

AIの導入が進むにつれ、データセンター、クラウドインフラストラクチャ、運用の各チームは、それぞれ独自のコンピューティング要件とデータ管理要件を持つ、新たなスタンドアロンAIと組み込みAIのワークロードに対応する必要がある。クラウド環境を購入または拡張する際には、パフォーマンス、市場投入までの時間、コスト、セキュリティを考慮する。

# 2025年のAIインフラストラクチャ：データセンターとクラウドに対する投資バランス

February 2025

Written by: Ashish Nadkarni, Group Vice President and General Manager, Worldwide Infrastructure Research

## はじめに

AIは、効率性、オペレーション、インテリジェントな意思決定を向上させることで、組織を変革している。企業はAIを活用して、デジタル製品やサービスを共同で開発しており、競合他社に比べて収益の成長を倍増させる可能性がある。現在のAI投資は、企業価値の創出に重点が置かれており、そのためにはITインフラストラクチャに対する新たな支出が求められる。AIは、モデルのファインチューニング、RAG (Retrieval-Augmented Generation : 検索拡張生成)、AIを活用したアプローチによるDevOpsの強化といった新たな手法の導入を促進している。

しかし企業は、インフラストラクチャに対する投資について、より賢明な判断を下すようになっている。IDCは、2026年までにGlobal 2000企業のCレベルの経営幹部の70%が、新しいAI対応インフラストラクチャに対する支出を正当化するために、ROI (Return on Investment : 投資収益率) と価値分析を組み合わせたテクノロジーとビジネス成果の連携を要求すると予測している。これまでの取り組みとは異なり、ほとんどの企業はAIワークロードに対してハイブリッドITの運用戦略を採用するであろう。これは、アクセラレーターとプロセッサのハイブリッド化や、パブリックとプライベートのクラウド環境のハイブリッド化、スタンドアロンAIと組み込みAIのワークロードのハイブリッド化を意味する。

以前の取り組みでも新規インフラストラクチャに対する投資につながったが、それだけでなく、AI主導の投資では、特にデータセンターとクラウドインフラストラクチャのチームを中心に、すべてのレベ

## AT A GLANCE

### 概況

IDCの予測：

- » AIシステムに対する企業の支出は、2022年から2026年にかけて年率27%で増加する。
- » 2024年には、Generative AI (生成AI) ソリューションに対する企業の投資は200億ドルを超えている。これには、オンプレミスインフラストラクチャおよびクラウドベースのIaaS (Infrastructure as a Service) ソリューションに対する支出が含まれている。
- » 2028年までに、エンタープライズAIワークロードの75%が、パフォーマンス、コスト、コンプライアンスを最適化しながら価値実現までの時間を短縮するために、目的に合ったハイブリッドインフラストラクチャに導入される。

ル、すべてのチームに渡って計画を策定する必要がある。AIは、インフラストラクチャの計画や支出のみならず、専門的なスキルの開発や人材確保に対する投資にも大きな影響を与えることになる。IDCは、2028年までに、エンタープライズAIワークロードの75%が、パフォーマンス、コスト、コンプライアンスを最適化しながら価値実現までの時間を短縮するために、目的に合ったハイブリッドインフラストラクチャに導入されると予測している。

したがって、企業のインフラストラクチャチーム（データセンターアーキテクト、プラットフォームエンジニアリング、I&O（Infrastructure and Operations：インフラストラクチャと運用）、CloudOpsとDevOpsチーム、および開発者が含まれる一体となったチーム）は、現在および将来の自社ビジネスのニーズを満たすために、AIに対する投資を推奨しなければならない。またこれらのチームは、企業ワークロードの要件を満たすためのインフラストラクチャスタックを設計し管理する際に、AIワークロードに固有な要件を考慮しなければならない。優れたインフラストラクチャ戦略は、オープンで柔軟性の高いハイブリッドIT戦略となるであろう。したがって、DevOpsチームは、特別な検証環境を構築する必要なく、迅速にトライアルとイタレーション（反復修正）を行えるようになるであろう。インフラストラクチャチームの決定事項には以下が含まれるであろう。

- ▶ 特定のAIモデル開発タスク向けにオンプレミスインフラストラクチャを強化するために、パブリッククラウドインフラストラクチャとPaaS（Platform as a Service）に投資する。
- ▶ AIモデルを本番環境に導入する際に、オンプレミスまたはコロケーションデータセンターのアーキテクチャを変更する。これはアクセラレーテッドコンピューティングまたはヘテロジニアスコンピューティング、高速メモリー、広帯域なネットワークング、フラッシュストレージを追加するための変更である。
- ▶ 複数の遠隔ロケーションに渡ってAI推論タスクを処理するために、電力とコンピューティングの観点で最適化されたプラットフォームを導入して、エッジのフットプリントを拡大する。
- ▶ 該当するタスクに対して、適切なリソースを使用することを保証するデータコンプライアンスやインフラストラクチャの調達および運用コストモデルを開発する。
- ▶ AIアプリケーション開発と最適化されたPIC（Performance-Intensive Computing：パフォーマンスインテンシブコンピューティング）インフラストラクチャ向けに、既存のI&Oと開発者のスキルおよびスタッフを強化する。
- ▶ AIアプリケーションの開発や、AI中心のビジネスのワークフロー、AIデータパイプラインを含むように、既存のDevOpsの方法論とプロセスを拡張する。

AIの実装は多面的でビジネス全体の取り組みである。インフラストラクチャチームにとって、これは、コストの最適化やデータコンプライアンス、およびセキュリティに関連する課題に対処しながら、潜在的な収益増を実現するために、AIワークロードを迅速に導入する戦略の採用を意味する。具体的で測定可能なROIを確保することは、あらゆるAIへの取り組みにおいて極めて重要な要素である。

## AIに対する投資- より賢明な方法

企業は、ビジネスの差異化と並行して収益源を拡大するために、AIに対する出資を増やしている。今までにない新しいアプローチで製品やサービスのイノベーションを加速することや、人間の思考や行動を強化すること、あるいは業務を合理化する、コストを削減するなど、企業がどれを重視するとしても、AIはビジネス推進の原動力になる。ビジネスにおけるAIの主なユースケースの中で、顧客エンゲージメント（営業、マーケティング、コミュニケーション、カスタマーサービス）、ITの最適化、コード生成、脅威インテリジェンスの強化といった分野にモメンタム（勢い）があるとIDCはみている。さらに、製造、小売、医療、金融サービスを含む、すべての主要な産業に、AIは価値を提供し続けている。IDCは、2022年から2026年までの間に、AIシステムに対する企業の支出は年率27%で増加するとみている。たとえば、2024年には、生成AIソリューションに対する企業の投資は200億ドルを超えている。

このうち、インフラストラクチャに対する支出は投資全体の半分以上を占めている。これには、オンプレミスインフラストラクチャおよびクラウドベースのIaaS（Infrastructure as a Service）ソリューションに対する支出が含まれている。しかし、この支出は経営幹部によって厳しく精査されている。ビジネスとITのエグゼクティブは、AIに対応するインフラストラクチャに対する新たな支出を正当化するために、テクノロジーとビジネス成果の整合性を確保すると共に、それらのROI算出や価値分析も求めている。つまり、ITチームは今こそ基本に立ち返るべき時なのである。IDCは、2027年までに、複数のAIプロジェクトの失敗を経験した後、国内のITチームの70%が基本に戻り、データロジスティクス、品質、ガバナンス、信頼性のためのAI対応データインフラストラクチャプラットフォームに重点を置くようになると予測している。

AIに適したインフラストラクチャのプロビジョニングは、単純ではない。IDCの調査では、多くのAIへの取り組みがインフラストラクチャに対する投資不足のために失敗するか、期待以下の結果を出していることが示されている。一方、インフラストラクチャに対する過剰投資は、コストがかかり、想定以上の環境への影響をもたらす可能性がある。インフラストラクチャに関する一連の正しい判断は、さまざまなチームの間での協力的なアプローチから始まる。AIインフラストラクチャの導入を検討する際には、以下を含める必要がある。

- » **ROIの管理**：インフラストラクチャチームは、AIワークロード向けにオンプレミスデータセンターまたはコロケーション施設に新しいハードウェアを購入するのか、あるいはAIワークロード向けにパブリッククラウドサービスに投資するのか、最適な投資戦略を決定する必要がある。それぞれのアプローチには、特定のコスト管理戦略が必要である。
- » **人材確保とスキルへの注力**：企業は、人材確保とスキルに関連する課題への対策を策定する必要がある。IDCは、2027年まで、適切なスキルを持つ人材の確保が困難な状況が続き、60%の企業ではROIの可能性が低下すると予測している。
- » **市場への迅速な参入**：AIの運用は複雑であり、あらゆる場合にうまくいくような万能のプランは存在しない。新しいタイプのハードウェアや専門知識の恩恵を受ける可能性があるAIプロジェクトもあるが、多くは、既存のデータセンターやクラウドプラットフォームに導入できる。ただし、最新世代のインフラストラクチャへのアップグレードが必要になる可能性がある。

- » **データガバナンスの維持**：多くのAIプロジェクトは、機密データ、または規制対象であるデータにグローバルにアクセスするためのアプリケーションを必要とする。インフラストラクチャチームは、データセキュリティを保ち、規制に準拠するために、特別な注意を払う必要がある。インフラストラクチャチームは、AIプロジェクトがサステナビリティの目標に与える影響を考慮する必要があり、学習データにおけるバイアスに関するどのような問題にも対処できる必要がある。

## ROIの最大化

データセンターとクラウドインフラストラクチャに対する投資のROIを向上させるには、現在割り当てられている予算内で、新しいAIアプリケーションとワークフローを統合する方法を考え出す必要がある。

### 汎用インフラストラクチャのアップグレード

新たなAIの取り組みにもかかわらず、予算は依然として一定のプレッシャーを受けている。より慎重にアップグレードを行うことで、チームはインフラストラクチャの価値を最大化することができる。旧式のサーバーをリプレイスし、ワークロードを最新世代のプロセッサが動作するインフラストラクチャに統合することで、チームは以下を実現できる。

- » **AIワークロードの大幅なパフォーマンス向上**：既存のCPUインフラストラクチャをリフレッシュすることによって、多くのAIワークロードが特殊なハードウェアなしで実行可能となる。たとえば、最新のCPUのいくつかは、20億パラメーター以下の大規模言語モデル（LLM：Large Language Model）に対して、100ミリ秒未満のトークンレイテンシーでリアルタイムなユーザーエクスペリエンスを提供して、SLAを満たすことができる。最新のサーバー用プロセッサは、データベース、暗号化、ストレージ、および他の一般的な企業ワークロードに対して、より高い性能も提供している。
- » **インフラストラクチャ全体のフットプリントの削減**：最新のプロセッサを搭載したサーバーでは、エネルギー効率が大幅に向上している。これによって、エネルギーコストを維持または削減しながら、電力およびスペースに制約があるデータセンターへの容量追加が容易になる。たとえば、IDCの調査では次の点が明らかになっている。
  - 最新のプロセッサを用いて5年前のサーバーを更新すると、同じパフォーマンスを得るために必要なサーバーの数が減り、電力コストが下がり、総所有コスト（TCO）を最大77%削減できる。
  - 新しいサーバーを購入する際に、AI機能強化が組み込まれた最新のプロセッサを選択すると、2世代前のプロセッサと比較し、特定のワークロードに対するパフォーマンスを5対1の比率に向上できる。
- » **ソフトウェアライセンスコストの削減**：多くのソフトウェアツールは、コア単位でライセンスが供与されている。新しいプロセッサにはより強力なコアが搭載されているため、インフラストラクチャチームはワークロードを統合し、必要なソフトウェアライセンス数を削減できる。
- » **開発者の支援強化**：最も人気のあるオープンソースAIモデルとフレームワークは、x86ベースのインフラストラクチャ上で動作するように最適化されている（これは、ほとんどの企業でデファクト

なアプローチである)。このインフラストラクチャで利用可能なツールに精通していると、DevOpsチームは、高度に専門化された開発者向けスキルセットの追加を必要とせず、AIアプリケーションを容易に導入できる。

### クラウドに対する投資の効率的な管理

CloudOpsチームは、ワークロード配置の最適化および自動化を行い、ワークロードのニーズと適切なクラウドサービスをよりうまくマッチングさせて、クラウドサービス向けの支出削減に役立てられる。これによって、サービス品質を向上させながら、ROIを高められる。

### 特殊なハードウェアまたはクラウドリソースに対する必要に応じた投資

企業は、現存するすべてのインフラストラクチャプラットフォームの利用方法を検討した後に、必要な場合に、必要な場所へ、特別なハードウェアの追加を検討すべきである。AIワークロードの中には、GPUやその他のディスクリートアクセラレーターからベネフィットを得るワークロードもある。企業は、必須なものだけを戦略的に調達することで、追加購入の種類と量について、より意図的に行動することができる。

### ROI拡大を目指した人材確保とスキルへの注力

IDCは、2027年まで、まったく新しいAIアプリケーションを活用するために必要なスキルセットが不足し続けると予測している。これには、ハイブリッドインフラストラクチャ環境の管理に必要なスキルも含まれる。高性能インフラストラクチャの構築と導入に必要なスキルセットが社内不足している企業では、構想段階からAIの取り組みの導入に成功する段階に移行するには、困難を伴う。従来のコンピューティングプラットフォーム上で実行するアプリケーションであれば、IT部門は一般的にトレーニングによってスキルギャップを克服できる。しかし、PICワークロードの場合、ITの専門家でも、特定のユースケースに合わせてデジタルインフラストラクチャをカスタマイズするには高度な知識を要する。そうしないと、スキルギャップどころか、さらに大きな障害に発展するリスクがある。このような深刻なスキルの欠陥が、戦略の大幅な変更によってしかROIを達成できない程度まで進捗を遅延、停滞させるリスクがあるとIDCは予測しており、実際に変革を実現できる企業は半分未満（40%）に留まるとみている。IT部門が今から準備できる対策をいくつか紹介する。

- ▶ テクノロジーパートナーを活用する。独自のテクノロジーソリューションの構築と導入を前提としている企業は、考え方を転換し、最も効果的なパートナーを慎重に見極めることが、業界独自のニーズを満たす上で極めて重要である。サービスパートナーとの関係から最大の価値を引き出す能力を確立することは、戦略的に不可欠である。
- ▶ 優れたテクノロジーパートナーは、最高のパフォーマンスとレジリエントなITサービスを提供するソリューションに注力していることを忘れてはならない。このようなパートナーは、ビジネス部門のニーズを特定して具体化でき、多くの場合、大規模なイノベーションを実現しつつ、IP（知的財産）とデータを保護するゲートキーパーの役割も果たす。また、事業部門のリーダーやサービスプロバイダーと緊密に連携し、イノベーションや差異化を推進するプラットフォームを構築する。
- ▶ 2028年までに必要とされるデジタル機能の概要を明らかにし、その目標を達成するために、事業部門とITチームを再編成する。組織が必要とする業界特有のイノベーションとプラットフォームの種

類を特定する。データをインサイトに変え、プロセスを変革し、イノベーションを加速させるために、業界特有のスキルセットと専門知識に投資する。

### 市場への迅速な参入で収益成長を加速

AIの迅速な本番環境への導入は、モデルのタイプ、データセットのサイズ、およびハードウェアの選択に左右され得る。

いくつかのシナリオを除き、ほとんどの企業は、ゼロからモデルを開発する必要はない。通常、企業は、事前学習済みモデルのファインチューニングまたは最適化によって、AIへのジャーニーを歩み始めることができる。多くの場合、オープンソースのこともある事前学習済みのAIモデルは、特定のビジネスニーズに合わせてカスタマイズできるため、開発者がモデルをゼロから学習する手間を省くのに役立つ。その結果、開発者は、AIアプリケーションをより速く作成でき、差異化した機能を用いてイノベーションを起こすために使える時間が増える。価値あるものにするために、これらのモデルを特別に大きくする必要はない。業界に特有なデータセット、またはユースケースに特有のデータセットについて学習した、より小さな「エキスパート」モデルは、コンピュータ要件を抑えながらも、企業のニーズを十分に満たし得る。

### 既存のデータセンターとクラウドプラットフォームに対する投資の活用

多くの企業は、AIを実行するにはGPUが必要であると考えている。しかしながら、GPUは必ずしも必要ではない。最新のCPUを搭載したサーバーは、推論に優れた性能を発揮し、より小さなモデル（最大約100億～200億個のパラメーターまで）の学習も可能である。

インフラストラクチャチームは、多くの場合x86アーキテクチャに基づいている既存のプラットフォームが、AIへの取り組みの一部に、またはすべてに適しているかどうかを評価しなければならない。さらに、以下の事項についても考慮すべきである。

- ▶ 学習時間が重要な要素でない場合は、企業のワークロード向けに既存のCPUベースプラットフォームは、学習、ファインチューニング、および推論の各タスクにも利用できる。
- ▶ すでにパブリッククラウドに投資しているインフラストラクチャチームは、推論タスクのために汎用コンピューティングインスタンスも評価しなければならない。これは、複数の地域に渡ってAIアプリケーションを拡張する際に大きなメリットを提供できる。
- ▶ すべてのワークロードを既存のプラットフォーム上に保持して、データの準備、取り込み、分析、推論などのワークフローを合理化できる。
- ▶ 既知のプロセッサアーキテクチャ（たとえばx86）に基づくインフラストラクチャは、DevOpsチームが複数のロケーションに渡って共通のツールセットを活用する上で役立つ。また、ハードウェアとクラウドエコシステム内のソリューションの可用性を大幅に強化できる。

既存のプラットフォームを活用すると、企業は、高価で特殊で汎用性のないインフラストラクチャを最初に調達することなく、POC（Proof of Concept）から本番稼働へと迅速に移行できる。したがって、チー

ムは、急速に進化するビジネス環境の中で、変化するビジネス要件に追いつくように、そのインフラストラクチャを拡張できる。

### コンプライアンスの維持

データセキュリティ、コンプライアンス、プライバシー規制の遵守に関しては、AIを使用する場合、他のワークロードと比較してはるかに高いリスクがある。AIを責任ある形で導入するには、企業のチームが「インフラストラクチャアップ（インフラストラクチャを最新に保つ）」のアプローチを取る必要がある。

### AIのグローバルかつセキュアな導入のためのコンフィデンシャルコンピューティングの使用

パブリッククラウドサービスは、多くのプロバイダーによってグローバルに展開されているため、多くのAI導入にとって魅力的なサービスである。残念ながら、一方では、これによって、企業がデータセキュリティの問題にさらされることにもなり得る。

ほとんどのインフラストラクチャチームは、インフラストラクチャとアプリケーションスタックを設計し実装する際に、保存されているデータの保護（オンディスク暗号化）とデータ転送時の保護（ネットワーク暗号化）のベネフィットを明確に認識している。しかし、多くのインフラストラクチャチームは、システムメモリーに対する低レベルの攻撃によってデータが乗っ取られる可能性があることをまだ認識していない。ここで、コンフィデンシャルコンピューティングが登場するのである。

コンフィデンシャルコンピューティング（一部のハードウェア、ソフトウェア、クラウドベンダーから入手可能）によって、インフラストラクチャチームは仮想マシン、コンテナ、およびアプリケーション全体をセキュアなコンピューティングエンクレーブに移動できる。これらは、パブリッククラウド、オンプレミス、またはコロケーション施設内に配備できる。コンフィデンシャルコンピューティングは、ハードウェアによるデータ使用中の暗号化によって、転送時および保存しているデータの暗号化を補完し、ゼロトラストセキュリティ戦略をサポートするもう一つの保護レイヤーを追加する。コンフィデンシャルコンピューティングによって、チームは以下を実現できる。

- ▶ オンプレミスインフラストラクチャ、コロケーション施設、マルチクラウド、およびエッジノードに渡って、よりセキュアに、大規模にAIアプリケーションを展開できる。
- ▶ 高度な攻撃や改竄、盗難から機密データと機密コンテンツを保護するために、アプリケーションに適した信頼境界線を設定できる。
- ▶ 連合学習（プライベートデータを公開することなく、分散したソースからAIモデルを学習させる）を採用できる。たとえば、医学の研究者は患者データを提供して、治療計画改善モデルの学習に役立てられる。

### 意思決定における透明性の向上

データのプライバシーと漏洩の問題が増加するにつれて、商用モデルまたは事前に構築されたモデルが精査されている。インフラストラクチャチームは、AIモデルの学習に用いられるデータセットおよび方法論を検討する方法を考え出すよう求められている。オープンソースのAIモデルを使用すれば、開発者

はモデルの学習に使用されるデータを「見る」ことができる。これによって、透明性が向上し、モデルのバイアスを特定でき、その原因の理解が容易になる。

## インテルの検討

インテルのプロセッサ、アクセラレーター、ソフトウェア、ネットワーキング製品のポートフォリオは、堅牢なデータセンターとクラウドのインフラストラクチャ戦略を実装するために広く用いられている。インテルのデータセンターとAIのテクノロジーは、企業がデータをタイムリーで実用的なインテリジェンスに効率的かつ安全に変換できるようにするソリューションの基盤を形成している。これらのソリューションは、広範なワークロード環境に対応し、予算との整合性を高めている。

予測型AI、生成AI、HPC（High-Performance Computing）のワークロードが複雑さを増すにつれて、そのパフォーマンスとエネルギー効率の要件も同様に増加している。インテルは、これらのワークロードのパフォーマンスとTCOの最適なバランスを達成するために、ホストCPUとディスクリットAIアクセラレーターを使用してAIアクセラレーテッドシステムを設計するアプローチを採用している。たとえば、E-Cores搭載のインテルXeon 6プロセッサは、計算密度の高いワークロードとクラウドスケールのワークロードに対応する高いコア密度によってデータセンターの電力効率を向上させる。また、P-Cores搭載のインテルXeon 6プロセッサは要求の厳しいクラウドソリューションに対し、コアごとに高いパフォーマンスを発揮し、優れたAIパフォーマンスを提供する。インテルによると、第5世代のプロセッサと比較してパフォーマンスが2倍以上向上し、ワークロードの容量倍増またはタスクの高速処理が可能になる。この技術の進歩を生かすことで、同じコア数で2倍のワークロードを実行したり、ワークロードの処理時間が半減することで、1ドル当たりのパフォーマンスを高められる。さらに、これらのプロセッサに対応するISVアプリケーションやOSツール、ライブラリー、フレームワークといったインテルの広範なエコシステムは、開発者の利便性を大幅に向上させる。

## パフォーマンス

インテルは、データセンター向けに業界トップクラスのコンピューティングおよびネットワークのハードウェアに関するポートフォリオを提供しており、あらゆるロケーションや導入形態において、広範なAIおよびエンタープライズワークロードを実行するために必要なパフォーマンスと効率性を、低いTCOで実現する。これには、汎用コンピューティングと推論向けのAIアクセラレーションを内蔵したインテルXeonプロセッサ、専用のAI学習向けのインテルGaudi AIアクセラレーター、高速接続を実現するインテルイーサネット製品が含まれる。インテルのポートフォリオによって、インフラストラクチャチームは、シームレスなCore to Edge to Cloud連携（コア、エッジ、クラウド間の連携）を用いて、コスト効率の高いハイブリッドAI戦略を実行できる。さらに、以下の事項についても考慮すべきである。

- ▶ 強力な製品ロードマップに裏打ちされ、オープンスタンダードに基づいているインテルのポートフォリオは、スムーズなアップグレードを可能にし、企業が既存のデータセンターやクラウドプラットフォームに対する投資を活用できるよう支援する。
- ▶ インテルは、パフォーマンスをさらに向上させるために、特定のワークロードタイプ（たとえば、AI学習や推論、SQLやNoSQLデータベースなど）向けに最適化したプロセッサの選択肢を提供している。

- » エネルギー効率の高いプロセッサと最適化された電源設定によって、パフォーマンスのバランスを取りながら消費電力を抑えられる。
- » インテルは、ハードウェアとソフトウェアのスタック全体に渡って、パフォーマンスの最適化に取り組んでおり、これには、最新のハードウェアパフォーマンス機能を活用するための大手のエンタープライズソフトウェアベンダーとの協力も含まれる。

## 選択

インテルのグローバルに分散したエコシステムによって、インフラストラクチャチームは、自社の好みのハードウェア/ソフトウェアベンダーやクラウドサービスプロバイダーから、業界をリードするコンポーネントによってフルスタックの環境を構築できる。この詳細を以下に述べる。

- » インテルのハードウェアとソフトウェアによって、実運用に対応し、相互運用可能な幅広いAIソリューションや、クラウドサービス、フレームワーク、AIモデルが実現する。
- » マルチベンダー、マルチアーキテクチャ環境向けのオープンで統一されたプログラミングモデルによって、DevOpsチームは開発を合理化し、ニーズに合わせてカスタマイズできる。インテルは、そのオープンソフトウェア環境の一部として、異なるハードウェアタイプ間で推論モデルを最適化するために、インテル oneAPIとOpenVINO ツールキットを提供している。
- » ソフトウェア製品とツール、特に「インテルデベロッパー・クラウド」から提供されているソフトウェア製品とツールは、最適なハードウェアまたはクラウドサービス上でのワークロードのテスト、最適化、配置を容易にする。

## 信頼

インテルのハードウェアベースのセキュリティ機能は、転送中、保管中、および使用中のデータの保護に役立つ。インテルのコンフィデンシャルコンピューティングエコシステムおよびトラストサービスによって、チームは、機密データを扱う場合でもデータとAIモデルを共有できる。チームは、セキュリティと責任を持って、AIモデルやワークロードをグローバルに展開できる。

インテルへの信頼は、データセキュリティに限定されているわけではない。数十年に渡るエンジニアリングの専門知識によって、インテルのハードウェアは、高いレベルのシステム信頼性、可用性、保守性を大規模に提供している。製品のサステナビリティに焦点を合わせた機能を実現するためのインテルの投資によって、企業は、その企業責任の目標に向かって前進しながら、ワークロードの需要に対応できる。

## インテルの課題と機会

ディスクリットアクセラレーターは、AIインフラストラクチャの選択肢の一つとしてかなりの勢いを得ているが、スケーラブルなAIに必要なインフラストラクチャスタックに関しては、さらに考慮すべきことがある。GPUなどのアクセラレーターは学習と推論に優れたパフォーマンスを提供するが、多くのAIワークロードにとっては過剰な場合もある。設計中および実装中の計画が不足していると、過剰投資につながる可能性がある。インテルと他のベンダーは、データセンターのアーキテクトとDevOpsのチーム

が適切な目的に合ったAIスタックに対する投資に伴う潜在的な課題と利点を把握するのを支援する必要がある。

インテルは、AI向けのハードウェアに関する十分なポートフォリオと強力な製品ロードマップを提供している。このロードマップによって、インフラストラクチャチームは同じプラットフォーム上でハードウェアの予測可能なアップグレードを行えるようにし、投資の将来性を確保できる。この詳細を以下に述べる。

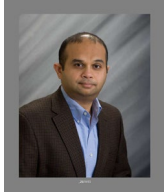
- » インテルXeonプロセッサは、推論のワークロードに対して良好なパフォーマンスを発揮する。これらは、一部のAIの学習やファインチューニングワークロード（たとえば、小さなAIモデルを扱う場合）にも適している。
- » インテルGaudiアクセラレーターでは、さらに少ないTCOで、一般的なベンチマーク（MLPerfなど）に対して、通常はより良い結果が得られると実証されている。
- » インテルプロセッサおよびアクセラレーターベースのクラウドサービスは、広く利用可能であり、AIワークロード向けに魅力的な価格対パフォーマンス比を提供している。
- » インテルは、インテルXeonプロセッサ上にコンフィデンシャルコンピューティングテクノロジーおよびサービスの包括的なポートフォリオを提供している。このポートフォリオは、機密データや規制対象データに依存するAIワークロードに対する顧客の多様なニーズに対応している。
- » インテルは、AIソリューションやコンフィデンシャルコンピューティング製品の選択肢を増やすことができるオープンなエコシステムを有している。

## 結論

インテルとそのパートナーエコシステムによって、データセンターのアーキテクトとDevOpsチームは、レジリエントで（回復力があり）、スケールラブル、そして信頼性がある長期的なAIインフラストラクチャを設計、実装できる。その結果、企業はTCOを低減し、データ保護を確保し、ほぼあらゆる場所と導入タイプに渡って、ワークロードを展開することができる。

データの力を活用できる企業は、インテリジェンス時代のリーダーになるであろう。

## アナリストについて



**Ashish Nadkarni**、グループバイスプレジデント兼ゼネラルマネージャー、*Worldwide Infrastructure Research*

Ashish Nadkarniは、IDCのWorldwide Infrastructure Research部門のグループバイスプレジデント兼ゼネラルマネージャーである。同氏は、AI、HPC、およびその他のエンジニアリングワークロードのために導入されるパフォーマンス集約型コンピュータインフラストラクチャを専門としている。

### スポンサーメッセージ

#### あらゆる場所でのAI活用のためのインテルテクノロジー

インテルは、クライアントとエッジからデータセンターとクラウドまで、あらゆるプラットフォームでAIコンティニュームを実現します。インテルのポートフォリオは、プロセッサとアクセラレーター、ネットワークング、およびソフトウェアなど多岐に渡ります。世界最高水準のハードウェア、ソフトウェア、クラウドベンダーのエコシステムと共に、インテルは企業のAIの導入を加速し、より迅速に価値を実現するためのソリューションを提供しています。

- » インフラストラクチャに対する投資の価値を最大限に引き出すための戦略の詳細については、[IDCとインテルのWebinar](#)に登録してください。このWebinarは、データセンターとクラウドのモダナイゼーションと最適化による費用削減に関するものです。
- » AIワークロードをサポートして収益を拡大する方法について詳しく知るには、[AIを使用したイノベーションに関するIDCとインテルのオンラインセミナー](#)に登録してください。
- » 新しいAIワークロードを導入しながらデータを保護する方法を探すには、[データ保護とリスク軽減に関するIDCとインテルのレポート](#)をお読みください。

 IDC Custom Solutions

本書の内容は、[www.idc.com](http://www.idc.com)で公開されているIDCの既存の調査結果に基づいています。

**IDC Research, Inc.**  
140 Kendrick Street  
Building B  
Needham, MA 02494, USA  
T 508.872.8200  
F 508.935.4015  
[blogs.idc.com](http://blogs.idc.com)  
[www.idc.com](http://www.idc.com)

本調査レポートはIDC Custom Solutionsが発行したものです。本調査レポートに記載する見解、分析、調査結果は、ベンダー スポンサーの記載がない限り、IDCが独自に行い、発行した詳細な調査と分析から導き出されたものです。IDC Custom Solutionsは、さまざまな企業による配布に対応するため幅広いフォーマットでIDCのコンテンツを提供しています。このIDCの資料は、外部使用を目的としてライセンスされたものであり、IDC調査結果の使用または公表は、スポンサーまたはライセンシーの製品または戦略に対するIDCの支持を示すものではありません。

International Data Corporation (IDC) は、ITおよび通信分野、消費者向けテクノロジー市場に関する調査・分析、アドバイザリーサービス、イベントを提供するグローバル企業です。世界中に1,300人以上のアナリストを擁するIDCは、110か国以上を対象として、世界規模、地域別、国別での市場動向の調査・分析および市場予測を行っています。IDCの分析と洞察は、IT専門家、企業経営者、および機関投資家が客観的にテクノロジー導入の意思決定を行い、主要な事業目標を達成するのに役立ちます。

©2025 IDC. Reproduction is forbidden unless authorized. All rights reserved. [CCPA](#).