

# Nutanix & NVIDIA の AI ソリューション

## AI を取り巻く環境の変化

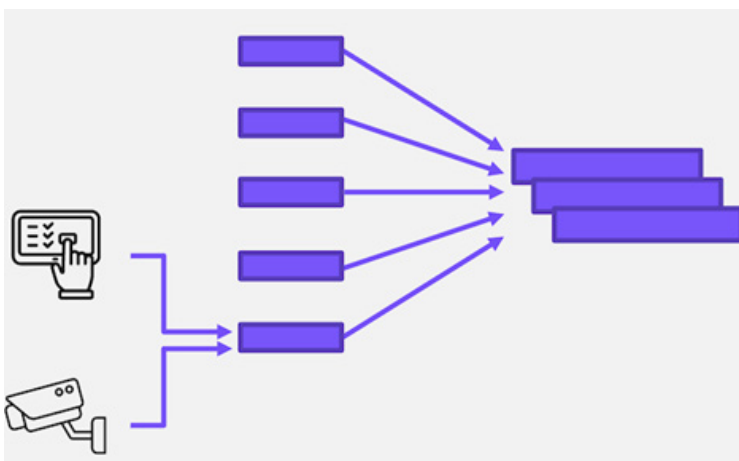
現在 AI (人工知能) は我々の日々の暮らしの中のほとんどあらゆる場所で利用されています。例えば、スマートフォンは画像や音声を理解し、言語を翻訳、楽曲名を特定するなど、様々な機能を提供しています。データセンターの中での AI はどうでしょうか？どのように皆さまを支援するでしょうか？更に最も重要なこととして、どうすればそれを活用できるようになるでしょうか？

企業におけるデータの爆発 — データベース、会話、ログ、計測値などのすべてが組織内で生成されています — AI がこれらのデータをビジネスにおけるより良い意思決定を行うための実践的な洞察として展開するという要望を加速させています。あらゆる業種において、AI が生成する実践的な洞察を活用することでビジネス価値を向上させることが可能です。2021 年まではほとんどのエンタープライズ向け AI アプリケーションは分類 (例：テキストまたはスピーチの認識、オブジェクトの検出等) と回帰タスク (例：売上の予測、株価の予想等) を行うことによって決定の境界を計算することにフォーカスがあり、識別を行うことがそのほとんどでした。2022 年以降、chatGPT と拡散 (diffusion) モデルの人気によって引き起こされた生成 AI ベースのアプリケーションの採用の拡大というエンタープライズの AI を取り巻く環境の抜本的な変化が起っています。識別を行う AI とは異なり、生成 AI はサンプルデータから分布データを生成します (新たなデータを生成するということです)。

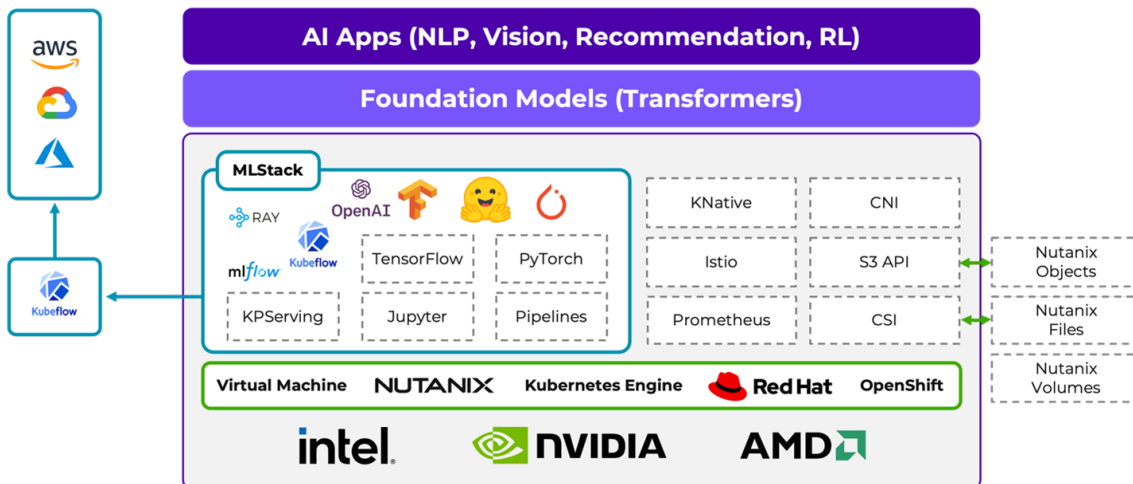
**Nutanix Cloud Platform** は非常に大規模な計算リソースを必要とする AI/ML (人工知能 / 機械学習) ワークロード、特に生成 AI アプリケーションを動作させるのに理想的なスタックを提供します。

## AI-on-Nutanix

エッジ、データセンターそれにパブリッククラウドを加えたインフラストラクチャーの必要性は日々増えています。例えば、リモートのエッジノード内でデータを生成、処理し、その後データセンターへ集約して意思決定やモデルのファインチューニングを行うなどです。



どのようなアプリケーションが展開されていたとしても、階層をまたがって一貫した運用モデルが必要とされます。Nutanix Cloud Platform はあらゆる場所 — シングルノードのエッジデバイス、データセンター内のマルチノードのクラスター、NC2 テクノロジーによってパブリッククラウド内でも — 動作します。これによってシンプルかつ一貫した AI/ML プラットフォームの運用が可能となる統合クラウド運用モデルが提供されます。



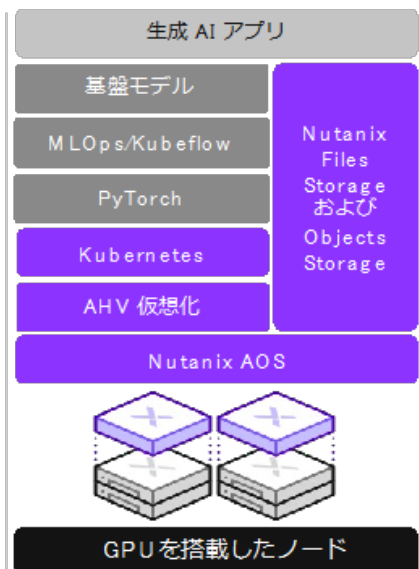
Nutanix Cloud Platformは様々なハードウェアと NVIDIA GPUカードをサポートしており、Nutanix Kubernetes Engine, Red Hat OpenShift、または仮想マシン上でワークロードを動作させることが可能です。

AI/MLワークロードを動作させる際に重要になるのは、機械学習のライフサイクルの管理です。MLOpsは機械学習と運用を組み合わせ、DevOpsと同じように自動化、オーケストレーション、展開モデルを一貫して効率よく行うことを意味します。Kubeflowは広く知られたMLOpsのソリューションで、Nutanix Cloud Platformは [Kubeflowで検証済み](#) であり、様々な階層における一貫したMLOpsを実現します。

AI/MLをNutanix上での稼働について詳しくは以下のリソースを参照してください：

- [Nutanix.com AI/MLソリューションページ](#)
- [Nutanix Validated Design for the AI-enabled Edge](#)
- [Nutanix.dev ブログ記事](#)
- [デモ:チャットボットをNutanix Cloud Platform上で動作させる](#)

## GPT-in-a-Box



**GTP-in-a-Box**は GPTの機能を実装したいと望む一方で、データとアプリケーションに対するコントロールを維持し続けたいと考える組織向けのターンキー AIソリューションです。AI 対応のインフラストラクチャーを構築するために必要な全てを含んでいます：

- **GPU**を搭載したサーバーノード上に構築された **Nutanix Cloud Platform**
- **GPT**モデルのファインチューニングを行うための **Nutanix Files** および **Object**
- **PyTorch**と Kubeflowを含む **AI**ワークロードを展開し動作させるためのオープンソースソフトウェア
- 大規模言語モデル (LLM) のサポート (**Llama2, Falcon**, および **MPT**等)

## Nutanix & NVIDIA の AIソリューションの提供価値

**Nutanix**は AIを活用するエッジからコアデータセンター、クラウドに至るまでを共通の運用モデルとデータ構造を利用する俊敏でレジリエントなインフラストラクチャーを提供し、AI/MLについての取り組みをあらゆる場所で一貫した運用で加速させます。

### 運用のシンプル化

インフラストラクチャーの複雑さと制限を取り払い、AI/MLについての取り組みをジャンプスタートさせる

- 単一のコントロールプレーンでのエッジ、データセンター、クラウド環境をまたがるシームレスなアプリケーション、ワークロード、そしてデータのモビリティを実現し、一般的なワークフローをシンプル化、合理化
- 開発者とデータサイエンティストのニーズと IT部門による統制をバランス、セルフサービスと自動化によって市場投入を加速
- リモートへのアプリの展開によってエッジでの推論を簡単に、リモートのインフラストラクチャー管理、アップグレード、トラブルシューティングによって、ネットワークに繋がっていない、または接続が限定的なエッジサイトの運営を実現
- エッジであっても、様々な AIアプリのニーズに合わせてファイル、ブロックおよびオブジェクトストレージが全て同じプラットフォームで利用可能で、ストレージのサイロを撤廃

### データとモデルを保護

AI/MLワークロードの統合性をネイティブのデータ保護とセキュリティ機能によって維持

- 高いレジリエント性、セキュリティ、自己治癒インフラストラクチャーとバックアップ、リストアの集約やエッジのローカルで行う機能を保持し、常にオンライン、データ喪失時もリカバリ時間を短時間に
- マイクロセグメンテーション、ファイル分析とプラットフォーム設計に組み込まれたセキュリティでエッジ、データセンター、クラウドにおける深層防御機構によってデータの流出、システムの汚染から保護
- ハイブリッド・マルチクラウドインフラストラクチャーをまたがるグローバルの視認性とレポート機能によってエンタープライズにおけるクラウドガバナンスフレームワークの構築をシンプル化、MLOpsの改善を支援
- 拡大し続けるビデオやセンサーデータのためにストレージを迅速に、停止を伴わずに拡張
- 堅牢なデータ保護とレジリエンスによって、AIが生成したコンテンツについてのデータ主権を保証
- すでに利用しているツールを利用して大規模言語モデル (LLM)を保護し、オンプレミス、パブリッククラウド上のモデルを圧縮、新たなセキュリティツールを追加する TCOを削減

## TCOを低減、多くの選択肢を

### TCOを低減しながら、それぞれのプロジェクトのニーズに答える AI を教育、展開

- コンテナ化されたアプリの拡張を様々な Kubernetes プロバイダーの選択肢とネイティブ K8s データサービスの組み合わせで迅速に、少ないリソースで実現、運用をシンプル化
- プラットフォームに組み込まれたネイティブハイパーバイザーによって仮想マシンベースの環境における効率性の改善
- ハードウェアとクラウドプロバイダー、そして展開モデル (1 ノード, 2 ノード, 3 ノード以上) から最適な選択が可能
- リソースと投資を直接ビジネスの要求と合致させ、予測と停止を伴わないオンデマンドなコンピューとストレージリソースの拡張で過剰割当と無駄を削減

## 多くの実績のあるテクノロジーをベースに構築、展開

### AI を活用するインフラストラクチャーから不確かさを削除し、価値が生まれるまでの時間を加速

- ミッションクリティカルな環境においてビジネスの要求に答えられるパフォーマンスが得られることを保証
- 10 年以上に渡るリモート及びデータセンターにおけるハイパーコンバージドインフラストラクチャー環境の成功を集約
- NVIDIA AI Enterprise 認定とエッジへ拡張された GPU のサポートを含む、NVIDIA とのパートナーシップを通じて構築された強力なソリューション
- Enterprise Edge with Artificial Intelligence についての Nutanix Validated Design で確認、検証済みの推奨構成で確信を持って環境を素早く展開

更に詳しくは <https://www.nutanix.com/jp/solutions/ai> をご参照ください。

**NUTANIX**

info-jp@nutanix.com | [www.nutanix.com/jp](http://www.nutanix.com/jp) | @nutanixjapan

©2023 Nutanix, Inc. All rights reserved. Nutanix、Nutanix のロゴ、その他の製品およびサービス名は、Nutanix, Inc. の米国その他国における登録商標または商標です。本書に記載したその他のブランド名は、識別目的に限定して使用するもので、それぞれは一般に各社の商標です。

sb-AI-on-Nutanix-jp 09/12/2023