフトウェアを提供します。また、デル・テクノロジーズは、最新鋭のNVIDIA GPU、SmartNICとDPU、NVIDIA AI Enterpriseソフトウェアと連携するサーバーおよびソリューションにも投資しています。NVIDIAとデル・テクノロジーズは、今までに想像しなかったレベルでAIを活用できるようお手伝いします。

[1] ESGインフォグラフィック、『AI主導の未来に向けてコンピューティングをモダナイズするDellサーバーとNVIDIA』、2022年。

[2] 「AI評価者」との比較。出典：デル・テクノロジーズとNVIDIA後援のIDCアナリスト報告、『AIのためのスキル拡張：早期導入からの教訓』、2022年8月。

[3] 「AI評価者」との比較。出典：デル・テクノロジーズとNVIDIA後援のIDCホワイト ペーパー、『AIを使用した生産現場のビジネスが出遅れた者に教えること』、2022年8月

[4] Dell Precisionデータサイエンス ワークステーションを使用。『DSW準備1日目ガイド』を参照。

[5] Dell Validated Designs for AIを使用。Forrester、『Dell Validated Designs For AIのTotal Economic Impact™』、2022年8月。

[6] Dell Precisionデータサイエンス ワークステーションを使用。デル・テクノロジーズの事例、『AIディープ ラーニングがデータサイエンスの地平線を拡張する』、2021年2月。

[7] パフォーマンス テストにおいて、デル・テクノロジーズとVMwareを使用した構成で、同じサーバーのペア メタルのパフォーマンスに対し最大で97.5%を達成。出典：Principled Technologiesのレポート、『Dell PowerEdge R7525サーバーで仮想GPUを使用し、ペア メタルに匹敵する画像分類ワークロードの推論スループットを達成』、2022年7月。

[8] Dell PowerEdge R750xaでNVIDIA H100構成とA100構成を比較すると、ワットあたりのパフォーマンスが66%向上。出典：デル・テクノロジーズ技術文書、『PowerEdge R750xaとNVIDIA H100 PCIe GPU：ワットあたりのHPCパフォーマンスが66%向上』、2022年。

[9] HPLベンチマークで、NVIDIA H100構成のPowerEdge R750xaが、NVIDIA A100構成と比較して67%のパフォーマンス向上を達成。出典：デル・テクノロジーズ技術文書、『PowerEdge R750xaとNVIDIA H100 PCIe GPU：ワットあたりのHPCパフォーマンスが66%向上』、2022年。

[10] H100は、第4世代TensorコアとFP8精度のTransformer Engineを備え、混合エキスパート(MoE)モデルにおいて、前世代と比較するとトレーニングが最大9倍高速化。出典：NVIDIA、NVIDIA H100 TensorコアGPU、2023年1月にアクセス。

[11] 前世代との比較。出典：NVIDIA、NVIDIA H100 TensorコアGPU、2023年1月にアクセス。

[12] NVIDIA H100 SXM GPU、構造的スパース性を有効化。スパース性なしの場合、仕様上の性能は半減する。出典：NVIDIA、NVIDIA H100 TensorコアGPU、2023年1月にアクセス。

[13] NVIDIA Webサイト、この時代の最も重要な業務を高速に行う、2022年6月にアクセス。

[14] NVIDIA Webサイト、NVIDIA NVLink、2022年6月にアクセス。

[15] デル・テクノロジーズ、主な情報、2022年。

[16] デル・テクノロジーズのCustomer Solution CenterとHPCおよびAIイノベーション ラボにて。詳細については、セールス担当者までお問い合わせください。

[17] 詳細については、dell.comを参照。