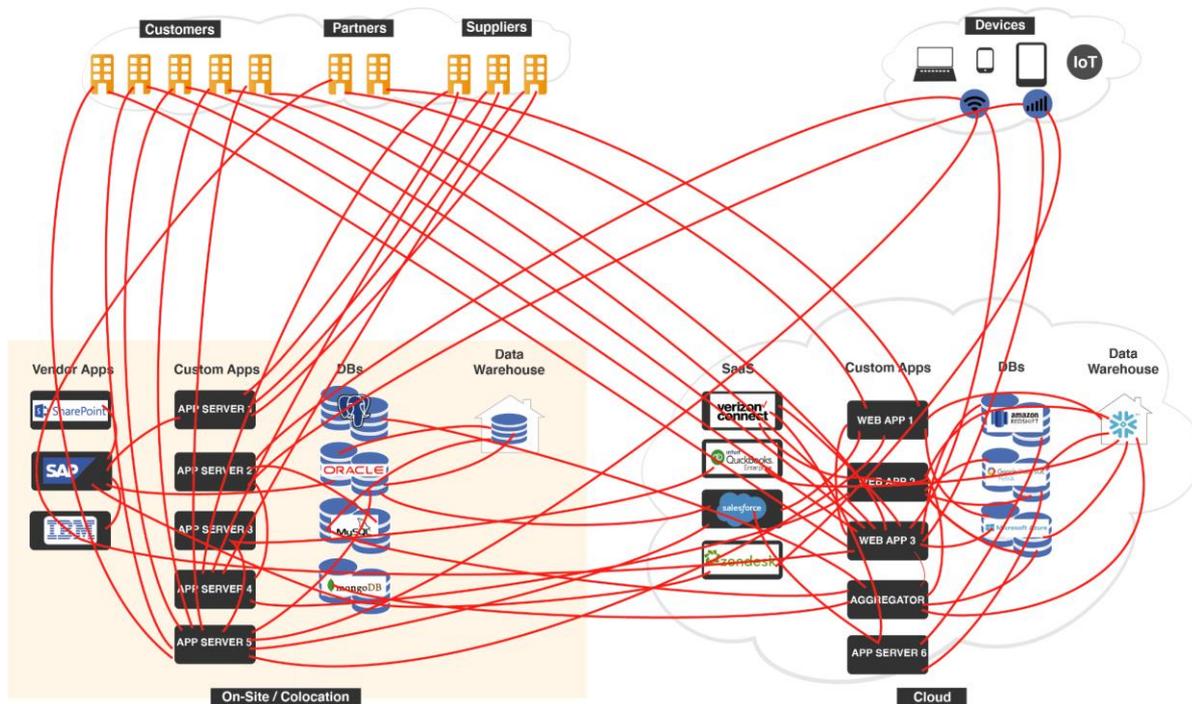


今なぜデータ・ユニフィケーションとファブリック？

企業が本当にデータ・ドリブンであるためには、ユニファイド・データ・インテグレーション戦略にコミット

はじめに

一般的に、システム統合は、個々の統合を別アプリとして扱ってきたサイクルを繰り返してきました。開発者は、その仕事に適切な技術を用いて努力してきました。しかし、繰り返す内に総合的データ統合戦略がなく、企業のデータ層インテグレーションの接続形態は以下と似たようなものになっています：



IDGによると少なくとも1,000人の従業員がいる200以上の企業の調査では、400以上の異なるデータ・ソースから線が引かれることが確認された¹

競争力を維持するためには、多くの企業はより一層データ・ドリブンでなければならないと理解していますが、多くのデータがアクセスできない状況にあります。理由は、データが異なるビジネス・ユニットで所有されており、企業のデータ層にわたりサイロ化しているからです。このため、企業データの極めて多くが利用されないままになっています（IDCの調査によると42%²）。

多くのシニアIT専門家は、より緊急のビジネス課題があるという理由でデータ・ユニフィケーションの取組みを中断したり、取り止めたりしました。もっともらしい理由は：

¹ King, Timothy. "Companies Are Drawing from over 400 Different Data Sources on Average." Solutions Review, 1 Nov. 2019, See <https://solutionsreview.com/data-integration/companies-are-drawing-from-over-400-different-data-sources-on-average/>

² Condon, Stephanie. "Enterprises Are Collecting More Data, but Do They Know What to Do with It?" ZDNet, ZDNet, 15 July 2020, See <https://www.zdnet.com/article/enterprises-are-collecting-more-data-but-do-they-know-what-to-do-with-it/>

-
- シニア・マネジメントが取組みを優先させず、その取組みを完成し維持するために必要なリソースをも想定していなかった -- これは、開発者や IT スタッフが取組みに参加しようとするインセンティブも与えていなかった
 - 見込みのない独自で高額のデータ・インテグレーション・ツールとプラットフォームにロックインしていった

分析の先へ

意思決定者とデータ・サイエンティストは、より優れた分析のためにデータ層にわたるスムーズなアクセスを必要とするだけでなく、より分析能力向上のために：

- 機械学習 (ML)
- IoT 用 AI と他の機械知能
- 検索 - ユーザは Google で Web サーチするように企業データを検索しようと思います

データ・ファブリック、データ・ユニフィケーション & ナレッジ・グラフの世界へ

ガートナーは、データ・ファブリック・アーキテクチャを以下のように説明しています：

“新たに出現したデザイン・コンセプトである”データ・ファブリック”は、絶え間なく出現するデータ・マネジメント・チャレンジへの確固たるソリューションになり得る。これまでの挑戦といえば、ハイコスト・ローバリュのデータ・インテグレーション・サイクル、初期のインテグレーションに対する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・シェアなどである³”

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、データは最初に作っ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⁴。このセマンティック・データは、ナレッジ・グラフ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。

³ Gupta, Ashutosh. "Data Fabric Architecture Is Key to Modernizing Data Management and Integration." Smarter With Gartner, Gartner, 11 May 2021, www.gartner.com/smarterwithgartner/data-fabric-architecture-is-key-to-modernizing-data-management-and-integration/.

⁴ "DBMS A Brief Look into Database Models, Data Warehousing, Mining, Trends and an Overview on Data Flow Diagram", Course Hero, Acharya Institute of Technology

データ統合 vs データ・ユニフィケーション

旧式のデータ統合では、通常はバッチで動作し（ETL など）、しばしばデータ・クォリティとビジネス・プロセス・フローがそれらに付随しています。旧式のデータ統合を使ったデータ層にわたる全般的なデータ統合レベルは、データ・ユニフィケーションと比較するときわめて低いものである。

データ・ユニフィケーションは、旧式の統合手法と比較するとデータ層全体により深い統合を提供し、データに付加価値を提供し企業をよりデータ・ドリブンにさせます。

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・システムの変更（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・サーバに変更を通知し、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SEMBL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合致したデータ・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）。詳しくは、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して、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（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・ツールがデータをクリーン、エンハンスとエンリッチします。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・ツールとインテグレート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。

統合 vs ユニフィケーションの比較

エンタープライズ・ワイド機能	旧式統合	フェデレイテッド・データ・ユニフィケーション
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	No	Yes
ガバナンス	No	Yes
データ・アクセス	No	Yes
データ・リネージュ	No	Yes
重複排除	No	Yes
データ・クウォリティ	おそらく	必要な場合

CIOの望むことを提供

別な見方をすると、CIOはすべてのマルチ・クラウドにわたるデータ層にシングル・ペイン・ビュー（一つの窓）を望んでいます。



「データ・ストレージ管理の挑戦へのキー・ソリューションは、ビジネス・オーナーが保存されたデータをどのように見るか。その答えは見る — そのすべてを — シングル・ペインビューでデータの民主化を超え、ストレージ・ユニフィケーションへ。CIOはシームレスな方法でマルチ・クラウド・エコシステム全体を見るべきです」とそのレポートは述べています⁵。

データ・ユニフィケーションとデータ統合のコストを比較

データ・ユニフィケーションに移行することの注目すべき成果の一つは、企業すべてにわたる全データ層の接続性を向上させながらもコストを削減できることです

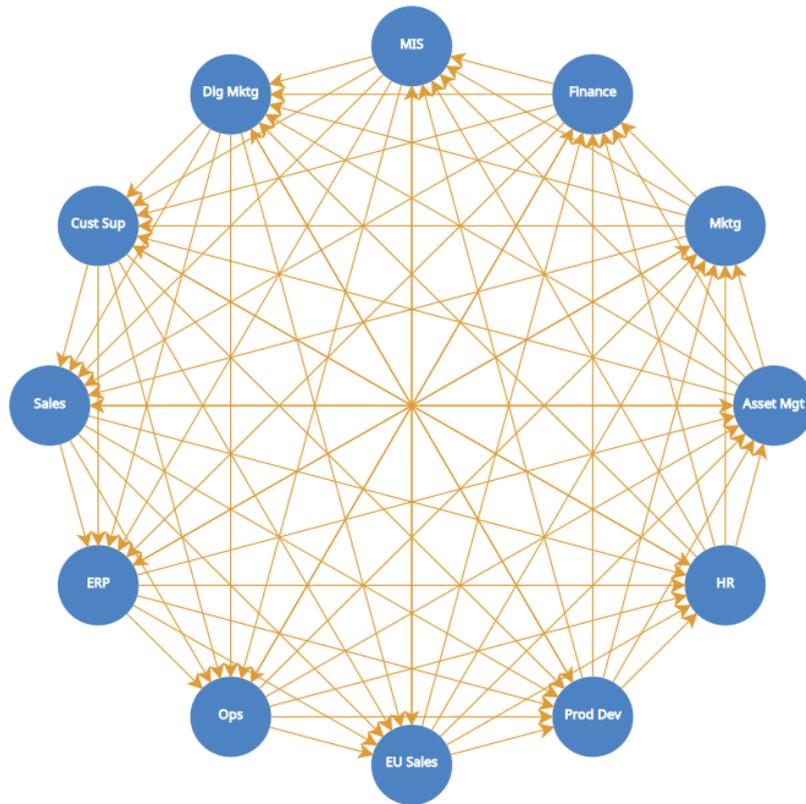
イベント・ドリブン・データ統合を設計するための特筆すべき重要なデザイン・パターンは、ハブの利用です。旧式の統合デザインは、ポイント・ツー・ポイント、またはハイブリッド・アドホックなアプローチです。多くの企業では、数十の統合が実施され、かつ今後数年間でさらに多くの統合が計画されています。データ層が統一されない限り、時代に即し競争力を維持するための企業の能力は、困難になるのみです。

12システムを統合することを計画している企業のデータ層を例に考えてみます。データ・タイプの違いは無視します。CIOは、すべてのシステムが互いに統合する機能を持つことを望んでいますが、控えめに考えて、条件を減らし、それぞれのシステムが他の半分(6)と接続することを考えましょう。⁶

⁵ Condon, Stephanie. "Enterprises Are Collecting More Data, but Do They Know What to Do with It?" ZDNet, ZDNet, 15 July 2020, <https://www.zdnet.com/article/enterprises-are-collecting-more-data-but-do-they-know-what-to-do-with-it/>

⁶ Reeve, April. "Canonical Modeling - Best Practice Message Modeling for Data Integration", DATAVERSITY, EMC Consulting, April 11, 2013, , <https://youtu.be/mRthhbAtj7o>

旧式データ統合



旧式データ統合では、統合数は急増します -- データ・ユニフィケーションが提供する機能のサブセットを提供するのみである。

システム数 : 12

半分のカバレッジに必要な統合数 : 66

データ・ユニフィケーション



データ・ユニフィケーションでは、すべてのシステムが互いに統合され、エンタープライズ・ワイドな機能すべてが上記のようにカバーされます。

システム数 : 12

フル・カバレッジに必要な統合数: 24

コスト比較

構築導入費用が\$12,000、年間保守費用が\$8,000と仮定します。5年間で12システム統合の費用概算は以下の通りです：

	旧式	データ・ユニフィケーション	節約
構築導入費用	\$792,000	\$288,000	\$504,000
年間コスト	\$528,000	\$192,000	\$336,000
5年間の総コスト	\$3,432,000	\$1,248,000	\$2,184,000

オープン vs ベンダ・ロックイン

最適な技術を選択する自由

歴史的に、もし企業全体の統合を望んだとき、非常に複雑で過度な機能のプラットフォームを持つ一つの独自技術のベンダに巧みにロックインされ、最新の技術を活用できない状況に追い込まれました。

データ・ユニフィケーションは、GraphDB、セマンティック・モデリング、データ・クリーニング、データ・クウォリティ、ビジネス・プロセス・マネージメント、データ統合のような多くの関連する技術と密接に連携しています — そして — これらの技術は、絶えず進化しており、一つのベンダでは顧客にすべてを含んだ最高のソリューションを提供することはできません。

多くのベンダは、独自のプラットフォームに顧客をロックインする画一的なアプローチを取っています。しかし、フェデレイテッド・データ・ユニフィケーションでは、顧客にデータ・ファブリック・アーキテクチャにインテグレーションしたい技術を選択する自由を与えつつ、核となるデータ統合、ガバナンスとデータ・アクセス機能を提供します。

最後に

企業が時代に即し競争力を維持するためには、全データ層にわたるすべてのデータで利用できるツールを意思決定者に提供しなければなりません。収集されるデータの量と企業内に増え続けるデータ・サイロの数に、旧式データ統合戦略ではこれらの要望を満たせないことを企業は認め、これに代わる手段となるデータ・ユニフィケーションの取組みが優先されなければなりません。