Red Hat で金融サービスへの AI の導入を加速

エンドツーエンドの
プラットフォームで
AI/ML ソリューションの
市場投入時間を短縮

AI モデルの複雑性の高まりにより導入の課題が増えている

金融機関は、人工知能 (AI) の導入によってもたらされる機会を活用する方法を探っています。ディープラーニング、会話型 AI、生成 AI などの分野では急速に開発が進んでおり、それにともなって AI ソリューションのスコープや適用性が大きく高まっています。しかし同時にモデルもより複雑なものになってきており、実行にともなう新たな課題が発生しているのに加え、既存の課題も浮き彫りになってきています。そうした課題には次のようなものがあります。

- **スタンドアローンの開発プロセス：**AI や機械学習 (ML) の開発やトレーニングは現在、ほとんどの場合は専用の環境で行われており、GPU を始めとするアクセラレーティング用ハードウェアなどの特別なリソースを必要とします。AI/ML 環境のプロビジョニングには長い時間がかかり、それが AI ベースの新しいサービスのロールアウトを阻害する要因となっています。
- **スケーリング、柔軟性、リソース最適化：**AI/ML ソリューションにはさまざまなコンポーネントが必要で、しかもコンポーネントごとに CPU、メモリ、ディスク、GPU、TPU、FPGA などのリソースのニーズが異なります。そのようなソリューションのスケーリングには、多くの場合ハイブリッドクラウドのアプローチが必要になります。
- **監視とドリフト：**AI/ML モデルは、継続的な監視と定期的な更新によりドリフトの検出と修正を行う必要があります。Red Hat® OpenShift® はアプリケーションベースのドリフト監視と AI/ML 開発パイプラインを接続できる標準ベースの監視インフラストラクチャを提供し、モデル更新の継続的インテグレーションを促進します。
- **モデルのサプライチェーンのセキュリティ：**AI/ML の開発者向けツールのエコシステムは、その大部分が、コミュニティが主導するオープンソースのフレームワークを基礎としています。この環境でソフトウェア・サプライチェーンを安全に保つことは課題であり、その難度も高まりつつあります。開発者は最新のツールを求めますが、組織はそれらのツールが安全で、セキュリティ強化されており、脆弱性や悪意のあるアーティファクトが含まれていないことを確認する必要があります。

メリットにより複雑性を大きく低減

Red Hat が提案する AI/ML ソリューションには、金融機関にとっての次のようなメリットがあります。

- モデルの開発、トレーニング、推論を行うためのエンドツーエンドのプラットフォーム。これによりパブリッククラウドとプライベートクラウドの全体を通じて一貫性がもたらされ、プロセスのフェーズ間の摩擦が減少します。
- ML 環境の価値実現までの時間を短縮する、セルフサービス機能。
- 一貫性のある最先端のオープンソース ML ツールおよびライブラリと、オープンソースおよびパートナーがサポートするテクノロジーの広範なエコシステム。
- ML モデルの迅速な開発およびデプロイと、デプロイしたモデルを最新の状態に保つ監視および高速反復機能。

事例：大規模言語モデル

金融機関にとっての課題とメリットの例に、GPT-4、BLOOM、BART、DOLLY などの大規模言語モデル (LLM) ベースのソリューションの実装があります。この種のソリューションは、オンボーディングまたは本人確認 (KYC) におけるドキュメントのデジタル化、ESG データレポートの分析、チャットボットなどの会話型ソリューションの実装などに使用されます。

そのようなソリューションでは、数億から数十億のパラメータを持つ大規模 ML モデルを使用するのが一般的です。そうしたモデルは、労力、複雑性、および必要となる演算処理能力のために、通常は事前学習済みのモデルや基盤モデルを基に作成されます。それらのモデルは基本的に汎用データセットを使用してトレーニングされるので、金融サービスのユースケースという特定のコンテキストに適用するためには、ファインチューニングや転移学習により、小規模なローカルデータを使用して業界または企業独自の追加トレーニングを行う必要があります。図 1 に、この種のソリューションのサンプル・アーキテクチャを示します。

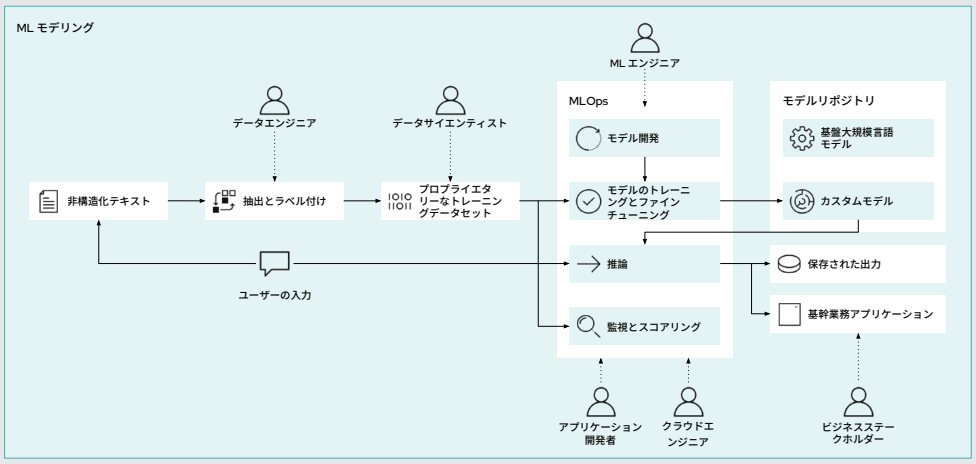

図 1. ML モデリングの概要

機能の概要

ソリューション・アーキテクチャ

Red Hat は開発から学習、推論まで AI/ML のライフサイクル全体を効率的かつ生産的にホストするプラットフォームを提供します。Red Hat のプラットフォームテクノロジーはベアメタル、オンプレミス仮想化、主要なパブリッククラウドなど、主な形式のインフラストラクチャで実行可能です。つまり、どのインフラストラクチャでも同じプラットフォーム、同じツール、同じ MLOps プロセスが使用できます。

Red Hat はオープンソースを熟知しており、ソフトウェア・サプライチェーンを守ることの重要性を理解しています。そのため、新しい優れたソフトウェアを開発し、信頼できる関係性を築くことができるようアップストリームのコミュニティと連携しています。その一環として、Red Hat は AI/ML 開発者が必要としている数多くのアップストリームツールの選定、サポート、認定を行っています。アップストリームのサプライチェーンの理解と、24 時間 365 日のサポート付きの頼れる製品の提供は Red Hat にお任せください。

プラットフォームのコンポーネント

オペレーティングシステム

Red Hat の AI/ML アーキテクチャの基盤は、オンプレミス、クラウド環境、ベアメタル、仮想マシンで実行可能なオペレーティングシステム (OS) である Red Hat Enterprise Linux® です。Red Hat Enterprise Linux は、極めて幅広いハードウェアエコシステムと、Amazon Web Services (AWS)、Google Cloud、IBM Cloud for Financial Services、Oracle Cloud、Microsoft Azure などの主要なクラウドプロバイダーでの使用が認定されています。この Linux プラットフォームは、セキュリティ、パフォーマンス、サポート、および Red Hat Ansible® Automation Platform を通じて世界最高級の自動化を提供します。さらに、Red Hat Enterprise Linux は GPU や FPGA など AI/ML モデルの開発向けの専用ハードウェアのサポートも提供します。

コンテナ・オーケストレーション

カスタムビルドのアプリケーションや商用アプリケーションに加え、AI/ML プロセスで使用されるオープンソースツールやライブラリの大部分はコンテナ化されています。事前トレーニングされた ML モデルやプロダクション環境で使用される ML モデルもコンテナイメージとしてパッケージ化されています。さらに、AI/ML のプロセスには、相互のインタラクションや弾力性に優れたスケーリングを必要とするコンポーネントが複数あります。そのようなコンポーネントには、大量の演算処理を必要とする新しいモデルのトレーニング、高スループットの推論エンジン、データサイエンティストが使用するモデル開発環境などがあり、これらはすべて柔軟で弾力性に優れたプラットフォームを必要とします。コンテナ化されたワークロードのデプロイとオーケストレーションで業界をリードしているのは、Kubernetes のディストリビューションの 1 つである Red Hat OpenShift です。 Red Hat OpenShift はサードパーティおよびオープンソースの AI 開発ツール向けプラットフォームの中で特に優れた人気を誇り、価値実現までの時間を短縮するために必要な AI/ML フレームワークへのアクセスを開発チームに提供します。また、Red Hat OpenShift では Operator テクノロジーを使用してコンポーネントのデプロイを自動化し、セルフサービスと運用コストを削減することを可能にします。

スケーラブルでセキュリティ強化されたストレージ

AI/ML プロジェクトで正確なモデルを構築するためには、大量のトレーニングデータが必要になります。このデータは履歴データの場合もあれば、市場データフィード、IoT (モノのインターネット)、可観測性などのソースからのライブデータの場合もあります。どのような場合であれ、それらのデータはユーザーフレンドリーかつ開発者が繰り返しアクセスできる方法で格納する必要があります。Red Hat は、Red Hat Ceph® Storage をベースとする Red Hat OpenShift Data Foundation を提供し、これによりオープンソースのソフトウェア・デファインド・ストレージをサポートおよび統合します。OpenShift Data Foundation は、Red Hat OpenShift と統合可能でペタバイト以上のスケーリングに対応するソフトウェア・デファインド・ストレージ・ソリューションです。ストリーミングデータは Apache Kafka をベースとした AMQ Streams によって消費し、開発者にストリーミングデータへの繰り返しアクセスを提供できます。OpenShift Data Foundation と AMQ Streams はどちらもコンテナにパッケージ化されており、Red Hat OpenShift で管理できるので、複数の開発チームがセルフサービスで運用できます。

プラットフォームの機能

セルフサービス

Red Hat OpenShift を使用すると、開発チームやプロジェクトは必要に応じてオンボーディングでき、リソースも状況に応じてスケールアップおよびスケールダウンが可能です。また、GPU を始めとする高価な専用ハードウェアはプールして共有できます。また、セキュリティ・コンプライアンスとソフトウェア・サプライチェーンの安全性があらゆるレベルに組み込まれています。

高度な監視と可観測性

Red Hat OpenShift には、オープンソースの Prometheus による業界標準の監視機能が含まれており、Splunk などのサードパーティ製監視ツールとの互換性も提供されています。これにより、MLOps パイプラインと、パイプライン全体で監視とアラートを提供する柔軟で一元化されたインフラストラクチャの統合が可能になります。モデルのパフォーマンスを追跡することで、スケーリングの自動化や、正確性が低下した場合のアラートなどが可能になります。

アジリティ

AI/ML モデルの作成は反復プロセスです。データエンジニアやデータサイエンティストはデータの示す道筋に分け入って探索します。そしてモデル開発の道のりには多くの開始点、終了点、予期しない脇道、落とし穴、行き止まりが存在します。よくある課題としては、さまざまなソース (データベース、ファイルシステム、ストリーム、API など) からの高品質なデータへのアクセスや、規制要件やセキュリティ標準へのコンプライアンスなどがあります。ツールの面での課題としては、広範なライブラリ全体でのバージョン管理、既存ツールの更新、新規ツールの受け入れなどがあります。Red Hat は AI/ML プロジェクトを加速できるようハイブリッドクラウド環境全体で一貫した体験を提供することで、作業者にとっての AI/ML パイプラインの単純化を支援します。

従来のアプリケーション開発と AI/ML アプリケーションの開発の違いの 1 つは、アプリケーションそのものやその核となる AI モデルを更新する必要性の大きさにあります。AI/ML の手法では、ML でモデルを初期トレーニングするだけでなく、モデルを継続的に更新していくことが可能です。そのため、モデルは従来のアプリケーションでは得られなかったメリットを提供することが可能です。しかしそれは、パフォーマンス強化のために定期的に「ループを閉鎖」してモデルを更新しなければならないということでもあります。Red Hat OpenShift を使用すると、アプリケーションチームは MLOps ツールチェーンのコンポーネントを透過的にスケールアップおよびスケールダウンすることができます。アプリケーションでモデルの更新が必要になったら、GPU やその他の専用コンポーネントなどの (コストの高い) トレーニング用リソースを手動で割り当てたり拡張したりすることが可能です。更新作業が完了したら、Red Hat OpenShift がそれらのリソースを必要な場所に再割り当てします。

トレーニングや推論のためのスケーラビリティと弾力性

AI/ML モデルのトレーニングフェーズは、MLOps パイプラインの中でも特にリソース使用量が多い操作です。このフェーズでは AI/ML ツールのインスタンスが最大限にスケールアウトされ、NVIDIA などの企業から提供される GPU、TPU、FPGA などの専用ハードウェアに対する需要が最も高くなります。個々のプロジェクトやチームは、トレーニングを行うために専用の環境にアクセスすることを望みます。Red Hat の AI/ML アーキテクチャは共有インフラストラクチャを提供できるので、効率性と経済性の面で大きな利点をもたらします。Red Hat OpenShift を使用すれば、開発者は高いコストのかかる専用リソースを大量に独占するのではなく、クラスタ全体にオンデマンドで仮想的にアクセスできます。Kubernetes はこのアクセスをオーケストレーションおよび仲介して、ビジネス的に最も必要とされる場所にそれらのリソースが提供されるようにします。

オープンなエコシステム

Red Hat の AI/ML プラットフォームは、他のあらゆる Red Hat 製品と同様、完全にオープンソースです。AI/ML の作業者が使用できるオープンソースのツールやテクノロジーのエコシステムには、次のようなものが含まれます。

- ML ライブラリ
- AI/ML ライフサイクル管理
- データアクセス、データ品質、メタデータ管理
- バイアス検知と説明可能性
- 事前トレーニング済みモデル

詳細　Red Hat で金融サービスへの AI の導入を加速

エコシステムのオープン性とプラットフォームの柔軟性により、これらのツールはソリューションの必要に応じて多様な組み合わせで使用できます。また、オープンなプラットフォームを使用すると先進的なテクノロジー、ツール、モデルを継続的にソリューションにプラグインできるので、継続的なイノベーションが可能になります。

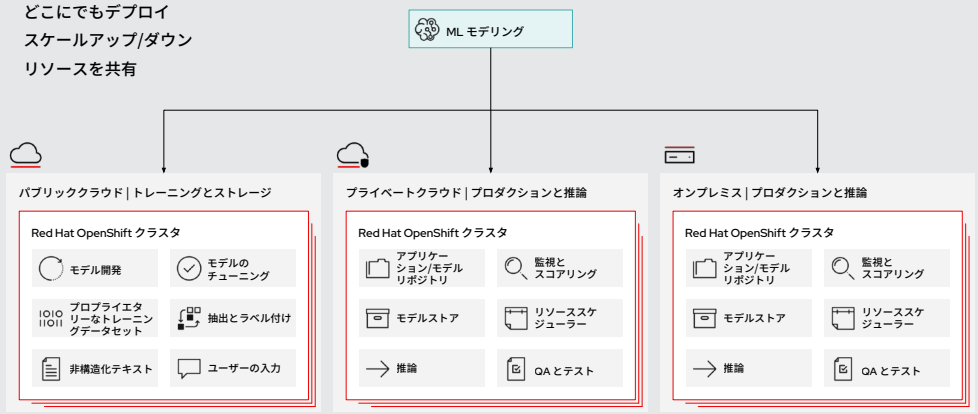

図 2. Red Hat を使用した MLOps の概要

Red Hat について

エンタープライズ・オープンソース・ソフトウェア・ソリューションのプロバイダーとして世界をリードする Red Hat は、コミュニティとの協業により高い信頼性と性能を備える Linux、ハイブリッドクラウド、コンテナ、および Kubernetes テクノロジーを提供しています。Red Hat は、クラウドネイティブ・アプリケーションの開発、既存および新規 IT アプリケーションの統合、複雑な環境の自動化および運用管理を支援します。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提供する Red Hat は、フォーチュン 500 企業に信頼されるアドバイザーであり、オープンな技術革新によるメリットをあらゆる業界に提供します。Red Hat は企業、パートナー、およびコミュニティのグローバルネットワークの中核として、企業の成長と変革を支え、デジタル化が進む将来に備える支援を提供しています。

アジア太平洋
+65 6490 4200
apac@redhat.com

オーストラリア
1 800 733 428

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
1 800 812 678

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

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

jp.redhat.com
#420450_0723