

MLの開発、トレーニング、テスト、およびハイブリッドクラウドへのデプロイ

ハイライト

インテリジェントなアプリケーションを、オンプレミスまたはパブリッククラウドに迅速に提供します。

AIのビジネスへの導入を単純化し、AIの導入を促進し、柔軟性をもたらします。

一貫したユーザーエクスペリエンスでチーム全体にAI/MLの運用上の一貫性を確立し、データサイエンティスト、データエンジニア、DevOpsチームが効果的にコラボレーションできるようにします。

ハイブリッドクラウドの柔軟性と、AI/MLワークロードをクラウド、オンプレミス、エッジなど、データが存在する場所の近くで構築、トレーニング、デプロイ、監視する機能を獲得します。

Open Data Hub は、**Red Hat OpenShift** で AI-as-a-Service (AlaaS) プラットフォームを構築するための青写真です。Jupyter や **Kubeflow** などのアップストリームの取り組みを継承しており、Red Hat OpenShift AI の基盤となっています。

人工知能と機械学習のデプロイを加速する

生成AIの導入が急速に進み、企業や組織にとって、人工知能(AI)、機械学習(ML)、ディープラーニング(DL)の重要性が急速に増えています。Gartnerでは、2026年には企業の85%が生成AIアプリケーション・プログラミング・インタフェース(API)を使用、または生成AI対応アプリケーションをデプロイしていると予測しています。¹しかし、このようなテクノロジーのデプロイは、複雑になりかねません。データサイエンティストがモデルの構築に取り組む過程で、急速に進化するツール間の連携が失われていることをよく体験します。このようなギャップは、データサイエンティスト、ソフトウェア開発者、IT運用担当者の生産性とコラボレーションに悪影響を及ぼすことがあります。AI/MLのデプロイをスケーリングするには、ハードウェア・アクセラレーションと生成AI向けの分散ワークロード用に高価なGPU(グラフィックス・プロセッシング・ユニット)リソースが必要とされる一方で、リソースが制限され、管理が複雑になることがあります。一般的なクラウド・プラットフォームはスケーラビリティと魅力的なツールセットを提供しますが、それらのツールはしばしばユーザーを縛り、アーキテクチャとデプロイの選択肢を制限します。

オープンソースの**Open Data Hub**プロジェクトに基づき、**Red Hat® OpenShift® AI**²ではデータサイエンティストが生成AIを含めたML/DLモデルを迅速にトレーニング、テスト、提供、監視できます。ユーザーはインフラストラクチャのプロビジョニングを待つことなく、すぐにモデリングとアプリケーション開発に集中できます。OpenShift AIは、Red Hat OpenShiftへのアドオンで、フルマネージド型のクラウドサービスとして、またはセルフマネージド型のソフトウェア製品として利用でき、Red Hatコンポーネント、オープンソースソフトウェア、およびテクノロジーパートナーの製品に、オンプレミス、クラウド、エッジインフラストラクチャでモデルを開発して提供できるという柔軟性をもたらします。

Red Hat OpenShift AI

OpenShift AIを使用すると、AI/MLモデリングを実行する一般的なオープンソースツールとサードパーティ製ツールの統合セットを効率的にデプロイできます。このプラットフォームは、個々のクラウドプロバイダーが提供する規範的かつ独断的なAI/MLスイートに代わるものです。導入すると、インフラストラクチャやパブリッククラウド固有のツールによるロックインを心配することなく、実験モデルを構築するためのコラボレーティブなオープンソースのツールセットとプラットフォームを取得できます。また、ベース・プラットフォームはパートナーツールを使用して拡張し、機能を強化できます。ハイブリッドクラウド環境とエッジ環境で一貫して、モデルをコンテナ対応形式で本番環境に提供できます。OpenShift AIは、セキュリティに重点を置いたスケーラブルで実証済みのプラットフォームでシンプルに構成でき、IT運用担当者は管理しやすい環境が得られます。

OpenShift AIは一般的な生成AI基盤モデルをサポートするので、これらのトレーニング済みモデルを固有のユースケース向けに固有のデータを使ってプロンプトチューニングおよびファインチューニングを行い、提供できます。ワークロードを、所在する場所にかかわらず、複数のRed Hat OpenShiftクラスタ上に分散させることもできます。このプラットフォームにより、中央処理装置(CPU)や、Nvidia GPUやIntel XPUなどのグラフィック処理装置(GPU)ベースのハードウェア・インフラストラクチャがサポートされ、AIハードウェア・アクセラレーションを簡単に利用できるようになります。データサイエンス・プラットフォームの日常的な管理を自分で行う必要はありません。

¹ Gartner プレスリリース、「[Gartner Says More Than 80% of Enterprises Will Have Used Generative AI APIs or Deployed Generative AI-Enabled Applications by 2026](#)」、2023年10月11日。

² 旧称 Red Hat OpenShift Data Science



Red Hat コンサルティングでは、組織が OpenShift AI の導入プロセスに着手し、既存のエンタープライズに統合する作業を支援するため、**OpenShift AI パイロット**の取り組みを提供しています。

モデルの実験からモデルのプロダクションへのデプロイに進める戦略の策定を検討している組織には、Red Hat コンサルティングでは **MLOps 基盤** コンサルティングサービスも提供しています。

アップストリームのオープンソース商用テクノロジー・パートナー・ツール

Red Hat OpenShift AI は、アップストリームの Open Data Hub プロジェクトにあるツールのサブセット (表 1) を提供します。あらゆるクラウド環境、フルマネージド型およびセルフマネージド型の Red Hat OpenShift でモデルを開発、テスト、デプロイし、そのパフォーマンスを一元的に監視できます。Red Hat はオープンソースツール (Jupyter、Pytorch、TensorFlow など) に定期的な更新を提供しているので、統合、テスト、メンテナンスの負担が軽減されます。この製品には、複数の AI/ML テクノロジーパートナーの製品も統合されています (表 1)。Red Hat OpenShift で製品を認定した 30 以上の AI テクノロジーパートナーを通じて、商用テクノロジーパートナーの製品を追加することもできます。

表 1. Red Hat OpenShift AI のエコシステム

AI/ML モデリングおよび可視化ツール	JupyterLab UI (事前構築済みのノートブックイメージと一般的な Python ライブラリおよびパッケージを利用可能)、TensorFlow、PyTorch、CUDA、複数のノートブックセッションを管理する Kubeflow ノートブック・コントローラー、Anaconda (Professional はオプション)、Intel の AI ツール
データエンジニアリング	Starburst (Galaxy と Enterprise はオプション)、Pachyderm (オプション)
データの取り込みとストレージ	Red Hat AMQ (オプションのアドオン)、Amazon Simple Storage Service (S3)
GPU サポート	NVIDIA (GPU オペレーター付き)、Intel XPU (Intel Xeon プロセッサ、Habana Gaudi、Intel Data Center GPU Flex シリーズを含む)
モデルの提供と監視	モデルの提供 (ユーザー・インタフェースを備えた KServe)、モデルの監視、OpenShift Source-to-Image (S2I)、Red Hat OpenShift API Management (オプションのアドオン)、OpenVINO ツールキットの Intel ディストリビューション
データサイエンス・パイプライン	データサイエンス・パイプライン (Kubeflow パイプライン) で、データ準備などのプロセスを連結、モデルを構築、モデルを提供



Red Hat について

Red Hat は、**受賞歴のある**サポート、トレーニング、コンサルティングサービスをお客様に提供し、複数の環境にわたる標準化、クラウドネイティブ・アプリケーションの開発、複雑な環境の統合、自動化、セキュリティ保護、運用管理を支援します。

f fb.com/RedHatJapan
twitter.com/RedHatJapan
in linkedin.com/company/red-hat

アジア太平洋
+65 6490 4200
apac@redhat.com

オーストラリア
1800 733 428

インド
+91 22 3987 8888

インドネシア
001 803 440 224

日本
03 4590 7472

韓国
080 708 0880

マレーシア
1800 812 678

ニュージーランド
0800 450 503

シンガポール
800 448 1430

中国
800 810 2100

香港
800 901 222

台湾
0800 666 052