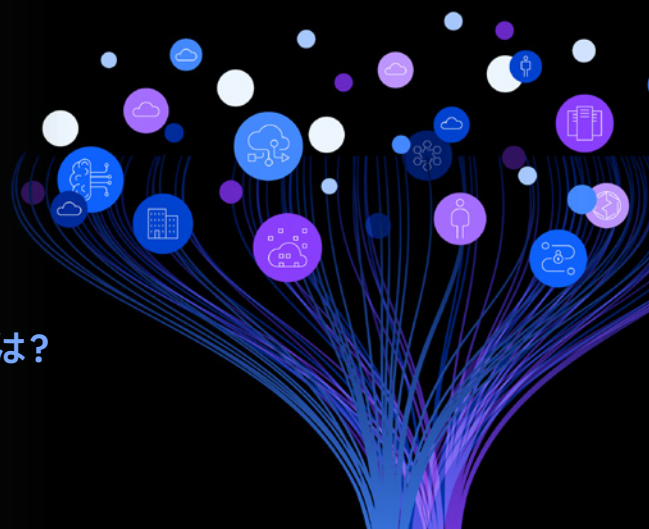
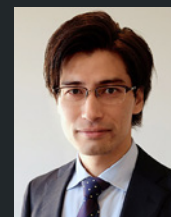


データ活用の 課題を解決する

IBMのインテリジェント・データ・ファブリックとは？



多くの企業がデータ活用を加速させ、データ・ドリブン(駆動型)な経営を推進しようとしている今、データ活用の環境はますます複雑化しており、これまでのアプローチは限界に直面していると言われていています。2019年頃からさまざまな環境に分散して存在する大規模なデータの課題に対処するための新しいコンセプトおよびアーキテクチャーとして、データ・ファブリックが提唱されるようになりました。企業はどのようにデータ・ファブリックに取り組み、データ活用をさらに推進することができるのでしょうか。日本アイ・ビー・エムでData and AIエバンジェリストを務める田中 孝にデータ・ファブリックの考え方および実現方法について話を聞きました。



ますます複雑化するデータ活用環境

ビジネスにおいて優れたパフォーマンスを発揮している企業は、データとAIを活用したビジネス変革を推進しています。しかし、データに基づく客観性の高い判断がビジネスに求められている一方で、データ活用環境の複雑化に伴い、さまざまな課題が出てきています。

「第一に、クラウド・アプリケーションの増加により、多くの企業で、ビジネスに使用するデータがオンプレミス、マネージドクラウド、パブリッククラウドなどに広く分散していること。次に、データ活用に向けたデータの移動や一元化は手作業に頼ることが多く、時間やコストがかかること。さらにはビジネスで扱うデータはどんどん増大しており、かつライフサイクルは短くなっているために、この状況は悪化し続けているということ」と日本アイ・ビー・エム株式会社 テクノロジー事業本部 Data and AIエバンジェリストの田中 孝は

データ活用環境の複雑化によって生じる課題を挙げます。「その結果多くの企業が『データ・ドリブンな経営を推進したいのに必要なデータがなかなか入手できない』という状況に陥っています」(田中)

データを活用するためにデータ基盤を整備する取り組みは、1990年代のデータ・ウェアハウスにさかのぼると言われています。「データ・ウェアハウスはデータの均質性、正確性が高く、構造化データを一元的に蓄積し分析する集中アプローチです。2010年頃から、非構造化データも含めた大量のデータを元の形式のまま蓄積し、分析するデータレイクがトレンドになってきました。データレイクは多様かつ大規模なデータをロー・データのまま扱うという点でデータ・ウェアハウスとは特性が異なりますが、集中アプローチという点はデータ・ウェアハウスと共通です」(田中)



図1 時代とともに変わるデータ基盤の要件

	1990	2010	2019
	データ・ウェアハウス	データレイク	データ・ファブリック
特性	均質性が高い	統合的で均質性が高い	疎結合、異種混在
データ量	中程度	大規模	超大規模
アーキテクチャー	集中	集中	分散
環境	オンプレミス、単一クラウド	オンプレミス、単一クラウド	ハイブリッド/マルチクラウド、オンプレミス
データの正確性	高(信頼できる)	低くなりがち(ガバナンス要)	低くなりがち(ガバナンス要)

しかし、データレイクに取り組んでいる企業では、「一部の部門からしかデータが集まらない」「複数部門がデータレイクを構築し、結局データが分散している」といった課題がでてきました。

そこで生まれたのがデータ・ファブリックです。ハイブリッド/マルチクラウドに分散して存在する超大規模なデータを、分散を前提としながら統合的に管理するアプローチです。「データ・エンドポイントに対する仮想的な接続を提供し、大量のデータ・コピーを回避することやデータに対するポリ

シー適用を自動化し、データ品質の向上を実現することなどがデータ・ファブリックの主な特徴です。データ・ファブリックのアーキテクチャーを適用することによって、より多くのデータを迅速に活用してビジネスの価値につなげ、コンプライアンスやセキュリティを確保したうえでデータを提供することで、ビジネス・ユーザーが自分でデータを活用できるようにして、企業内での広範なデータ活用が可能になります」と田中はデータ・ファブリックのアーキテクチャーがもたらすメリットを説明します。(図1)

IBMのデータ・ファブリックへのアプローチ

「IBMのデータ・ファブリックのアプローチでは、分散するデータの課題に対してAI駆動型の自動化テクノロジーで対処しようとしています」と田中は説明します。「IBMが提供するデータ・ファブリック・テクノロジーでは、既存のアーキテクチャーを使用しながら、データの複雑性を排除したり、ソースを問わず、あらゆる場所からのデータをつなげ、理解し、保護し、提供したり、また一元化されたデータとAIのライフサイクルの自動化を実現することが可能になります」と田中。IBMは、データとAI活用のための統合プラットフォーム製品IBM Cloud Pak for Data (以下、Cloud Pak for Data) を通じてデータ・ファブリック構築の実現を支援します。Cloud

Pak for Dataはデータの収集、整理、分析、活用というステップを支えるためのさまざまなコンポーネントを同一のプラットフォーム上で提供します。これらのコンポーネントはRed Hat OpenShift上でコンテナ化されているため、オンプレミス、IBM Cloud、他社のクラウドなどさまざまな環境で稼働可能です。

Cloud Pak for Dataのデータ・ファブリックを構成する重要な要素は4点あります。それは、インテリジェントなデータ・カタログ、スマートなデータ・アクセスとオーケストレーション、自動化されたポリシー・ベースのデータ・プライバシー、そしてバイ・デザインのAIとハイブリッドクラウドです。(図2)



図2 Cloud Pak for Dataのデータ・ファブリックを構成する重要な要素

1

インテリジェントなデータ・カタログ

散在するデータを自動的に検出し、関連付け、意味付けを行い、ビジネスに使用できるデータとしてカタログ化

2

スマートなデータ・アクセスとオーケストレーション

データの移動を最小化するために、分散するデータに対して、仮想的なダイレクト・アクセスを提供し、データ統合を高度化

3

自動化されたポリシー・ベースのデータ・プライバシー

分散してデータが存在する環境全体に対して、プライバシーと使用ポリシーの適用を自動化

4

バイ・デザインのAIとハイブリッドクラウド

ハイブリッドクラウド環境での使用、AIを活用した自動化を想定して構築されているアーキテクチャー・デザイン

前述したIBMのデータ・ファブリックのアプローチは、4つのテクノロジーを中心として実現されます。それは、AutoSQL、AutoCatalog、AutoPrivacy、そしてAutoAIです。(図3)

AutoSQL

分散した環境に各々の技術で格納されているデータに対して、単一のクエリー・エンドポイントを提供します。

AutoCatalog

データ・カタログを強化したもので、AI 技術を活用して自動化、効率化する機能。工数を大幅に削減しつつ、それ

まで存在に気づいていなかったデータを発見したり、その業務的な意味や品質も含めて探索、理解したりして、データ活用促進につなげます。

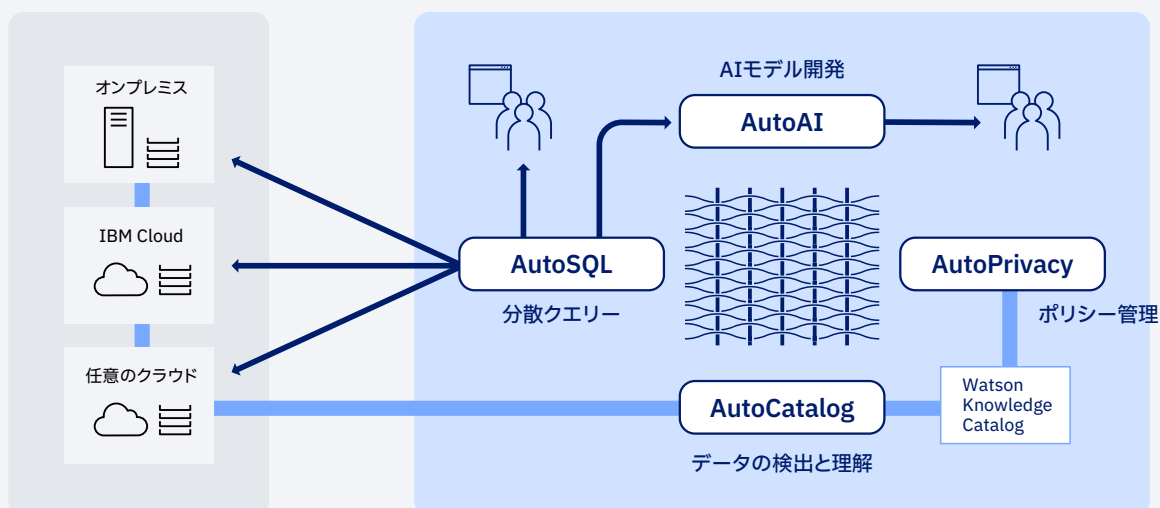
AutoPrivacy

セキュリティーやプライバシーに関する分類、ポリシー適用を自動化し、組織全体のデータ・プライバシー管理を強化します。

AutoAI

AIモデルを自動開発するツールで、開発をスピードアップします。

図3 インテリジェント・データ・ファブリックを実現するCloud Pak for Dataのテクノロジー





Cloud Pak for Dataによるデータ・ファブリックの全体像

「AutoSQL、AutoCatalog、AutoPrivacy、および AutoAI の4つのテクノロジーを主軸としてデータ・ファブリックを実現するCloud Pak for Dataでは、さまざまな環境に分散して存在しているデータの中からデータ利用者が自分自身で必要なデータを見つけ出して、活用できるようにするための次の機能*を統合プラットフォーム上で一貫して提供します」と田中は前置きし、Cloud Pak for Dataによるデータ・ファブリックの全体像について説明します。(図4)

*今後提供予定の機能も含まれます

拡張ナレッジ

さまざまな環境に分散して存在する大規模なデータをデータ利用者が十分に活用できるようにするためにデータを可視化する役割を担います。データを探索し、カタログ化して公開する機能、メタデータを付与し、データの意味を明らかにする機能、オントロジーやナレッジ・グラフによりデータの関連性も含めて視覚化する機能など、データを見つけ、理解し、活用するための機能群を提供します。

スマートなデータ統合

データ利用者に対して、さまざまな環境に分散しているデータ・ソースからデータを取り込み、変換して提供する機能や、仮想的なデータ・アクセスの機能を提供します。データ特性

やデータ・アクセスに関わる要件に応じて、データ仮想化と ETL、データ・レプリケーションなどから最適な技術を選択して、データ利用者にデータを提供します。

データ・セルフサービス

AI、機械学習モデルの開発やBIレポートの開発の際、データ利用者がデータを探し、取得するためのセルフサービスのデータ・アクセス機能を提供します。「スマートなデータ統合」機能と「拡張ナレッジ」機能によって、整備され、提供されたデータに対して、データ利用者が自分自身でデータを探し、選択することができるようになります。

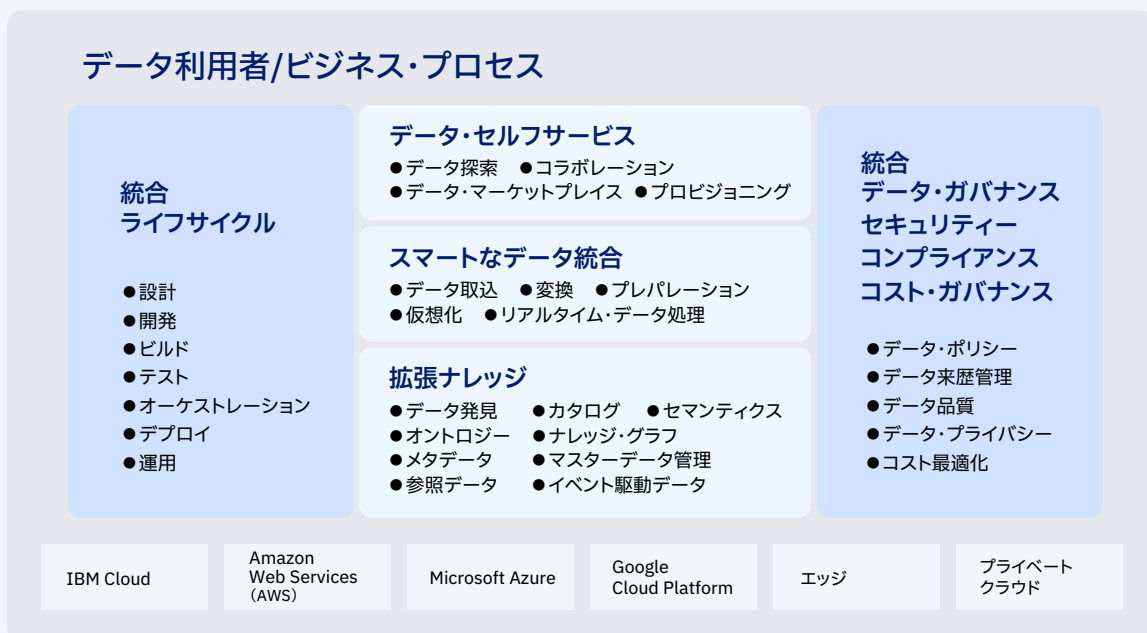
統合データ・ガバナンス、セキュリティ、コンプライアンス、コスト・ガバナンス

データを整備し、提供する際に一貫して、データ・ガバナンス、セキュリティ、コンプライアンスを確保するために、データに対するポリシー適用、来歴管理、データ品質の確認、データ・プライバシー管理などを提供します。

統合ライフサイクル

提供されたデータを活用し、BIやAI、機械学習のモデルを開発、テスト、デプロイ、運用するための機能群を提供します。

図4 Cloud Pak for Dataによるデータ・ファブリックの全体像





「4つの主軸テクノロジーのうち、AutoSQLは主に『データ・セルフサービス』と『スマートなデータ統合』の領域をカバーします。AutoCatalogは『統合データ・ガバナンス』、『データ・セルフサービス』、および『拡張ナレッジ』の一部の領域を

担います。AutoPrivacyは『セキュリティ』、『コンプライアンス』の領域をカバーします。また、AIモデルの自動開発を支えるAutoAIは『統合ライフサイクル』を加速させるためのツールとなっています」(田中)

コンポーネント化されたプラットフォーム

データ活用のプロセスは一連のステップから構成されます。主な流れは「データの発見、理解」、「ガバナンス、カタログ化」、「変換」、「アクセス、編集」、「分析、モデル化」、「組み込み、管理」で、これらのプロセスを、データの提供者と利用者がコラボレーションしながら進めていくのが一般的です。「Cloud Pak for Dataは、データの提供者と利用者がシームレスにデータ活用のプロセスを進めていくための統合プラット

フォームを提供します。一方でそれぞれのコンポーネントはモジュール化されており、他社のツールと組み合わせることも可能で、連携などに考慮しつつ、既にある環境やツールを生かして取り組むこともできます。また、大規模にデータ・ファブリックを展開する場合には、国や事業部単位で、公開範囲に合わせて複数のデータ・カタログを持ち、全体でのデータ・カタログと階層化して構成する方法もあります」(田中)

データ・ファブリックのアプローチは、さまざまなビジネス上のメリットを生み出します。技術部門にとっては、データ・デリバリー・プロセスの簡素化や工数の削減になり、事業部門は、より迅速かつ広範なデータ活用に取り組めるようになります。

「データの蓄積、管理、活用といったさまざまな側面で統合的な管理を実現するレイヤーを持ち、企業全体でのデータ活用を促進することがIBMのデータ・ファブリックの目指す姿で

あり、その構築を支援する製品が、Cloud Pak for Dataです。データ・ファブリックの実現には、データのカタログ、ガバナンス、活用といった領域全体でエンドツーエンドに統合して管理することが不可欠です。これを製品、サービスの両面から一貫通貫にご支援できるのがIBMの強みです。各領域の専門家も多数在籍しており、国内はもちろんのこと、グローバル規模でのご支援経験も豊富です。お客様の全社規模でのデータ活用について、ぜひIBMにご相談ください」(田中)

参考情報

Data Fabricのより詳しい情報やセミナー情報はこちらをご覧ください。

<https://www.ibm.com/jp-ja/analytics/data-fabric>

お問い合わせ

お問い合わせフォームへご記入ください

1営業日以内にご連絡いたします。



ibm.biz/contact_analytics_jp



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
©Copyright IBM Japan, Ltd. 2021 All Rights Reserved

IBM、IBM ロゴ、ibm.com、IBM Cloud、IBM Cloud Pak、および IBM Watson は、世界の多くの国で登録された International Business Machines 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 商標リストについては www.ibm.com/legal/copytrade.shtml をご覧ください。

Microsoft は Microsoft Corporation の米国およびその他の国における商標です。

Red Hat および OpenShift は、Red Hat Inc. または子会社の米国およびその他の国における商標または登録商標です。