



スタートアップのための 機械学習完全ガイド

このプレゼンテーションの対象者:

デベロッパー



アプリとサービスに機械学習を取り入れたい

データサイエンティスト



生産性を向上したい

機械学習の
エキスパート



最先端であり続けたい



AWS による機械学習スタートアップの成長支援



利用促進

無料機械学習トレーニング
無料機械学習ワークショップ
クレジット
オフィスアワー
専任チーム
リファレンスアーキテクチャ



構築

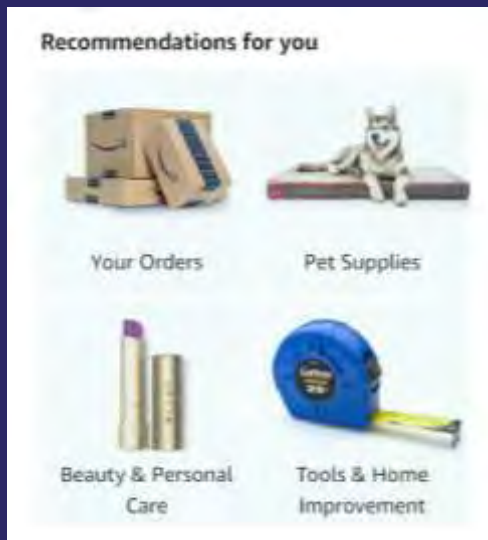
最も幅広く最も奥深い一連の製品
技術サポート
ベータ製品とロードマップへのアクセス
コスト最適化のレビュー
スピーキングエンゲージメント
共同マーケティング



コミュニティ

資金調達 (VC)
エンタープライズ顧客
パートナー (APN)
AWS Marketplace
業界での豊富な経験

Amazon における大規模な機械学習イノベーション



毎分 4,000 個の商品を
Amazon.com で販売



毎日 160 万個の
パッケージ



毎週数十億回の
Alexa とのやり取り



2016 年 12 月 7 日に
最初の Prime Air 配送

Amazon における大規模な機械学習イノベーション

- Amazon は、今日あるような機械学習 (ML) 主導の企業になるべく、何年にもわたって機械学習に取り組んできました。パーソナライズやサプライチェーンなどの分野での機械学習の応用は 20 年以上にのびります。また、その独自のパーソナライズモデルを時間をかけて大幅に改良し、Amazon プライムなどの他の製品に移行させています。
- 世界中で販売している各商品の適切な需要を予測できるシステムの開発など、フルフィルメントプロセス全体で機械学習を使用し、利便性、コスト、配送スピードに対するお客様の期待に応えています。
- Amazon は、自然言語処理テクノロジーを開発し、Alexa を介してテクノロジーと対話するまったく新しい手法をお客様に提供しています。また、Prime Air ドローンによる自律飛行と、フルフィルメントセンターにおけるロボット工学を使用した画期的な技術を開発し、お客様により迅速に荷物をお届けしています。これは大きなカルチャーの変化でしたが、どのような組織でもできることです。

AWS がスタートアップの 機械学習に選ばれた理由

AWS がスタートアップの機械学習に選ばれた理由



最も幅広く最も奥深い AI と機械学習のサービス

2020 年だけで 200 超の新機能と
新サービスを開始

すべての人のためのソリューション

3 つの主要な
機械学習フレームワークの
すべてをサポート



SageMaker で機械学習の 導入を迅速化

機械学習ワークフロー全体を
サポートする単体 IDE

TCO を 54% 以上削減

データラベリングのコストを
最大 70% 削減

マネージドスポットトレーニング
コストを最大 90% 削減



最も包括的なクラウド プラットフォームに構築

高い安全性、信頼性、充実した
機能を備えたデータストア

強力なコンピューティング、
ストレージ、セキュリティ、
データベース、
分析機能一式を搭載

数千のスタートアップが AWS の機械学習を活用

coinbase

 cookpad

Aurora

common
edits

STUDIO71

stripe

 dely

CONVOY

Root
Insurance Co

DELHIVERY

 grammarly

 Betterment

 TransferWise

affirm

 fiddler

DEEPMAP

cravelabs

TECTON

 splice

 ABACUS.AI

AWS の AI/機械学習

イノベーション、豊富な選択肢、柔軟性



100,000 以上

AWS で機械学習 (ML) を利用する
お客様

250 超

過去 12 か月における
機械学習と AI の新機能

92% クラウドの深層学習 (DL) のうち
AWS で実行されている割合

91% クラウドベースの PyTorch のうち
AWS で稼働している割合

AWS の機械学習ソリューション

トレーニング時間を 50% 短縮
90% のスケーリング効率性を提供
3 倍速いネットワークスループットを実現
料金とパフォーマンスを 25% 改善

AWS の AI/機械学習

イノベーション、豊富な選択肢、柔軟性

- AWS は、100,000 を超えるお客様が選択するベンダーであり、AI/機械学習トレーニングと推論のための最も幅広く最も奥深いコンピューティングインフラストラクチャの提供におけるリーダーです。
- 最新の CPU、GPU、およびカスタムアクセラレーターに基づく幅広い Amazon EC2 インスタンスを選択でき、お客様の予算のニーズとモデルのパフォーマンス要件に合わせ、業界をリードするネットワークとストレージを採用できます。
- AWS はお客様のニーズに最適なオプションを選択できるよう、ハイパフォーマンスで低コストの幅広いインフラストラクチャを実現するための投資とイノベーションを行っています。
- 例えば、AWS は、機械学習の推論パフォーマンスに最適化されたシリコンを最初から構築しています。Graviton CPU と Nitro システムのイノベーションから得られた学びを活かし、AWS Inferentia チップを構築しました。これは、クラウドでの推論で最低コストを実現したカスタムチップです。
- AWS で機械学習ワークロードを実行することにより、機械学習と DL モデルをトレーニングしデプロイするためのハイパフォーマンス、低コスト、使いやすいインフラストラクチャサービスにオンデマンドでアクセスできます。

一般的なスタートアップの課題



デベロッパー

データサイロ、準備とクリーニング、
全体的なデータワークフロー



スキルギャップ

人材も経験も不足



ビジネスケースの利用

機械学習を活用できる適切なビジネスユースケースを見つける必要がある



機械学習オペレーション

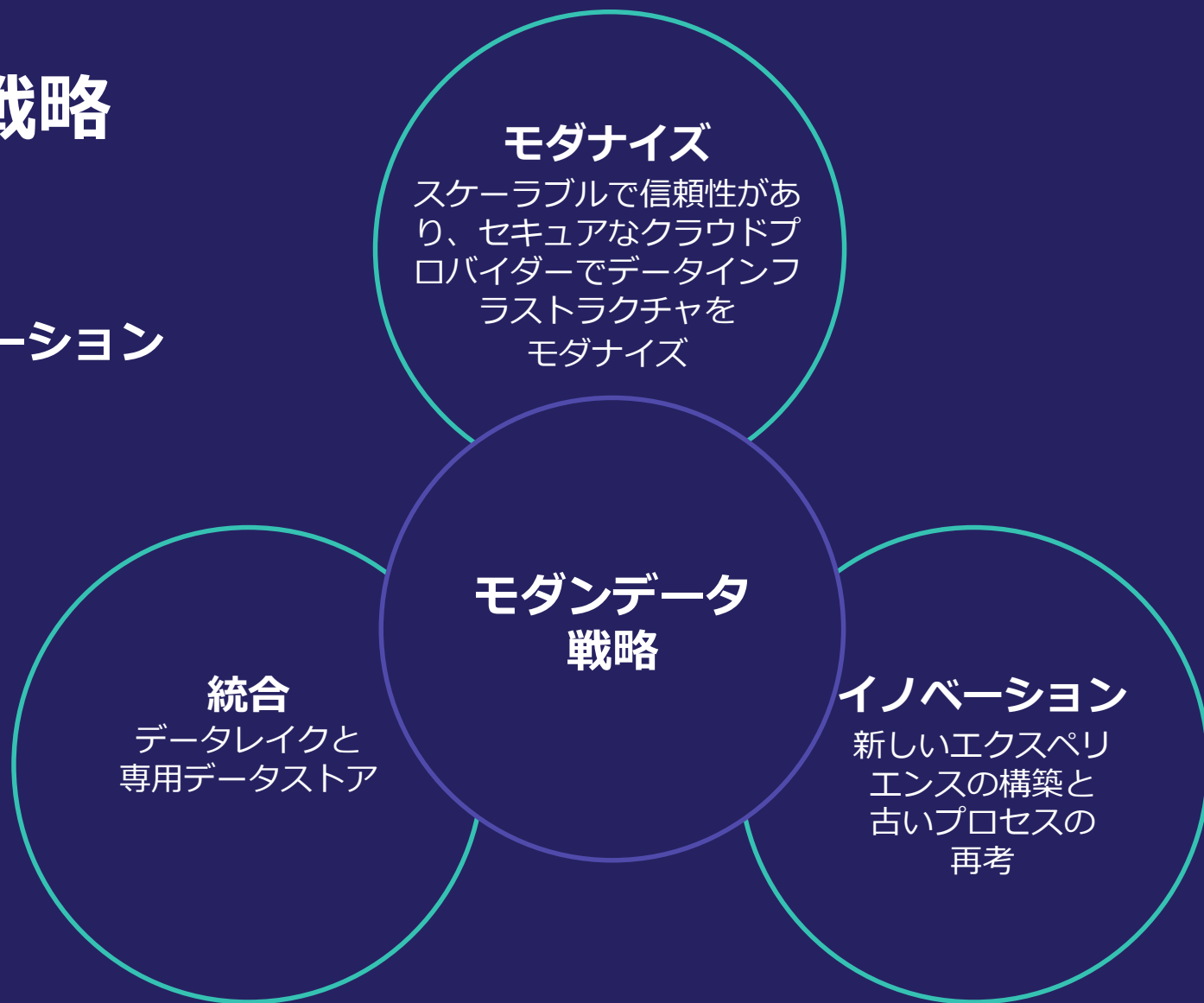
機械学習ワークフローの作成と管理は、時間がかかり複雑

AWS のモダン データ戦略



AWS のモダンデータ戦略

モダンデータ戦略の過程における
モダナイゼーション、統合、イノベーション



AWS のモダンデータ戦略

データの活用はデジタルトランスフォーメーションの次の波

1

AWS によるスタートアップ支援策:

スケーラブルで信頼性が高く、セキュアなクラウドプロバイダーでデータインフラストラクチャをモダナイズします。オンプレミスでレガシーのデータストアを運用している、またはクラウドで自己管理している組織は、データベースのプロビジョニング、パッチ適用、設定、バックアップなどの管理タスクを継続しなくてはなりません。これらのタスクを AWS にお任せいただくことで、私たちの比類のない経験、成熟度、信頼性、セキュリティ、パフォーマンスなどによるメリットをお客様の最も重要なアプリケーションでも享受できます。

AWS のモダンデータ戦略

データの活用はデジタルトランスフォーメーションの次の波

2

AWS によるスタートアップ支援策:

統合: データへのアクセスを安全かつ適切に管理することでデータを活用します。すばやく決定を下すためには、ビジネスニーズの変化に合わせてスケールし成長する新しいデータストアが必要です。また、データレイク、データウェアハウス、専用データストアを含むすべてを安全で十分に管理された一貫性のあるシステムにまとめて接続する必要があります。AWS は、データレイクと専用データストアでこれを実現します。

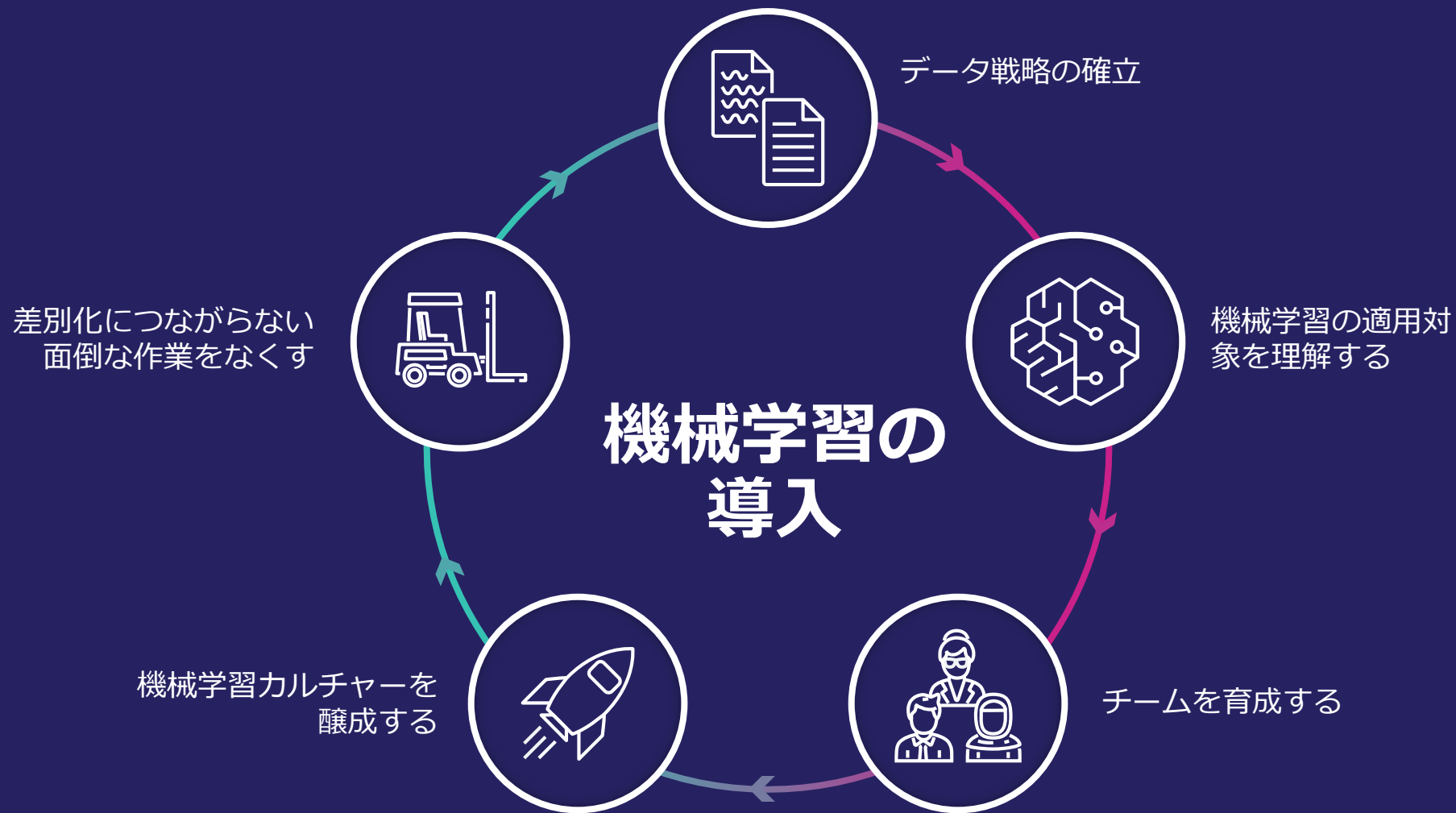
AWS のモダンデータ戦略

データの活用はデジタルトランスフォーメーションの次の波

3

AWS によるスタートアップ支援策:

AI/機械学習で新しいエクスペリエンスを創造し、古いプロセスを再考します。機械学習は、この 25 年間における最も革新的なテクノロジーの 1 つです。まったく新しい収益機会の創出、迅速で優れた意思決定、運用効率の改善を行います。AWS は、あらゆるスキルレベルのビルダーに向けた最も広く、最も包括的な一連の機械学習と AI サービスを提供しており、導入段階を問わず利用できます。



機械学習の導入を スタートする



機械学習の導入をスタートする



データ戦略の 確立

トランスフォーメーションと
イノベーションの下地を作る



ビジネス上の課題から スタート

ビジネスのニーズに基づいて
適切なユースケースを見つける



協力して成果を 上げる

複数のプログラムと
トレーニングオプションを
利用して過程を支援する

機械学習の導入をスタートする



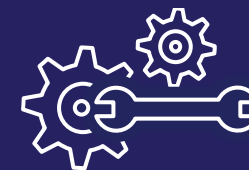
業界横断的な ユースケースへの対応

デベロッパーがどのアプリケーションにも簡単にインテリジェンスを追加できるようにします



業界特化 ソリューションの発見

ニーズに適合する、業界独自のサービスを発見します



独自モデルの 構築

Amazon SageMaker で機械学習を迅速化し、容易化します

開始方法: 一般的な機械学習ユースケース

機械学習 (ML) で現実世界の問題を解決する

カスタマーエクスペリエンスを向上する



パーソナライ
ゼーション



コンタクトセン
ターのインテリ
ジェンス



メディアイン
テリジェンス

運用コストを削減しながら
顧客満足度を高める

ビジネスを最適化する



インテリ
ジェント検索



インテリジェン
トドキュメント
処理



不正利用の
検出



ビジネスメト
リクスの分析

生産性を向上させ、
ビジネスプロセスを最適化する

イノベーションを加速する



機械学習のモダ
ライゼーション



次世代
DevOps

機械学習でイノベーションをスピード
アップしスケールアップする

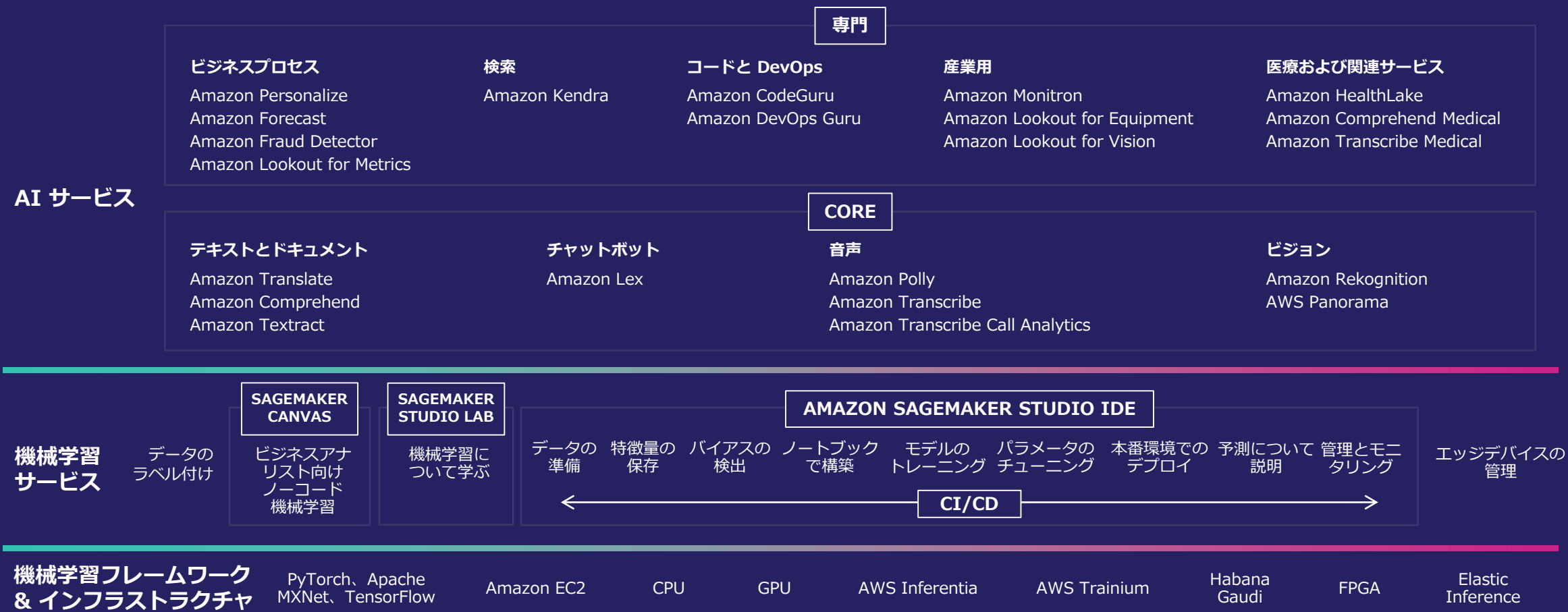
開始方法: 一般的な機械学習ユースケース

機械学習 (ML) で現実世界の問題を解決する

- 機械学習は、さまざまなビジネス機能や業界において多くの一般的な課題に適用できます。カスタマーエクスペリエンスの強化からテクニカルチームの生産性向上まで、これらの機械学習ユースケースは、複雑で高コストな人的オーバーヘッドを要するタスクの最適化を支援します。AWS は、複数の使いやすい AI サービスを実現してこれらの一般的なユースケースに取り組めるようにします。事前の機械学習の経験は不要です。クラウドで機械学習ワークフローを実行し、独自の機械学習モデルとアルゴリズムを使用する場合は、Amazon SageMaker があります。これは、データサイエンティストと機械学習デベロッパーが機械学習モデルをすばやく構築、トレーニング、デプロイするのに役立つフルマネージドサービスです。

AWS の機械学習スタック

最も幅広く多彩な機械学習機能のセット



AI サービス: アプリケーションにインテリジェンスを簡単に導入

機械学習のスキルは不要です



ビジョン



チャットボット



ビジネスツール



検索



医療および関連サービス



音声



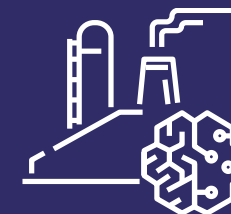
テキスト



コンタクトセンター



コードと Dev Ops



産業用

AI サービス: アプリケーションにインテリジェンスを簡単に導入

機械学習のスキルは不要です

- 私たちの事前にトレーニング済みの AI サービスでは、あらかじめ用意されたインテリジェンスをアプリケーションとワークフローに提供します。AI サービスをお使いのアプリケーションに簡単に統合して、一般的なユースケース (パーソナライズされたレコメンデーション、コンタクトセンターのモダナイゼーション、安全とセキュリティの強化、カスタマーエンゲージメントの向上など) に対応できます。
- Amazon.com および Amazon の機械学習サービスを強化するのと同じ深層学習テクノロジーを使用しているため、API を継続的に学習することで品質と精度を得ることができます。何と言っても優れているのは、AWS の AI サービスには機械学習の経験が必要ないということです。

Amazon Fraud Detector

不正の特定を迅速化



機械学習で
不正検出を強化



機械学習の知識のレベルを問わず機械学習の詐欺モデルを構築できます



Amazon 体験とエンリッチメントからの機械学習ブースト



誤検出とマニュアルレビューの減少



不正検出スタッフのセルフサービスにより脅威への対応を迅速化



TCO の削減と TTV の早期実現

Amazon Fraud Detector

不正の特定を迅速化

Amazon Fraud Detector を使用したお客様の利点

- **機械学習で不正検出の取り組みを強化します。** Amazon Fraud Detector を使用すれば、既存のルールベースの不正検出ソリューションを強化または入れ替えられます。
- **機械学習の専門知識がどのレベルのユーザーでも機械学習モデルを構築できます。** Amazon Fraud Detector を使用すると、機械学習のエキスパートではない不正検出オペレーションスタッフでも、機械学習ベースの不正検出モデルを構築して利用できます。その結果、フルタイムのデータサイエンティストが不正検出の取り組みにかかわる必要がなくなるか、組織のその他のニーズに対応できます。
- **Amazon のエンリッチメントと不正検出の経験により機械学習モデルの結果を向上します。** 大部分のお客様は、Amazon Fraud Detector で生成したモデルは、自分たちで構築したモデルより優れていると報告しています。
- **誤検出と全体的なマニュアルレビューの処理数が減少します。** Amazon Fraud Detector でモデルを生成することにより、カスタマーサービスと不正検出オペレーションチームが処理しなければならない作業量が大幅に減少することがわかります。
- **不正検出オペレーションチームが、脅威に対応するための新しい対策を実装できます。** 不正検出チームは、Amazon Fraud Detector を使用して、新しい攻撃に対応する新しいモデルを瞬時に定義してデプロイし、IT と連携する必要を回避します。
- **総保有コスト (TOC) を削減し、価値実現までの時間 (TTV) を加速します。** AWS コンソールにログインして、Amazon Fraud Detector を使用して直ちに不正検出ソリューションの開発とデプロイをスタートでき、前払いコストやコミットメントは不要で、ソフトウェアをインストール、設定、または保守する必要もありません。

Amazon Personalize

顧客満足度を高め、カスタマーエクスペリエンスを向上する



高品質のレコメンデーションを提供



顧客の意図の変化にリアルタイムで適応する



レコメンデーションモデルを数クリックでトレーニングする



ほぼすべての製品やコンテンツのレコメンデーションを生成する

Amazon Personalize は、基本的なユースケースにすぐに使えるソリューションを提供します。
これらの機能を使用して、幅広い、パーソナライズされたユーザーエクスペリエンスを生み出すことができます。

Amazon Personalize

仕組み

1. データを処理して検査します
2. 何が有意義かを特定します
3. 適切なアルゴリズムを選択します
4. データに合わせてカスタマイズした、パーソナライズされたモデルをトレーニングして最適化します


**ユーザーイベント/
インタラクション**
(ビュー、サインアップ、
コンバージョン、など)

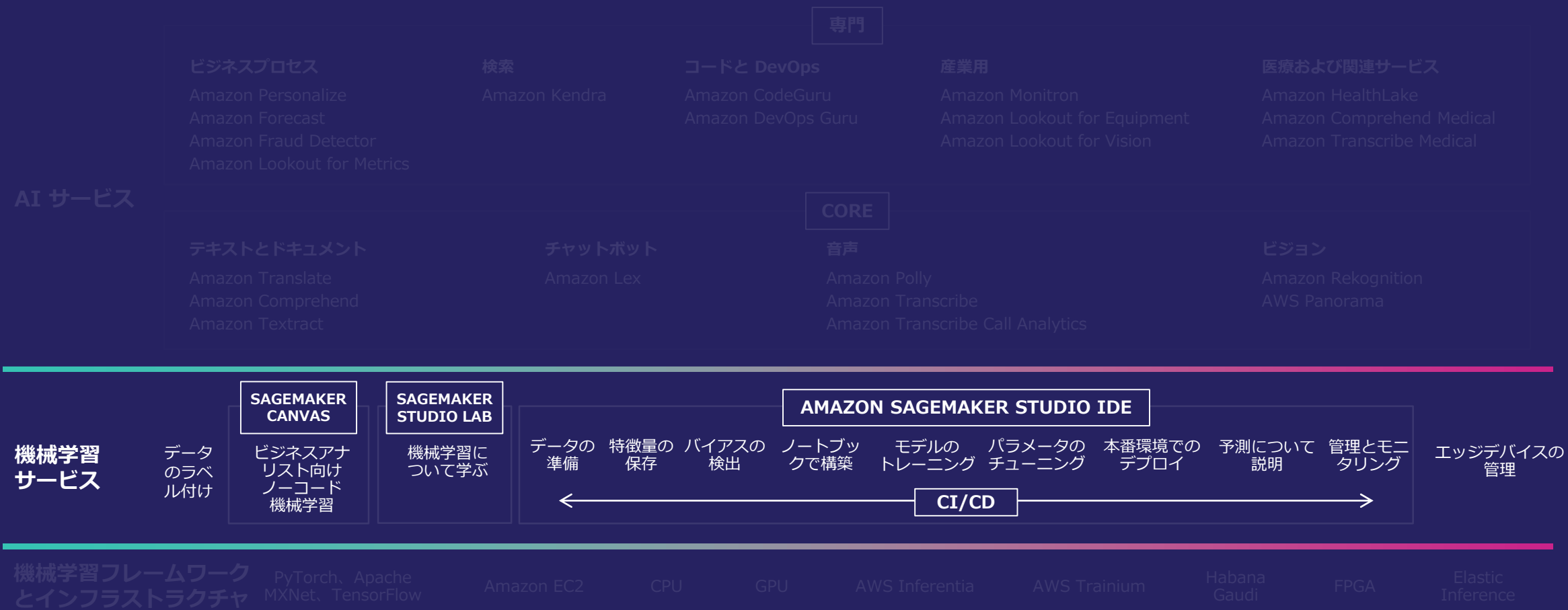

項目メタデータ
(記事、製品、動画、
などの詳細)


ユーザーメタデータ
(年齢、場所、など)



AWS の機械学習スタック

最も幅広く多彩な機械学習機能のセット



Amazon SageMaker



統合ワークベンチ

機械学習、データの準備、実験の管理、パイプラインのために特別に設計された IDE

マネージドインフラストラクチャ

超低レイテンシーと高スループット、オートスケールリング、分散トレーニングのために設計されています

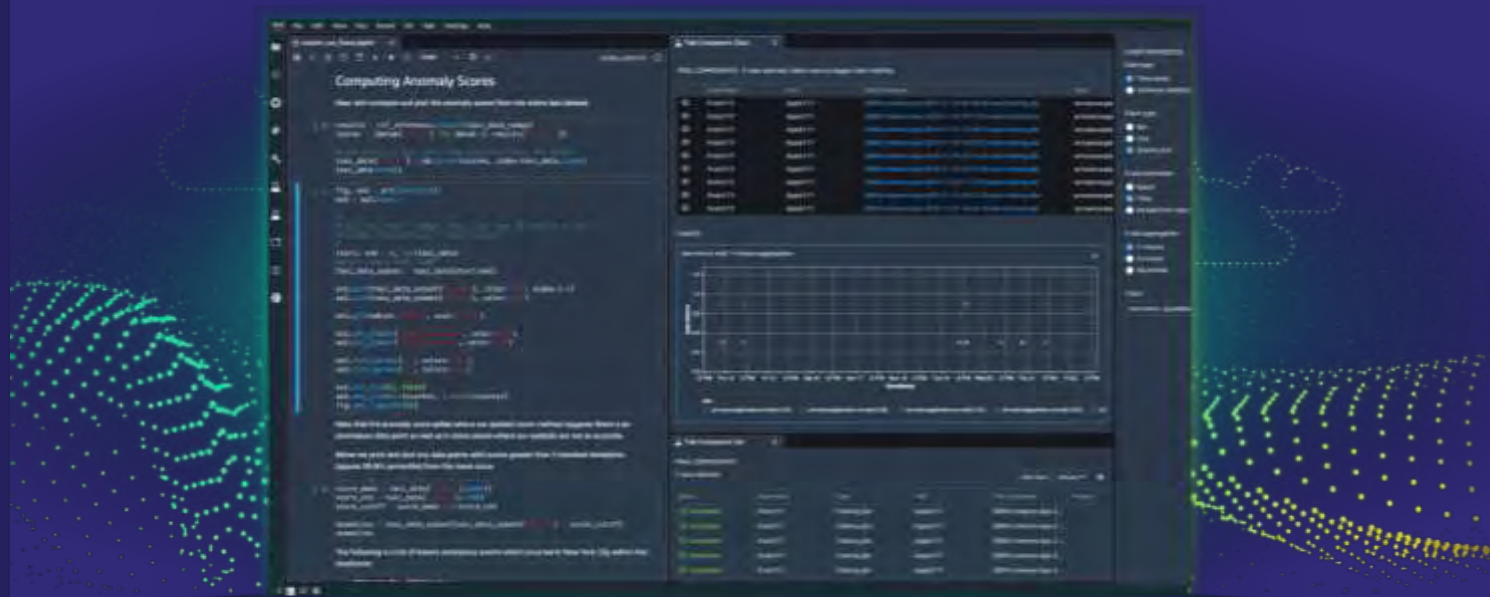
マネージドツール

SageMaker Autopilot、コラボレーション、Jupyter ノートブック、Experiments、Debugger、Model Monitor などは、連携するように最初から専用に構築されています

<https://aws.amazon.com/sagemaker>

Amazon SageMaker

最も完全なエンドツーエンドの
機械学習サービス



Amazon SageMaker

最も完全なエンドツーエンドの機械学習サービス

- Amazon SageMaker は、最も完全なエンドツーエンドの機械学習サービスです。データサイエンティストと機械学習オペレーションチームのためのマネージドサービスとして、機械学習に関連付けられている画一的で面倒な作業を取り除き、ビジネスに集中するための時間、リソース、エネルギーを持てるようにするために役立ちます。SageMaker には多数の機能と便利なワークベンチがあり、SageMaker の存在意義とともに使いみちを体現する主要な 3 本の柱があります。
- まず、SageMaker Studio は統合されたツールのワークベンチを提供します。例えば、SageMaker Studio から Jupyter ノートブックと JupyterLab 環境を瞬時に起動できます。SageMaker は、完全な実験管理、データの準備、パイプラインオートメーションとオーケストレーションを提供して、データサイエンティストの生産性向上を支援します。

Amazon SageMaker

最も完全なエンドツーエンドの機械学習サービス

- 次に、これまでに Jupyter ノートブックを使用したことがある場合は、コンピューティング環境で実行する必要があることをご存じでしょう。SageMaker は、クラウドでフルマネージドサーバーを提供して、データサイエンティストやデベロッパーが簡単に実行できるようにしています。ノートブック以外にも、SageMaker はマネージドインフラストラクチャ機能を提供しています。分散トレーニングジョブ、データ処理ジョブから、モデルホストまで、SageMaker はモデルの構築、トレーニング、ホストに関連して、すべてのスケーリング、パッチ適用、高可用性などに対応します。
- 最後に、データサイエンティストの生産性をさらに高めるために、マネージドインフラストラクチャにあるこの統合ワークベンチは、すべて機械学習専用で、連携する前提で設計されている膨大なツールのエコシステムにより充実しています。

Amazon SageMaker: 機械学習をより身近に



Amazon SageMaker:

機械学習をこれまでよりも身近に

- Amazon SageMaker は、最も完全なエンドツーエンドの機械学習サービスで、向上した俊敏性、生産性、コスト効率によりお客様を支援します。
- 当社は開発当初から、すべてのデベロッパーとデータサイエンティストが、機械学習モデルの構築、トレーニング、デプロイをすばやく低コストで実現できるように SageMaker を構築しました。これを達成するため、AWS は機械学習開発ライフサイクルのすべてのステップに必要なツールを 1 つに統合したフルマネージドサービスで提供しています。実際、私たちは昨年だけで 50 を超える機能をリリースし、そのすべてがお客様にとってこのプロセスを容易にすることを目的としています。
- 昨年、1 か所ですべてのツールにアクセスできるように、Amazon SageMaker Studio をリリースしました。

Amazon SageMaker の概要

Amazon SageMaker

準備

SageMaker Ground Truth
機械学習用にトレーニングデータをラベル付け

SageMaker Data Wrangler
機械学習のデータをすばやく集計および準備

SageMaker Processing
組み込み Python を使用するか自分の R/Spark を持ち込む (BYO)

SageMaker Feature Store
特徴量を保存、更新、検索、共有

構築

SageMaker Studio notebooks
伸縮自在なコンピューティングと共有を備えた Jupyter ノートブックを使用する

組み込みと BYO アルゴリズム
数多くの最適化されたアルゴリズムを使用または独自のアルゴリズムを使用

ローカルモード
ローカルマシンでのテストとプロトタイピング

SageMaker Autopilot
完全な可視性を備えた機械学習モデルを自動的に作成

トレーニングとチューニング

ワンクリックトレーニング
分散インフラストラクチャの管理

SageMaker Experiments
すべてのステップをキャプチャし、整理して、比較

モデルの自動チューニング
ハイパーパラメータ最適化

SageMaker Debugger
デバッグのトレーニング実行

マネージドスポットトレーニング
トレーニングコストを 90% 削減

デプロイと管理

ワンクリックデプロイ
フルマネージド、超低レイテンシー、高スループット

Kubernetes と Kubeflow の統合
Kubernetes ベースの機械学習を簡素化

マルチモデルエンドポイント
インスタンスごとに複数のモデルをホストしてコストを削減

SageMaker Model Monitor
デプロイされたモデルの正確性を維持

SageMaker Pipelines
ワークフローのオーケストレーションとオートメーションを実装

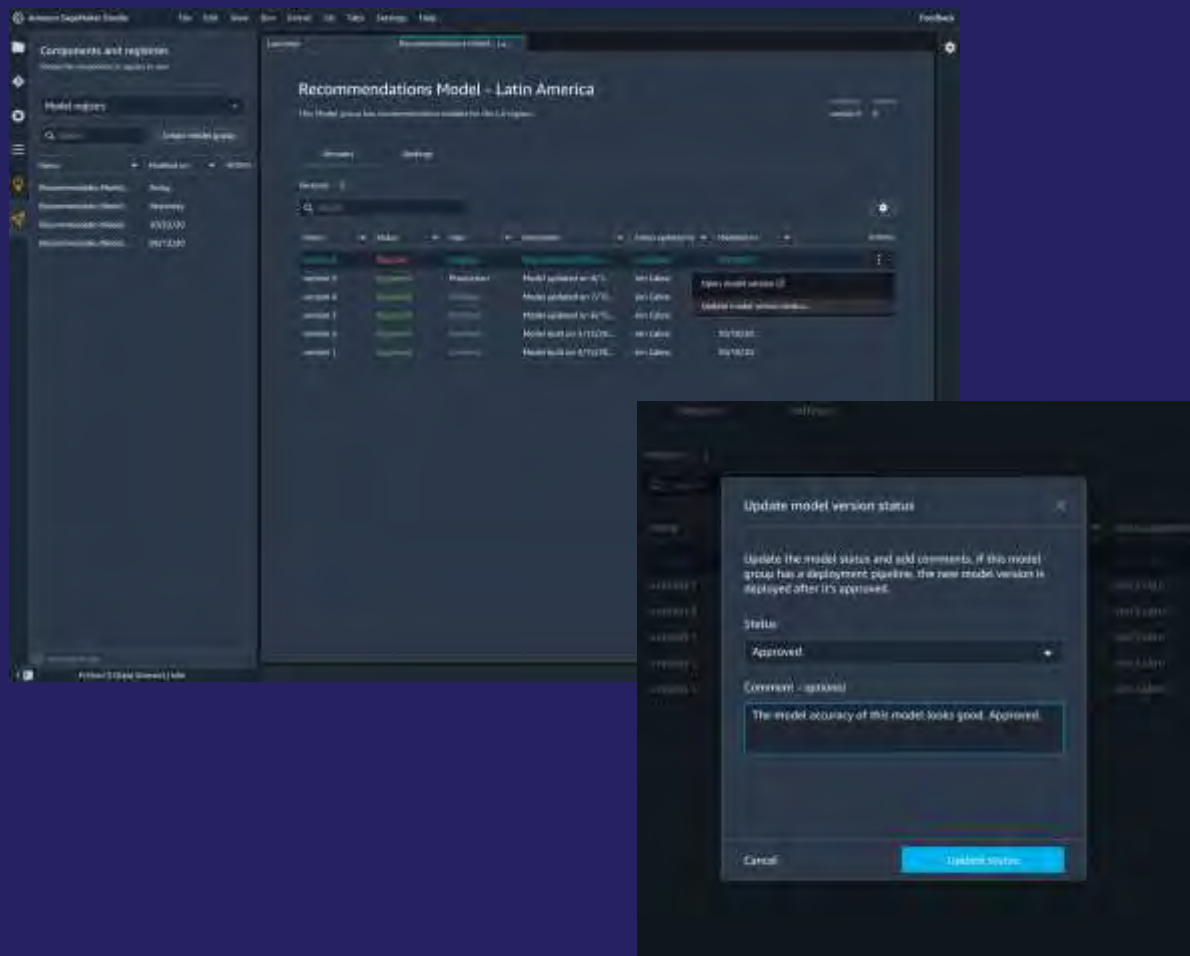
SageMaker Studio
機械学習用の統合開発環境 (IDE)

Amazon SageMaker は 組織による機械学習の 活用を支援します

- Amazon SageMaker は、ビジネスアナリスト、データサイエンティスト、MLOps エンジニアが、機械学習の専門知識に関係なく、どのようなユースケースに対しても機械学習モデルを構築、トレーニング、デプロイできる、包括的な機械学習サービスです。
- ビジネスアナリストは、SageMaker Canvas でビジュアルインターフェイスを使用して機械学習の予測を行えます。
- データサイエンティストは、SageMaker Studio で簡単にデータを準備し、機械学習モデルを構築、トレーニング、デプロイできます。
- MLOps エンジニアは SageMaker MLOps により大規模にモデルをデプロイして管理できます。

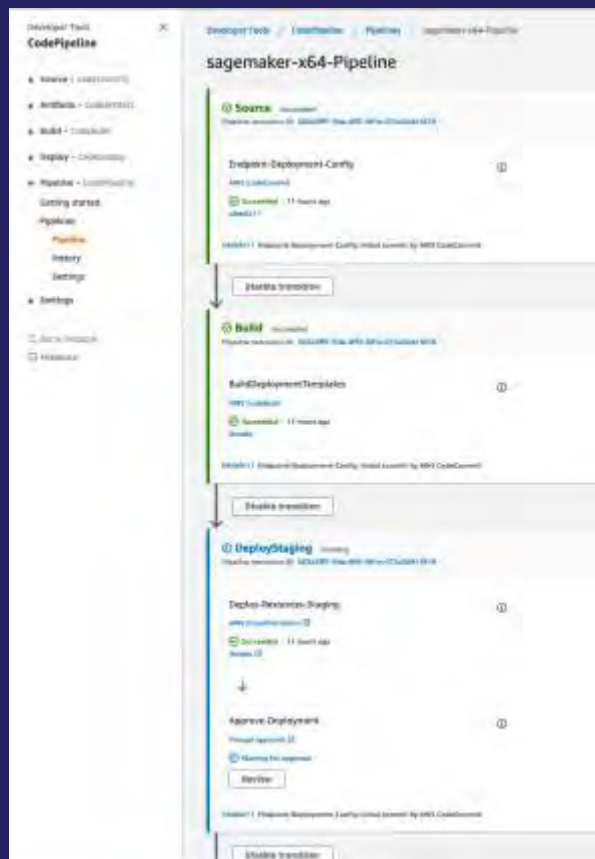


本番稼働用モデルの承認



これらがデータサイエンティストに提供されると、モデルレジストリ内で「承認」して、データサイエンティストが自分たちのモデルをCICD パイプラインに通します。

フルマネージド CI/CD パイプラインを使用するデプロイ

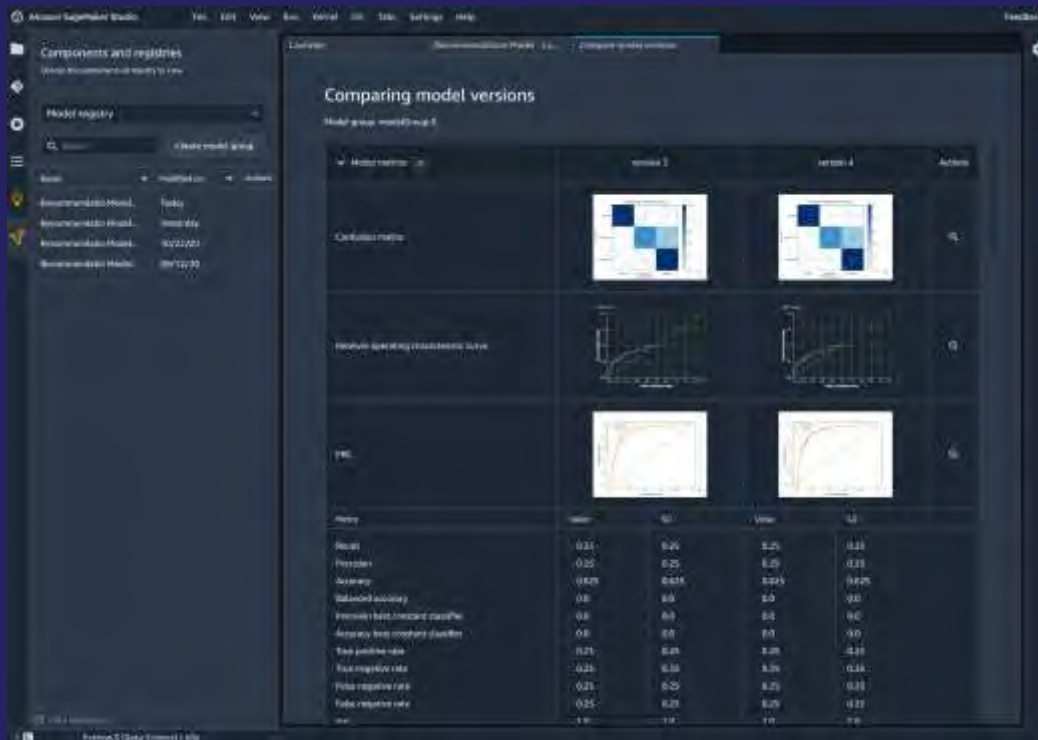


話題を変えて、モデルをトレーニングした後、および、モデルレジストリに追加された後にどうなるのかについて説明しましょう。CICD ワークフローを開始するためにどのように SageMaker Pipelines を使用するのでしょ

まず、CICD ワークフローの作り方を説明します。その中で、新しいペルソナ、DevOps エンジニアについて説明します。

このユースケースでは、DevOps エンジニアは実際に慣れている AWS ツールを継続して使用します。例としては、AWS CodePipelines があります。CICD パイプラインを定義して、モデルアーティファクトをさまざまな環境と統合されたテストにどのように通すかをコード化できます。

トレーニングステップの 評価メトリクスを表示して比較する



この統合の優れている点は、SageMaker Studio が提供する単一の画面を利用できることです。モデルが本番稼働にデプロイされた後、それらのモデルの結果は Studio 内で確認できます。

Amazon SageMaker の主なメリット

最も完全なエンドツーエンドの機械学習サービス



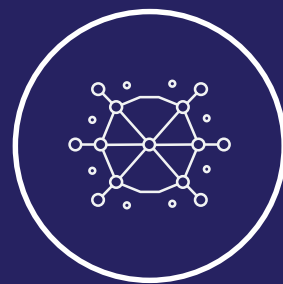
機械学習イノベーションを 誰もが利用できるようにする

ビジネスアナリストを含め、
より多くの人々のグループを支援



機械学習ライフ サイクルを加速化

トレーニング時間を数時間
から数分間に短縮



大規模なデータの 準備

構造化データと非構造化
データにアクセスし、ラ
ベル付けして処理する



機械学習プロセスの ストリーミング

MLOps プロセスの
自動化と標準化

Amazon SageMaker の主なメリット

最も完全なエンドツーエンドの機械学習サービス

1. ビジネスアナリストを含むさらに多くのグループの人たちが、独自のポイントアンドクリック、ノーコードビジュアルインターフェイスを使用して機械学習モデルを作成できるようにします
2. 機械学習の膨大な構造化データ（例：表形式データ）と非構造化データ（例：写真、動画、オーディオ）にアクセス、ラベル付け、処理します
3. 最適化されたインフラストラクチャでは、トレーニング時間を数時間から数分間に短縮します。専用ツールでチームの生産性を最大 10 倍向上します
4. 組織全体で MLOps プロセスを自動化および標準化し、数十万の本番稼働環境でモデルをデプロイして管理しています



Amazon SageMaker は DevOps 対応



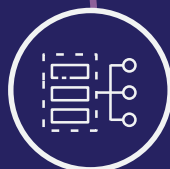
セキュリティ

機械学習ワークロードの厳格なセキュリティ要件に適合するように支援するセキュリティ機能



コンプライアンス

PCI、HIPAA、SOC 1/2/3、FedRAMP、ISO 9001/27001/27017/27018 のコンプライアンスに準拠



機械学習ワークフロー

数分間で自動化されたワークフローを作成して数千のモデルをサポート



スケーラビリティ

膨大なデータセットで複雑なモデルをトレーニング



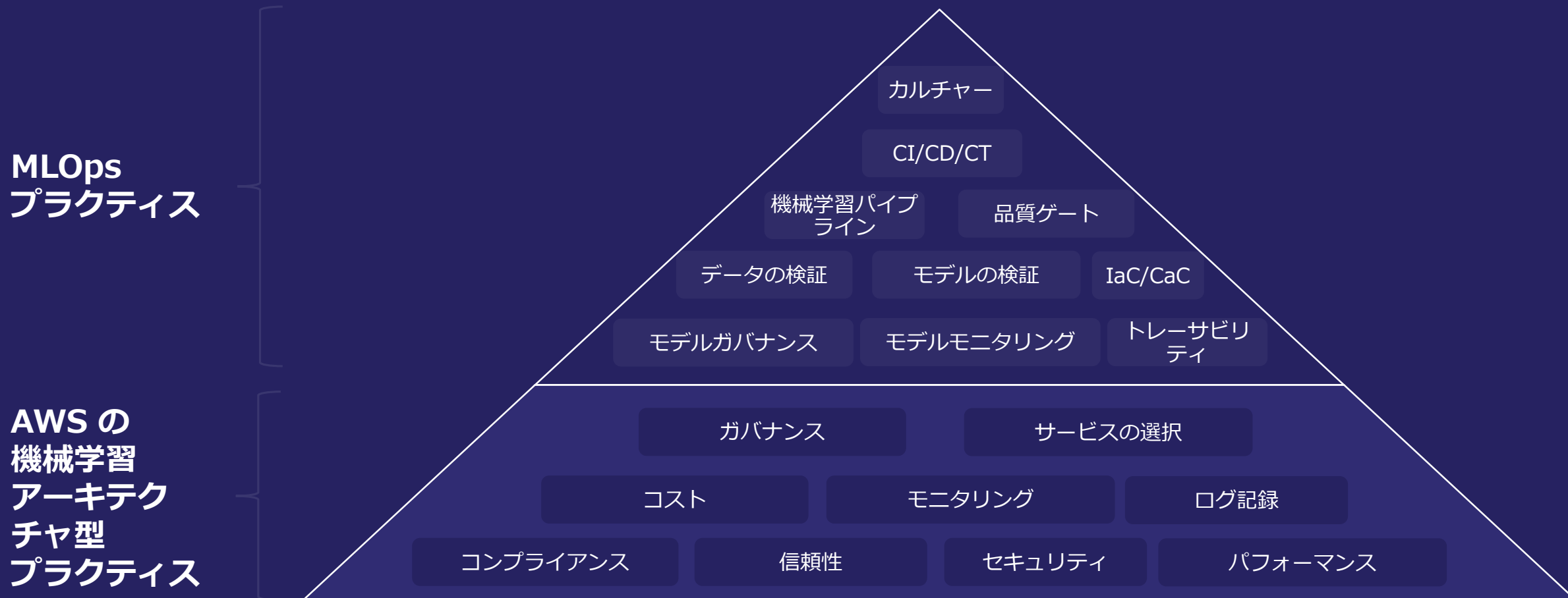
オーケストレーション

マネージドインフラストラクチャでジョブを自動的にスケジュールして実行



機械学習運用加速化フレームワーク

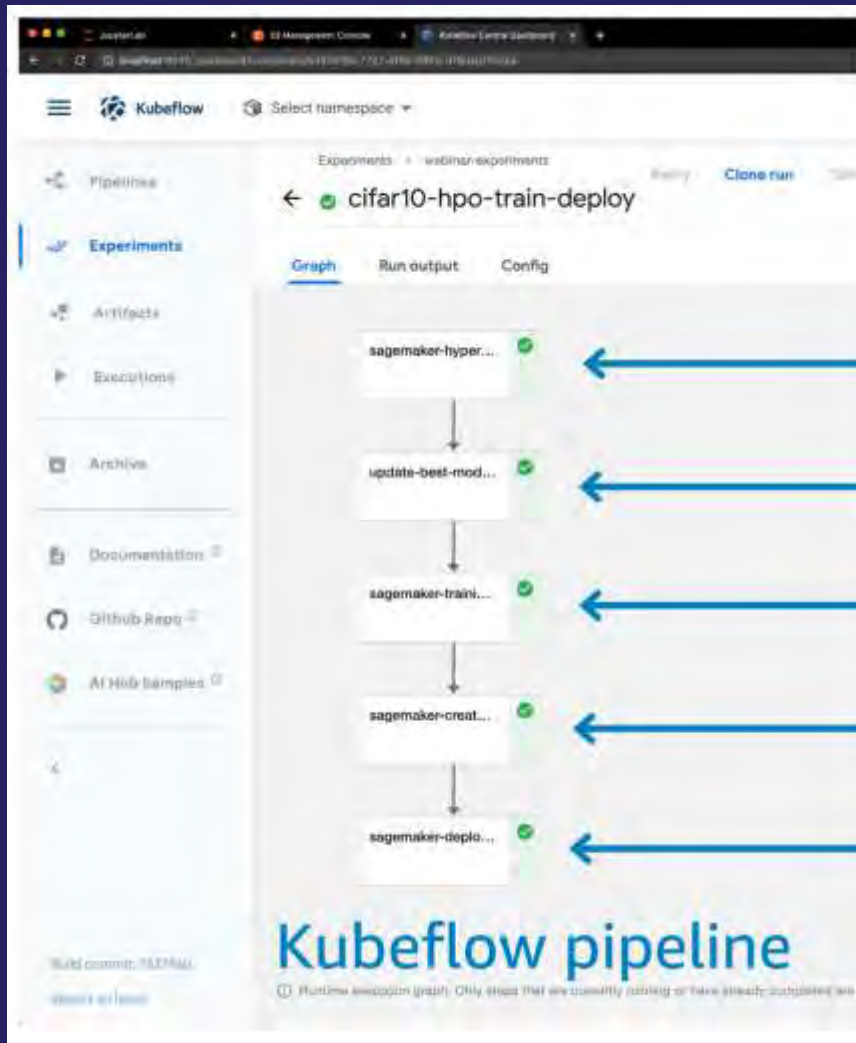
基本的な AWS 機械学習アーキテクチャと MLOps の練習



MLOps の定義

MLOps の成熟度モデル

	人材	データ	トレーニング	デプロイ
初期	- 切り離されたデータサイエンスと IT チーム - 制限されたクロストレーニング	アドホックなデータ収集と準備	- マニュアルトレーニングと再トレーニング - デプロイへの明確なパスがない	マニュアルデプロイ
繰り返し可能	- ステークホルダーとの共同作業の向上 - 共有プロジェクト目標	自動化されたデータパイプライン	- 実験のために定義されたパス - 自動化されたトレーニングパイプライン - マニュアルモデル検証	- 自動デプロイパイプライン - 制限されたモニタリング/測定
高信頼性	- 部門の枠を超えたプロジェクトチーム - ある程度のクロストレーニング	- 自動化された機械学習パイプライン - データガバナンス	- 実験の管理 - 自動化された機械学習パイプライン - モデルガバナンス - 自動化されたモデル検証	- 自動化された機械学習パイプライン - モニタリングとロギング (モデル、ワークロード、パイプライン)
スケーラブル	- 部門の枠を超えたプロジェクトチーム - クロストレーニング	- CI/CD - Policy-as-Code - Configuration-as-Code - 自動検証	- CI/CD - ポリシー/Config-as-Code - 自動化されたモデル検証 - 自動統合検証	- CI/CD - ポリシー/インフラ/Config-as-Code - モデルモニタリング - ダッシュボードと透明性



Amazon SageMaker の機能

ハイパーパラメータチューニング
ジョブ

epoch を更新するカスタム関数

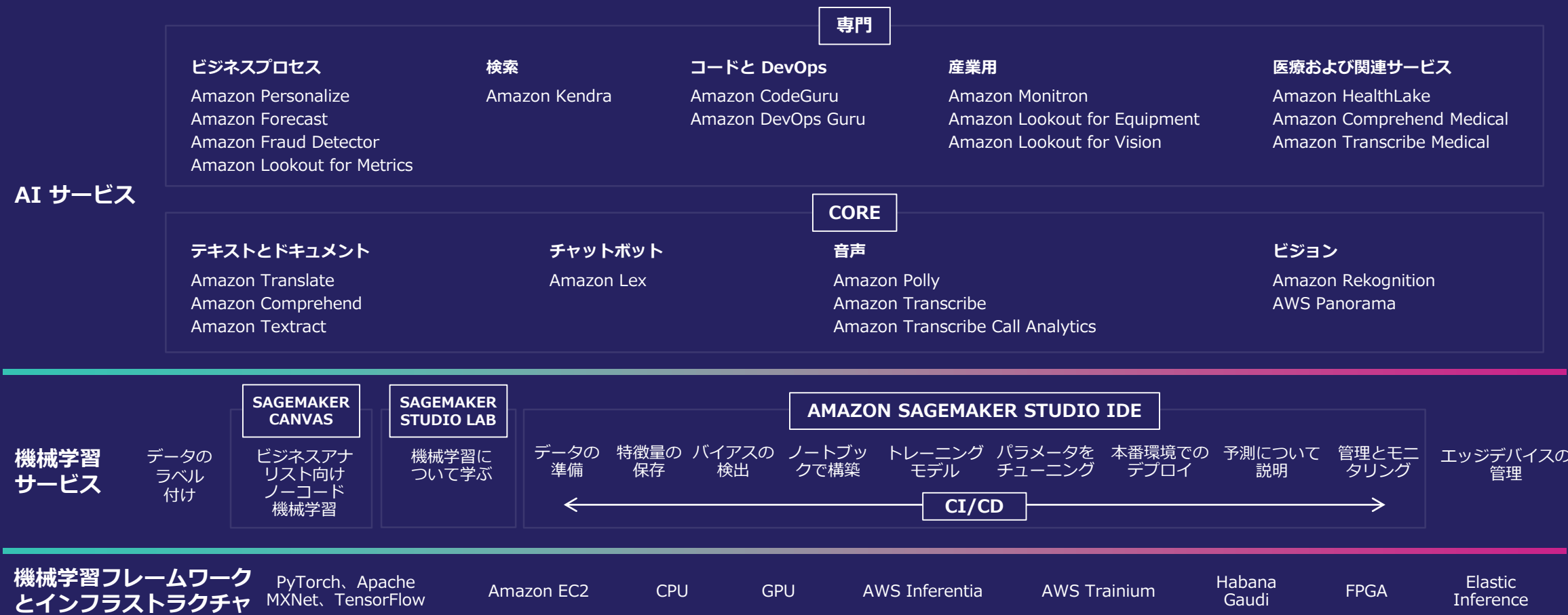
トレーニングジョブ

モデル作成

推論エンドポイントのデプロイ

AWS の機械学習スタック

最も幅広く多彩な機械学習機能のセット



AWS の機械学習スタック

最も幅広く多彩な機械学習機能のセット

- マクロレベルでは、ML には 3 層のスタックがあると考えています。
- 最下層には、専門的な ML プラクティショナーのためのフレームワーク、インターフェイス、インフラストラクチャがあります。
- フレームワーク層には、主に使用される 3 つのフレームワーク、TensorFlow、PyTorch、MXNet が存在します。
- 現在、クラウドベースの TensorFlow の 92% およびクラウドベースの PyTorch の 91% は AWS で実行されています。
- さまざまなフレームワークで新しいアルゴリズムが常に関開発されています。企業は、それらのアルゴリズムから借用するか実行したいとは思いますが、異なるフレームワークに移植することは望みません。その結果、AWS はすべての主要なフレームワークの専任チームを作り、業務に合った適切なツールをお客様に提供できるようにしました。

AWS の機械学習スタック

最も幅広く多彩な機械学習機能のセット

- AWS は機械学習インフラストラクチャサービスの最も幅広く最も奥深いポートフォリオを提供しており、お客様独自のパフォーマンスと予算のニーズに合わせてプロセッサとアクセラレーターを選択できます。[Amazon EC2 P4d インスタンス](#)は、初めてクラウドで使用される 400 Gbps インスタンスネットワークと結合された最新の NVIDIA A100 Tensor Core GPU を利用して、クラウド上でハイパフォーマンスな機械学習トレーニングを提供します。P4d インスタンスはハイパースケールクラスター (EC2 UltraClusters と呼ばれます) にデプロイされ、複雑な機械学習トレーニングジョブに対してスーパーコンピュータクラスのパフォーマンスを提供します。
- 推論に関しては、AWS Inferentia チップを搭載する [Amazon EC2 Inf1 インスタンス](#)がクラウドでハイパフォーマンスと最も低いコストを提供します。

AI/機械学習の最も幅広く最も奥深い コンピューティングインフラストラクチャ

パフォーマンスと予算のニーズに合わせた CPU、GPU、アクセラレーターの選択

従来の機械学習 (ML)

深層学習 (DL)

トレーニングと推論

推論

トレーニング



Cascade Lake CPU、Skylake CPU
Habana Gaudi アクセラレーター



EPYC CPU



Graviton CPU
Inferentia チップ



A100、V100、T4 GPU



AI/機械学習の最も幅広く最も奥深い コンピューティングインフラストラクチャ

パフォーマンスと予算のニーズに合わせた CPU、GPU、アクセラレーターの選択

- AWS は、AI/機械学習トレーニングと推論のための最も幅広く最も奥深いコンピューティングインフラストラクチャの提供におけるリーダーです。クラウドでの機械学習に GPU ベースのインスタンスをリリースしたのは AWS が最初でした。最新の CPU、GPU、およびカスタムアクセラレーターに基づく幅広い Amazon EC2 コンピューティングインスタンスから選択でき、お客様の予算のニーズとモデルのパフォーマンス要件に合う、業界をリードするネットワークとストレージを採用することができます。従来の機械学習を実行している場合、私たちの CPU ベースのインスタンスを選択して、小から中規模の機械学習モデルのトレーニングと推論を実行できます。深層学習モデルのトレーニングと推論を実行するためのハイパフォーマンスインスタンスを探している場合、AWS の GPU とアクセラレーターベースのインスタンスを活用できます。
- AWS の EC2 P4d インスタンスは、クラウドで深層学習モデルをトレーニングするための最もハイパフォーマンスなコンピューティングインスタンスです。前世代の GPU ベースの P3 インスタンスより 2.5 倍優れたパフォーマンスと 60% 低いコストが得られます。大規模な DL モデルのマルチノード分散トレーニングのために設計されています。P3 インスタンスは、中から大規模の深層学習モデルのトレーニング、および単一ノード機械学習トレーニングのためのハイパフォーマンスで、コスト効率が高いインスタンスです。AWS はお客様のニーズに最適なオプションを選択できるよう、ハイパフォーマンスで低コストの幅広いインフラストラクチャを実現するための投資とイノベーションを行っています。

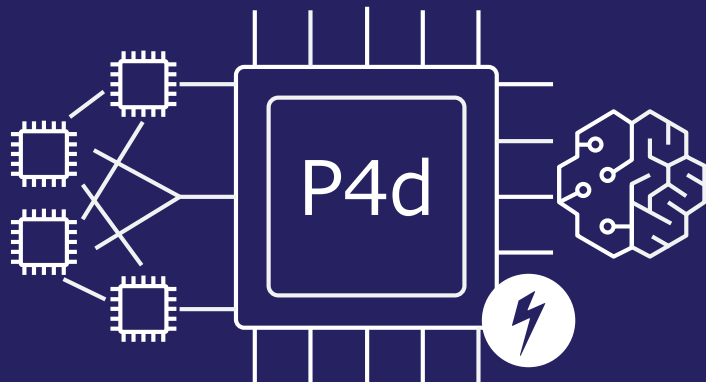
AI/機械学習の最も幅広く最も奥深い コンピューティングインフラストラクチャ

パフォーマンスと予算のニーズに合わせた CPU、GPU、アクセラレーターの選択

- 例えば、Habana Labs (インテルの子会社) の Gaudi アクセラレーターを搭載した EC2 DL1 インスタンスは、深層学習モデルのトレーニングで競合他社の GPU ベースのインスタンスより 40% コストパフォーマンスに優れています。AWS は、機械学習の推論パフォーマンスに最適化されたシリコンも最初から構築しています。Graviton CPU と Nitro システムのイノベーションから得られた学びを活かし、AWS Inferentia チップを構築しました。これは、クラウドでの推論で最低コストを実現したカスタムチップです。
- ハイパフォーマンス GPU ベースの推論には、G4d インスタンスを使用できます。AWS で機械学習ワークロードを実行することにより、機械学習と DL モデルをトレーニングしデプロイするためのハイパフォーマンス、低コストと使いやすいインフラストラクチャサービスにオンデマンドでアクセスできます。

Amazon EC2 P4d インスタンスの紹介

P4d インスタンス



クラウドでの最も強力な
GPU インスタンスの 1 つ

機械学習モデルでトレーニングコストが 最大
60% 減少、平均 2.5 倍の深層学習パフォーマン
ス、25% 多い GPU メモリ容量

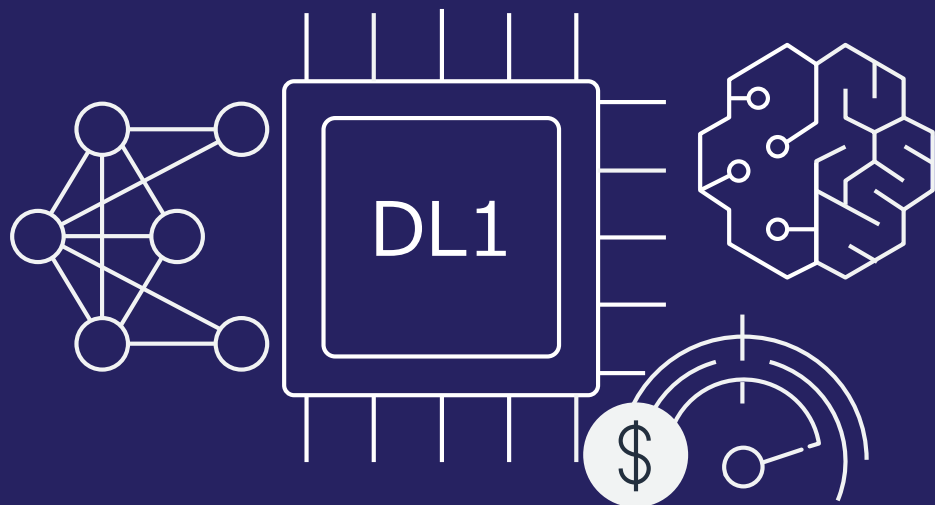
8 基の NVIDIA A100 GPU と 400 Gbps のネッ
トワーク帯域幅を備え、2.5 ペタフロップスのパ
フォーマンスが可能

数千の緊密に結合された GPU で構成される
UltraClusters でデプロイされており、機械学習
トレーニングと HPC に最適

Amazon EC2 DL1 インスタンスの紹介

深層学習モデルをトレーニングするための優れたコストパフォーマンス

DL1 インスタンス



Habana Labs (インテルの子会社) による最大 8 つの Gaudi アクセラレーターを搭載

深層学習モデルのトレーニング専用として構築

最新の GPU インスタンスと比べて最大 40% 優れた料金パフォーマンスを実現

TensorFlow と PyTorch にシームレスに統合されたカスタムソフトウェア

DLC、DL AMI、または Amazon SageMaker を使用して簡単に開始可能

コンテナ化された機械学習アプリケーション用の Amazon ECS や Amazon EKS を介して DL1 インスタンスを起動

AWS を利用する大規模な NLP スタートアップ

- CA スタートアップは、世界最大の NLP モデル (2T を超えるパラメータ) を構築することを目標にしています
- POC の後、50B パラメータモデルを 200 個の P4d (800 個の A100 GPU) でトレーニングするため AWS での稼働を決定
- トレーニングワークロードは現在 500 を超える P4d (4000 GPU) で本番稼働
- EKS オーケストレーションとネットワークのスピードアップのための EFA を併用し直接 EC2 で稼働
- シングル AZ でデプロイされ、複数の AZ にまたがるようにアーキテクチャの拡張を検討
- **システムの主なインフラストラクチャ要素:**
 - オーケストレーションのための EKS
 - 大規模な高速ノード間通信をサポートする EFA
 - 500 を超え、増加を続ける P4d のノード
 - S3 ストレージレイヤー
 - AWS ベースの DLAMI から構築されたカスタム DLAMI
 - PyTorch + DDP + NCCL



セルフマネージド機械学習

進化する開発ニーズに応える幅広い選択肢

ツール

データ処理とラベル付け				DL フレームワーク		開発、トレーニング、チューニング (OSS)				デブロイ		サーバーレス		ML OPS
														コードビルド コードコミット CDK ECR Jenkins*
Kinesis	Glue	EMR	SageMaker Ground Truth	DLAMI	コンテナ	ノートブック	Horovod	Katib	Kubeflow オペレーター	KFServing*	AWS Greengrass	AWS Lambda	Nuclio*	
OSS				OSS PYTORCH				DASK		TF Serving*、NVIDIA Triton*				

オーケストレーション

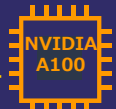
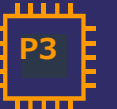
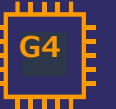
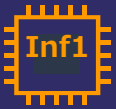
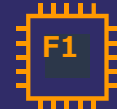
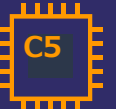
AWS のサービス

				
EKS	ECS	Fargate	Batch	ParallelCluster

ワークフロー

OSS	AWS/パートナーサービス			OSS	
					
Kubernetes	Step Functions	Cloud Formation	Terraform	Kubeflow Pipelines	Airflow

インフラストラクチャ

コンピューティング	ストレージ	データ転送	ネットワーク
(プレビュー)       アクセラレーター	 S3  EFS  Lustre	 Direct Connect  Snowcone  Snowball  Snowball edge	 EFA



開始方法: 次のステップ



開始方法: 次のステップ



コラボレーション

AWS は
機械学習のエキスパート



ワークショップを見つける

ユースケースの特定



トレーニング

AWS ML Embark Program
AWS の Deep Devices

開始方法: 次のステップ

最も幅広く奥深い一連の機械学習サービスにより、AWS は機械学習の導入をガイドできます。お客様と協力して、機械学習を使用して解決できるビジネス上の課題を解明し、適切なユースケースの特定を支援します。

AWS にはスタートを支援するためのトレーニングやプログラムがあります。AWS ML Embark プログラムおよびお客様のデベロッパーのための無料オンライントレーニングと認定プログラムを通じて、テクニカルとビジネス両方のステークホルダーに対面でのトレーニングを提供します。さらに、AWS の Deep Devices ポートフォリオ (AWS DeepRacer、AWS DeepComposer、AWS DeepLens) は、デベロッパーが実践的に楽しく、関心を持って機械学習を学べるように設計されています。





ありがとうございました。

[お問い合わせ](#)

[詳細はこちら](#)