



YOUnite フェデレイテッド・データ・ユニフィケーションと ナレッジ・グラフでネクスト・ジェネレーション のデータ管理データ・ファブリックを実現

1. はじめに

我々は、ビジネスとイノベーションのスピードに前例のない、ディスラプティブ(過去の常識を打ち消し、画期的な新しいことを生み出す)な世界に生きています。このような世界を生き抜くには、データを如何に活用できるかがキーとなることは既知のことです。データは、企業が競争上の優位を獲得し、繁栄するための重要な資産であり、昨今の顧客やビジネス・ニーズを満たすにはデータをリアルタイムで活用できることは不可欠となります。

しかし、企業において収集するデータが爆発的に増え続ける中、データの種類やソースも数多くあり、多くの企業ではそれらデータがサイロ化しているのが現状です。また昨今では、それらデータが自社内だけでなく、パブリック/プライベート・クラウドなどにデータを配置しハイブリッド環境になっている企業も多く見られます。これらのデータをうまく処理しようと、70%以上の企業が何らかのデータ・インテグレーション・ツールを使用していますが、データの迅速なインジェスト、統合、分析、共有が課題となっています。IDC の調査によると、データ・プロフェッショナルは、メインのデータ分析以外のタスクに 75%の時間を費やし、企業がタイムリにデータから最大限の価値を得ることの障害となっています。

ディスラプティブな世界を生き抜くには、具体的なデータ・マネジメント戦略がなければイノベーションを起こすことはできません。

このような状況で、これまで効果がなかったデータ・インテグレーション・ツールに代わり、次世代 (Next Generation)データ・マネジメントで重要なものがデータ・ファブリック(Data Fabric)です。

2. データ・ファブリックとは

2021年5月11日にリリースされた"Data Fabric Architecture is Key to Modernizing Data Management and Integration"には以下のように記載しています。

「エマージング・デザイン・コンセプトといわれる"データ・ファブリック"は、高コスト低バリュのデータ・インテグレーション・サイクル、初期の統合での頻繁なメンテナンス、リアルタイムでイベント・ドリブンなデータ共有など常に取り組みされてきたデータ・マネージメント・チャレンジへの確固たるソリューションになり得る」と Gartner の著名な VP アナリストである Mark Beyer 氏は述べ、データ・ファブリックをデータと連結プロセスの結合レイヤ (ファブリック) を提供するデザイン・コンセプトです。

また、データ・ファブリックは、これまでに引き出されたメタデータ資産の継続した分析を通して、統合され再利用可能なデザイン、デプロイ、利用を、ハイブリッドとマルチ・クラウドを含む全環境でサポートするものと定義しています。

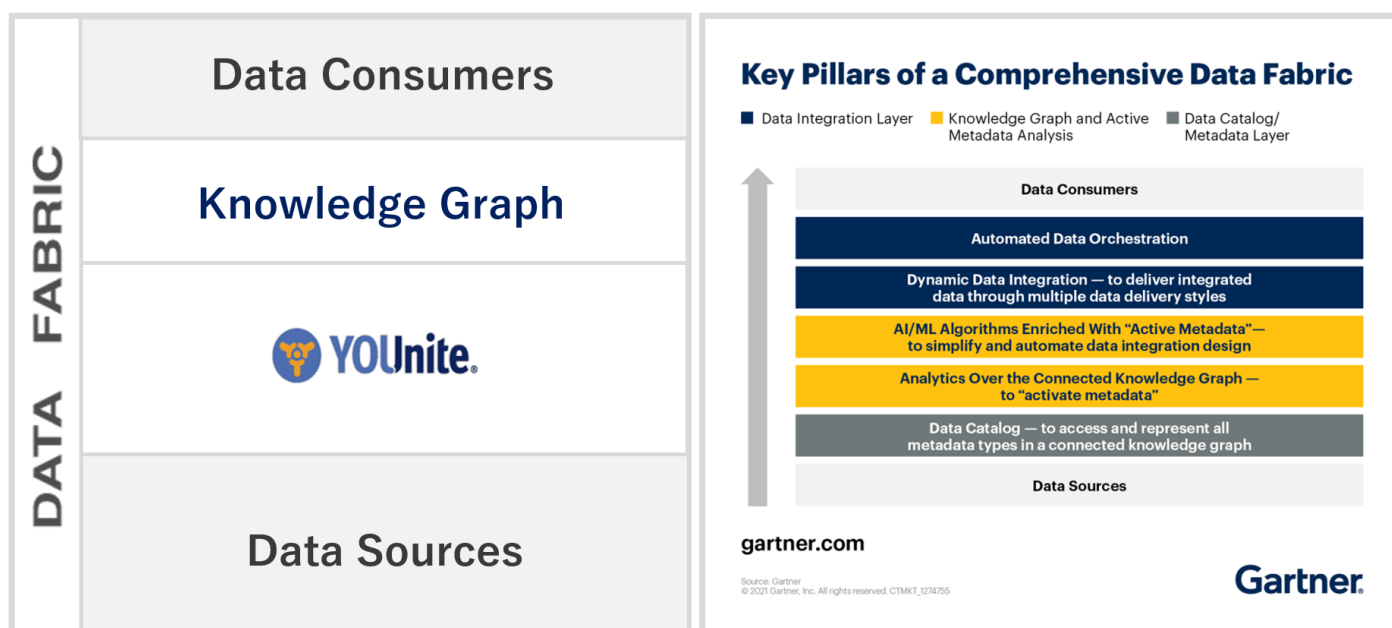
データ・ファブリックは、データへの適切なアクセスや必要であれば統合をサポートします。データ・ファブリックは、いろいろなアプリケーションからのデータを継続して認識し接続することで、データ・ポイント間でユニークな、ビジネスに関連した関係を発見します。

さらに、インサイトとは、リエンジニアリングされた意思決定をサポートし、従来のデータ管理手法より迅速なアクセスと包括的な理解を通してより大きな価値を提供します。

データ・ファブリックの究極の目的は、企業や組織にあるすべてのデータを利活用して企業により良いインサイトを獲得し、より良いビジネス意思決定を行なえるようにすることです。

Marketsandmarkets.com レポートによると、データ・ファブリックの市場規模は、2026年には42億ドル(4,600億円)、CAGRは26.3%になると言われています。日本を含むAPAC市場はさらに高いCAGRになると予想されており、全マーケットにおいてデータ・ファブリックの採用が加速されると思われます。

Gartnerは、データ・ファブリックを今後数年間で重要視するデータ管理に関連する10のトレンドの一つに挙げており、日本を含む世界各地で開催されている2021年Gartner Data & Analytics Summitにおいてもトップ10トレンドの一つとなっています。



データ・ファブリック (Data Fabric)

データ・ファブリックは、今後データ・マネジメントのデファクト・スタンダードとなるものです。

3. ナレッジ・グラフ

3-1. ナレッジ・グラフとは

リレーショナル・データベースが生まれて久しく、今日では扱うことができないデータが多くなった現状では、柔軟性と複雑でデータ相互の関連をも扱うことができ、そのパフォーマンスによって急速に人気が高まっているのが、ナレッジ・グラフです。(Graph Database として知られています。) 企業は、顧客、製品やサービス、機能、マーケットなどの企業にインパクトを与える繋がりを理解できるようにデータを管理します。グラフは、これらの繋がりをストレートに表しており、ビジネスをドライブする関係を分析し理解することができます。知識(ナレッジ)は、会社にとって何が重要であり、そしてそれらが互いにどのように関連しているかといった背景を提供します。ビジネス知識(ナレッジ)の明確な表現は、異なるデータ・セット同士の参照を可能にします。ナレッジ・グラフが、ビジネス・データとビジネス知識を組み合わせることで組織のデータにより完全で総合的なエクスペリエンスを提供します。

総合的なデータ・エクスペリエンスにおいては、問題はエンタープライズ・データの管理方法にあります。企業では、個々のビジネス・ニーズに独自のデータ・ソースがあることが多く、その結果、やり取りがほとんどない独立して管理される”サイロ”となっています。この状況を打破したいと企業はもがき、ベンダからの甘い提案もあり、これまで間違った方法で対処してきました。それは、データ統合を 1 回限りのアプリケーションとしてアプローチしてきました。しかし、データ統合に問題が発生するたびに新しい統合アプリケーションを構築することは到底できず破綻してしまいました。

むしろ、企業内のデータへのアプローチ方法を再考する必要があります。データ資産は、特定のアプリケーションから離れて、それ自体の価値と考える必要があります。

企業がインサイトを得たい場合は、データについて全く異なる考え方を採用し、特定のアプリケーションから独立した、データ自体をリソースと考えなければなりません。データの利用は、企業全体(セールス、製品、顧客)および業界全体(規制、材料、マーケット)にわたるデータを紡ぐプロセスになります。この種のデータ・アーキテクチャをデータ・ファブリックと呼びます。

データ・ファブリックを実現する方法は、ナレッジ・グラフを展開して、企業データを意味のある方法で紡ぐことです。これにより、データ資産は、過去・現在・未来の複数のアプリケーションにも対応でき、データ・アーキテクチャを真にスケーラブルにすることができます。

ナレッジ・グラフは、企業全体でデータを効果的に共有できる基盤とアーキテクチャに貢献します。

3-2. ナレッジ・グラフと機械学習

Google 社で AI 開発の先頭に立つ AI の世界的権威であるレイ・カーツワイル (Ray Kurzweil) 氏がその著書 “The Singularity Is Near” (2005 年) で「2045 年、AI が人類の知性を上回り、ついに私たちは生物の限界を超えてシンギュラリティへ」と予言したときから AI ブームが到来した感があります。

2045 年、AI が人類の知性を上回るようになるかについては様々な議論がありますが、現時点で AI と呼ばれているものは、コンピュータが人間に取って代わるようなものではなく、過去に蓄積された膨大なデータと今流れているデータ (Fast Data) を駆使して学習分析する Machine Learning (ML) や Deep Learning (DL) だと考えられます。今いる情報経済時代では、このようにデータを利活用して正しい判断を如何に迅速に行うことができるかが、TCA (Total Competitive Advantage) につながり、企業や社会の存続に重要な意味を持ちます。

Dell Technologies の創業者であるマイケル・デル (Michael Dell) 氏も、"It's all about Data. Data is fuel for AI." (データがすべて。データは AI の燃料である) といってデータを活用する必然性について熱弁しています。日本においてもデータの利活用の重要性が認識され始め、さまざまな企業においてデータ・サイエンティストを育成する取り組みを始めたところもあり、また大学においてもデータ・サイエンティスト学部を新設するところも出ています。

エンタープライズ・データは、世界で最もビジネスで重要な資産ですが、現状はサイロ状態で、またデータは自社内(オンプレミス)だけでなく、パブリック/プライベート・クラウドなどにあるハイブリッド環境になっています。サイロやハイブリッド環境が、アプリ開発、データ・サイエンスとアナリティクス、レポートニング、コンプライアンス、AI の取組みの妨げになっています。

機械学習は、増え続ける膨大な情報のデータからより正確に予測を生成ようにデザインされたセルフ・ラーニング・アルゴリズムで、しばしば膨大かつ様々なデータ・セットからデータを取得する必要があります。

ナレッジ・グラフは、データと情報の統合されたコレクションで構成され、異なるデータ間の膨大な数のリンクが含まれています。

データ・ファブリックは、社内の全てのデータをサポートするプラットフォームです。このプラットフォームは、エンタープライズ・ナレッジ・グラフから形成され、統一統合されたデータ環境を作成します。

従って、データを利活用して正しい判断を迅速に行うことができ、TCA (Total Competitive Advantage) に繋がる環境を実現するには、ナレッジ・グラフで次世代のデータ管理であるデータ・ファブリックを構築することが必須となります。

以降、どのように YOUNite フェデレイテッド・データ・ユニフィケーションとナレッジ・グラフで次世代のデータ管理データ・ファブリックを実現するかをご説明いたします。

4. YOUNite フェデレイテッド・データ・ユニフィケーションとナレッジ・グラフで真のデータ・ファブリックを実現

データ・ファブリックに重要となるのは、グラフ技術と組み合わせたイベント・ドリブン・フェデレイテッド・データ・ユニフィケーションです。イベント・ドリブン・フェデレイテッド・データ・ユニフィケーションは、ビジネス・ユニットが、ユニット自身のデータを管理する必要があるという感性を持ちながら、データ層にわたり統合、ガバナンス、データのアクセスを行います。グラフ・データベースは、データ自体が重要であるように、データ間の関係を扱うように設計されています。それは、事前に定義されたモデルに制限を受けることなくデータを保持します。むしろ、データは最初に作られたようにデータは保管され -- 個々の項目が如何につながり、他と関連しているかを示しています。このセマンティック・データは、ナレッジ・グラフ (Knowledge Graph) を構築するために使用され、組織に対しトピックスについて、エンタープライズ・データ層すべてにわたる構造化データを使うサーチ・エンジン (Google のような) 構築を手助けします。

イベント・ドリブン・フェデレイテッド・データ・ユニフィケーション・プラットフォームの上にデータ・グラフ技術を置いて、企業は分散コンピューティング・エコシステムという現実に順応しながら、データ・

ドリブンの取組みを可能にするデータ・ファブリック・アーキテクチャを作り上げます。

4-1. データ・ユニフィケーションの特徴

イベント・ドリブン、フェデレイテッド・データ・ルーティング

フェデレイテッド・データ・ユニフィケーションは、企業の繋がりのないシステム間のイベント・ドリブン・インターフェースとなります。アダプタは、ソース・システムの変化（データ・イベント）を察知し、フェデレイテッド・ユニフィケーション・サーバにその変化を通知し、データ層で通信しているシステム間にデータを配信（ルーティング）するタスクを実行します。

リクエストは、ユーザのユニークな要求に対して様々なソース・システムのデータ・レコードを検索し、一つのフェデレイテッド・データ・レコードにアSEMBルされます。そして、どれが最善のソース・オブ・トゥールズであるか、またそのリクエストに満足したデータ・ガバナンスであるかに基づいています。

ガバナンス

ガバナンスは、データ層のソース・システムとソース・システム間のデータ管理プロセスです（すなわち、誰がロールやアプリケーションなどに基づいてデータ・セットにアクセスするか）。具体的には、ガバナンスはコンシューマ（API コンシューマまたは他のソース・システム）がデータにアクセスするかどうかを説明するために使用されます。ガバナンスは、どのようなデータの動きがアプリケーション間で許されているかを含みます。

また、特定のデータ・レコードに詳細なアクセス・コントロールを施し、それは、データ・レコード属性に許可/不許可を以下に設定：

- 一つのシステムから他のシステムへ
- API でフェデレイテッド・データ・アクセス・リクエストに応答

エンタープライズ-ワイド・データ・アクセス

開発者とデータ・ラングラは、全データ層にアクセスする必要があります。一つのインターフェースを通い、アプリケーション、ML、AI、分析とルシオンや GraphQL のような最新のインターフェースを使った REST API エンドポイントによるデータ・グラフ取組みを動かします。

データ・アナリストは、ビジネス・インテリジェンス(BI)のようなツールに供給するために一つの JDBC コネクタでエンタープライズ・ワイドにアクセスが必要です。

リネージュ

データ・リネージュ（Lineage、追跡）には、データのオリジン、データに何が起こり、データ層通って時間とともにどこに移動したかが含まれます。データ・リネージュは、データ分析過程で根本原因までエラーを追跡する能力を極力簡略化しながらもデータ・コンプライアンスを可視化します。

重複排除

重複排除（Dedupe）は、フェデレイテッド・データ・ユニフィケーション・プラットフォームの役割を言い、企業全体にわたり重複データ・レコードを特定し、レポートないしリンクします。企業全体にわたる共有

データの中央のクリーニング・ハウスのようなものとして、フェデレイテッド・システムは重複データをクリーン・アップし、データ・スチュワードによって手動で曖昧データを解決するケースの報告をする唯一の場所です。

データ・クウォリティ

フェデレイテッド・ユニフィケーション・プラットフォームにデータ・レコードをマッピングする前に、データ・クウォリティ・ツールによりがデータをクリーン、エンハンスとエンリッチできます。ほとんどの企業は、いまあるデータ統合とデータ・クリーニングに大幅に投資をしていますので、フェデレイテッド・データ・ユニフィケーション・プラットフォームは、他のベンダのデータ・クウォリティ・ツールと統合して使用できることが必須となります。従って、企業はこれまでのデータ・クウォリティ投資を失うことはありません。

4-2. データ・ユニフィケーションでデータ・ドリブン企業に

フェデレイテッド・イベント・ドリブン・データ・ユニフィケーションは、企業の構造化データ層を統合し、データ・ドリブンを可能にします。

データ・ドリブン・エンタープライズ

企業が競合優位であり続けるためには、ビジネス・パフォーマンスを改善し、顧客やパートナーとの持続可能な価値を生むデータ・ドリブンになるべきです。

ケース：データ・ドリブン企業は、Google で Web サーチするようにビジネス全データ層を検索したい
ゴール：グラフ・データがエンタープライズ・サーチ、AI、ML の取り組みを強化する

フェデレイテッド・データ・ユニフィケーション

課題：それが一回限りのアプリケーションであるとき、一回の統合で得られる価値は限られます —
幾度もそれが繰り返されれば、数多くの個々の統合が企業全体にばらまかれる結果となります
ゴール：ゴールは、マルチクラウド・エコシステムにわたるすべてのデータ層をシームレスに見ること
e.g. シングル・ペイン・ビュー (Single Pane View)

解決策、エンタープライズ・ワイド：

- フェデレイテッド、イベント・ドリブン・データ・ルーティング
- データ・ガバナンス
- データ・リネージュ
- データ・アクセス

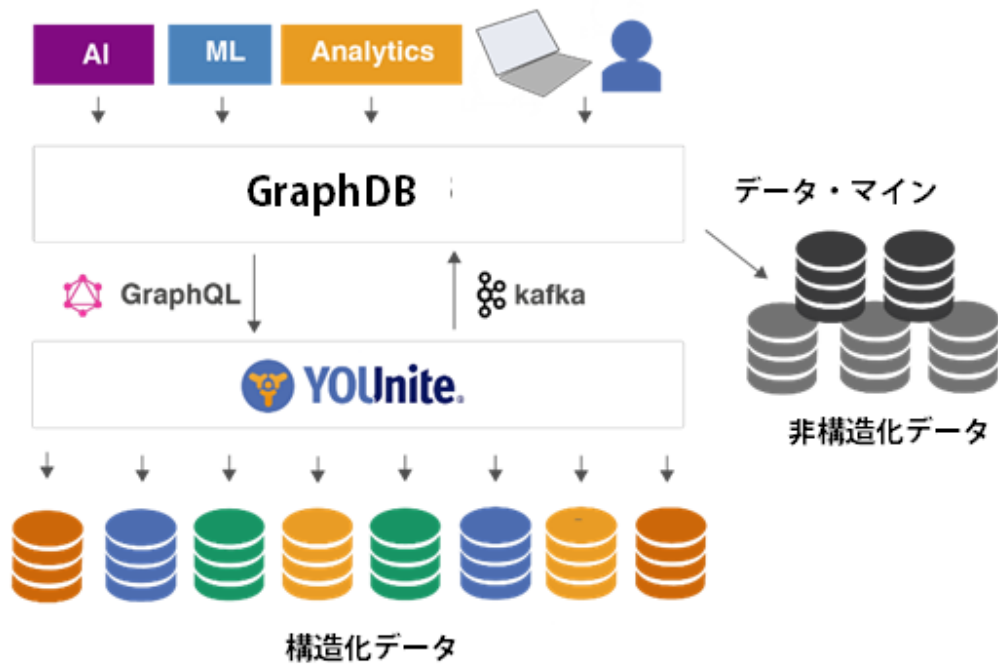
以下を推進：

- フェデレイテッド・データ・アクセス e.g. GraphQL
- エンタープライズ・ワイド・データ・ファブリックの構築 (Graph DB)
- データ・ファブリックが、エンタープライズ・ワイド AI や ML にフィードする

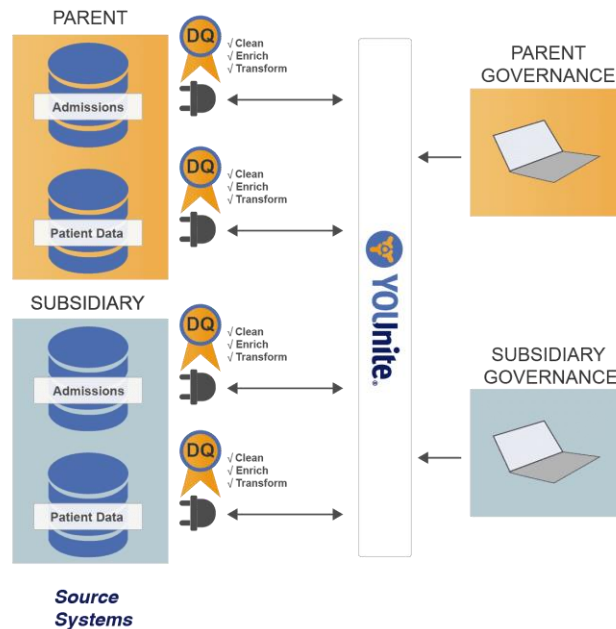
データ・ユニフィケーションは、データ統合コストを削減、かつ生産性全般を改善して、データ統合の幅 / データの関連性を高めます。

データ・ファブリックの例

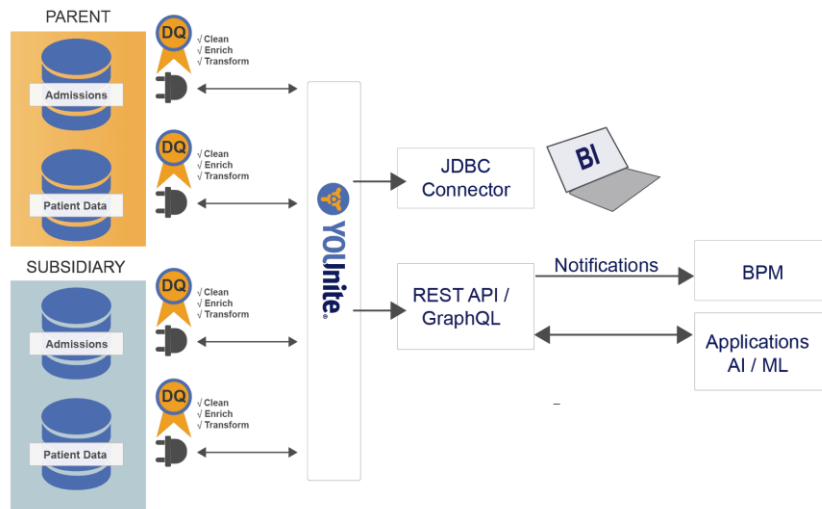
データがより統一されるようになると、企業全体のナレッジ・グラフを構築できます



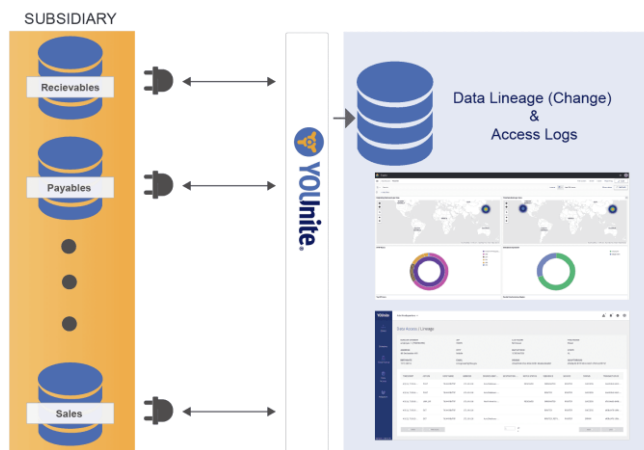
- 企業のデータ層にわたる構造化データのフェデレイテッド・イベント・ドリブン同期
- リアルタイムで適用されるガバナンス -- ビジネス・ユニットは、ガバナンスをコントロール



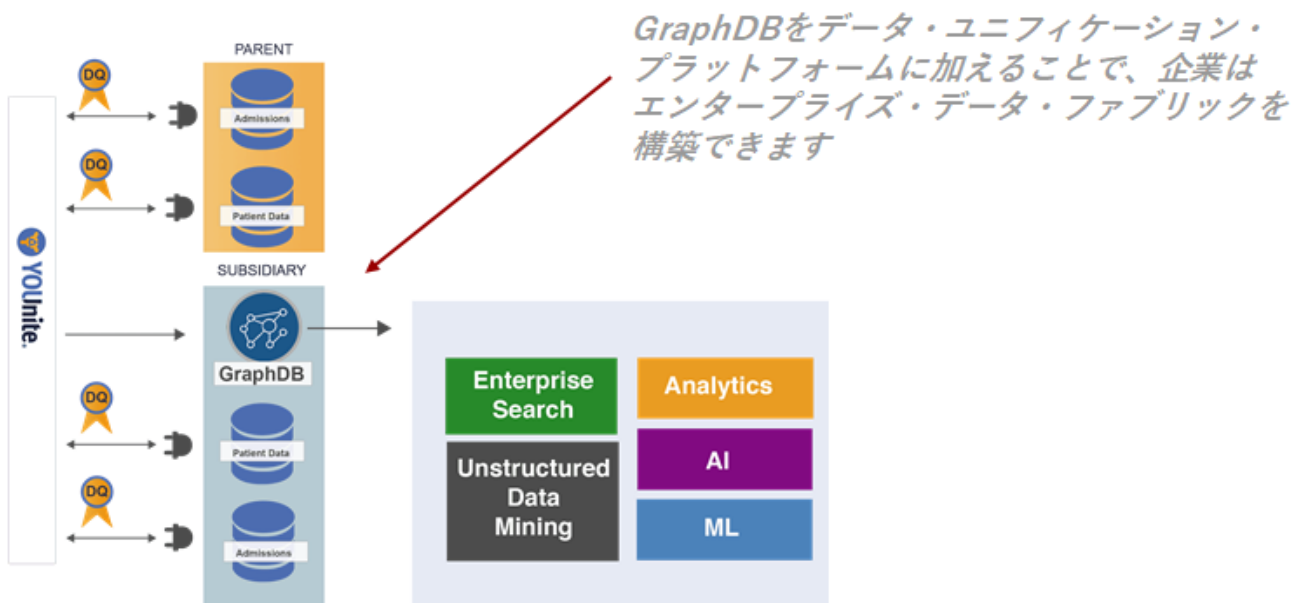
- フェデレイテッド・データ・アクセスは、全データ層にわたるデータのシングル・ペイン・ビューを提供



- データ・リネージュ(履歴)とデータ・アクセスをシングル・ペイン・ビューで



- ビジネス・ユニットは、自身のデータ・ドリブンの取組みを共有、購入または構築できる



5. 最後に

データ・ファブリックは、次世代(Next Generation)のデータ・マネジメントとして 2021 年に入り、これに関する記事を数多く目にするようになっていきます。

Gartner が、データ・ファブリックを今後数年間で重要視するデータ管理に関連する 10 のトレンドの一つに挙げているだけでなく、IBM もデータ・ファブリックは、企業の資産であるデータをより有効に活用することを可能し、担当者が適切なデータに適時アクセスできるようにすることで、データ管理にかかるコストやリスクを削減することができるとデータ・ファブリックを強く勧めています。

これまでのデータ管理というと、サイロ化しているデータから ETL などのツールを使って中央にデータを集めて管理する方法でした。しかし、企業が管理しているデータ・エコシステムは非常に複雑になり、複数の地域にわたりオンプレミス/クラウドのシステムが分散していることも珍しくありません。

ソーシャル・メディアやブログ、センサー、IoT デバイスなど多数のソースからデータが集まり、業務ごとに専用ソフトウェア利用しています。さらに、GDPR (General Data Protection Regulation : 一般データ保護規則) に代表されるデータ保護やプライバシーに関する規制が各国・各地域で施行されるようになり中央型のデータ管理では対応できないのが現状です。

データを中央に集めて管理するのではなく、自社内、パブリック/プライベート・クラウドにあるデータを結びつけ、エコシステム全体を一つの窓(シングル・ペイン・ビュー)から管理できる手法が求められるようになっていきます。データは、各部門・地域のソース・システムに置いて、ガバナンスを効かせながらデータのリアルタイム同期、データ・リネージュとともに最新のデータを使って分析が行えるとように必要なところからデータをアSEMBルできる機能やデータが生まれた、またはアップデートされたことをトリガとしてビジネス・プロセスを連動させる機能が求められるようになっていきます。

YOUnite イベント・ドリブン・フェデレイテッド・データ・ユニフィケーションは、ユーザとともに作り上げたもので、ユーザがいま求める機能をすべて実現させています。また、ユーザの声を取り入れて常に進化し続けています。ロックインさせない、常に best of breed を取り込む。

YOUnite は、製品をリリースした当初どのカテゴリの製品といえ良いのかを IDC カナダのスチュワート・ボンド氏(ディレクタ、インテグレーション・ソフトウェア)に相談しました。「YOUnite は、現時点で世の中にあるデータ管理製品が必要とすることを成し遂げているが、より先進的にリアルタイム同期にフォーカスした方法で行い、エコシステム全体の構造化データを統一している。さらに、強固なガバナンスとパーミッション・モデルを使ってビューの統合を提供している」と述べ、当時のデータ管理製品のかなり先を走っているとの評価をいただきました。当時は、データ・ファブリックという言葉が世の中に存在していませんでしたが、その後ユーザの声とともに進化を続け、イベント・ドリブン・フェデレイテッド・データ・ユニフィケーションとナレッジ・グラフでデータ・ファブリックを実現しています。YOUnite は、ネクスト・ジェネレーション・データ・マネージメントであるデータ・ファブリックの先駆者です。

さらに詳しくお知りになりたい方は、「今なぜデータ・ユニフィケーションとファブリック?」、「YOUnite カスタマバリュ」、「What is DU? Why you need it?」の日本語資料も併せてご参照ください。

その他、ご不明な点やご質問などがございましたら、お気軽に YOUnite 担当営業または info@younite.us までお問合せください。