Eckerson Group | Report

# データファブリック

## データアーキテクチャの進化に向けた次のステップ

JAY PISCIONERI著

2023年1月

InterSystems®
Creative data technology

## 著者紹介

**Jay Piscioneri** は 25 年以上にわたり、データウェアハウス、ビジネスインテリジェンス、データ品質、データガバナンスなどのデータテクノロジーに携わり、さまざまな業界の企業で次世代データプラットフォームの計画・構築を支援してきました。多くのイニシアチブのリーダーとして、データソリューションの導入のコツや、新しい優先順位、プロセス、ツールの採用に伴う組織の課題について、豊富な経験を持っています。

# Eckerson グループについて

Eckerson Group は、企業がデータからより多くの価値を引き出せるよう支援する、グローバルリサーチ、コンサルティング、アドバイザリー企業です。当社の専門家は、データ分析について批判的に考え、明確に記述し、説得力のあるプレゼンテーションを行います。専門分野はデータ戦略、データアーキテクチャ、セルフサービスアナリティクス、マスターデータ管理、データガバナンス、データサイエンスです。データとアナリティクスを解明し、データの力を活用したビジネス主導の戦略を策定することで、多くの企業から信頼をいただいています。

Eckerson Group のサービスについて

# 当レポートについて

本レポートはソフトウェアベンダーへの調査に基づくもので、スポンサーは Nexla、InterSystems、TIBCO です。各社は本レポートの内容の独占的な配布許可を受けています。

# 目次

# レポート概要

企業は分析などのビジネスニーズに対応するために長年、企業データの一貫した統一ビューを実現する、アーキテクチャアプローチを数多く試みてきました。しかし電子商取引、ソーシャルメディア、クラウドコンピューティング、そしてユーザーニーズの高まりによる急速な変化で、従来のアプローチでは対応できない新たなデータ課題が生まれました。

1990 年代、データウェアハウスは、企業データを次元的に構造化したデータを一元管理し、ビジネスインテリジェンスの要件をサポートするアーキテクチャとして登場しました。2010 年代には、データレイクが登場し、構造化データだけでなく、あらゆる種類のデータを、データサイエンスの新しい用途に活用するために、一元管理するようになりました。しかし、その柔軟性の高さゆえ、データ品質の問題や分析作業の重複を生むことになりました。

最近、異種分散データの統一ビューを提供し、ビジネスインテリジェンスからアドホック分析、データサイエンスまで、あらゆるタイプのワークロードをサポートするデータファブリックが登場しました。データファブリックは、メタデータと人工知能/機械学習（AI/ML）を利用して、新しいデータソースのオンボーディングやメタデータの管理といった、データ管理機能を自動化します。また異なる種類や異なる場所にあるデータの結合を簡素化し、企業データへのシームレスなアクセスを提供します。

**ポイント**

- データファブリックは、メタデータ、機械学習、自動化を用いて、形式や場所に関係なく企業データの統一ビューを提供するアーキテクチャのアプローチです。
- データファブリックは、データウェアハウスやデータレイクに取って代わるものではなく、それらを包含してビジネスインテリジェンス、データサイエンス、組み込み分析をサポートするものです。
- データファブリックは、インサイトまでの時間を大幅に短縮することを目指しています。
- データファブリックは、発見、カタログ化、準備、検証、モニタリングなど、データ管理の多くの側面を自動化する AI 主導のプロセスを組み込んで、絶え間ないデータの需要に対応できるよう支援します。
- データファブリックは、単一の製品として購入できるものではありません。データファブリックは、1 つのベンダーから統合済みのツールを購入するか、複数のベンダーから最高のコンポーネントを購入し、自分でそれらを統合するかを選択することになります。
- データファブリックは、データ環境が急速に拡大し、さまざまなデータ形式が複数の場所に保存され、多くの分析要求を満たすためにデータアクセスを民主化する必要がある組織に最適です。

**推奨項目**

- データファブリックの導入を検討している企業は、他の新しいテクノロジーと同様に、ニーズを慎重に評価し、現実的な期待にもとづき、本格導入前に小規模な導入で試す必要があります。
- 組織のデータ管理の成熟度を評価し、どの分野を改善すればビジネスに最も大きな影響を与えるかを判断します。これらの分野に対応するデータファブリックコンポーネントに重点を置くことが重要です。

- データ管理ツール（データのカタログ化、準備、品質管理など）に、まだ大きな投資をしていない組織では、統合済みのコンポーネントスイートを検討しましょう。このアプローチでは、より迅速に価値を提供することに集中できますが、ひとつのベンダーの製品にコミットすることになります。
- うまく機能しているデータ管理ツールがある場合は、組織のニーズに対応するコンポーネントを購入し、自動的な統合機能を自前で構築するのがよいでしょう。
- データファブリックの実装には、データエンジニアリング、データサイエンス、ソフトウェアエンジニアリング、ネットワーキング、セキュリティ、構成管理など、さまざまなスキルが必要です。そのため、必要なスキルがあるかどうか、どのようにチームを構築するかを判断する必要があります。

# データアーキテクチャの進化

近年データアーキテクチャは、めざましく進化しています。しかしデータアーキテクチャの範囲や複雑さは変わっても、その基本的な目的は変わっていません。データアーキテクチャは、ビジネスニーズをシステムデザインに変換し、現在および将来の要件をサポートするために企業全体にデータを提供します。

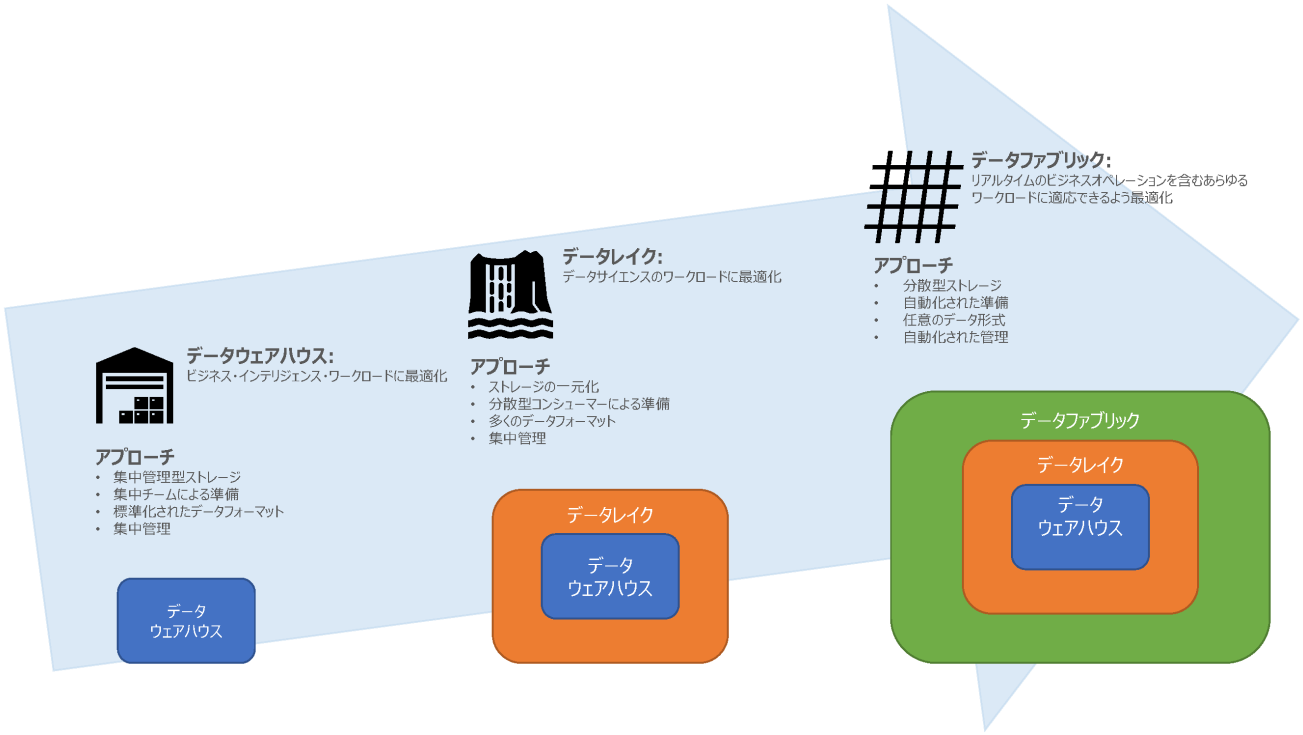
データアーキテクチャは、ビジネスニーズをシステムデザインに変換し、現在および将来の要件をサポートするために企業全体にデータを提供します。

## データアーキテクチャの世代

データウェアハウス、データレイク、データファブリックという三世代のデータアーキテクチャは、企業がデータを管理・活用する上で大きな役割を果たしています（図 1 参照）。データウェアハウスは、企業データの統一ビューを作成するための最初のアプローチとして登場しました。その後 e コマースやソーシャルメディアが登場し、データフォーマットの量と種類が爆発的に増え、ウェアハウスが想定していなかったデータニーズが生まれました。データレイクは、データウェアハウスが取り残した部分を補うように登場し、新しい形式のデータに対応するための柔軟性を備えています。しかし、その柔軟性ゆえに多くの解釈や結論が生まれ、真の一元管理は実現しませんでした。

このような課題に基づきデータファブリックは生まれました。データファブリックの目的は、従来のアーキテクチャの長所を組み合わせて、現代のデータに関する課題に対処することです。進化の各段階を振り返ってみましょう。

**図 1. データアーキテクチャの進化**




## データウェアハウス

データウェアハウスアーキテクチャは、現在と同じような問題を解決するために 1980 年代に生まれた、一元的なストレージと統一フォーマットを持つデータベースです。その目的は、異なるシステムからデータを集めて統一的なデータ表示を行い、経営判断を支援することです。現在でも多くの組織がこの目的のためにデータウェアハウスを使用しており、おもにビジネスインテリジェンスレポートやダッシュボードを作成するアナリストが利用しています。

データチームは、あらかじめ定義された問い合わせのカテゴリに答えるためにデータを変換し、構造化することによって、データウェアハウス内のデータを統合します。これは分析要件が事前に理解されている場合には有効ですが、アドホックなデータ探索には不向きです。また中心となるチームがデータを利用するためのモデリングと準備を行う必要があるため、データへのビジネス上の要求を満たすまでに時間がかかる傾向があります。

## データレイク

データレイクは、現代のデータ量と多様性に対応し、かつデータサイエンスのユースケースをサポートするために生まれました。分散した場所からデータを集め、変換せずにソースシステムのそのままの形式で一元的に保存します。これによってデータサイエンティストが複数の仮説を検証し、予期せぬ洞察を得るために必要な柔軟性を提供します。またアプリケーションに統合可能な分析機能をサポートします。

特定の用途のためにデータを準備する、つまりデータを必要とする側が用途に応じて準備することになるため、データ利用者には高度なデータエンジニアリングのスキルが求められます。また文書化されていないソースデータに対して様々な解釈がなされるため、結果的に矛盾した結論が導き出されることになってしまいます。

### データファブリック

データウェアハウスとデータレイクは、どちらも企業のデータに関する課題をすべて解決するものではありませんが、それぞれ得意分野があります。データファブリックは、ビジネスインテリジェンス、データサイエンス、アプリケーションとの統合をサポートし、組み込み分析を可能にします。双方のアーキテクチャを置き換えるのではなく、包含することが目的です。

データファブリックは、構造化データと非構造化データの両方に対応します。データファブリックは、一元的なストレージを構築するのではなく、データをそのまま残して、分析とリアルタイムのビジネスオペレーションに必要なデータに接続することに重点を置いています。もう一つの大きな違いは、データ管理の負荷を軽減するために、データファブリックが採用している自動化の程度です。例えばデータの準備は依然として分散化されていますが、技術的なバックグラウンドのないデータ利用者は、ローコード/ノーコードのツールを使って複雑な変換を行い、データを分析に対応する形に整えています

## データファブリックの定義

データファブリックは、メタデータ、機械学習、自動化を用いて、あらゆる場所にあるあらゆる形式のデータをまとめ、人々やシステムが簡単に見つけて利用できるようにするアーキテクチャアプローチです。データファブリックは、統合、準備、カタログ化、セキュリティ、発見といったデータ管理の個別の機能を、インテリジェントな自動化によって一貫したプロセスに統合するものです。

データファブリックは、メタデータ、機械学習、自動化を用いて、あらゆる場所にあるあらゆる形式のデータを編み込み、人やシステムが簡単に見つけて利用できるようにします。

データマネジメントの機能は新しいものではありませんが、データファブリックが規定する自動化と統合の度合いは新しいものです。データファブリックには、以下のデータ管理機能が含まれています。

- 統合：データソースに接続し、データを元のフォーマットで利用できるようにする。
- 準備：特定の目的に合わせてデータをクリーニング、正規化、再編成（変換）する。
- カタログ化：データ資産に関するメタデータの作成、表示、管理。
- アクセス管理：データが何であるか、誰がそれを必要としているかに基づいて、適切なアクセス権を提供する。
- 発見：名称、定義、概念、または関連する用語によってデータを検索できるようにする。

データファブリックは、新しいデータソースの取り込み、メタデータの管理、異なるソースからのデータの結合などの機能を自動化します。データファブリックは抽象化レイヤーを作成し、分散データや異種フォーマットの処理に伴う複雑さを解消します。企業データへのシームレスなアクセスを実現し、データチームが膨大な量のユーザーリクエストに対応するのを支援します。データファブリックは、静的なデータと移動中のデータに対応しています。アナリティクスのユースケースが一般的であることから、実装のほとんどは、静的なデータに重点を置いています。

## データファブリックとは？

データファブリックは、身近なデータ管理機能を取り入れたものですが、データ仮想化ツール、データカタログ、論理データウェアハウス、データレイクハウス、データメッシュなどのコンポーネントとは区別されます。また DataOps の分野とも異なります。それぞれの類似点と相違点を考えてみましょう。

- データ仮想化ツール：分散した場所や多様な形式のデータを橋渡しするもので、データをソースから中央へコピーする必要はありません。後述するように、データの仮想化は、データファブリックの重要な機能＝データの抽象化の一手法です。データの抽象化により、分散したソースや多様なフォーマットによる複雑さを解消し、利用者がデータを見つけやすく、使いやすくすることができます。データファブリックには、データパイプラインを通じてソースデータを物理的に抽象化する機能も含まれています。
- データカタログ：メタデータを保存・管理します。データカタログはより多くの機能を備えて進化していますが、ビジネス分野・技術分野のメタデータの参照用に保存するのが主な役割でした。データカタログはデータファブリックに不可欠な要素ですが、データファブリックはあらゆる種類のメタデータを積極的に利用します。参照用だけでなく、データチームやデータアナリストの作業負荷を軽減するための自動化を促進します。データファブリックは、ストリーム、API、ファイル、その他のデータシステムなど、従来のカタログでは必ずしもカバーされていないデータをカタログ化します。
- 論理的データウェアハウス：異なるソースやロケーションのデータを分析用に利用できるようにするもので、データ仮想化に似ています。トランザクションアプリケーションやリアルタイムのストリーミングデータも扱えます。
- データレイクハウス：データを抽象化したもので、さまざまな形式のデータを扱えるだけでなく、データアナリストがデータを使いやすくするための構造化データセットも提供します。データファブリックの一部となり得ますが、データファブリックの自動化機能は備えていません。
- データメッシュ：データファブリックもデータメッシュも、あらゆるデータのユースケースのサポートを加速・拡大させることを目的としています。データメッシュは組織の変化を重視するのに対し、データファブリックはデータソリューションを大規模に設計・実装するために必要な技術に重点を置いています。データメッシュの分散型アプローチでは、データを作成するシステムのチームが、データ消費者のニーズを満たすデータ製品を構築するよう求めています。データファブリックは、データ製品の生産と消費を自動化するためのアーキテクチャを提供できます。
- DataOps：データファブリックと多くの点で同じ目的を持っています。しかし DataOps には、データファブリックが対応していないコードの管理、環境の設定、構築、実装が含まれます。データファブリックは、その大量のアクティブなメタデータを提供して、DataOps の監視、検証、および運用機能を強化することができます。(DataOps の詳細については、Eckerson Group のレポート「DataOps Deep Dive」を参照してください）。DataOps プラットフォームへのさまざまなアプローチ)をご参照ください。

前述の構成の多くはデータファブリックの一部ですが、それだけでデータファブリックが定義されるわけではありません。コンポーネント間の統合と、データ管理の負荷を軽減する自動化によって、データファブリックの全体像が完成します。

# データファブリックの潜在的メリットと落とし穴

## データファブリックの利点

データファブリックの目的は、データの量、ソース、フォーマット、用途が増加し続ける中で、データから引き出す価値を高めることです。ビジネス部門はデータ駆動型の意思決定を迅速に行う必要があり、データ部門はデータの需要に対応するための新しい方法が必要です。データファブリックの利点は、これらの課題に対処できることです。

- **洞察までの時間を短縮：**洞察までの時間とは、データマネジメントの有効性を示す指標で、ビジネス行動につながる気づきの瞬間までに、かかるデータの使用時間のことを指します。データファブリックは、洞察までの時間を大幅に短縮することを目的としています。インテリジェントな自動化と発見力の強化に加え、企業データの複雑さを隠し、データがどこに、どのようなフォーマットで保存されているかを利用者が知る必要がないようにします。このように詳細な情報を取り除くことで、データ利用者は必要なデータの収集よりも、分析に集中することができます。
- **データ管理の作業負担を軽減：**データファブリックは、AI と機械学習を用いて、新しいデータソースのカタログ化やデータ準備の支援など、多くのデータ管理作業を自動化します。自動化により、データ部門がデータの需要に対応できるよう、時間のかかる手動ステップを排除できます。
- **効果的なデータ発見：**最新のデータカタログと高度な検索機能により、利用者はどのようなデータが利用可能かを知ることができます。企業データ用の Google を持つようなものです。キーワード検索機能とデータ資産に関する豊富なドキュメントにより、ビジネスアナリスト、データサイエンティスト、データチームはデータリソースを見つけ、特定のユースケースに対するデータの価値を評価することができます。

## データファブリックの落とし穴

データファブリックは、企業のデータ管理機能の全領域をカバーし、上記のようなメリットが実現すれば、データマネジメント全般に革新的な影響を与えます。同時に多くの人、プロセス、テクノロジーが影響を受けるため、リスクも大きくなります。

データファブリックがもたらす影響は、変革をもたらす可能性があります。
しかし、多くの人、プロセス、テクノロジーが影響を受けるため、リスクも大きくなります。

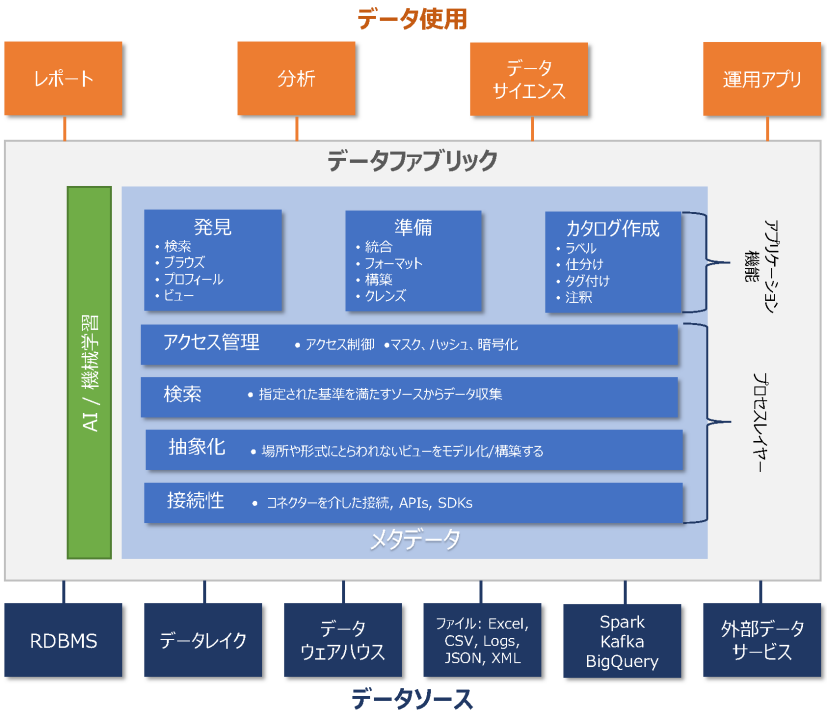
- **広範囲のため影響大、従ってリスクも大：**データファブリックは多くのデータ管理機能に影響を与えるため、一気に取り組むと、実現までに時間がかかりすぎる可能性があります。またユートピア的なデータの未来を約束しているため、関係者の期待をコントロールしなければ失望したり、最終結果へのコミットメントが揺らぐ場合があります。
- **データファブリックは製品ではない：**データファブリックは製品ではありません。データファブリックは一連のツールで構成され、そのうちのいくつかは既に持っていたり、既存ツールの統合かもしれません。ツールの評価や機能の比較、統合の計画や設定など、様々な準備が必要です。
- **多くのベンダー、多くのアプローチ：**データ管理ソフトウェア会社は、データファブリックのコンセプトに沿った製品の提供を急いでいます。ある会社は実績のある自社の専門分野の、最も成熟したコンポーネントを含むツール

群を提供しています。また限定的な役割を持つ一つの製品に焦点を当てる企業もあります。このような複雑な製品群は、データチームのソリューション設計をより困難にしています。しかし製品市場は成熟しつつあり、データファブリック機能の、より多くの側面が製品スイートに統合されることを期待します。

# データファブリックの特徴

データファブリックを構築する方法は一つではありませんが、データファブリックが、データを見つけやすく、使いやすく、管理しやすくするという目的を達成するためには、以下のようなコアコンポーネントで構成される必要があります（図 2 参照）。

**図 2. データファブリックのコアコンポーネント**




## メタデータ

メタデータは、データを記述するデータでデータファブリックの基礎となるものです。データファブリックはメタデータを使用して抽象化レイヤーを作成し、分散データや異種フォーマットでの作業の複雑さを軽減します。またメタデータを使用して、データのカタログ化や準備（プリパレーション）などの機能を自動化することもできます。データファブリックは、4 つの大きなカテゴリにまたがるメタデータを使用し、生成します。

メタデータは、データファブリックの基盤として機能します。メタデータは、分散したデータや異なるフォーマットを扱う複雑さを隠し、データのカタログや準備などの機能を自動化するために使用されます。

- **技術的メタデータ：**形式、タイプ、長さ、場所など、システムがデータを扱うために必要なデータの特性が含まれます。

- **ビジネスメタデータ：**アプリケーション、レポート、ダッシュボードに表示されるデータのコンテキストを提供するために、ビジネス言語を使用します。用語、定義、分類などの要素が含まれます。
- **運用メタデータ：**データがどのように使用され、使用された時に何が起こるかについての情報を提供します。実行ログ、ルールエンジン、エラーログ、監査登録など、様々なソースからの情報が含まれ、イベントの日付と時刻が記載されています。
- **ソーシャルメタデータ：**ユーザーとのコラボレーションから生まれる豊かさを表現するものです。タグ、評価、コメントなどの要素が含まれます。

## アプリケーション機能

ディスカバリー、準備、カタログの双方向的な機能により、ビジネスユーザー、パワーユーザー、データチームがデータを見つけて扱えるようになります。

- **ディスカバリー（発見）：** データを見つけやすくするためのデータファブリック機能です。効果的なディスカバリーがなければ、必要なデータは、テラバイトやペタバイトのデータの中に紛れ込んでしまいます。ディスカバリーは、ユーザーが利用可能なデータセットをブラウジング・検索し、異なるソースのデータをプレビューして、それが分析に役立つかどうかを理解することを可能にします。ブラウジングでは、ビジネスの対象やデータソースなど、データのカテゴリをナビゲートすることができ、多くの場合、階層構造で表示されます。検索機能は、基本的なキーワード検索から、自然言語処理（NLP）を用いた AI アシスト検索まで、様々な種類があります。
- **プリパレーション（準備）：**パワーユーザーが異なるフォーマットや分散した場所に対処することなく、一つまたは複数のソースからデータにアクセスできるようにするためのデータファブリック機能です。例えばデータサイエンティストは、統計モデルに必要なデータを収集・整理するためにプリパレーション機能を使用し、データチームはレポート作成のためにデータを適合・集計するために使用します。
- **カタログ作成：**カタログは、ユーザーがメタデータを作成、表示、管理できるようにするデータファブリック機能です。発見時にデータリソースに関する情報を表示し、ユーザーがタグ、コメント、評価、認定などでメタデータを充実させるためのインターフェイスです。また、カタログ機能は、テーブルへのアクセス頻度、クエリでの使用状況、使用者などのデータの使用状況も把握することができます。これは、消費者が利用可能なデータリソースの信頼性を評価するのに役立ちます。

## レイヤー処理

ファブリックの運用では、データソースへの接続、ワークフローの管理、データを抽象化して統一的なビューを作成し、データへのアクセスを適切に制御することが必要です。

- **接続性：**コネクティビティ層は、データソースとデータファブリック間の通信を可能にするアプリケーション・プログラミング・インターフェース（API）を通じて、データソースへのアクセスを自動化します。データファブリック製品には、以下のようなさまざまなプラットフォーム向けのソース固有のコネクタが多数含まれています。Snowflake, Salesforce, AWS S3, MongoDB,等々。

- **抽象化：**抽象化レイヤーは、データソースの異なるフォーマットや分散した場所を消費者から隠し、消費者がエンタープライズデータを利用しやすくします。データファブリックでは、ユーザーはデータソースと直接対話することはありません。ユーザーはソースデータを抽象化したもの（抽象化データオブジェクトまたは ADO）を使用するため、データがどこにあるか、どのプラットフォームから来たかを気にする必要がありません。ADO は仮想的なもので、ソースに保存されているデータを参照できます。ソースに保存されているデータを参照する仮想的なものと、データをファブリック内の別の場所にコピーする物理的なものがあり、通常はより速く検索できるようになります。
- **検索：**検索層はクエリーエンジンで構成され、指定された目的のために、指定された条件を満たすデータをソースシステムから収集します。ユーザーがレポートを作成する時、データサイエンティストがモデルで使用するデータを準備する時、データパイプラインプロセスが物理的な ADO を更新する時に、検索層はデータの場所とその配信を最適化する方法を決定します。
- **アクセス管理：**アクセス管理層は、ユーザーの役割と要求されるデータの分類に基づいて、特定のデータを使用できる人を決定します。データソースのセキュリティポリシーは、データにアクセスできる人を最低限、管理します。データチームは、マスキング、難読化、暗号化など、アクセス制御の追加レベルを定義することができます。

## 人工知能（AI）機械学習（ML）機能

AI と ML は、データチームが手作業で追いつくのが難しいデータ管理のワークロードを、インテリジェントに自動化することで、データファブリックの目的達成を支援します。ファブリックのメタデータを使用して、AI と ML は処理とアプリケーション機能の多くの側面を自動化することで、スケーラビリティを改善します。表 1 の自動化の例は、データチームとユーザーの作業負担を軽減します。

*AI と ML は、データチームが手作業で処理するのが困難なデータ管理のワークロードを*
*インテリジェントに自動化することで、データファブリックの約束を実現するのに役立ちます。*

**表 1. データファブリックアプリケーション機能における AI/ML の例**

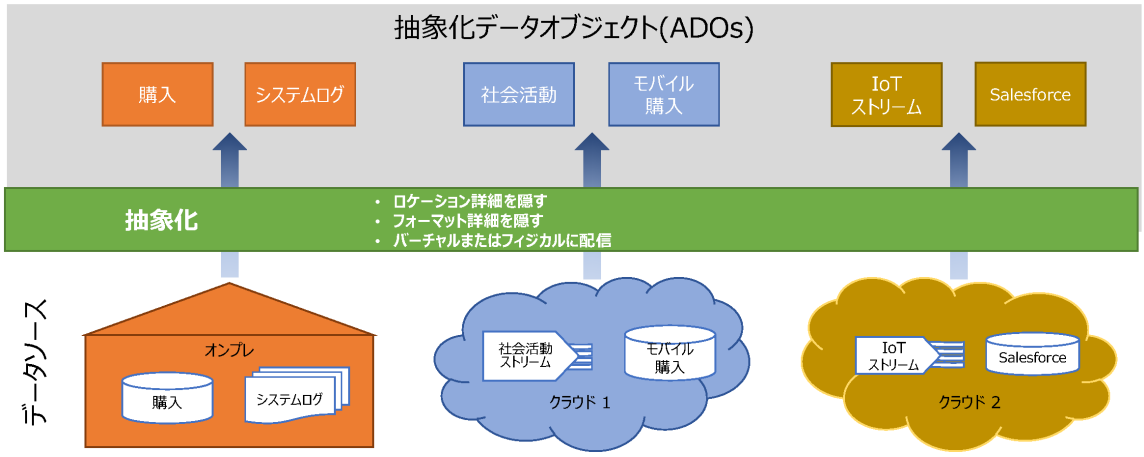
| 機能 | AI/Machine Learning Process |
|---|---|
| 発見 | **NLP 検索：**「今後 3 ヶ月で解約しそうな顧客は何人いるか」といった自然言語の文章を入力すると、その質問に答えるためのデータリソースを検索することができます。 |
| 準備 | **プログラムコード生成：**適切な結合、フィルター、ソート、グループ化、小計を備えたプログラムコードを自動生成する AI や機械学習プロセスで準備をサポートします。また AI/ML は、自己制御による検証やエラー検出により、データ運用を改善します。 |
| カタログ作成 | **カタログ作成の自動化：**AI/ML の機能により、新しいデータソースを搭載する際の技術、ビジネス、運用のメタデータの生成を自動化し、データの使用方法に関する運用メタデータを取得します。 |

## データの抽象化

データファブリックにおける抽象化の重要性は言い尽くせません。データをユーザーに届ける機能であり最も重要です。そこで言葉を定義してみましょう。抽象化とは、ある分析に関連する詳細を強調するために、ある文脈では関連性のない詳細を削除するプロセスのことです。データファブリックは様々な方法で抽象化を利用します。まずソースデータの形式や場所などの詳細を取り除き、ソースが提供するデータの内容を強調します（図 3 参照）。

抽象化とは、分析に最も関連するものに焦点を当てるために、いくつかの詳細を削除するプロセスです。

**図 3. データファブリックにおける抽象化データオブジェクトの生成**



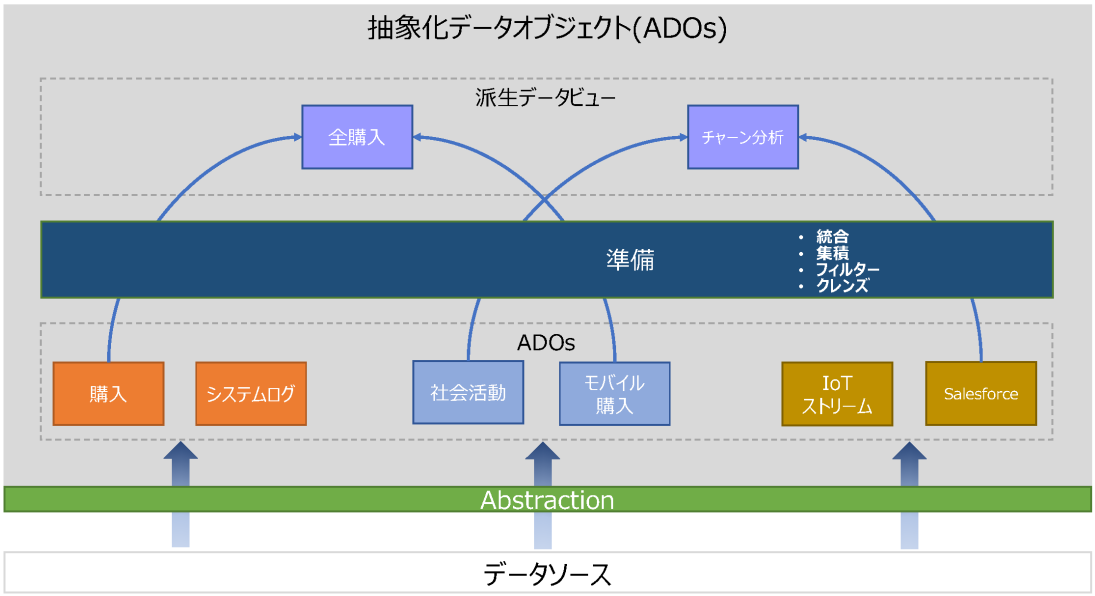

## 抽象化データオブジェクト向けソース

抽象化レイヤーは、物理的または仮想的な形でデータを提供します。データパイプラインは物理的な抽象化データオブジェクト（ADO）を作成し、データ仮想化技術は仮想 ADO を作成します。仮想化は、データのコピーを作成しないことでコストとリスクを削減できるのが利点ですが、パフォーマンス上の問題やセキュリティ上の制約から、データをコピーして移動する必要がある場合もあります。従ってデータファブリックはその両方を行う必要があるのです。

## ADO と仮想データビュー

抽象化されたデータは、様々な分析用途をサポートします。ソースオブジェクトをミラーリングする ADO は、柔軟なデータ探索と、マッシュアップというデータサイエンスのニーズをサポートする論理的なデータレイクを作成します。さらにパワーユーザーは、準備機能を通じて ADO にアクセスし、特定の分析用途に最適化された派生データビューを作成できます。これらの派生データビューは、複数のシナリオをサポートすることができます。例えば、ビジネスインテリジェンス用の適合データセットとして使用したり、予測モデル用の準備データ入力として使用したりすることができます（図 4 参照）。

**図 4. 派生データビュー**




## データファブリックアプローチ

データファブリックの設計には様々なアプローチがあります。組織が直面する課題と機会によって、データファブリックのどの側面を重視すべきかが決まります。ここでは異なるビジネス上の問題に焦点を当てた 3 つのアプローチを紹介します。

- **ビジネスパーソンのエンパワーメント：** ビジネスパーソンがデータにアクセスし、質問に答え、情報に基づいた意思決定を行えるようにすることは、データファブリックの重要な目的です。このことを第一に考えれば、発見と準備の容易さが重要です。例えば人工知能の一種である自然言語処理（NLP）を活用すると、通常の文章やフレーズを取り込んで、単語をインテリジェントに解析し、関連するデータリソースを見つけることができます。
- **全方位型アナリティクス：** データサイエンスのサポート、組み込みアナリティクス、データ管理の効率化は、データファブリックより広範かつ高度な目的です。データサイエンティストは、AI/ML モデルのためのデータの整理と洗練、およびアプリケーションに組み込まれたリアルタイム分析のために、ADO（上記図 3 参照）の変更されていないデータを必要としています。大規模で複雑なデータ環境を持つ企業では、カタログ作成、データ品質、コンプライアンスに対応するため、自動的なデータ管理機能へのニーズが高まっています。
- **運用データ管理：** これまで、データファブリックがアナリティクスのユースケースをどのようにサポートするかについて述べてきましたが、データファブリックは、運用のトランザクションの世界でも役割を果たします。データサイロを埋めるためのアプリケーション間のリアルタイム統合の管理や、IoT デバイスが存在するデータランドスケープのエッジへの到達は、データファブリックの最も高度なアプローチを必要とするトランザクションのユースケースです。

# 実装オプション

「特徴」の項目で説明したように、データファブリックはデータカタログ、変換ツール、クエリーエンジンなどのコンポーネントで構成されています。データファブリックを実現するためには、これらのコンポーネントが必要です。問題は、データファブリックで連携するように設計された単一の「スイート」ベンダーから、これらの大部分または全部を購入するかどうかです。あるいは複数のベンダーからベストオブブリードのツールを購入し、それらを自分で統合してデータファブリックを構築することもできます。

問題は、データファブリックのコンポーネントを連携するように設計した単一のベンダーから購入するか、複数のベンダーからベストオブリードのツールを購入し、データファブリックに自分で統合するかということです。

以下の基準は、スイートアプローチとベストオブブリードのどちらを選ぶかの参考になります。どちらのオプションを選択するかは、組織のニーズ、現在のデータアーキテクチャの構成、利用可能なリソースによって異なります。

- **組織の課題：** まず組織の目的を達成する能力に影響する、データの問題を理解する必要があります。データの強み、弱み、機会、脅威は何ですか？組織はカスタムソリューションを必要とする特別な課題を抱えていますか？どのような問題を、どのような優先順位で解決する必要があるのかを把握してください。
- **エンタープライズ・データアーキテクチャ：**データカタログ、データ統合ツール、データ仮想化ツールなど、データファブリックの構成要素の一部には、すでに投資しているかもしれません。これらのツールは、現在の環境にどの程度深く組み込まれているのでしょうか。また、それらのツールはどの程度機能しているのでしょうか。データアーキテクチャのうち、維持したい要素と交換する必要がある要素を判断することで、意思決定が可能になります。
- **利用可能な製品：**データファブリック製品の市場は急速に発展しており、多くのデータファブリック機能に対応する統合製品群を提供している企業もあります。あるいは 1 つまたは 2 つのコアファブリック機能に焦点を当て、他の製品と統合するための API を提供している企業もあります。すべてを網羅した純粋なデータファブリック製品は市場に存在しません。ほとんどの製品は、データ管理の既存企業によるものです。各社が提供するデータファブリックは、自社が得意とする機能に重点を置いています。
- **開発リソース：**複数のベストオブブリードのコンポーネントからデータファブリックを構築するには、開発に多大な労力が必要です。またデータエンジニアリング、データサイエンス、ソフトウェアエンジニアリング、ネットワーキング、セキュリティ、構成管理など、様々なスキルが必要になります。ベストオブブリードのアプローチを選択する前に、必要なスキルがあるかどうか、または、どのようにチームを構築できるかを判断してください。
- **予算：**いずれのアプローチでも費用はかかります。データファブリックの導入には、ライセンス費用、スタッフの増員、インフラのアップグレード、クラウドストレージやコンピューティングの費用などが考えられます。またデータストレージ、プロセッシング、ソフトウェアライセンスの増加に伴う支出の増加も計画する必要があります。

データ管理ツール（データのカタログ化、準備、品質管理）に、まだ大きな投資をしていない組織では、統合ツールを導入することが良い選択肢になります。このアプローチでは、早く価値を提供することに集中できます。一方で、それは一つのベンダーの製品にコミットすることを意味します。

もし複数のツールに投資しており、それらがうまく機能している場合や、組織独自の必須要件がある場合は、個々のコンポーネントを選択し、自動的な統合を構築することが望ましいでしょう。このアプローチは既存のアーキテクチャを強化し、特定の要件を満たすことができますが、実現するには十分な開発リソースと適切なスキルが必要です。

いずれの場合も、データファブリックを段階的に実装する計画を立て、まず組織の最も差し迫った問題を解決する機能に焦点を当てる必要があります。

## まとめ: データファブリックとその先

数十年前にデータウェアハウスアーキテクチャが導入されて以来、私たちは真のバージョンを一つ持つこと、データ駆動型の意思決定と行動を取ること、そしてデータの民主化という、捉えどころのない同じ目的を追求してきました。データファブリックは、今日の膨大で複雑なデータ環境において、目的を達成するための最新のアプローチです。データファブリックは、手作業によるデータ管理では、現代のデータ需要に対応する拡張性がないという現実に対処しています。

データファブリックは、広範なメタデータと AI/ML の自動化を用いて、データ管理の作業負荷を軽減し、チームが際限のない需要に対応できるようにします。高度な発見力を通じて、ユーザーが必要なデータを見つけ、完全に入力されたデータカタログの豊富な文書を通して、その価値を理解することができます。データファブリックは、異種分散したエンタープライズデータの複雑さを取り除くことで、洞察までの時間を短縮できるため、ユーザーはデータの整理よりも分析に集中できるようになります。

データファブリックは、かつてのデータアーキテクチャと同様に進化し続けるでしょう。製品は成熟し続け、機能と統合のギャップを埋めていくでしょう。データファブリックの範囲は、プログラムコードの管理、データ環境の構成、自動テスト、データ品質の監視を含む DataOps など、データ管理の他の側面を包含するように拡大すると思われます。またドメイン指向のデータ所有権や製品としてのデータといったデータメッシュの組織的革新と、データファブリックのアーキテクチャパターンが融合していくことも予想されます。このような機会を実現するために、企業のデータチームはデータファブリックの基本を学び、それが自分たちの環境にどのように適用されるかを判断する必要があります。

# Eckerson グループについて

著者、講演者、コンサルタントとして世界的に知られる Wayne Eckerson は、組織がデータとアナリティクスからより大きな価値を得ることを支援するために Eckerson グループを設立しました。Eckerson グループの目標は、データおよびアナリティクスを活用するためのあらゆる段階において、専門家によるガイダンスを組織に提供することです。

Eckerson グループは、次の 3 つの方法で組織を支援します。

- **Eckerson グループのソート・リーダー**は、データ分析分野の最新トレンド、テクニック、ツールを常に把握できるよう、実用的で魅力的なコンテンツを発行しています。
- **Eckerson グループのコンサルタント**は、注意深く耳を傾け、深く考え、ビジネス要件を説得力のある戦略やソリューションに変換するためのオーダーメードのソリューションを構築します。
- **Eckerson のアドバイザー**は、ソフトウェアベンダーが市場参入戦略を向上させるために、競合情報および市場ポジショニングのガイダンスを提供します。

Eckerson グループは、データとアナリティクスに特化したグローバルなリサーチ、コンサルティング、アドバイザリー企業です。データガバナンス、セルフサービスアナリティクス、データアーキテクチャ、データサイエンス、データマネジメント、ビジネスインテリジェンスを専門分野としています。

私たちは勤勉で、洞察力に優れ、謙虚であるとお客様から評価されています。それはすべて、私たちのデータへの愛と、組織が洞察を行動に移すのを支援したいという願いからきています。私たちは継続的に学習する家族であり、データとアナリティクスの世界をあなたのために解釈します。

データからより多くの価値を得る。専門家を味方につける：

Eckerson グループのサービスについてはこちらをご覧ください

Eckerson Group
RESEARCH CONSULTING ADVISORY
GET • MORE • VALUE • FROM • YOUR • DATA

# スポンサーについて

1978 年に設立されたインターシステムズは、金融、製造、サプライチェーン、ヘルスケアの各分野において、企業のデジタル変革を支援する次世代ソリューションのリーディングプロバイダーです。同社のクラウドファーストのデータプラットフォームは、世界中の組織が抱える重要課題である、相互運用性、スピード、スケーラビリティの問題を解決します。インターシステムズは 80 カ国以上の顧客とパートナーに対して、24 時間 365 日のサポートを通じて、卓越したサービスを提供することに努めています。

インターシステムズは、データサイロを解消し、ビジネス全体のデータ資産へのアクセスを迅速化、簡素化するスマートデータファブリックアーキテクチャのアプローチを支持しています。InterSystems IRIS® データプラットフォームは、スマートデータファブリックの中核であり、組織は、複数のソースからのデータにオンデマンドでアクセス、変換、調和、分析し、様々なビジネスに利用可能な実用的なデータにすることができます。データ探索、ビジネスインテリジェンス、機械学習、自然言語処理などの広範な組み込み分析機能により、組織は新しい洞察を得て、インテリジェントな予測・処方サービスやアプリケーションを、迅速かつ容易に実行することができます。

インターシステムズのテクノロジーは、クリーンでアクセスしやすく、すぐに行動に移せるデータを顧客に提供することで、あらゆる課題に立ち向かい、ビジネスと世界を前進させることができます。

詳しくは [InterSystems.com/jp](InterSystems.com/jp) をご参照ください。