

データ・ファブリック・ アーキテクチャー からすぐに得られる 3つのメリット

社内のあらゆるデータを統合・活用して、
ビジネスの成果を向上

目次

- 3 データに関する現在の課題
- 5 データ・ファブリックがもたらす
ビジネス上のメリット
- 6 IBM のデータ・ファブリックの特徴
- 7 データ・ファブリックの活用
- 7 企業に変革をもたらす
データ・ファブリック

データに関する現在の課題

データは、企業のDX(デジタル・トランスフォーメーション)推進に欠かせない要素です。しかし、企業がデータを活用しようとする、データのソースや種類、構造、環境、プラットフォームが多様なために生じる課題に直面します。この幅広く奥深いデータの問題は、組織がハイブリッド・アーキテクチャーやマルチクラウド・アーキテクチャーを採用すると、さらに複雑になります。現在、多くの企業では、業務データはサイロ化されて可視化されておらず、膨大な量のダーク・データとして有効活用されていない状況が多発しています。

北米のあるエネルギー企業を例にとりましょう。同社は、データ・ドリブン型に組織を再構築し、複数の部門でデータサイエンスを容易に活用できる企業になることを目指していました。しかし同社はすぐに、サイロ化されたデータ、一貫性のないツール、異なるスキル・レベルなどの課題が発生し、当初予定との重大なギャップによりDXが妨げられていることに気が付きました。

同社が直面した問題は、同社固有のものではありません。それは即ち、データ管理アーキテクチャーのレベルの問題ではなく、多くの企業で見られるようにデータ環境全般に関わる問題でした。

最終的にこのエネルギー会社をDX成功の道へと導いたのは、データ・ファブリックと呼ばれる、新しいデータ・アーキテクチャーのコンセプトの採用でした。

データ・ファブリックとは一体どのようなものであり、これまでのアーキテクチャーと何が違い、ビジネスに何をもたらすものなのか、そしてデータ・ファブリック導入におけるIBM が果たす役割とはどのようなものなのか？ このホワイトペーパーを通して、これらの疑問にお答えします。

データ・ファブリックとは？

これまで、組織はデータ・アクセスの問題に対処するための方法として、ポイント・ツー・ポイントの統合やデータ・ハブの導入を試みてきました。しかし、データが高度に分散し、サイロ化している場合には、どちらも適切な対処方法であるとは言えません。ポイント・ツー・ポイント統合は、接続が必要なエンドポイントが増える度に、コストが指数関数的に増加するため、拡張性に欠けるアプローチとなります。データハブは、アプリケーションやソースの統合を容易にしますが、ハブ内のデータの品質と信頼性を維持するためのコストと複雑さは逆に増大します。

データ・ファブリックは、ハイブリッドなデータ環境から生じるデータの課題に対処するための新しいアーキテクチャーです。データ・ファブリックの基本的な考え方は、データのエンドポイント間の仮想的な結合組織として機能することで、分散化とグローバル化のバランスを取ることで(4ページ参照)。

データ・ファブリック・アーキテクチャーは、統合の自動化と拡張、フェデレーテッド・ガバナンス、メタデータの有効化などのテクノロジーにより、分散化された環境全体でデータのダイナミックかつインテリジェントなオーケストレーションを可能にし、ビジネスに必要な情報を即座に入手できるネットワークを構築します。

データ・ファブリックは、展開するプラットフォーム、データ処理、地理的な場所、アーキテクチャーの方法に左右されることはありません。これにより、企業資産としてのデータを活用しやすくなります。データ・ファブリックによって、企業はさまざまな種類のデータを効率的かつ効果的に組み合わせ、アクセスして、管理できるようになります。

データ・ファブリックの機能と原則

データ・ファブリックの中核となるアーキテクチャーは、ディスカバリー、ガバナンス、キュレーション、オーケストレーションなど、幅広い統合データ管理機能を可能にするデータ管理プラットフォームです。

ただし、データ・ファブリックは、プラクティスの確立にのみ焦点を当てたDataOpsのような従来のデータ管理の概念から進化・発展を遂げており、データの運用レベルを向上させるものになります。データ・ファブリックは、データ資産の極端な多様性と分散から生じるニーズに対応可能な分散型のアーキテクチャーと先進的なテクノロジーで構築されています。

データ・ファブリックは、論理的に4つの機能(またはコンポーネント)に分けることができます。

ナレッジ、洞察、セマンティック

- データ・マーケットプレイスで買物をするようにデータを選ぶ体験を提供します
- ナレッジとセマンティックにより、発見されたデータ資産を自動的に強化し、データ利用者がデータを検索し理解できるようにします

ガバナンスとコンプライアンスの一元化

- メタデータのローカル管理とガバナンスを可能にする一方、統一されたグローバル・ビューとポリシーの実行をサポートします
- グローバルおよびローカル・ルールに基づいて、データ資産にポリシーを自動的に適用します
- 高度な機能を活用して、データ資産の分類とキュレーションを自動化します
- データの有効化を強化するために、カタログ化された資産に対して照会可能なアクセスルートを実験的に確立します

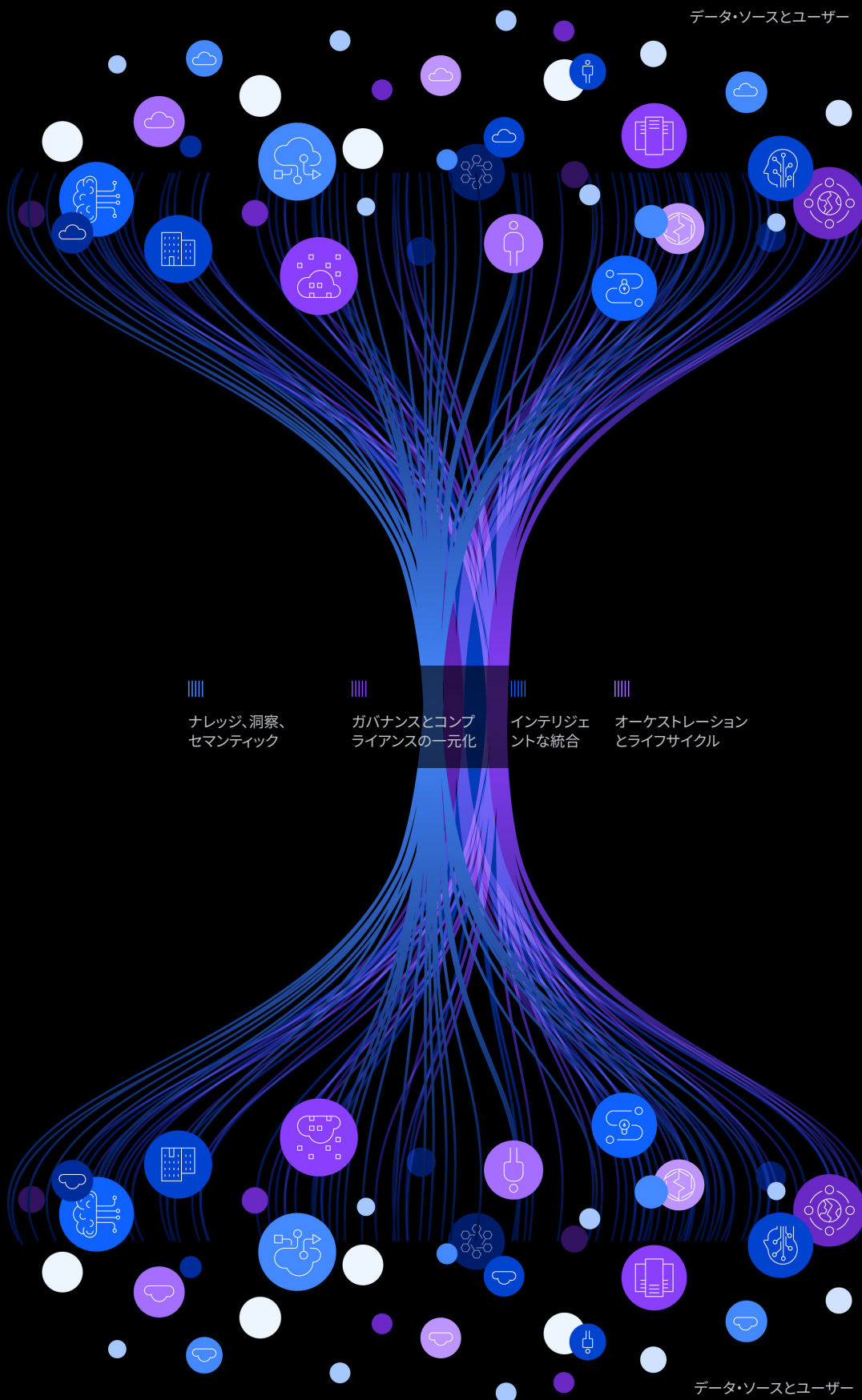
インテリジェントな統合

- 分散したデータソース全体に対応したフローとパイプラインの自動生成により、データ・エンジニアの業務を高速化します
- データ保護ポリシーをローカルおよびグローバルに徹底することで、あらゆるデータの取り込みとデータ・アクセスのセルフサービス化を実現します
- ワークロード配分の最適化、スキーマ・ドリフトのセルフチューニングと修正により、最適な実行を自動的に判断します

オーケストレーションとライフサイクル

- データ・パイプラインの構成、テスト、運用、モニターを可能にします
- データ・ライフサイクルにAIの機能を加えることで、タスクの自動化、自己調整、自己修復、ソース・データの変更の検出などを行い、自動更新を促進します。

データ・ソースとユーザー



データ・ファブリックがもたらす ビジネス上のメリット

データは、コンテキスト化され、組織内のどのユーザーやアプリケーションからもアクセスできるようになって初めて、ビジネス上の価値を発揮します。データ・ファブリックが正しく実装されていれば、最も効率的かつ自動化された方法で、組織全体でこれらの価値を利用できるようになります。これにより、データ・ファブリックは、主に3つのメリットを提供します。

1. セルフサービスでのデータ利用やコラボレーションの実現
2. アクティブなメタデータによる、ガバナンス、保護、セキュリティの自動化
3. データ・エンジニアリング・タスクの自動化と、ハイブリッドクラウド環境全体を通じたデータ統合の推進

セルフサービスでのデータ利用とコラボレーションの実現

複数のソースからのデータを統合し、日々生成される膨大な量のデータをより多く分析することで、組織はより良い洞察を得て、変化するビジネスの要求に迅速に対応することができるようになります。データ・ファブリックは、データを必要とする人に素早く届けることができます。セルフサービスにより、組織全体で適切なデータを素早く入手し、そのデータを使って具体的な洞察を得るためにより多くの時間を費やすことができます。

セルフサービスでのデータ利用を可能にするデータ・ファブリックのメリット

- ビジネス・ユーザーは、単一のアクセスポイントから、組織全体のデータの検索、理解、作成、利用が可能です。
- データ・ガバナンスとリネージュを一元化することで、ユーザーはデータの意味、データ・ソース、他の資産との関連を理解することが容易になります。
- 広範でカスタマイズ可能なメタデータ管理は、簡単に拡張でき、API を介してアクセス可能です。
- 信頼性が高く管理されたデータにセルフサービスでアクセスすることで、他のユーザーとの業務上のコラボレーションが可能になります。

Forrester 社の Total Economic Impact 調査¹ は、これらの機能が意味することを明らかにしています。

580万 459%

米ドルの利益

ROI

アクティブなメタデータによる、ガバナンス、保護、セキュリティの自動化

すべてのデータ関連イニシアチブに対応する分散型のアクティブなガバナンス・レイヤーが、信頼性と透明性を提供することで、コンプライアンスと規制のリスクを低減します。あらゆるデータ・アクセスに自動でポリシーを施行でき、レベルの高いデータ保護とコンプライアンス要件への準拠を実現します。

AIや機械学習の技術を活用することで、データ・ファブリックのユーザーは、規制文書の文言や定義に基づいてデータ・ガバナンスのルールを自動的に抽出するなど、自動化のレベルを高めることができます。これにより、企業は、ほんの数分で業界固有のガバナンス・ルールを適用ことができ、高額な罰金の支払いを回避し、データの格納先に関わらず倫理的な利用を保証することができます。

データ・ファブリックによる管理された仮想化へのメリット:

- データ・エンジニア、データ・サイエンティスト、ビジネス・アナリストの俊敏性、セキュリティ、生産性が向上します。
- 複数のグローバル・データ・ソースが、1つのデータベースとして表示されます。
- 個人情報 (PII) や重要なデータエレメントについて、業界最高レベルの新たな発見を大規模に行うことができます。

これらの機能は、次のことを意味します。

240万 430%

米ドルの利益¹

効率化²

データ・エンジニアリングのタスクの自動化と、データ統合の強化

高度なデータ・エンジニアリングとは、ほぼ全てのデータ・アクセスや配信プロセスが自動化され、面倒でエラーが発生しやすいコーディング・プロセスを必要としないことを意味します。統合の強化は、メタデータ情報を利用し、データの配信とアクセスを最適化します。

データ・ファブリックによるデータ・エンジニアリングと統合へのメリット:

- 自動的に最適化されたデータ統合により、データの提供が加速されます。
- 自動ワークロード・バランシングと柔軟なスケーリングにより、ジョブは、あらゆる環境、あらゆるデータ量に対応できます。
- レジリエンシーと CI / CD 自動化が組み込まれています。
- 変化をリアルタイムに捉える自動化されたプロセスが、ビジネス・プロセスのための高い品質のデータ提供をサポートします。
- 機械学習が、カスタム・データの発見、分類、キュレーションのプロセスを自動化し、拡張することで、成果が出るまでの時間を短縮できます。
- 継続的な分析を、データが格納される場所であればどこでも、自動的に実行できます。

ある大手小売業者では³以下の効果を実現しています。

60倍 20倍

データデリバリー時間の短縮

カスタマー・アフィニティ分析の高速化

IBM によるデータ・ファブリックの提供方法

IBM Cloud Pak® for Data が、このデータ・ファブリックの概念を可能にします。IBM Cloud Pak for Data は、ビジネス全体にAIを導入するのに最適なプラットフォームです。

データの収集、整理、分析を簡素化して、自動化しAI活用に必要な洞察を導くことができます。

あらゆる場所でデータを接続してワークロードを実行し、ハイブリッドクラウド環境でAIを大規模に構築、展開、管理する機能を備えたIBM Cloud Pak for Data は、ビジネスのデジタル・トランスフォーメーションを実現します。

このプラットフォームは、ハイブリッド・なエンタープライズ全体での、シームレスな統合を実施します。

- IBM Cloud Pak for Data サービス
- 外部アプリケーションとデータ・ソース
- データ管理とガバナンスのための高度なAI ベースの機能

この基盤は、コスト、パフォーマンス、コンプライアンスの最適なバランスを保ちつつ、ワークロード、データのローカリティ、データポリシーに基づくデータ処理を調整し最適化するインテリジェンスを備え、厳選されたデータを利用者に提供します。

IBM Cloud Pak for Data は、以下の機能を活用して、データ・ファブリックが求める、ビジネスに役立つデータを提供します。これらの機能はすべて、データ・ファブリック・アーキテクチャーを支える役割を担っています。

1. メタデータベースのナレッジ・コア

データソースやカタログの分析を容易にし、データ資産を充実させ、AIを使って、自動化を推進するための洞察を抽出するための分析を行います。ナレッジ・コアは、セマンティック検索でマーケットプレイスを強化す

るために使用されます。

2. セルフサービス・データ・マーケットプレイス

ビジネス・アナリストなどのデータ利用者が、企業内のデータ環境からデータを検索できるようにするための次世代データ・カタログです。

3. スマートな統合

データの抽出、仮想化、変換、ストリーミングによるデータ活用を可能にします。データ統合を自動化するために、ナレッジ・コアと統合されており、ワークロードとデータポリシーに応じて最適な統合方法を判断するインテリジェンスを備えています。また、データ・エンジニアリングの作業の一環として、またはデータの準備やデータ製品の作成にも使用できます。最後に、データ製品のアップデートを公開する機能も備えています。

4. ガバナンス

メタデータのカタログ化とキュレーション、プライバシー用のデータ・ポリシーの定義、データのキュレーション、データ・リネージュの取得、セキュリティとコンプライアンスに関連するその他のタスクを実行します。この層は、データの形式（構造化と非構造化）やデータの重要性（公開されているか保護されているかなど）を理解し、データと可能性のあるユーザーそれぞれに適切なポリシーを適用します。データに基準やルールを手動で適用するのではなく、この統合プラットフォーム機能は、組織レベルで適用され、必要に応じてさまざまなデータリソース全体に伝播できます。異なるツールの分析モデルが相互に連携でき、緻密なレベルでのデータ・ポリシーの実行は、ほぼ自動化されています。

5. 統合された開発と運用

本番環境のデータ・プラットフォームのすべての側面を構成し、運用する







データ利用者 / ビジネス・プロセス					
統一されたライフサイクル - 設計 - 構成 - 構築 - テスト - オーケストレーション - 展開 - 運用 - モニター	セルフサービスのデータ利用 閲覧、コラボレーション、データ・マーケットプレイス、プロビジョニング	統合されたデータ・ガバナンス、セキュリティ、コンプライアンス、財務ガバナンス - データ・ポリシー - データ・リネージュ - データ品質 - データ・プライバシー - コスト最適化			
	スマートな統合 データの取り込み、変換、準備、仮想化、リアルタイム・データ				
	拡張されたナレッジ ディスカバリー、カタログ、セマンティック、オントロジー、グラフ、メタデータ、マスター・データ、参照データ、イベント・ドリブン・データ				
IBMパブリック・クラウド 	AWS 	Microsoft 	Google 	エッジ 	Private 

図 1. IBM Cloud Pak for Data プラットフォームの機能
これらはすべて、AI開発とデータ・ファブリックをサポートしています。

ための統一されたライフサイクルを実現します。

データ・ファブリックの活用

データ・ファブリックの価値を理解するには、分析プラットフォームにデータ・ファブリックを実装した企業の実例を見るのが確実です。

エネルギー企業

本ホワイトペーパーの冒頭の「課題」のセクションで、デジタル・トランスフォーメーションを進めている北米のエネルギー会社の事例を紹介しました。

このお客様は IBM Cloud Pak for Data をベースにしたデータ・ファブリック・アーキテクチャーを実装する IBM と提携されました。

柔軟で統合されたデータ・ファブリックを導入したことで、同社は部門の枠を超えて、以下の幅広く重要なデータプロジェクトを実施することができました。

- eMobility
- ガス部門における運用文書の分析 (手書きの抽出を含む)
- 電気部門の顧客のセグメンテーションと負荷予測
- 資産管理
- 新型コロナウイルス感染症による負荷の影響
- 職場復帰に関するリスク・モデル

データ・ファブリックで同社が実現したこと

- セルフサービスで使える洞察獲得用のプラットフォームを通じた、複数部門によるデータへの直接アクセス
- 数十億行のデータに対して効率的なモデルを実行するために、目的に応じた計算能力を活用
- モデルの共同開発と、それらのモデルを簡単に展開して、会社全体に洞察を浸透

移動を抑え監視体制を強化

大手製造業サービス・プロバイダーは、データ・ガバナンスを改善し、規制遵守をさらに進めました。

この企業は、大量のデータをクラウド・ベースのデータレイクに移すことに苦労していました。同社には、データアクセスのすべてのインスタンスに強力なガバナンス層を適用すること、またビジネス・ユーザーにアクセスを提供する前にデータ品質を判断するという2つの要件がありました。

この企業は、IBM Cloud Pak for Data上に構築されたデータ・ファブリックを導入することで、データ・ガバナンス、データ準拠、データ変換プロセスを大幅に改善しました。データ・ファブリックを導入したことで、クラウド・ベースのデータレイクやオンプレミスの SAP データ・ソースにある何百ものデータソースに、ユーザーが簡単かつ安全にアクセスできるようになりました。

また、業界の規制確認ツールを活用して、データ・ソースをスキャンし、GDPR や CCPA の規制対象となる可能性のある個人情報データを探しています。自動データ・フラグ機能により、オンプレミスとクラウドの両方で、個人情報の取り扱い管理が必要なデータ・ソースを特定することができます。

データへのアクセスが改善され、個人情報のスクリーニングが迅速に行われるようになったことで、ユーザーは、データへのアクセスに時間がかかったり、保護されたデータが公開されるリスクを負うことなく、重要な洞察を得るために関連データのマイニングを行うことができます。

データ・ファブリックを用いることで、この企業が達成したこと

- 全社のデータを活用しながら、適切なデータ・ガバナンスを確保
- 信頼できるデータを取得し、準備するデータの量を削減
- GDPR や CCPA などのプライバシー規制への準拠を支援

データ・ファブリックで企業変革の次の一歩を踏み出す

IBM Cloud Pak for Data テクノロジーに基づいて構築されたデータ・ファブリックを使うことで、ハイブリッドおよびマルチクラウドのデータ環境で、データ・ディスカバリー、データ・ガバナンス、データの利用を高度に自動化することができます。データ・ファブリックを採用することで、ビジネス・ユーザーの価値実現までの時間を短縮し、データ・エンジニアリングと運用の生産性を向上させ、ガバナンスとコンプライアンス要件への対応の精度を高めることができます。

まずは、IBM Cloud Pak for Data を無料でお試ください。

ビジネスでのAI 活用のご相談がございましたら、お気軽にお問い合わせください。



© Copyright IBM Corporation 2021

日本アイ・ビー・エム株式会社
〒103-8510
東京都中央区日本橋箱崎町19-21
Produced in Japan
2021年8月

IBM、IBMロゴ、IBM Cloud Pak は、米国やその他の国における International Business Machines Corporation の登録商標です。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/trademark をご覧ください。

本資料は最初の発行日時点における最新情報を記載しており、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。

記載されている性能データとお客様事例は、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。IBM 以外の製品とプログラムの操作またはサービスの評価および検証は、お客様の責任で行っていただきます。本書に掲載されている情報は現状のまま提供され、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。

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、破壊もしくは悪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、また単一の製品またはセキュリティ対策が、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 のシステムおよび製品は、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。

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し、また、IBM のサービスまたは製品が、お客様においていかなる法を遵守していることの裏付けとなることを表明し、保証するものでもありません。

- 1 A Forrester New Technology: Projected Total Economic Impact™ IBM からの委託調査、2020年、2月
- 2 効率測定値は、IBM データ仮想化を用いて、各種の 100TB のデータ・ソースに対して、IBM シリコンバレー・ラボにて、管理されたテスト環境の中で収集されました。2019年5月に取得された測定値と効率の向上は、IBM フェデレーションと比較されています。
- 3 IBM お客様による体験に基づいています。