

データとAIを活用し変化に迅速に対応する 「DataOps」と「MLOps」、そしてこれらを実現する仕組みとシステムとは？

データの提供者、利用者、分析者が協働してデータの価値とその活用を促進する「DataOps」と「MLOps」という手法がある。この手法を採用するメリットとは何か。これらを実現するためにはどうすればいいのか。

2020年初頭の新型コロナウイルス感染症の拡大は、デジタルトランスフォーメーション（DX）への取り組みに拍車を掛けた。感染症の拡大防止のために、対面でのコミュニケーションが難しくなった。人が介在しなくても業務が回るようにするためのさまざまなテクノロジーが注目されるようになった。その一つに、AI（人工知能）テクノロジーを使った自動化や最適化の仕組みを業務に取り入れることが挙げられる。

勤と経験に頼らずデータ分析に基づいて迅速に意思決定し、ビジネスの課題を解決することは、コロナ禍であるか否かにかかわらず必要だ。また今回のような不測の事態にビジネスが素早く対応することの重要性も明らかになった。しかしながらデータを使った意思決定やAIテクノロジーを活用した自動化の取り組みを、実際にどう実現すればいいのか分からないという企業も少なくない。そこで鍵となるのが「DataOps」「MLOps」という手法だ。これらが意味するものは何なのか、具体的に何を実現すべきなのか。次のページで紹介する。

変化に迅速に対応するためのシステムとは

データとAIテクノロジーをビジネスに活用するとはどういうことか。一つがさまざまな業務アプリケーションから生成されるデータを集めて分析し、洞察を見出すことだ。もう一つはその洞察を機械学習などでモデル化して業務アプリケーションに組み込み、自動的に業務に適用される仕組みをつくることだ。この2つを実現するプロセスの構築に加え、市場や事業に変化が生じたときにこのプロセスを高速に進めることがDXを目指す企業に求められている。

DataOpsとMLOpsは、開発（Development）と運用（Operations）の密接な連携によってシステムの改善サイクルを高速化する開発手法「DevOps」を、それぞれデータ活用と機械学習に応用した概念だ。

DevOpsはアプリケーションの開発と本番運用、運用からのフィードバックに基づく再開と再運用というサイクルで、アプリケーションの継続的なデリバリーを実現する。DataOpsもMLOpsも、データと機械学習モデルの迅速な提供とフィードバック、フィードバック内容を基にした機械学習モデルの改善をサイクル化し、質の高いデータと機械学習モデルを迅速にビジネスに組み込む仕組みだ。DataOpsはデータ（Data）とOperationをつなげた造語で、データの品質とガバナンスを保ちながら、データ提供者とデータ利用者間のワークフローを自動化する仕組みを指す。人とプロセス、テクノロジーの三位一体で、データの準備時間を短縮してデータ活用のサイクルを速める。MLOpsは、機械学習（Machine Learning）とOperationを組み合わせた造語で、データの加工と機械学習モデルの開発、モデルの本番適用、さらには本番稼働におけるモデルのパフォーマンスのモニタリング、モデルの再トレーニングと本番への再適用といった、機械学習モデル活用のサイクルを円滑に進める仕組みを指す。

企業は以下の3つのサイクルを取り入れてさらに高速化することで、データやAIテクノロジーを活用して変化に迅速に対応することが可能になる（図1）。

- DevOpsで必要に応じてアプリケーションの開発と更新を迅速にできるようにすること
- DataOpsでさまざまなデータを収集、加工して迅速に分析に活用できるようにすること
- MLOpsで開発した機械学習モデルの業務への実装と改善、再実装ができること

データの準備時間を短縮するDataOpsの実践によって、どんなメリットがあるか。代表的な例が経営価値の向上だ。欧州のある小

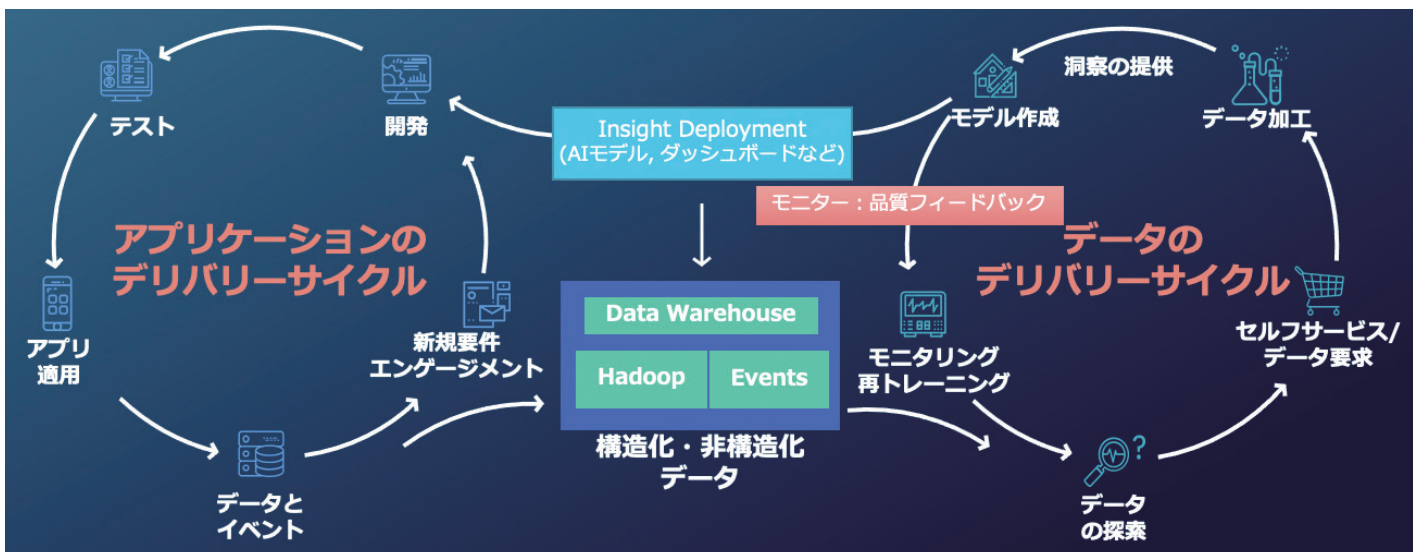


図1 DevOpsとDataOps、MLOpsのデリバリーサイクル（出典：日本アイ・ビー・エム）

売り手は、DataOpsを導入することで顧客属性分析の時間を20日間から1日未満に、在庫管理分析の時間を24時間から4時間未満に短縮した。これは、データを活用して素早い経営判断が可能になったことを意味する。

データとAIの活用を加速する「IBM Cloud Pak for Data」

DataOps、MLOpsを実現するために、システムはどうあるべきか。今までのシステムは「業務系システム」と「情報系システム」の2つに大別されていたが、AIテクノロジーを活用するにはもう一つの「分析系システム」が必要となる。情報系システムはデータウェアハウスやBI（ビジネスインテリジェンス）ツールなどを指す。これらのツールは正確な情報を決められた時間内に利用可能にしなければならないなど、一定の性能要件が求められる。新しいデータやレポートを追加しようとすると、ジョブやテーブルの設計、データマートやレポートの開発などに相当の時間が必要となる。

一方の分析系システムは、あらゆるデータを素早く分析することを重視し、情報系システムに比べて性能要件や精度よりも迅速性を重んじたシステムだ。分析系システムは、変化する分析要件に対応するため、さまざまなデータを試行錯誤しながら分析することを目的とする。さらにそこで新しく得た洞察を素早く業務系システムにフィードバックし、情報系システムでその効果を測定することで3種のシステムの補完関係が成り立つ。

新しい概念である分析系システムに求められる要件は3つある。1つ目は迅速性だ。素早くデータを提供してさまざまな分析要件に対応できることを指す。2つ目は拡張性だ。データ量の増大などシステムの対応に加え、データの配置場所や分析環境も変化する要件に合わせて拡張できるようなインフラ・テクノロジーの進化に対応できることが必要となる。3つ目は柔軟性で、必要な機能や分

析テクノロジーを必要なタイミングで利用できること、進化の速い分析テクノロジーを素早く取り入れられること、エッジコンピューティングのようにモデルをさまざまなロケーションに実装できることなどが含まれる。

IBMはこうした観点から、分析系システムのプラットフォーム「IBM Cloud Pak for Data」の提供を通してDataOpsとMLOpsの実現を支援する。データの提供者と分析者、利用者が同サービスを共同で利用することで、迅速なデータ活用と一貫したデータ、AIモデルの管理が可能になる。

「IBM Watson」ブランドでAI関連製品/サービス群を展開し、さまざまなAIプロジェクトを実施してきたIBMは近年、「AI Ladder」を提唱している。AI Ladderは、AIを業務に活用していくために、地道ではあるが、散在しているさまざまな種類のデータに簡単にアクセスできるようにし、利用しやすい形に加工して整理し、分析して業務に適用するために「はしご（Ladder）」を登るという考え方だ。このはしごは一度登ったら終わりではない。業務に適用したAIモデルに問題があればすぐさま対応が必要となる。何度もこのサイクルを速く回す必要があり、システム全体をモダナイズする必要がある。IBMはAI Ladderで提唱する以下の4つのステップごとに今までも製品やサービスを提供し、どこでも簡単に利用できるようにそれらのアップデートを進めてきていた。

- 「COLLECT」（データをシンプルでアクセスしやすいものにする）
- 「ORGANIZE」（データを分析可能に整理・整備する）
- 「ANALYZE」（信頼と透明性を備えたAIモデルを構築、スケールリング）
- 「INFUSE」（ビジネス全体でAIを運用する）

各ツールの狭間にタスクが落ちて分析のサイクルに時間がかか

るのを回避するために、IBM はこれまで個別に提供してきたツールの機能をマイクロサービス化し、Red Hat のコンテナ管理ツール「OpenShift」で動作するように最適化した。これらサービス群の集合体が IBM Cloud Pak for Data だ。データ分析者やデータ提供者、システム管理者など、さまざまな役割を持つユーザーが、同じプラットフォーム上で、同じ UI で、それぞれの作業を実行できる。

IBM Cloud Pak for Data のうち、COLLECT ではさまざまなデータソースへのアクセスや、データがどこにあるかを意識させずに利用可能にするためのデータ仮想化の機能を提供する。ORGANIZE の領域ではデータ品質管理やデータカタログ、データガバナンスのための機能を提供する。ANALYZE と INFUSE の領域では、GUI ベースのツールでの分析、プログラミング言語「Python」を使った AI モデルの開発、AI モデルの業務アプリケーションへの展開などが可能だ。さらに AI モデルを自動的に作成してデータサイエンティストのスキルを補完する「AutoAI」や、AI モデル群を一元管理して本番実行時のパフォーマンスをモニタリングしたり AI の説明性や透明性を確保したりする「Watson OpenScale」などの機能を提供する。これらの中から必要な機能を必要なタイミングでダウンロードし、アドオンの形で Cloud Pak for Data に追加することができる（図 2）。

アップデートで より高度な AIOps が可能に

IBM Cloud Pak for Data は、2018 年に提供が開始された。2020 年 6 月にリニューアルした最新の V3.0 が 8 回目のバージョンアップとなるなど、頻繁なアップデートで進化を続けていることも

特徴だ。

V3.0 では、UI にオープンソースのデザインフレームワーク「Carbon 10」を採用し、各機能の操作性を統一した。AI モデルを自動生成する AutoAI も強化され、さまざまなパターンの統計手法を分析してより精度の高いモデルを作成できるようになった。その他、データに基づく予算策定や事業計画などをシミュレーションする機能「Planning Analytics」やガバナンス違反をチェックすることでリスクを管理する機能「InstaScan」も追加された。IBM は今後も企業を取り巻くデータプラットフォームの課題解決に向けて、IBM Cloud Pak for Data の機能を迅速に拡張する。

IBM は AI テクノロジーを活用して負荷軽減や自動化を実現するサービス「Watson AIOps」も発表済みだ。アプリケーションの運用中に発生するイベントや通知、システム状況の使用率、リソースの使用状況などの構造化データや、システムログやチケットといった非構造化データから障害を自動的に予測・検知する。アプリケーションの稼働状況はチャットで素早く知らせるとともに障害解決策の候補を提示し、必要であれば API と連携して復旧作業を自動化する。日本では 2020 年内に IBM Cloud Pak for Data に Watson AIOps の追加が可能になる予定だ。

データに基づいた意思決定を迅速に業務に取り入れる組織をつくるには、DataOps と MLOps が重要な鍵を握る。データを提供する側とデータを使う側が同一のサービスを使うことで、迅速で一貫したデータ活用が実現する。

IBM は IBM Cloud Pak for Data を 1 週間無償で利用できるトライアルサービスを用意している。「まずは体験し、自分の目で確認したい」というユーザーの声に応えたい考えだ。

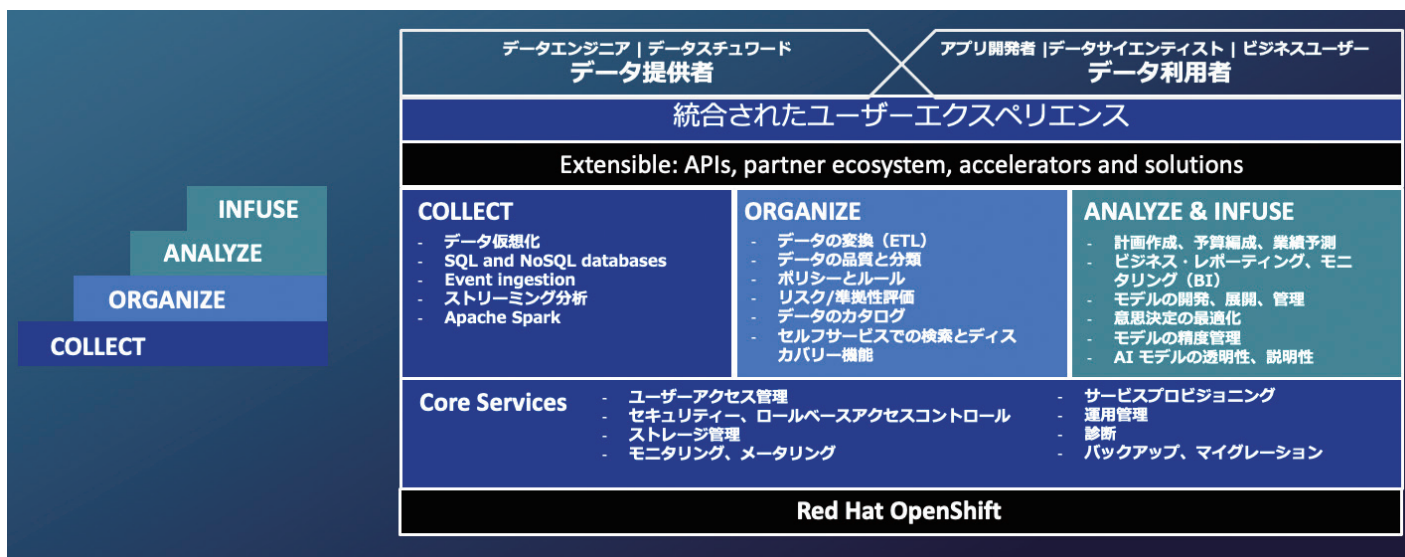


図 2 IBM Cloud Pak for Data の機能（出典：日本アイ・ビー・エム）

● お問い合わせ

日本アイ・ビー・エム株式会社

お問い合わせフォーム：<https://ibm.biz/otoiawase>

〒103-8510 東京都中央区日本橋箱崎町 19-21 IBM、IBM ロゴ、ibm.com、IBM Watson は、世界の多くの国で登録された IBM Corp. の商標です。
他の製品名およびサービス名等は、それぞれ IBM または各社の商標である場合があります。現時点での IBM の商標リストについては、ibm.com/legal/copytrade.shtml をご覧ください。

※この冊子は、TechTarget ジャパン (<https://techtarget.itmedia.co.jp/>) に 2020 年 10 月に掲載されたコンテンツを再構成したものです。
<https://techtarget.itmedia.co.jp/t/news/2010/20/news02.html>

copyright © ITmedia, Inc. All Rights Reserved.