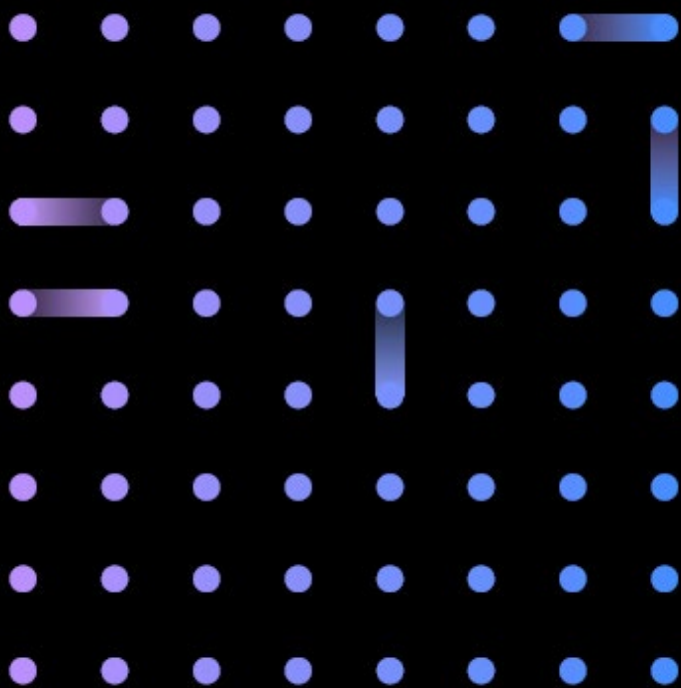


DataOps による、
ビジネスにすぐに
利用できるデータの迅速な提供

IBM DataOps の方法論と実践



目次

概要	3
DataOps の定義	3
目標の比較	4
IBM DataOps プログラム	7
IBM Cloud Garage 方法論における DataOps の活用	7
優れた DataOps の実践による効果	10
まとめ	11
付録: DataOps のパイロット・プログラム・テンプレート	12

ハイライト

- DataOps は、信頼できる高品質なデータをデータ・シチズン(data citizen, データ市民)とよばれるデータ利用者に素早く提供するための、人、プロセス、テクノロジーのオーケストレーションです。
- 自動化を原動力とする DataOps は、データへのアクセス、準備、統合、利用可能にする際の非効率性に関する課題を解決します。
- IBM DataOps は、高度な自動化を活用し、データ・キュレーション・サービス、メタデータ管理、データ・ガバナンス、マスター・データ管理、セルフ・サービスでのアクセスといった機能において非常に重要な効果をもたらします。
- IBM は、規範的な方法論、人工知能(AI)を活用した自動化、IBM DataOps Center of Excellence により、DataOps 実戦への道筋を提供します。
- DataOps ワークショップは、DataOps ロードマップにおいて重要な役割を果たします。ワークショップを通して、自社の DataOps の成熟度を評価し、パイロット・プロジェクトの実行を計画するのに役立ちます。

概要

データはイノベーションを推進し、競争上の優位性を維持するための原動力です。また、アナリティクスを推進し、ビジネスのトレンドや機会を把握するための重要な要素です。データの価値を新たな方法で引き出すことで、企業の AI 活用への道をさらに加速させることもできます。

しかし、データ関連のプロジェクトが期待された投資収益率(ROI)を達成できないと、関係者はその理由を理解しようとします。Experian 社の「2019 Global Data Management Research」レポートによると、89%の企業はデータの管理に苦労していると報告されています。主に、洞察を獲得するまでのスピードの遅れや、基礎データの信頼性の欠如に悩まされています。¹

組織のビジネス目標を理解することは、アナリティクスと AI のための効果的なデータ戦略を策定するうえで不可欠です。ビジネス・モデルが機能するには、顧客のニーズを満たすものでなければなりません。成功の鍵を握るのは、統合されたビジネス対応のデータ・パイプラインによってデータの処理作業を効率化し、いつでもビジネスを完全かつ一貫して把握できるようにすることです。これにより、常に、ビジネスに一貫した見通しがもたらされます。

より迅速に結果を出すことへの期待は、高まり続けています。あらゆる企業は、組織内のサイロを解消し、最良の意思決定を行うために、運用の効率と効果を改善する方法を探求しています。この2つの要因により、ビジネス・リーダーは、1つの枠組みの下で最大の課題に取り組む新たな道を模索しています。

データ運用の変革を目指す企業に、自動化テクノロジーは競争上の優位性をもたらすかもしれません。信頼性が高く、ビジネスにすぐに利用できるデータが、差別化を実現する洞察や優れた運用を企業にもたらすことで、データは価値あるものとなります。

このホワイト・ペーパーの目的は、DataOps の方法論、実践、ロードマップのメリットをご案内することです。

DataOps の定義

DataOps とは、信頼できる高品質なデータをデータ利用者に素早く提供するための、人、プロセス、テクノロジーのオーケストレーションです。DataOps の実践は、組織全体のコラボレーションを可能にし、俊敏性、スピード、新しいデータ・イニシアチブを大規模に推進することに重点を置いています。自動化の力を活用する DataOps は、データへのアクセス、準備、統合、利用可能にする際の非効率性に関連した課題を解決するよう設計されています。

DataOps の潜在的なメリットには、個人に情報とデータを提供する際の大幅な生産性の向上と、プロセスを改善して効率性と最適化を図ることなどが挙げられます。AI データを活用した取り組みを含むデータ運用の自動化は、以下のような成果をもたらします。

- すぐに利用可能な統合データをビジネスに提供し、アナリティクスと AI を大規模に推進
- 運用効率の達成
- データ・プライバシーとコンプライアンスの遵守

89%

の企業はデータの管理に苦労しています。¹ 組織のビジネス目標を理解することは、アナリティクスと AI の効果的なデータ戦略を策定するために不可欠です。



DataOps と DevOps は異なる

大多数の組織は、開発分野において、ある程度の DevOps を実装しています。DevOps の実践は広く知られており、新たに登場した DataOps とネーミングも似ています。よく比較される両者は、どちらも運用のベスト・プラクティスを推進するための方法論ですが、それぞれ組織内での領域は異なります。

以下の表に、目標と、組織に対するメリットという観点から、2つのプラクティスを比較してみました。

目標の比較

	DataOps	DevOps
主な焦点	迅速に活用でき、ビジネスにすぐに利用可能で、信頼できる高品質データ	アプリケーションおよびソフトウェアの開発
変革の目標	<ul style="list-style-type: none">- すべてのデータ利用者が信頼できる高品質データにセルフ・サービスでアクセスできるようにすることで、ビジネスの継続的かつ迅速なイノベーションを推進- データ・ガバナンスおよび統合の自動化により、データの継続的な提供を実現すると同時に、規制上の考慮事項にも対応- データ・パイプラインのモニタリングおよび最適化により、すべてのデータ利用者から継続的に学ぶためのフィードバック・ループを提供	<ul style="list-style-type: none">- コラボレーションによる開発とバリュー・チェーン全体でのテストにより、アイデアの継続的なイノベーションを推進- ソフトウェアの提供プロセスを自動化して無駄を排除することで、規制上の考慮事項に対応することで、アイデアのイノベーションの継続的な提供を実現- ソフトウェア主導のイノベーションのモニタリングと最適化により、顧客からの継続的な学習のためのフィードバック・ループを提供
効率性の目標	<ul style="list-style-type: none">- IT システム・サポート、運用、ビジネス間の関係の強化により、人と目標の不整合を是正- データ提供のサイクルを通して自動化を導入することにより、デリバリーの迅速化と品質向上を実現- 結果を利用して最適化を推進することにより、メタデータとデータの真の価値に対する洞察を改善	<ul style="list-style-type: none">- 開発者、運用、ビジネス間の関係強化により、人と目標の不整合を是正- 開発サイクルを通して自動化を導入することにより、高速化を実現し、変更のデリバリーのエラーを排除- 顧客のフィードバックを利用して最適化を推進することにより、アプリケーションの真の価値に対する洞察を改善

DataOps とは、人、プロセス、テクノロジーのオーケストレーションであり、DataOps の実践に向けた取り組みを継続するには、すべての機能を横断する密接なコラボレーションが必要です。また、アナリティクスのスピードと精度を向上させるためのデータ管理の手法とプロセスの確立にも取り組む必要があります。

人とプロセス

DataOps は、自動化テクノロジーにより、プロジェクトのアウトプットと提供にかかる時間を短縮することで効率化を図り、生産性の高いチームをサポートします。しかしメリットを享受するには、社内の文化を真にデータ・ドリブンなものへと進化させる必要があります。より多くの部門がコンテキストに基づく洞察を得るためにデータを管理することを必要としている場合、以下を行う適切なタイミングであると考えられます。

- 組織に流れ込むデータの質と速度を向上させる。
- 企業全体でデータ・ドリブンのビジョンを支持し、これを維持するために、経営層からコミットメントを得る。

この種の変革は、ビジネスの真の目標を理解することから始まります。データは顧客に影響を与える意思決定やサービスにどのように役立つのでしょうか？市場での競争力を維持するために、データはどのように役立つのでしょうか？データが解決に役立つ財務上の優先事項にはどのようなものが挙げられるのでしょうか？

DataOps リーダーは、すべてのデータ利用者が果たす役割を定義し、企業文化と DataOps の実践を推進するために、すべてのデータ利用者が果たす役割を定義する必要があります。各組織には独自のニーズがあり、IT、データ・サイエンス、ビジネス分野の関係者は、ビジネスの成功を推進するための価値を高めていく必要があります。また、既存のデータ・ガバナンス委員会を活用したり、長年のデータ・ガバナンス・プログラムから得られた知見を利用したりすることで、この文化とコミットメントを確立することができます。ガバナンスは DataOps をサポートするために必要な原動力の 1 つとなります。

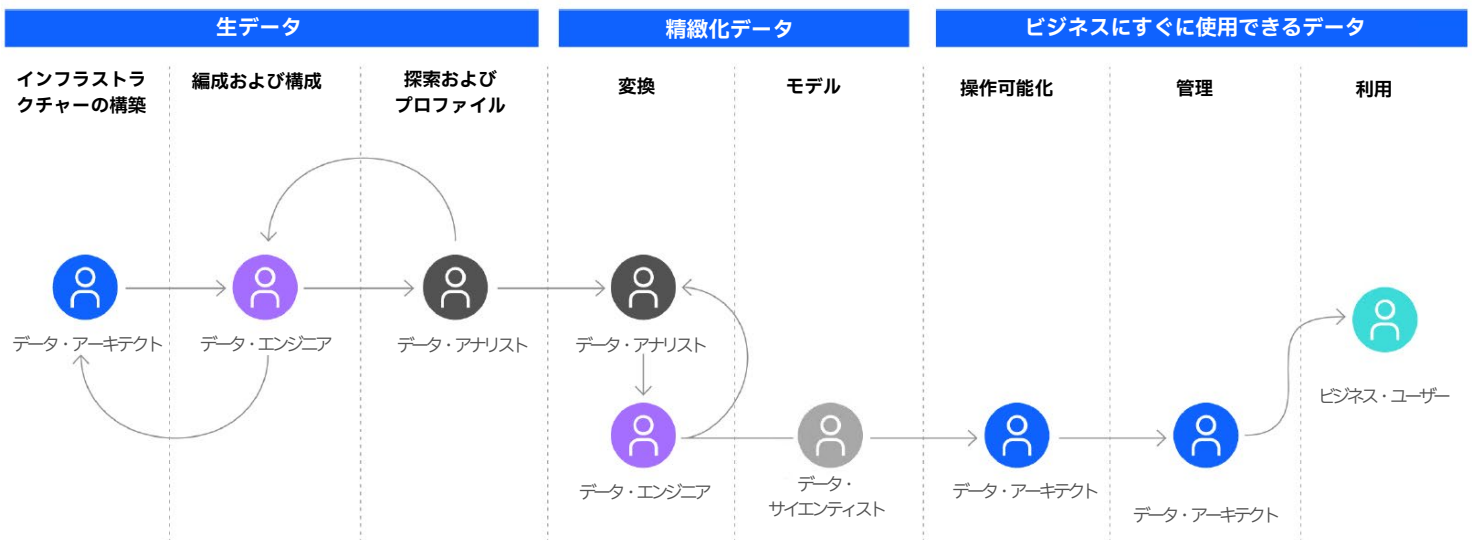


図 1: 役割別 DataOps ワークフローの例

テクノロジー

DataOps の中核となるのは、組織の **情報アーキテクチャー** です。あなたはデータを把握していますか？ データを信頼していますか？ エラーを素早く検出できますか？ データ・パイプライン全体を「中断」することなく、段階的に変更を加えることができますか？ これらの質問に答えるためには、まず、**データ・ガバナンスとデータ統合** のためのツールや実践状況を棚卸しする必要があります。自動化に依存する実践をサポートするためにはツールが必要です。

組織内で DataOps の実践をサポートするツールを検討する際には、次の 5 つの重要領域での自動化によりデータ・パイプラインをどのように変革できるか考えてください。

1. データ・キュレーション・サービス
2. メタデータ管理
3. データ・ガバナンス
4. マスター・データ管理
5. セルフ・サービスでのアクセス

DataOps が提供する機能

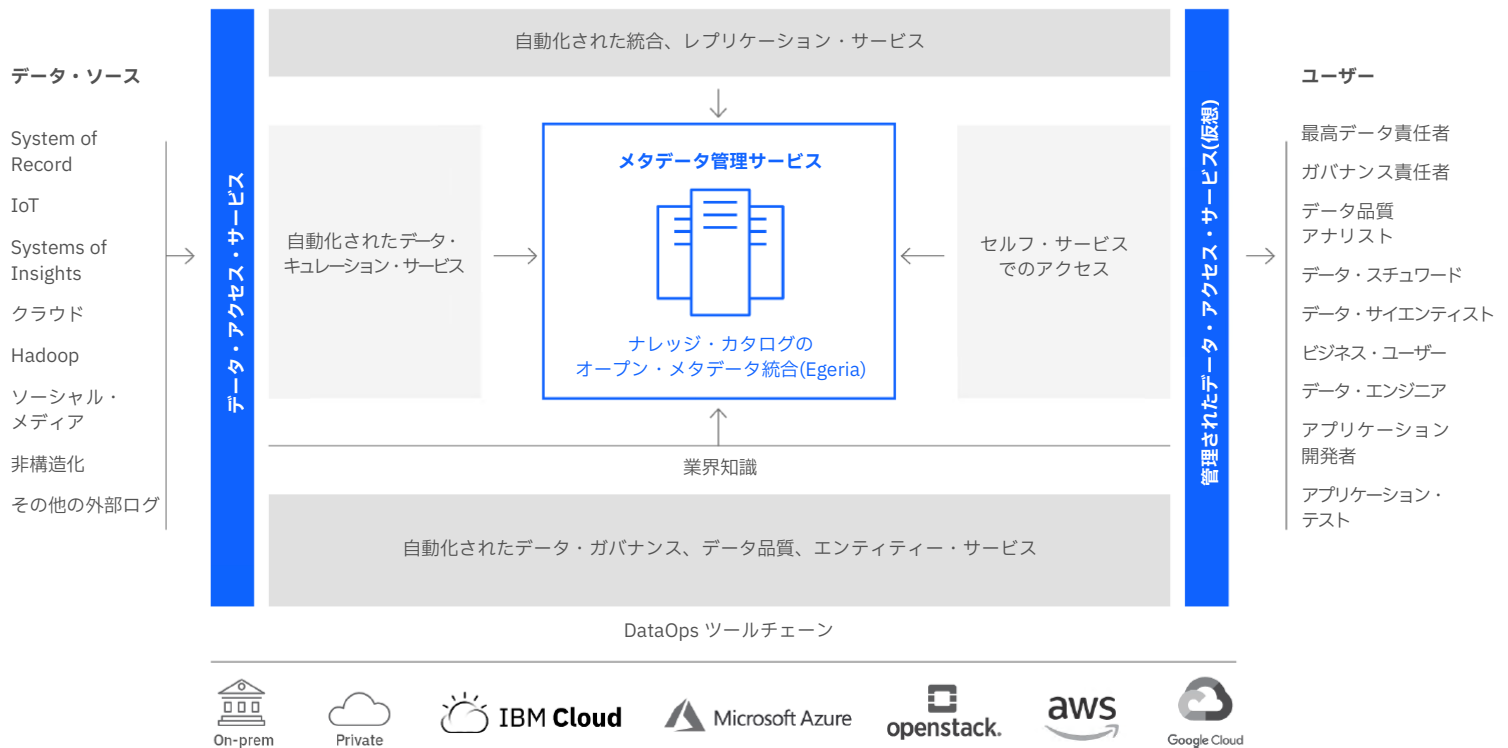


図 2: DataOps をサポートする情報アーキテクチャーの分析

ビジネスにすぐに利用できるデータの提供にはこれら全ての側面が含まれており、DataOps を実践するには、5 つの側面すべてを取り入れた全体的なアプローチが必要です。データ・パイプラインの 1 つの要素に焦点を当てて他をなおざりにする組織は、DataOps の実践から得られるメリットを実現することが難しくなります。テクノロジーの検討と実装を、人やプロセスに関する継続的な計画から切り離して考えるべきではありません。ツールは、文化を支え、維持するのに役立ちます。

結論

データ・スチュワードの役割の定義やデータ検証ルールの作成など、基本的な問題にまだ苦労している組織にとって、DataOps は困難なものに思えるかもしれません。しかし、DataOps の実践は、企業がデジタル変革の取り組みで経験して

きた多くの失敗に対する解決策を提供します。

組織が認識している最も一般的な失敗例は、データレイクに関するものです。多くの組織は、技術的な成功を収めるまでに 2 回、3 回、あるいは 4 回と試行を繰り返し、そのたびにリーダーシップ・チームの活性化を行い、必要な文化的変更を取り入れています。しかし、データレイクの実装は、これまでなぜ失敗してきたのでしょうか？

これらのプロジェクトの多くは、クレンジングやガバナンスが行われていないデータをデータレイクに取り込むことだけにフォーカスしていました。おそらく、失敗の原因は、人、プロセス、テクノロジーの問題に効果的に対応できていなかったことです。

IBM DataOps プログラム

実際、DataOps 採用への転換が起きています。最近の調査によると、[DataOps への投資を計画している企業は 73%に上ります](#)。² IBM は、規範的な方法論、先進的なテクノロジー、IBM DataOps Center of Excellence (CoE)により、DataOps 実戦への道を提供します。CoE では、IBM の専門家が組織と連携して、ビジネス目標に基づいてアプローチをカスタマイズし、関係者に価値をもたらすための適切なパイロット・プロジェクトを特定します。

IBM DataOps の各機能は、AI に対応した自動化、ガバナンスの適用、強力なナレッジ・カタログを実装し、継続的かつ高品質なデータを企業全体で使用できるようにする業界最高レベルのテクノロジーを通して、ビジネスにすぐに利用できるデータを提供できるよう支援します。DataOps は、効率性、データ品質、検索性を向上させ、ガバナンス・ルールを組み込むことで、基本的にあらゆるソースから適切な人に適切なタイミングでデータ・パイプラインを提供するセルフ・サービスを実現します。

データレイクの管理を容易にするソリューションから、アプリケーションの開発や法規制遵守を支援するソリューションまで、IBM DataOps は、意思決定とタイミングの最適化におけるデータの価値を企業に示すことができます。高品質な企業データを提供して AI を実現させることは、手の届くところにあります。ただしそれには、企業がデータを把握し、信頼し、活用して、クラウドや重要な環境で価値を高めることが必要です。

IBM Cloud Garage 方法論における DataOps の活用

IBM Cloud Garage 方法論は、ビジネス、開発、運用が、新たな機能を継続的に設計、提供、検証できるようにするためのアプローチです。その実践、アーキテクチャー、ツールチェーンは、製品の初期段階から、顧客のフィードバックや市場の変化の確認と対応まで、製品のライフサイクル全体をカバーしています。オープン・ツールチェーン・アーキテクチャーは、オープンソースと先進的なサード・パーティー・ツールによる IBM Cloud™ Platform サービス(継続的デリバリー (CD)など)を、DataOps プラクティスに合わせた統合ツールチェーンに容易に構築できるように設計されています。これらのパターンは、組織全体での DataOps の適切な採用を促進するテンプレートとして、チーム間で共有できます。

IBM は、DataOps の実践を成功させるための、DataOps ライフサイクルにおける 6 つのフェーズに加え、必要な文化面の考慮事項を特定しました。これは、IBM の変革のジャーニーの一環として社内で DataOps を採用したことにも基づいています。

IBM Cloud Garage 方法論では、これらの 6 つのフェーズを次のように記述しています。

- **Think:** 機能の概念化、詳細化、優先順位付け
- **Code:** 機能の生成、拡張、最適化、テスト
- **Deliver:** オファリングの自動生産と提供
- **Run:** 実行に必要なサービス、オプション、機能
- **Manage:** オファリングの進行中のモニタリング、サポート、復旧
- **Learn:** 実験の結果に基づいた継続的な学習とフィードバック



図 3. IBM Cloud Garage 方法論の 6 つのフェーズ

Think: DataOps の成熟度を継続的に評価し、ビジネス目標との整合性を図る

DataOps は、既存の組織や確立したプロセスにとっては変革的なものかもしれません。DataOps の目的は、既存の多くの手作業を自動化し、データ・パイプライン作成プロセスを効率化することにあります。基本的な DataOps の実践を始めるにしても、維持するにしても、ビジネスにすぐに利用できるデータを素早く提供し、ビジネス価値の創出に向けた改善計画を立てるチームの能力を評価することが重要です。

DataOps の成功は、スプレッドシートや分断された知識、あるいはハンド・コーディングではなく、データ資産を **カタログ化** することから始まります。これは、メタデータを収集してデータ・クラスにポリシーを割り当てること、データ品質の評価とスコアリングを行うこと、データの統合のためのツールを活用することから始まります。チームの成熟度レベルが定義されたら、できるだけ多くの DataOps の側面で、能力を向上させることを目標とします。

DataOps チームは、ビジネスに価値をもたらすために必要なデータの提供に注力する必要があります。質問してみてください。「この情報をもっと素早く利用できるようになれば、いくらか節約したり、儲けることが出来るでしょうか?」

Code: バージョン管理システムの使用—ソース・コントロール管理

データ・パイプラインは、未加工のコンテンツを有用な情報に変換する役割を担うソース・コードです。このパイプラインは、データ・アナリティクスの中核となるもので、エンドツーエンドで自動化して、再生可能な形式で取り込むことのできるソース・コードを生成することができます。アナリティクスに関連付けられた様々なファイル、構成、パラメーターが、統制管理されることなく組織内のさまざまな場所や環境に分散されると、一貫性のないデプロイメントにつながります。GitHubなどのリビジョン管理ツールは、コードや構成に対するすべての変更を保管・管理するのに役立ちます。また、一元化されたリポジトリは、企業が環境全体で一貫性があり信頼性の高い情報を常に保有すること、また確実なリカバリーを必要とする不測の事態や災害への対応を可能にします。リビジョン・コントロールでは、ブランチとマージを使用することで、チームの開発作業を並列化することもでき、デリバリー・パイプラインのアジャイル化を実現します。

データ・アナリティクス・パイプラインが必ず正しく機能するようにするには、テストの実施が必要です。継続的インテグレーション/継続的開発(CI/CD)をパラメーター化することで、完全に自動化された方法でデプロイとテストを行うことができます。データ・アナリティクス・パイプラインの各段階で入力、出力、ビジネス・ロジックをテストし、その正確性や逸脱の可能性と、エラーや警告を併せてチェックし、一貫した品質を保証してリリースする必要があります。手動によるテストは、エラーが起りやすく、時間と労力を要するため、ハイパフォーマンスの組織では選択されません。CI/CDの達成における主要な要素であり、オンデマンド・エコノミーに不可欠なのは、堅固で自動化されたテスト・スイートです。

Deliver: DataOps プロセスとワークフロー自動化—データ・テクノロジー

DataOps 方法論が成功するためには自動化が必要不可欠であり、実行時の柔軟性を考慮して設計されたデータ・アナリティクス・パイプラインが必要です。信頼できるデータを提供するための重要な要件は、メタデータとデータ・サンプリング技法によるデータ・キュレーション、データ統合、カタログ、データ分類に依存する、管理された一貫性のあるデータ・パイプラインです。

信頼性があり、管理されたデータの提供を実現する、反復可能で堅固なデータ・パイプラインでは、以下のアクティビティを実行するメカニズムが必要です。

- データ・ガバナンスとデータ・プライバシーのポリシーを一貫して定義・実施する。
- 効率的なデータ移動をサポートする。
- 修正に着手する、あるいは業界固有のベスト・プラクティス、および事前定義されたビジネスの用語を含むテンプレートを用意する。

このプロセスでは、各種プラットフォームに、ソース・コードや構成を変更することなく、管理されたデータ・パイプラインを一貫してデプロイしたり、完全に管理された信頼できるデー

タを提供したりすることができます。また、DataOps プロセスでは、例外処理と管理をサポートするため、適切なツールで修正を行う必要があります。変更履歴の追跡と監査能力が、管理されたデータ・パイプラインでは必要不可欠です。

IBM は、組み込みの機械学習(ML)、AI 自動化、適用されたガバナンス、強力なデータ・カタログにより、企業全体にわたって継続的で高品質なデータを使用できるようにする、新しい革新的な機能を提供します。DataOps の効率性は、データ・パイプラインに使用されるデータ・テクノロジー・コンポーネントの高度な自動化により向上します。

- IBM Watson® Knowledge Catalog (WKC) を含み IBM Cloud Pak for Data®は、効率のかつ堅固で、自動化された反復可能な方法でこれらの要件に対応できます。
- IBM Cloud Pak for Data Server は、データ品質とポリシー実施を確保しつつ、データ・パイプライン内でのデータの移動、公開、利用のニーズに対応します。効果的なソース・コントロール管理により、CI/CD パイプライン内から効率的に自動化して実行することができます。
- IBM Cloud Pak for Data 向けの IBM Watson Knowledge Catalog 内の組み込み ML は、自動化プロセスを補完し、反復ごとに最適化して、堅固なパイプラインを実現します。
- IBM Cloud™ DevOps Insights は、データ・パイプラインの運用上の洞察と可視化を提供する上で役立ちます。高度な自動化と、IBM Cloud Pak for Data とのカスタマイズされた統合に基づき、セキュリティと品質の指標の継続的なモニターと強化、予期しない変化の検出、運用上の統計の生成ができます。
- Apache Airflow と NiFi は、ワークフロー設計とそのオーケストレーションに役立ちます。
- REST エンドポイントを使用する高度な自動化と、パラメーター化を併せて利用することで、特定のデータ・セットまたは環境を動的に選択し、パイプラインのソース・コードに影響を与えずに動作を変更し、データ分析の専門家の日々のニーズに対応することができます。

Run: 継続的インテグレーションとデプロイメント継続的インテグレーション

データ・パイプラインのエンジニアやオーナーは、いつでもパイプラインの更新や変更を行うことができ、開発ブランチまたはプライベート・ブランチ内のプライベート・コピーとして、リビジョン・コントロール・システム内にそれを保持することができます。複数のエンジニアが並列して作業し、開発ブランチやプライベート・ブランチを同時に変更することで、多方面で生産性が向上します。パイプラインの変更がブランチ内で完了し、テストされる場合、ソース・コードをメインのコード・ベースかトランクにマージして、本番ラインに提供することができます。マージされたコードが機能しない場合、データ・パイプラインは常に以前のパイプライン・ソース・コードの機能しているバージョンにフォールバックできます。ブランチとマージによって、データ・アナリティクス・チームは、独自のテストを実行し、変更を加え、試してみて、一連の変更が成功しないと分かった場合は破棄することができます。

継続的デプロイメント

データ・アナリティクスの専門家は、ソース・コードのプライベート・コピーとは別に、パイプラインで使用する関連データと、これらのパイプラインを実行するための環境を必要とします。本番のデータベースや環境で直接作業を行うことは効率的ではなく、しばしば矛盾による問題につながります。矛盾や依存関係による問題をなくすためには、データ・パイプラインにおいて以下が必要です。

- 効率的なソース・コントロール管理
- 柔軟な環境でのデプロイメント・オプションの利用
- データの振る舞いのテスト

Jenkins Pipeline は、IBM デリバリー・パイプラインを補完するツールです。Red Hat® OpenShift® は、実行時に提供される各種の値でデータ・パイプラインの同時インスタンスを検証するために使用される、反復可能で一貫性のあるデプロイメント・プラットフォームを提供します。

Manage: 一貫性のある頻繁なデプロイメントの処理

データ・アナリティクスの専門家は、現行のデータ・パイプラインを中断するような変更を本番環境に導入することは避けたいと考えています。このような状況では、次の 2 つの重要なワークフローで対応できます。

- **バリュー・パイプライン:** 本番環境へデータ・フローを取り込むことにより、今すぐ組織の価値を創出します。
- **イノベーション・パイプライン:** 開発を経て、現行パイプラインに取り込むことにより、新たなアナリティクスの将来につなげます。

この 2 つのパイプラインは、DataOps 組織がデータのオーケストレーションと、完璧な品質を維持しながら新しい機能のデプロイメントを習得することにより、本番環境で交わります。データと新しい開発パイプラインをモニターする統計プロセス制御などのデータ・パイプラインの品質コントロールにより、開発チームは本番環境のシステムの中断を心配することなくデプロイできます。アジャイル開発と DevOps によって、新しいアナリティクスは最大限に高速化されます。ビジネス・ニーズを分析的なアイデアに変えるための時間と労力を最小限に抑え、反復可能で再現可能なプロセスとしてリリースすることができます。

Learn: コミュニケーションとプロセス管理

効率的で自動化された通知は、DataOps 方法論におけるコミュニケーションと修正プロセスの中核を成します。任意のソース・コードに変更が行われた場合、あるいはパイプラインが起動、失敗、完了、デプロイされた場合に、通知することができます。障害が発生した場合、通知と共に、その問題に対処するための情報をプッシュできます。

修正後のプロセスは自動的に起動し、パイプラインの検証、次のステージへのデプロイ、最新の情報とデータ品質によるダッシュボードの更新が可能です。関係者間のコミュニケーション、コラボレーション、フィードバックの収集と共有を促進するために、DataOps ツールチェーンの一部として、一般的に、Slack、Apache Kafka、PagerDuty、Trello のようなツールが使用されます。

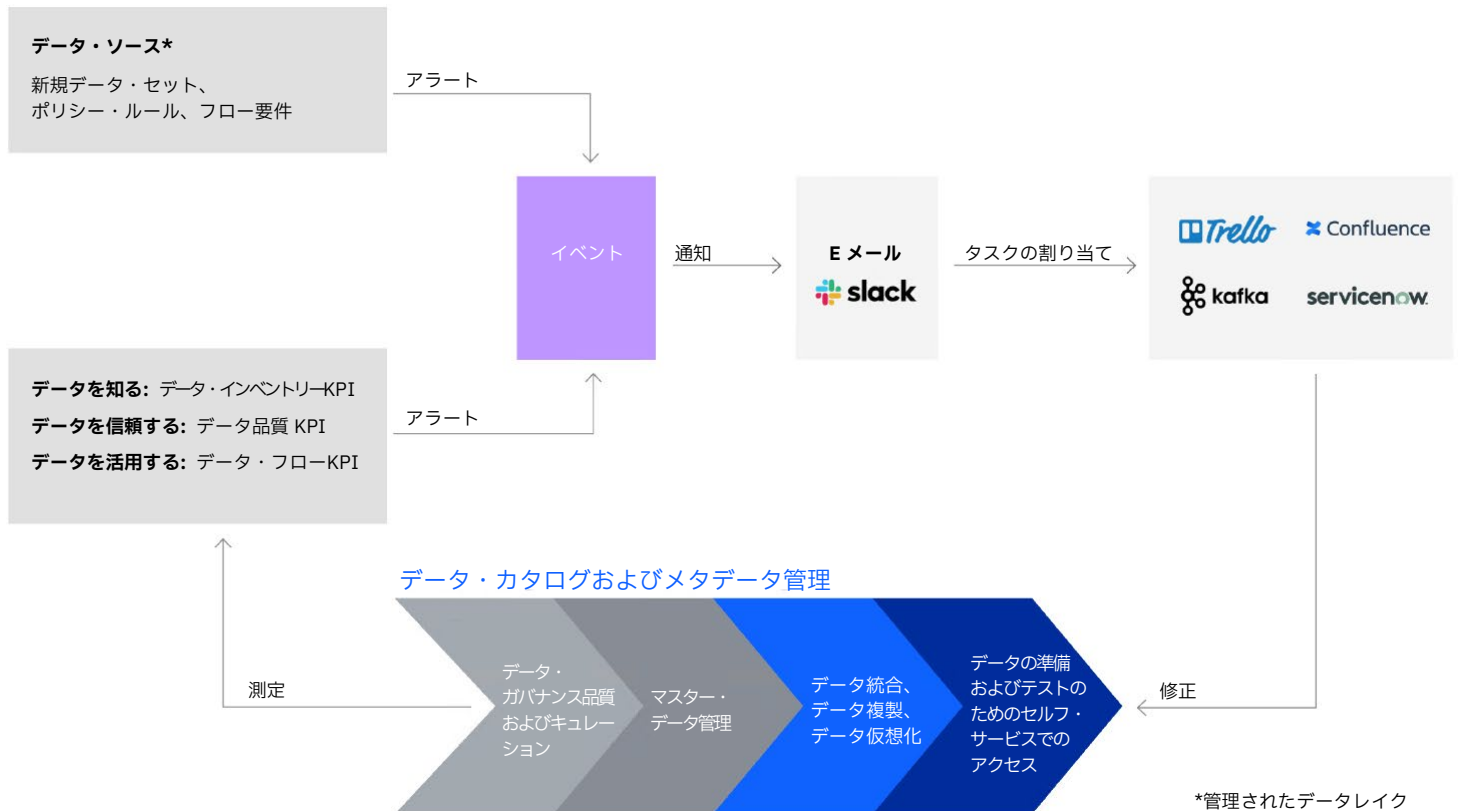


図 4: 管理されたデータレイク環境内でのコミュニケーションとプロセス管理の視覚化

優れた DataOps の実践による効果

ある小売業者は、DataOps を実践することで、データ・パイプライン全体を改善し、従来 3 週間かかっていたデータ変更を 2 分以内で組織全体に適用できるようになりました。その結果、ビジネスにすぐに利用できるデータを活用して、従来 20 日かかっていた顧客の親和性分析を 1 日以下で行い、さらに、在庫状況の報告に必要な時間を従来の 6 分の 1 に短縮することが可能になりました。

優れた DataOps の実践の基準には、以下が含まれます。

1. **データ・オフィスの確立:** このプロセスには、利益につながるリソースとしてのデータを提供する役割の範囲の定義、主要経営層の利害関係者の特定、データ・パイプラインのすべての利害関係者のコラボレーションによる運用と文化に対するコミットメントの理解などがあります。
2. **ビジネス目標との整合:** 競争力を維持するには、市場における新たな機会に迅速に対応する必要がありますが、それには十分な情報とデータ主導のアプローチが不可欠です。リーダーは、ビジネスとデータ提供の間に優れたコミュニケーションがなければ組織は生き残れないということを認識する必要があります。
3. **データの成功を拡大:** データを活用した取り組みでは、リーダーは、生成されたデータを継続的に利用/再利用するたびに価値を高めていく必要があります。これは、データが一元的に共有され、検索可能であり、ビジネス用語に沿ったものでなければ達成できません。



結論

DataOps の導入に成功した組織は、どのようなデータ資産にアクセスできるかを把握し、データの意味とその品質を信頼し、データの可能性を最大限に活用しています。信頼性の高いビジネスにすぐに利用できるデータが、差別化された洞察力、卓越した運用、コラボレーション、競争上の優位性の促進に役立つとき、データは価値を持ちます。

DataOps 実践の確立には以下が必要です。

- *パイロット・プロジェクトを実施することで、組織固有の能力や課題を理解するための投資を行う*
- *パイロット・プロジェクトの成功をもとに、DataOps のスキルと組織の拡張と育成を図る*
- *成功をプロモートし、DataOps プラクティスに参加するチームをさらに募集する*
- *教訓を共有し、DataOps CoE の構築を開始する*

データ・カタログ、データレイク、またはマスター・データなどの取り組みを既に行なっている組織は、DataOps の採用に適していると考えられます。市場をリードするテクノロジーによる DataOps サポートについて詳しくは、ibm.com/DataOps をご覧ください。



© Copyright IBM Corporation 2021
〒103-8510 東京都中央区日本橋箱崎町 19 番 21 号
日本アイ・ビー・エム株式会社

法務・知的財産 知的財産権ライセンス渉外
2021 年 9 月

IBM、IBM ロゴ、ibm.com、IBM Cloud、IBM Cloud Garage、IBM Watson、および InfoSphere は、世界の多くの国で登録された International Business Machines Corporation 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、<http://www.ibm.com/legal/copytrade.shtml> をご覧ください。

Microsoft、Windows、Windows NT および Windows ロゴは、Microsoft Corporation の米国およびその他の国における商標です。

本書の情報は最初の発行日の時点で得られるものであり、予告なしに変更される場合があります。すべての製品が、IBM が営業を行っているすべての国において利用可能なものではありません。

記載されているお客様事例は、例として示す目的でのみ提供されています。実際の結果は特定の構成や稼働条件によって異なります。本書に掲載されている情報は特定物として現存するままの状態を提供され、第三者の権利の不侵害の保証、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任なしで提供されています。IBM 製品は、IBM 所定の契約書の条項に基づき保証されます。

お客様は自己の責任で関連法規を遵守しなければならないものとします。IBM は法律上の助言を提供することはいたしません。また、IBM のサービスまたは製品が、お客様がいかなる法規も遵守されていることの裏付けとなると表明するものでも、保証するものでもありません。

適切なセキュリティの実施について: IT システム・セキュリティには、企業内外からの不正アクセスの防止、検出、および対応によって、システムや情報を保護することが求められます。不正アクセスにより、情報の改ざん、破壊、悪用もしくは誤用を招くおそれがあり、またはシステムの損傷や、他のシステムへの攻撃での使用を含む悪用につながるおそれがあります。完全に安全と見なすことができる IT システムまたは IT 製品は存在せず、また単一の製品、サービスまたはセキュリティ対策が、不正使用や不正アクセスを防止する上で、完全に有効となることもありません。IBM のシステム、製品およびサービスは、合法的で包括的なセキュリティの取り組みの一部となるように設計されており、これらには必ず追加の運用手順が伴います。また、最高の効果を得るために、他のシステム、製品、またはサービスを必要とする場合があります。IBM は、何者かの悪意のある行為または違法行為によって、システム、製品、またはサービスのいずれも影響を受けないこと、またはお客様の企業がそれらの行為によって影響を受けないことを保証するものではありません。

- 1 「2019 Global data management research: Taking control in the digital age」 Experian, 2019.
- 2 Jarah Euston, 「The DataOps Trend is Real: 73% of Companies Plan to Invest in DataOps to Manage Data Teams in 2018」, Nexla

10028810JAJP-01

付録: DataOps のパイロット・プログラム・テンプレート

プロジェクト名:

日付:

部門名:

パイロット・プログラムのリーダー:

氏名	役割	Eメール	電話

ビジネス全体の利害関係者:

氏名	役割	Eメール	電話

問題の記述:

根本原因のチェックリスト:

課題	該当するかどうか? Y/N	追記

成功のための指標:

開始日:

スプリント終了日:

評価:

実装	現在の状態	パイロットで期待する成果	期待する成果を達成するためのアクションのステップ
データ資産の取り込み、自動検出、および分類			
データ品質の評価と修正			
ビジネス用語の割り当て			
データ・プライバシー、法規制遵守、企業ポリシーの定義と実施			
データ利用者の要件定義の要求処理			
データ要求のコミュニケーションと通知(例外とエラーの処理および修正を含む)			
キュレート済みのデータのカタログへの公開			
データ・リネージュ(来歴)および報告			
コラボレーション、フィードバック、監査			

実装監査で受ける質問の例:

データ資産の取り込み、自動検出および分類

- 事業継続のために大容量、低遅延の複製を行っているか?
- リアルタイムで低遅延の分析を行うために高度なストリーミング・アナリティクスを使用しているか?
- あらゆるデータ・ソースに容易にアクセスできて、複雑なデータの変換と統合を実行できるか?
- ソーシャル・メディア、気象データ、その他パブリック・クラウドのデータ・ソースからデータを提供できるか?
- データ利用者はあらゆるデスクトップ・アプリケーションから社内のメタデータ・リポジトリにリアルタイムでアクセスできるか?
- データ利用者が自分の業務に関連するデータ・セットを見つける際に、社内のデータ・カタログにリアルタイムにセルフ・サービスでアクセスしたり、サポートを受けることができるか?
- データの把握、データの値、カラム、テーブルの関係の検証、異常の検出と分析にデータ・プロファイリング・ツールを使用しているか?
- ビジネス・ルールの管理がメタデータ管理インフラストラクチャーと統合されているか?