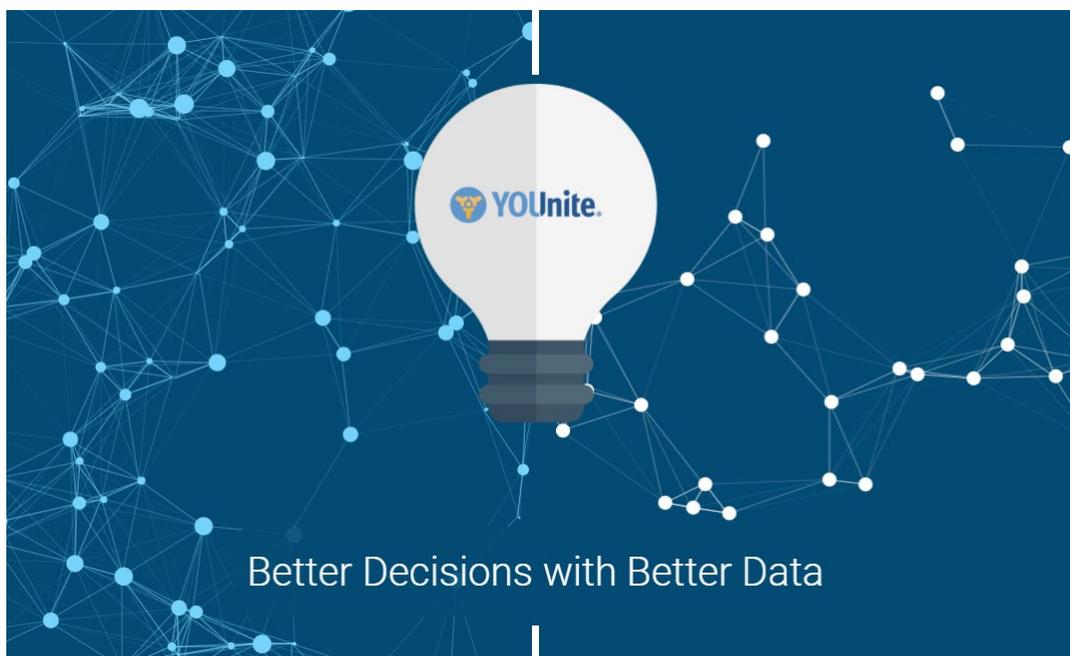


What is Data Unification? Why you need it?



版数	発行日	改定内容（概要）
Version 1.0.0	2021年2月02日	初版発行
Version 1.1.0	2021年2月14日	項目追加（DQ 関連、Cross Reference）など
Version 1.1.1	2021年3月22日	一部記述更新、図の差替、レイアウト調整
Version 2.0.0	2021年5月18日	MDM から Data Unification に
Version 2.1.0	2021年5月22日	加筆、修正、調整
Version 2.1.1	2021年6月6日	修正、調整

目次

1. DU (Data Unification)とは	2
ビジネス・バリュ	5
2. YOUnite Federated DU.....	11
2.1 データ・ドメイン（Data Domains）	15
クロス・ドメイン（Cross Domains）	18
2.2 Zone.....	20
2.3 Adaptor.....	21
2.4 YOUnite Federated DU のしくみ（サマライズ）	22
2.5 データ・イベント・ノティフィケーションズ	23
2.6 データ・リネージュ	24
2.7 データ・クウォリティ（Data Quality）	25
3. 今後に向けて	34

1. DU (Data Unification) とは

我々が現在いる経済は、農業経済、工業経済ではなく、情報経済です。これまでの経済でもそれぞれ鍵になっていたものがあるように、情報経済においてはデータが最も重要な鍵となります。データをいかに活用してビジネスにつなげるかが、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 Michael Dell 氏も、"It's all about Data. Data is the fuel for AI."（データがすべて。データは AI の燃料である）といてデータを活用する必要性について熱弁を振るっています。

すなわち、データは情報経済のコアコンピタンスです。

現状は多くの企業で数多くのシステムのデータがサイロ化され、同期すべきデータが同期されないため、どのデータが正しく、どのデータが最新であるかまったくわからない状況になっています。企業においては、データは爆発的に増えつづけ、その膨大なデータをインタラクティブに処理し、ビジネスに活かそうとするニーズが高まっていますが、BI、アナリティクス、ML/DL、AI においてデータ・サイエンティストの全作業に占めるデータの準備作業は 80% にものぼると言われています。そのような時間の余裕はありません。

このようなデータ準備作業にかかる時間やコストは膨大なものになります。企業は迅速な意思決定をすることは必然ですが、意思決定するうえでのデータが正しいデータでなければ、正しい意思決定は不可能です。また、リアルタイムにデータを利用できれば、より正確な判断につながります。1ヶ月前のデータを使って分析している企業と、今流れているデータ（Fast Data）を使って分析している企業とではビジネスに大きな差がでてきます。

また、Harvard Business Review 誌によると、「企業のデータのわずか 3% しか基本的なデータ品質基準を満たしていない（Only 3% of Companies' Data Meets Basic Quality Standards）」、「平均して、新しく作成されたデータ・レコードの 47% に、少なくとも 1 つの重大な（作業に影響を与える）エラーがある（On average, 47% of newly-created data records have at least one critical (e.g., work-impacting) error）」と述べています。

このような状況を改善するために、DU のビジネス・バリュ（詳細は 5 頁から記述）が認識され始めました。しかしながら、そもそもデータ基盤となる手立てがなく、各企業はその場しのぎの対策として、それぞれのシステム間のデータをそのつど連携させるように対処してきました。日本のある大手企業では、社内に 100 以上のシステムが稼働し、使われているデータがサイロ化されているような状況では、とても対処できるものではありませんでした。

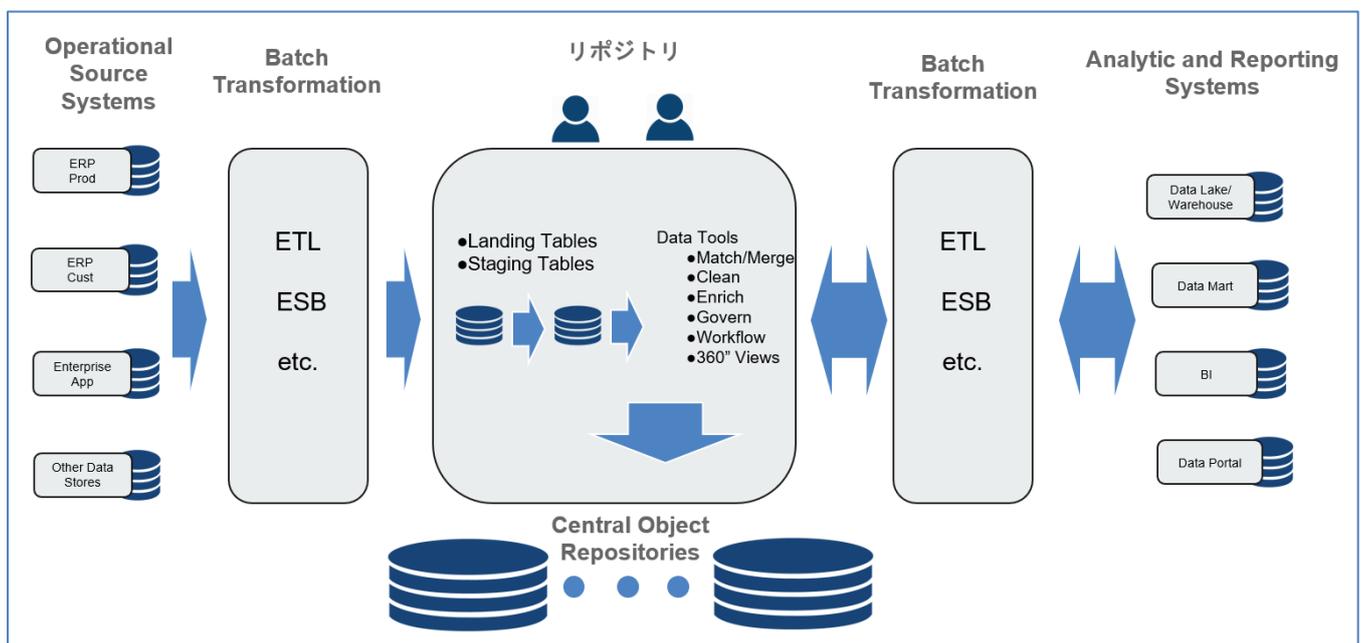
さらに、IT ベンダは MDM (Master Data Management) 市場に目をつけ、MDM とはさらさら言えないものをお客様に提案し、お客様の混乱を招くアプローチをとってきました。それは、ETL などのツールを使ってデータを中央に集め(データ統合)、その集めたデータにアクセス・コントロールを施すことが MDM であるかのように、お客様をベンダのソリューションにロックインさせるビジネス手法でした。

この手法は、お客様が望むリアルタイムにデータを利用したいという要望に ETL を勧めた時点で対応できません。さらに、GDPR をはじめとするデータ保護法にも抵触するものでした。

そこで、ある IT ベンダは ETL を使った手法ではデータ保護法などに抵触することから、ETL を用いない方法として、ソース・データを参照して分析できる手法となるデータ仮想化を提案してきました。

しかしながら、この方法ではデータやビジネス・フローそのものに目を向けたものではなく、データの更新性やデータの正確性、一貫性を保証するものではなく、ビジネスとの関わりはできません。

[Fake MDM]



本来的な MDM とはいかにあるべきでしょうか？

Gartner は、「マスタ・データ・マネージメント (MDM) は技術が可能にした規律であり、企業の正式な共有マスタ・データ資産の均一性、正確性、スチュワードシップ、セマンティックを確実なものにするもの」と述べています。

正式な共有マスタ・データ資産とは、どのようなものでしょうか？

マスタ・データは、しばしばデータ・ドメインにおいて、情報のゴールデン・レコードと呼ばれます。データ・ドメインは業界によって変化します。例えば、製造業であれば、顧客、製品、サプライヤや材料といったものになります。銀行であれば、顧客、口座情報、商品というものになるかもしれません。医療機関では、患者、医療機器、消耗品などになるかもしれません。これらの情報が各システムでガバナンスが効いたかたちで同期される必要があります。したがって、マルチ・データ・ドメインが必然となります。

マスタ・データは様々なデータ・ドメインで処理されるトランザクションは含みません。したがって、信頼できる情報源が常に最新の状態でなければならないのは当然のことです。

これまでの MDM ベンダが言っていたデータ・インテグレーション (Data Integration : データ統合) ではなく、これからはデータ・ユニフィケーション (Data Unification : データ統一) が必要となります。

これはユーザの声をもとにユーザとともに作り上げたもので、データ・インテグリティ (Data Integrity : データ健全性) を実現するものです。

つまり、DU とは、データを組織全体にわたりビジネスで一貫して利活用するためのデータ基盤でなければなりません。

適切なデータ基盤がビジネスと相まって動くことがゴールとなりますが、依然として、ほとんど認識されていないのが現状ではないでしょうか？ というよりも、これまで適切なデータ基盤が IT ベンダにより提供されてこなかったことが現実です。

データ基盤をしっかりとしたものにするために、以下のようなことが必要になります：

- ・信頼性の高い最新のデータを使い、ビジネスの正確な分析と優れた予測を行うことができる
- ・バッド・データによるビジネスへの悪影響を避ける
- ・複数のビジネス・ユニットが正確なデータを利用できる
- ・GDPR などのデータ保護法を順守できる
- ・正確なデータを使うことで、問題の発生を避けコストを低減する
- ・ビジネス・プロセスを改善するために正確なデータの準備
- ・企業の合併や組織の変更で、各組織が利用するデータの一貫性を保ち、統合された組織でのデータ運用を最適化できる
- ・データのサイロ化を避ける

加えて、データ基盤をインプリメントするにあたっては、既存環境に影響を与えることなく、かつ既存システムを止めることなく、段階的実装（フェイズド・インプリメンテーション）できるものであるべきです。また、GDPRをはじめとする各国のデータ保護法に抵触することがないようにしなければなりません。ソース・データはその地域や組織内で管理されるべきで、それらのデータの同期をとる際も中央にデータを集めることなく、ガバナンスされたものが必要となります。

ビジネス・バリュ

経済産業省が2018年に発表した研究資料「DXレポート～ITシステム『2025年の崖』の克服とDXの本格的な展開～」において、日本でDigital Transformation（デジタル・トランスフォーメーション、DX）が進まなければ、2025年以降に年間12兆円の経済損失が発生する可能性があるとして警告しています。

この大きな経済損失を『2025年の崖』と呼び、各企業にDXを速やかに推進するように求めています。しかしながら同省は1年後に発表したレポートにおいて“多くの企業ではDXを進めるために、デジタル部門を設置する取り組みが見られる”と評価しながらも、“実際のビジネス改革には繋がっていない”とDXが進んでいない現状を嘆いています。

では、何故DXが進まないのでしょうか。それには、いくつかの理由があると考えます。

データのサイロ化

DXが進まない理由の一つに、各企業の組織が縦割りで部門間でのデータが連携されていないことにあります。顧客情報一つをとっても、営業部門、マーケティング部門、カスタマ・サポート部門など各部門でそれぞれ別々のシステムで管理・運用しています。データ自身をそれぞれの部門で管理することは良いことですが、データの同期がなされていないので、どのデータが最新で正確なものかがまったくわかりません。このようなデータのサイロ化を解消しない限り、DXを本格的に展開することはできません。

測定可能なKPI（Key Performance Indicators）を策定できない

DXを進めるにあたっては、全社で取り組まなければなりません。目標や望ましい結果の策定やどの部門がイニシアティブを取るのか、進捗状況をどのように測定すればよいかなど、各部門の思惑などもあり、前に進まない状況となっています。既存のITシステムがビジネス・プロセスと密結合していることが多いため、既存のシステムの問題を解消しようとすると、ビジネス・プロセスそのものの刷新が必要となります。

経営陣が理解できていない

経営層は、プロジェクトを承認するにあたって ROI を求めますが、ROI だけでは DX を推進する上での正確な絵を描くことはできません。Gartner は、「MDM だけでは、ワールド・クラス、またはグローバルに通用するビジネス・ストラテジを作成するには十分ではありませんが、MDM のサポートなしではそのようなストラテジを作成、維持することは困難です」とも述べています。

企業は、爆発的に増加するデータを活用しきれずに、DX を実現できず、デジタル競争の敗者となる恐れがあります。

『2025 年の崖』を飛び越え、DX を推進する鍵となるのは“データ”であり、データを利活用できるしっかりしたデータ基盤を構築することが急務となります。

しかしながら、IDC Japan の 2020 年のデータ管理についての調査において、およそ 80% の企業のデータ管理が未成熟という結果がでています。また、同じく 2020 年に公表された日本オラクルの IT 環境のサイロ化の調査においても、73 % の企業が「サイロ化が原因で、ステークホルダが必要とするデータを提供できていない」という結果がでています。得意先の取引状況を求められても、そのデータの提出に何日も要するようでは、データをビジネスに活かしているとは言えません。

データを利活用できるデータ基盤を早急に構築しなければなりません。それは、これまで MDM だけではなく、データ・ユニフィケーションを導入すべきです。

データ・ユニフィケーションという言葉、技術にまだ馴染みがない方も多いのではないかと思います。データ・ユニフィケーションについていまい少し記述してみたいと思います。

データ・ユニフィケーションは、さまざまなシステムで扱われているデータをそれらの単一のソースに統一する非常に複雑なプロセスであり、多くの大規模な組織が直面している根本的な課題の 1 つです。ここに優れたデータ基盤があれば、ユーザはこの複雑なプロセスを自由に利用することができます。データ・ユニフィケーションは、IT エコシステムには不可欠であり、また信頼できるデータ基盤に、意味のあるビジネス洞察や高い生産性への波及が期待できます。時代遅れのアプローチである ETL によるデータ統合では、このようなビジネス・メリットを期待することができません。

ETL などを利用したデータ統合は、データが静的で比較的少ない場合には有効であったかもしれませんが、情報経済の時代に企業は膨大な量のデータにアクセスすることになります。

2014年のEMCのレポートによると2013年に作成されたデータ量は4.4ゼタバイトで、2020年には世界のデジタルデータの年間量が44ゼタバイトに達するとの記載があります。このようなデータ爆発は、あらゆる業種の全ての企業の環境が動的に変わりますが、誰しものデータから得られる機会を最大化できるわけではありません。データがあるだけでは何も得られません。データが存在していることと、そのデータを活用してビジネス上の意思決定を行うことには大きな違いがあります。

データを利用して意思決定を行いたいとETLによるデータ統合の手法を導入した試みも上手くいったところはありませんでした。このような試みの背景には、データ・レイク（Data Lake）という言葉が流行ったこともあるかもしれません。しかし、データ・レイクがデータ・スワンプ化（Data Swamp 泥沼化）しているところがほとんどではないかと思われます。

データ・ユニフィケーションは、ETLによるデータ統合とは異質なものです。

データ・ユニフィケーションとは、組織が所有する複数のシステムとフォーマットからなるあらゆるデータを統合し、一つのソースとして扱うことができるように基盤を作ることです。同期すべきデータをデータが追加、変更された時点（データ・イベントが発生）で、ガバナンスのあるかたちでリアルタイムに同期させ、それらをSingle Pane View（シングル・ペイン・ビュー（単一の窓））から確認できるものです。

サイロ化されたデータを持つシステム間では、そのデータから得られる結果は不透明です。統一されたデータではサイロ化が全て解消されます。

データ・ユニフィケーションは、顧客にかかわるデータを縫い合わせてクロスセルやアップセルの機会を創造し、顧客を繋ぎ止めることやサプライヤの統一した見通しを生み出しコスト削減を実現したり幅広く活用できます。たとえば、金融サービスの企業では、データ・ユニフィケーションによって顧客の統一したビューを構築し、迅速にセールス・チャンスを見逃さず可視化し、潜在的なリスクを特定することもできます。製造業では、データ・ユニフィケーションを活用してサプライヤをネットワークし、掌握し、そしてサプライヤ、部材、部品を分類することにより調達業務を最適化できます。

データ・ユニフィケーションでは、既存環境やワークフローを変更することなく統一されたデータ、クリーニングを実現できますので、あらゆる分野において大きな効果が期待できます。

データ・ユニフィケーションは、既存システム環境をあるがままに、IT全体の投資を最大化することでお客様が望む環境を実現することができます。ビジネス・パフォーマンスを低下させる

データの冗長性や不整合をなくし、真に統一することでビジネス・プロセスを極限まで効率化することができます。

データを活用してビジネス・インサイトを得る企業では、毎年平均 30%以上の成長が見込まれています。また、データをうまく利用できない競合他社から年間\$1.8T（約 200 兆円）のビジネスを奪い取ることができると予測されています（Forrester より）。

ゼネラル・エレクトリックでは、グローバルに 80 もの異なるシステムがあったために、購買システムのデータ・ユニフィケーションに何年も費やしました。

このような状況で、うまい案件を狙おうとしている購買担当者は、過去の記録からのデータしか見ることができず、すべての GE グループ企業内の最新の取引価格の情報を見ることができませんでした。データ・ユニフィケーションに時間を要しましたが、結果的に年間\$1.0B（約 1,100 億円）の節約になったと見積もっています。

データ・ユニフィケーションによる無形のビジネス価値は以下のようなものが挙げられます。

GDPR に代表されるデータ保護法

[Major GDPR fine totals]

	€
2018	400,000
2019	440,515,407
2020	142,426,406
TOTAL	€ 583,341,813

GDPR が 2018 年に施行され、すでに 3 年が経過しましたが、これまでのところ約 5 億 8 千万ユーロ、日本円で 700 億円を超えるペナルティが課せられています。データ保護法やコンプライアンス違反による企業への打撃は相当なものになります。しかしながら、日本の企業は対岸の火事と捉えているように見えます。日本に飛び火してから対策を講じても手遅れです。

イノベーションの進化

ある調査会社によると、DU を利用している組織がデータの検索に使う時間が 1.3 時間であるのに対し、DU を持たない組織は 6.2 時間を費やし、約 5 倍にもなると報告しています。DU を導入することにより、データのサイロ化が解消され、各部門のデータが同期され、最新の状態を保ちます。

より良いデータの流れを実現

企業が持つデータは膨大で、日々増大しています。多様化する社会において顧客ニーズの変化を察知して対応しなければなりません。そのためにも、企業に蓄積された膨大なデータと

いま流れているデータ（Fast Data）を利活用し、対応できる基盤が必要になります。データ基盤を持たない企業では、即座に対応ができません。データ基盤を持つ企業と持たない企業との差は歴然としています。

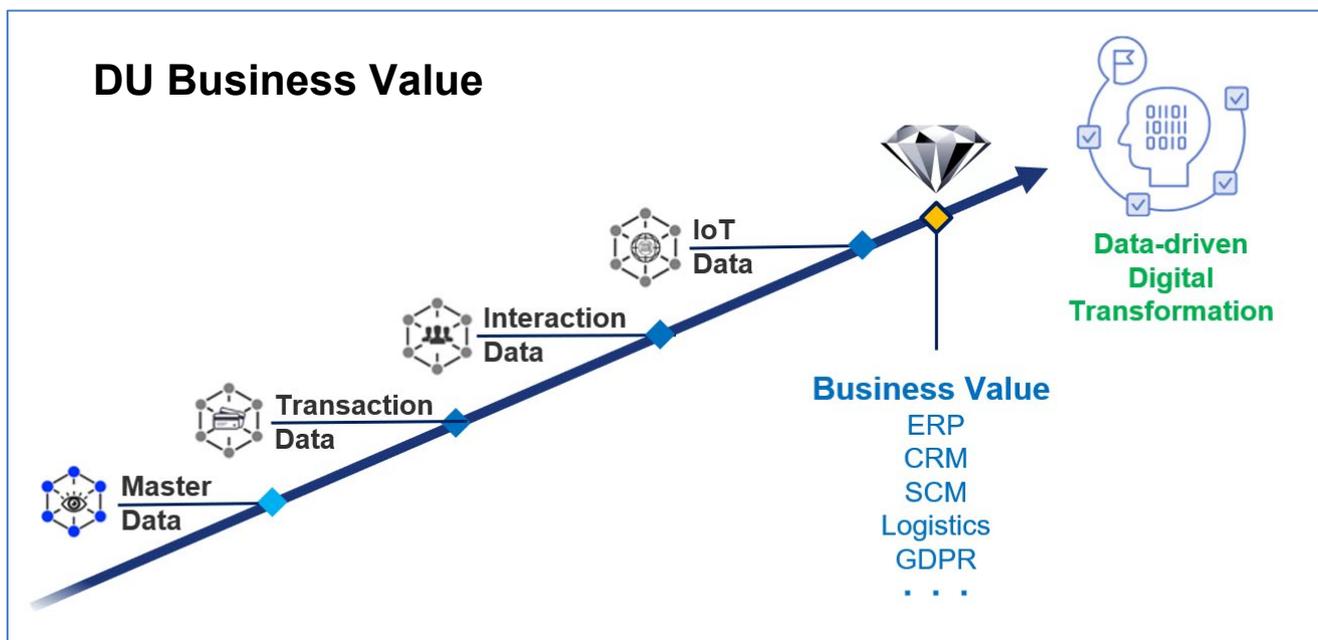
収入の急増

多様化する社会において絶えず変化する顧客ニーズをタイムリに捉えることができれば、他社より優位にたつことができ、ビジネス・バリュの向上につながります。また、ビジネス価値を高めるために行ったM&Aにおいても、データ連携ができないために、期待した効果があがらないケースも多々見受けられます。データ基盤をしっかりと構築しさえすれば、データの同期・共有が即座にでき、期待した効果を実現することにつながります。

先行者優位性

先行者は例外なく優位性を得られます（ネットワーク外部性）。80%の企業のデータ管理が未成熟という状況において、他社が対応を先延ばしにしているうちに、しっかりしたデータ基盤を持つことで大きな優位性を獲得できます。

[DUによるビジネス・バリュ]



*EnterpriseZine（翔泳社）「マスターデータを制する者がすべてのデータを制す！」より一部引用

DU(データを組織全体にわたり一貫して利活用するためのデータ基盤)を導入することによって、どのデータが正しいかといった終わりのない不毛の議論や、データ・サイエンティストの全作業の80%がデータの準備作業であるといったことに終止符を打つことができます。

DUのメリットがITインフラの近代化をはるかに超えているにもかかわらず、多くの組織がまだそれを認識できていないのが現状です。

ここで、DUを導入することにより、データ保護法に抵触することなくガバナンスもあり、経営者に最新のデータを使って分析した結果を即座に提供できるようになります。

データがどこかで作成・更新され、リアルタイムに必要なところにガバナンスされて同期され、その最新データを使ってデータ・サイエンティストが分析を実施し、経営者が経営判断できる基盤が実現します。また、データが作成・更新されたタイミングで、必要なアプリケーションを起動するトリガになれば(DUとビジネスの関わり)、どれだけ人員を他の創造的な仕事に振り向け、人が介在することによるヒューマン・エラーをなくすことができます。

DXを推進しようとするするとビジネス・プロセスも刷新する必要があるので、現場の抵抗が大きく進まないという勘違いについて記載しました。YOUnite Federated DUを導入すると、既存のITシステムやビジネス・プロセスを刷新することなくDXを推進することができます。

ROIに現れないDUのビジネスへの効果は計り知れないものになります。TCO(Total Cost of Ownership)ではなく、TCA(Total Competitive Advantage)を考えるべきではないでしょうか。

『2025年の崖』から転落しないためにも早急な対応が必要です。

※経済産業省が2018年に発表した研究資料

「DXレポート～ITシステム『2025年の崖』の克服とDXの本格的な展開～」

2. YOUnite Federated Data Unification

YOUnite Federated DU は、上記のような顧客が必要とするものから生まれたもので、顧客の課題をすべて解決できるものとなっています。YOUnite Federated DU について詳しく見ていきましょう。

YOUnite, Inc.は、カッティング・エッジ (Cutting Edge)のエンタープライズ データ・ユニフィケーション ソリューションを開発する目的で設立されました。その後、カリフォルニア・コミュニティ・カレッジ(CCC)から、このコミュニティで利用する DU 構想について相談されました。CCC 内のデータを安全にセキュアにやり取りしたいということで検討が始まり、最終的に CCC テクノロジ・センタと YOUnite の間でパートナーシップが締結され、YOUnite Federated DU プラットフォームがカリフォルニア州の資金をもとに、共同で将来に渡って使えるシステムを検討し、開発を行なっています（一部運用中）。

CCC は、検討するにあたり、当初 Gartner にも相談しましたが、CCC が望むことのできるものは市場にはないとの意見もあり、YOUnite との共同開発が始まりました。つまり、YOUnite Federated DU は、ユーザとともに作り上げたもので、現在ユーザが必要としたものを具現化したものです。

CCC システムは、カリフォルニア全体で 116 の大学キャンパスで利用されているシステムであり、54 億ドル（約 5,700 億円）の予算で運営されています。これは、世界で 3 番目に大きい大学システムです。CCC では、116 の大学でデータを安全にセキュアに、且つデータ保護法に抵触することなく同期するために、まずウィッシュ・リストを作成しました。そのウィッシュ・リストは以下のようなものでした。

1. 各大学のデータは、各大学で管理ならびに維持しながら、他大学と同期が必要なデータのみを同期し、データ品質を保証しつつ、プライバシー、コンプライアンス（ガバナンス）を順守するというものです。つまり、中央型システムを望んでいませんでした。データは元のソースに保持しつつ、エコシステム全体で利用できるようにするために、フェデレイテッド・データ・マネージメント・モデルが必要でした。
2. リアルタイム更新とバッチ更新の両方を望んでいました。また、各大学の管理者は、いつ、どのような頻度で更新を走らせるかを決める機能が欲しいという要望がありました。
3. データ・レコードを要素レベルまで詳細に制御するという要望がありました。たとえば、PII (Personal Identifiable Information : 個人を特定できる情報)を要素レベルでコントロールすることで、データ保護法に違反することなく、データを共有する必要がありました。

4. DU エコシステム全体を単一のダッシュボードからコントロールしたいという要望もありました。アドホックなポイント・トゥ・ポイントのデータ交換では、管理や更新が困難です。単一のダッシュボードでは、システムの統合、同期、ガバナンス、通知のためのエコシステム全体をコントロールすることが可能になります。すべてのデータ・イベントの履歴の確認も単一のダッシュボードから行うことができます。
5. 既存システムを止めることなくフェイズド・インプリメンテーションでいることが必要でした。CCC は、一度に 116 の大学すべてに導入すれば、運用に大きな混乱をもたらす可能性を恐れていました。混乱を最小限に抑えるために、時間をかけて小規模なグループごとに導入していきたいと考えていました。段階的な導入を実施することにより、各組織の望むペースでインプリメントができました。つまり、全システムを止めることなく、必要なときに必要なだけサービスの導入・追加ができるようになります。
6. カutting・エッジ(各組織にとって常に最善のもの)を利用したいという要望がありました。各大学（116 校）には、現在使用している独自の会計システム、学生情報システム、従業員/教師システムなど現在利用しているシステムがあり、それらの変更は考えられませんでした。すでに使用していて慣れ親しんでいるツールやアプリケーションを使用したいと考えていました。また、明日（将来）利用できるかもしれない最善のアプリケーション、またはツールを選択できるようにしておきたいとも考えていました。将来的に効果のない、または古くなったツールやアプリケーションにロックインされたくありませんでした。
7. イベント通知 (DU とビジネスの関わり)
CCC は、DU システムがデータ・イベントを他の大学やソース・システムに自動的に通知し、さまざまなビジネス・プロセスのトリガとなることを望んでいました。この機能は、現在行われているようなアドホック・ソリューションでは、実現にはかなりの問題がありました。

「ウィッシュ・リスト」の作成後、CCC は Gartner に、CCC が望む DU システムを提供できるベンダはありますかと助言を求めました。

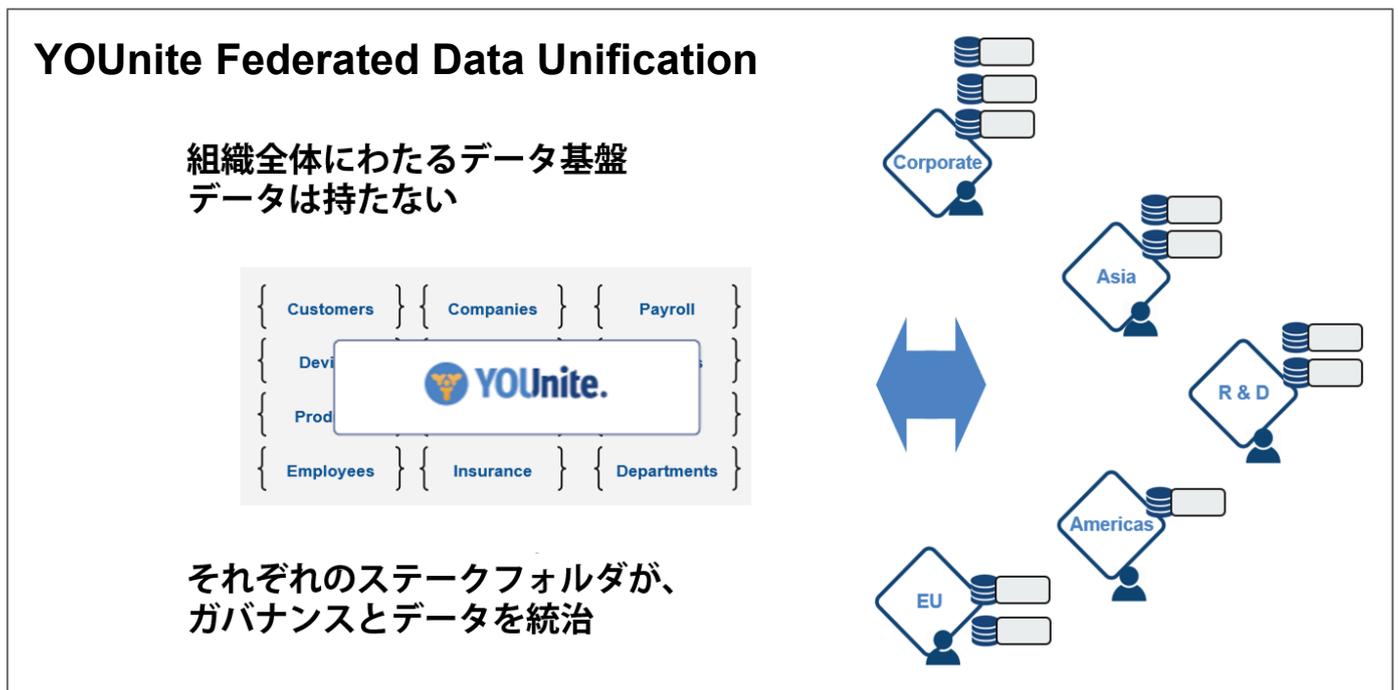
Gartner は、CCC が望む機能の一部を備えたベンダを見つけましたが、多くは MDM 機能を持っていると主張するもののデータ・クリーニングや ETL ツールでしかありませんでした。

Gartner は、DU には新しいデータの追加や更新などのデータ・イベントの通知や、システムが過負荷や停止した場合でも、データ・イベントをキューに入れる機能をサポートすべきなのに、

このような特長を持つ製品、技術はどこにも存在せず、それらすべてが実現できるものを見つけることができれば、それは素晴らしいものになると述べていました。

CCC への報告は、Gartner はユーザが望み使える DU は、これまで世の中に存在しなかった、ということを確認していることとなります。

[YOUnite Federated DU]



YOUnite Federated DU は、CCC の要件をすべて満たすシステムを作り上げ、またユーザが望むシステム要件をも満たすことができる唯一の製品となりました。

YOUnite Federated DU は次の 3 つのレイヤからなっています。

Layer 3 : フェデレイテッド・データ・アクセス

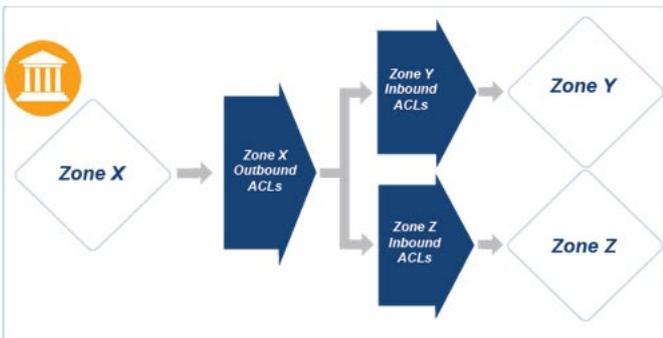
Layer 2 : ガバナンス

Layer 1 : データ・シンクロナイゼーション

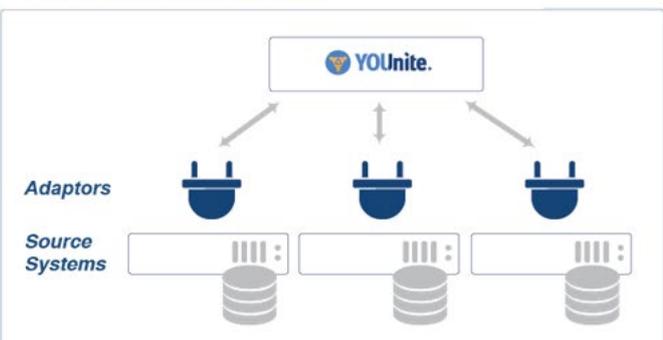
[YOUnite DU レイヤ]



Layer 3 : フェデレイテッド・データ・アクセス



Layer 2 : ガバナンス



Layer 1 : データ・シンクロナイゼーション

Layer 3

フェデレイテッド・データ・アクセスは、データ・イベント・ノティフィケーションズ、アSEMBル、データ・リネージュ、グローバル・デリートのような、アプリケーション・レイヤの機能を持ちます。

Layer 2

ガバナンスでは、Zone 間でリアルタイムでガバナンスをコントロールできるように制限を設けます。ファイル・アクセス制御のようなファイル・アクセス・コントロールではありません。

Layer 1

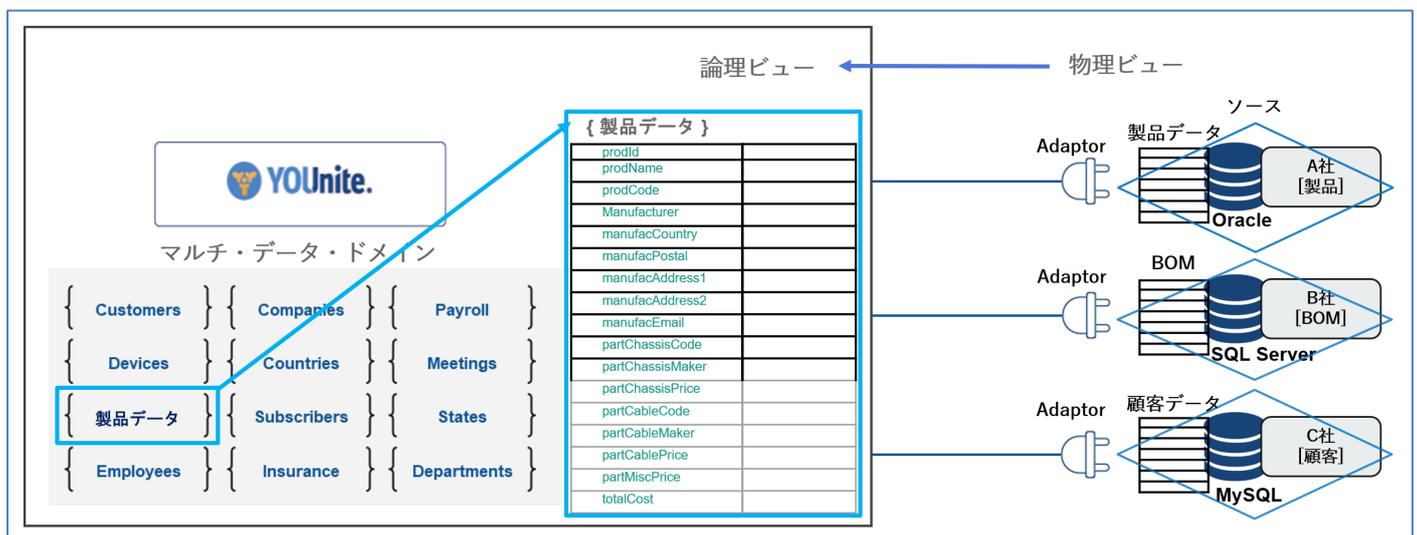
データ・シンクロナイゼーションでは、Adaptor がソースのデータ・イベントを探知し、YOUnite Server に通知し、ルーティングなどを行います。

2.1 データ・ドメイン (Data Domains)

データ・ドメインは、複数のソース・データ間における「共通言語」として、複数のソース・データでお互いに同期させたいデータを定義します（スキーマ、YOUnite Server）。このようなデータ・ドメインは、たとえば製品データや顧客データ、企業データなどとして、複数定義することができます（マルチ・データ・ドメイン）。

YOUnite においては、データ・ドメインは YOUnite Server で JSON スキーマ (JSON Schema) で構造定義され、YOUnite Federated DU エコシステム全体のデータ基盤として機能します。

[Data Domains]



データ・ドメインは、フィールドの集まりで、それぞれに色々な属性があります。

ソース・データは YOUnite Server には保存されませんが、各ゾーンにあるソース・データをリアルタイムで参照できます。例えば・・・

- ・実際のデータが存在するソース・システムのレジストリ・リンク
- ・データが最後に更新された日時
- ・どの Zone に属しているか
- ・グローバル ID (DR キー)

Data Domains の詳細について YOUnite Knowledge Base に記述があります。

Knowledge Base

Generated on Jan 26, 2021

Table of Contents

YOUnite Guides

- [Intro to YOUnite](#)
- [Zone Admin Guides](#)
- [Data Steward Guides](#)
- [Adaptor Guides](#)
- [API Developer Guides](#)
- [Implementation Guides](#)
- [Downloads](#)

Welcome to the YOUnite knowledge base. This page contains links to YOUnite guides, reference documentation and frequently asked questions.

After the introductory guides, the remaining content is grouped by common roles within an IT organization.

Data Domains - An Introduction to YOUnite

Data Domains are the heart of MDM. In traditional database parlance, data domains are a collection of fields (values) that are encompassed by an attribute (database column). For example, using a Customer table example below, the timeZone attribute has a data domain of A, P, M, C, E, or null, which represent Alaska, Pacific, Mountain, Central and Eastern time zones. In other words, the data for timeZone is limited to this data set, or data domain.

Domain Types: YOUnite Data Store vs Federated - An Introduction to YOUnite

Each record within the domain is referred to as a data record. Data records can be either stored in a YOUnite Data Store domain (inside YOUnite) or, linked to using a federated model. With the federated model, YOUnite keeps a registry of links to the source systems of where the actual data resides. For more on Domain Types see the Data Domains page.

Domains Stored In a Federated Domain - An Introduction to YOUnite

If a data domain is defined as Federated, its associated data records are NOT stored inside YOUnite but are created in real-time by referencing the properties (data elements) as they are stored in the company's multiple systems.

For each data record associated with a given federated domain, YOUnite stores: * Registry links of where the data is stored * When it was last updated * Which zone it is shared with * The unique global ID(s) known as the Data Record or DR Key.

Domain - Glossary

A domain refers to a data model, such as student or customer, and is defined by the parties responsible for data governance. The goal is to have universally-adopted domains across the organization's digital ecosystem. Domains are defined via versioned JSON schemas that are specific to the domain version. Schemas constitute the design for Data Records (DRs). Domains can be configured in any way that is necessary to meet the business needs of the organization when it comes to sharing data. For instance, a domain may be a "Student" that includes data definitions (properties) such as student name, address, etc. See Domain Version and the Data Domains document for more.

Federated – Glossary

YOUnite's Federated (similar to Gartner's Registry MDM model) refers to an MDM model in which the data handled through MDM is not stored centrally in the YOUnite MDM Server but instead references to the data are stored that point to the source systems where YOUnite can retrieve the data (source entities).

Federated vs YOUnite Data Store Access Patterns – Accessing Data Records

YOUnite Data Store - Data Store Data Records are stored in YOUnite's database and can be requested for a single data record or for all data records in a given data domain version.

Federated - Federated Data Records can also be requested for a single data record or for all data records in a given data domain version. These Data Records are NOT stored in YOUnite's database and the data must be requested from adaptor(s) and assembled. When YOUnite receives the request for federated data record(s), it:

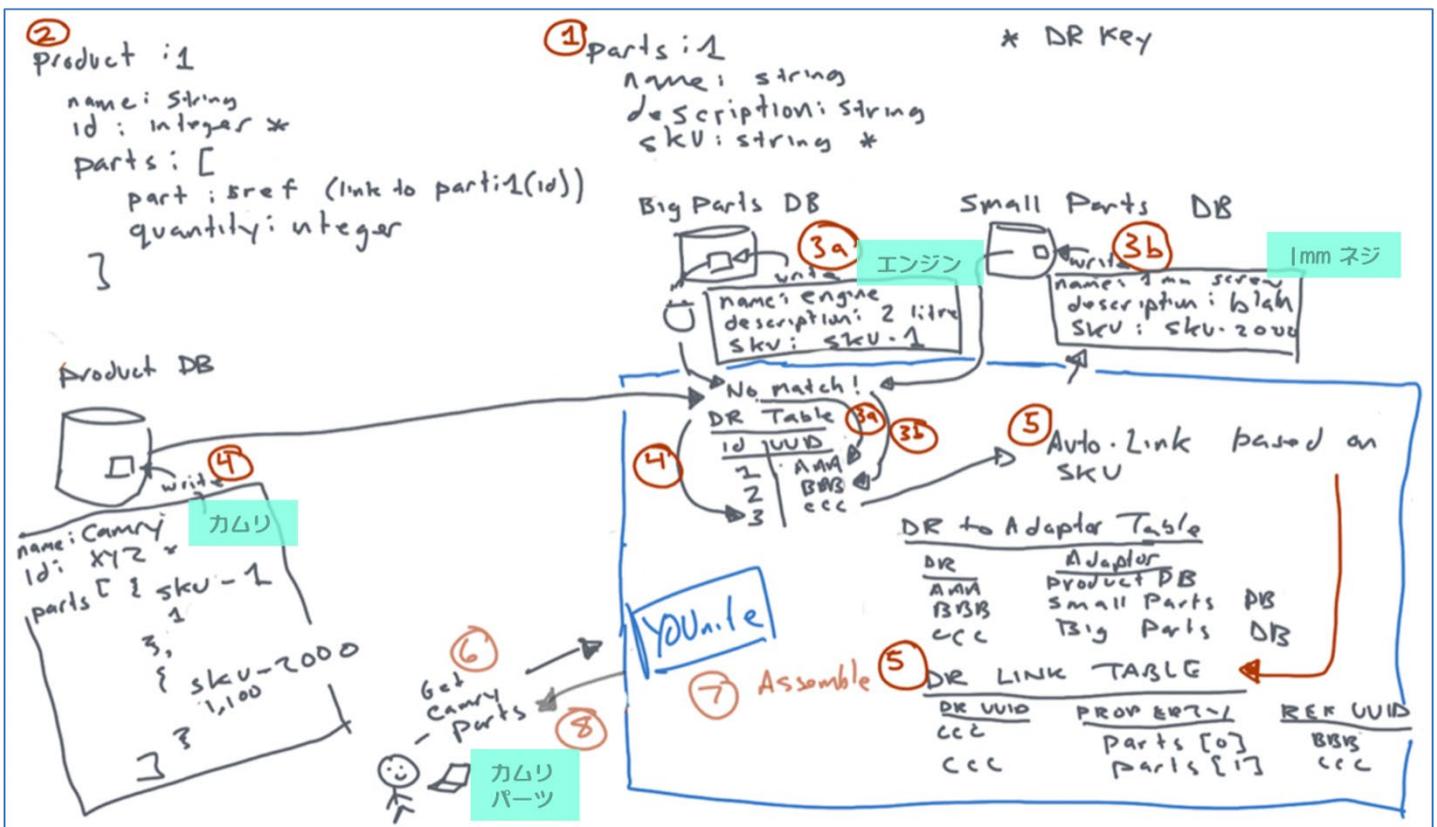
1. Looks up what systems/adaptors contain the data record(s)
2. Consults the appropriate inbound/outbound ACLs (Governance)
3. Makes a request for the data from the appropriate systems/adaptors
4. Assembles the data based on Gold/Silver adaptor status
5. Returns the assembled data to the requestor

クロス・ドメイン(Cross Domains)

Data Domainをクロス・リファレンスすることができます。これをクロス・ドメインと言います。

フェデレイテッド・レコードのデータはYOUnite Serverに保存されません。フェデレイテッド・データ・レコードは、UUIDとそのデータ・レコードにかかわるAdaptorによって識別されます。データ・レコードにかかわるAdaptorにクエリを投げ、結果をコンパイルすることで、データ・レコードをアセンブルできます。フェデレイテッド・データ・ドメインにクロス・リファレンスがある場合、クロス・リファレンスもアセンブルされます。

[Cross Domains]



Cross Domains の詳細について YOUnite Knowledge Base に記述があります。

<https://younite.us/resources/Cross-Domains.html>

Federated vs MDM Data Store

Both Federated and MDM Data Store domains support cross-references, however, the process for establishing a cross-reference differs between the two.

MDM Data Store Cross-references

MDM Data Store records are created by POSTing to the /drs endpoint. Cross-references are established for MDM Data Store records when their data is POSTED.

Federated Cross-references

Data for Federated records is not stored in YOUnite. Federated Data Records are identified by a UUID and a list of adapters that have that Data Record is stored. This allows "assembling" data records by querying the adapters that have the data record and compiling the results. If a Federated domain has a cross-reference, when it is assembled, the cross-reference will be assembled as well. See Accessing Data Records for more information regarding assembling data records.

Cross-references for data records in a Federated domain are established manually by a PUT operation on the /drs/{dr-uuid}/link endpoint and specifying the UUID of the record to link to: PUT /drs/{dr-uuid}/link

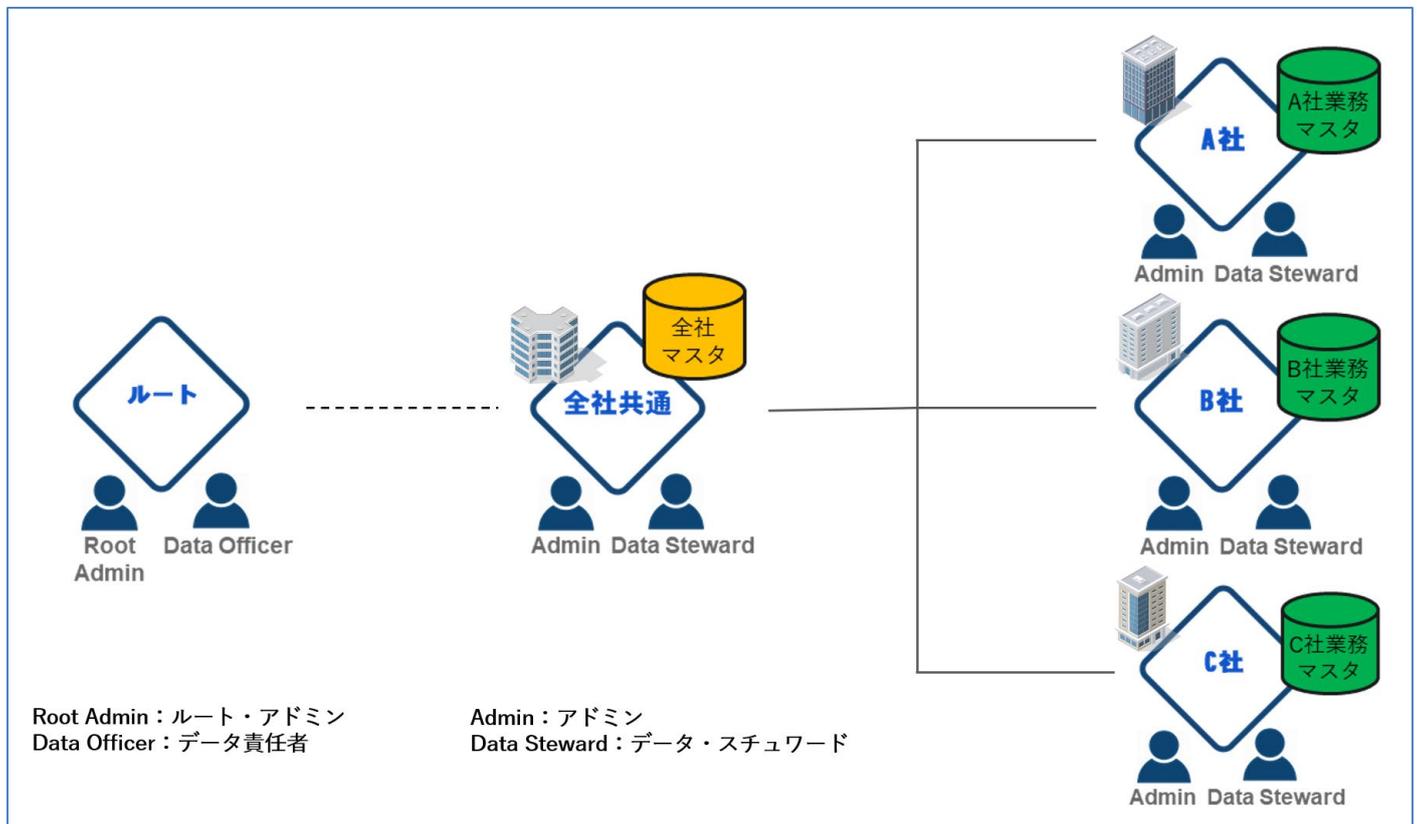
2.2 Zone

Zone は、YOUnite がガバナンスをインプリメントするために導入したものです。ガバナンスは制限を設けることであり、ファイル・アクセス制御のようなファイル・アクセス・コントロールではありません。

YOUnite では、Zone は物理的な「境界（バウンダリ）」を意味し、すべてのソース・データは Zone に存在します。たとえば、国、地域、企業、組織が Zone となります。

ガバナンスには、Zone 間のアウトバウンドと、同一 Zone 内のインバウンドのコントロールがあります。

[Zone]



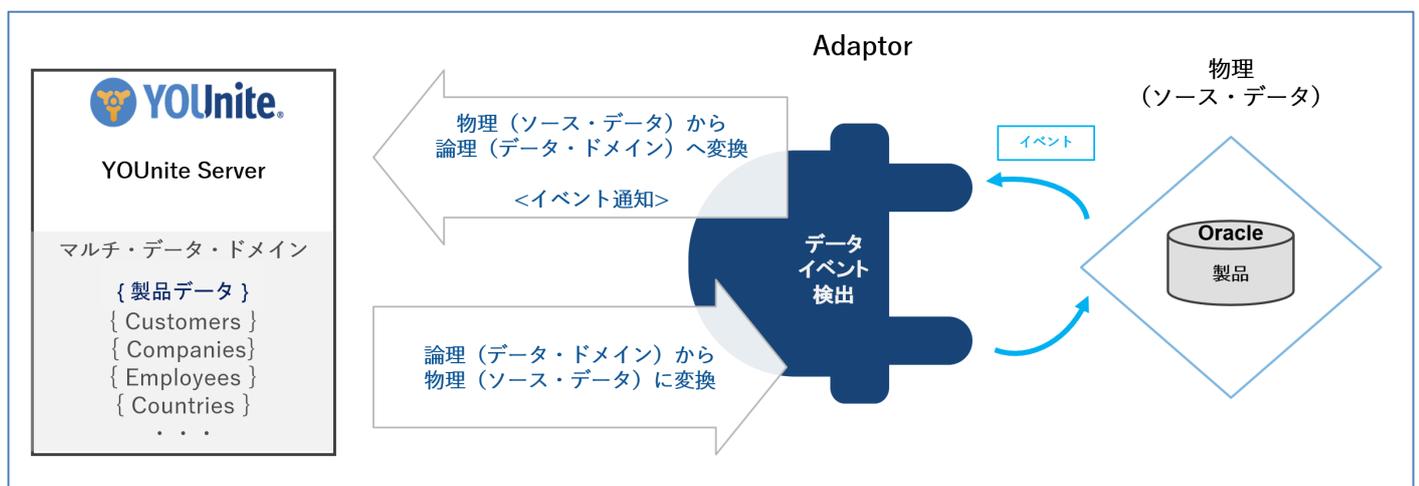
Zone 間のコントロールは、YOUnite Server に設定された ACLs (Access Control Lists、ファイル・アクセス・コントロールとは異なります) に従いルーティングされます。ガバナンスは、PII (Personally Identifiable Information、個人を特定できる情報) のコントロールなど、詳細について設定することができます。

2.3 Adaptor

Adaptor は、ソース・データのイベントを察知して、必要に応じて他のソース・システムに同期する機能を持ちます。

- ・ 接続されているソース・データの変更などのイベントを検出
- ・ 物理（ソース・データ）と論理（データ・ドメイン）の変換（マッピング）
- ・ ソース・データで発生したデータ・イベントを、YOUnite Server に通知
- ・ YOUnite Server からの指示に従い、該当するソース・データに通知

[Adaptor]



Adaptor は、ソース・データごとに用意します。また、新たなソース・データに対応した Adaptor を YOUnite が準備した SDK を利用して作成することが可能です。

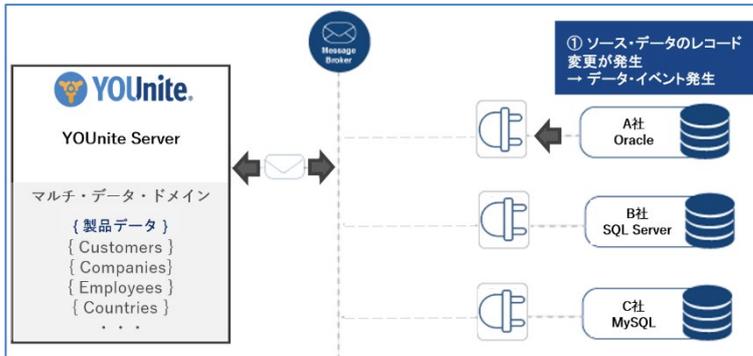
実装されている各 Adaptor は、Zone で設定されるガバナンスに従ったルーティングで YOUnite Server とメッセージ・バスを通じてメッセージのやり取りを行います。

Adaptor は、コンフィグレーション DB (Configuration DB) を持っており、MongoDB が使われています。その情報としては、マッピング情報などがあります。

MongoDB : MongoDB はリレーショナル・データベース (RDB) ではなく、NoSQL と呼ばれるカテゴリーのデータベースです。ドキュメント型と呼ばれる構造的データを使い、その「ドキュメント」の集合を「コレクション」として管理します。Adaptor のコンフィグレーションに利用するのに向いています。

2.4 YOUnite Federated DU のしくみ (サマライズ)

ここまでの説明[2.1]から[2.3] (Layer 1, 2) でソース間のデータ同期を行うことができます。A 社システムの Oracle [製品 DB]内のレコードの変更が、B 社システムの SQL Server [BOM DB] 内のレコードに反映される流れを説明します。

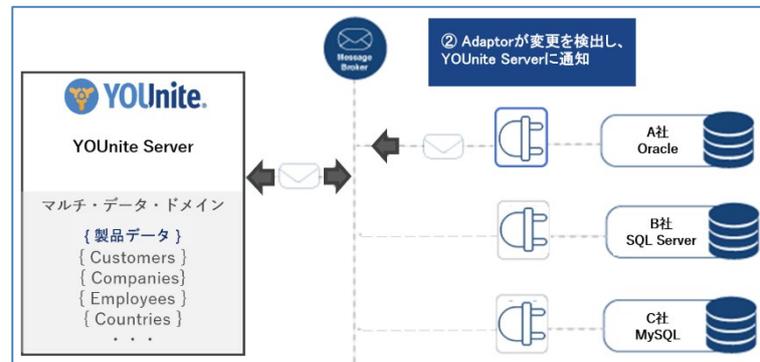


A 社ソース・データ :

- ・レコード変更
- ・データ・イベント発生

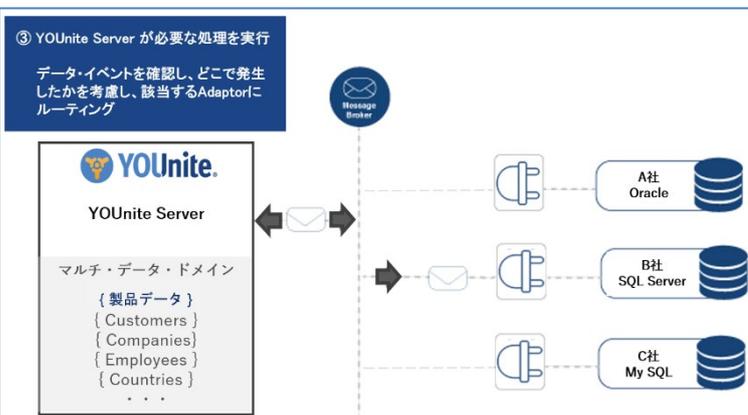
Adaptor (A 社) :

- ・A 社ソースデータのデータ・イベント検出
- ・物理→論理の変換
- ・YOUnite Server に通知



YOUnite Server :

- ・データ・イベントを確認し必要な処理
- ・該当する Adaptor (B 社) にルーティング



YOUnite Server :

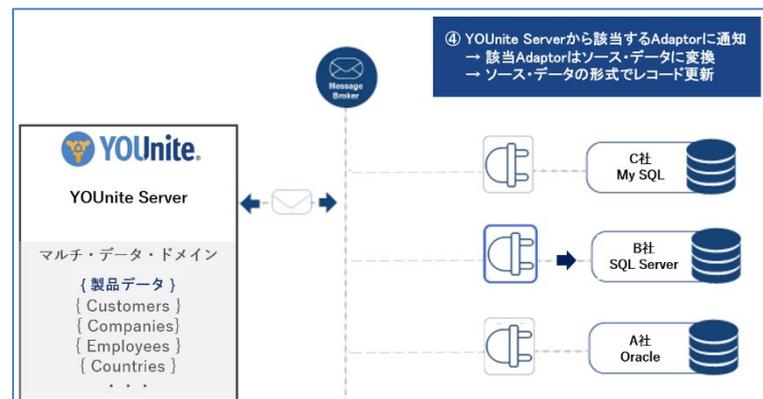
- ・Adaptor (B 社) に通知

Adaptor (B 社) :

- ・ソース・データの形式に変換

B 社ソース・データ :

- ・レコード更新

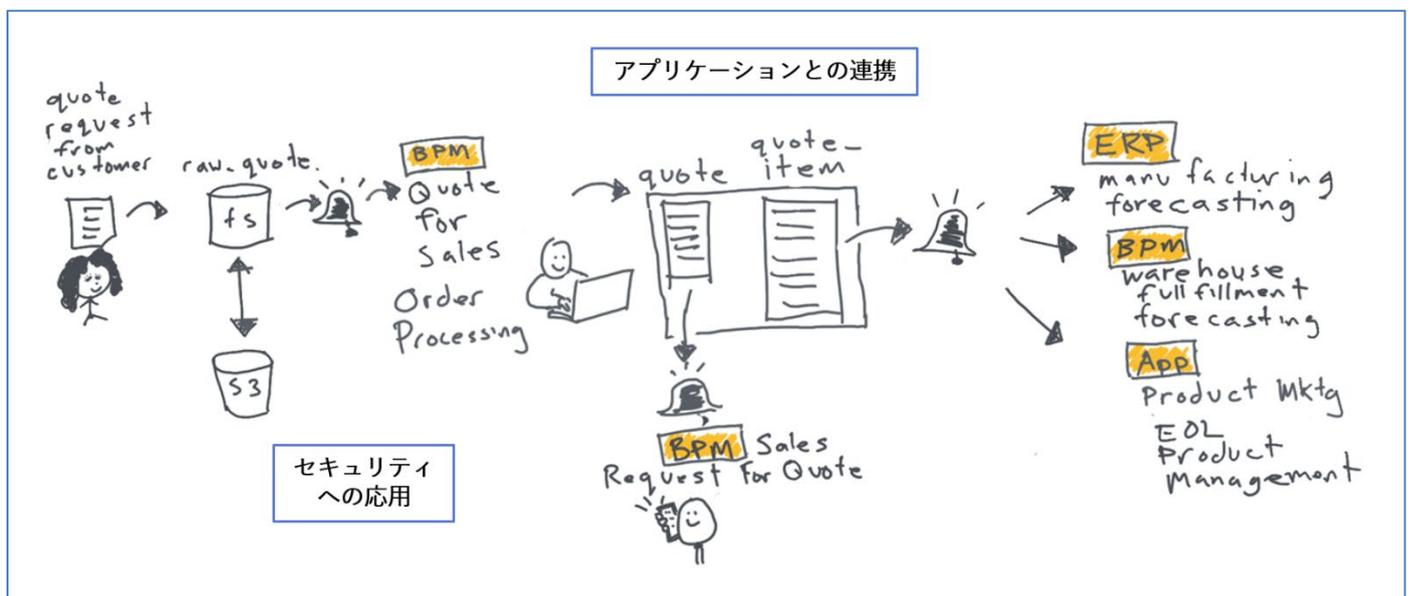


2.5 データ・イベント・ノティフィケーションズ

データ・イベント・ノティフィケーションズ (Layer 3 の機能) は、エコシステム内のアプリケーションに YOUnite API で「通知」を行い、ビジネス・プロセスを起動します。例えば、ビジネス・プロセス・マネージメント (BPM) にデータ・イベント・ノティフィケーションズで通知し、それをトリガとして他のアプリケーションと連携し、活用する方法が期待されます。

また、すべてのデータ・イベントを理解していることから、セキュリティ・ソフトウェアに、ソース・データに対する不正アクセスを通知する、などといった応用方法もあります。

[データ・イベント・ノティフィケーションズの応用例]

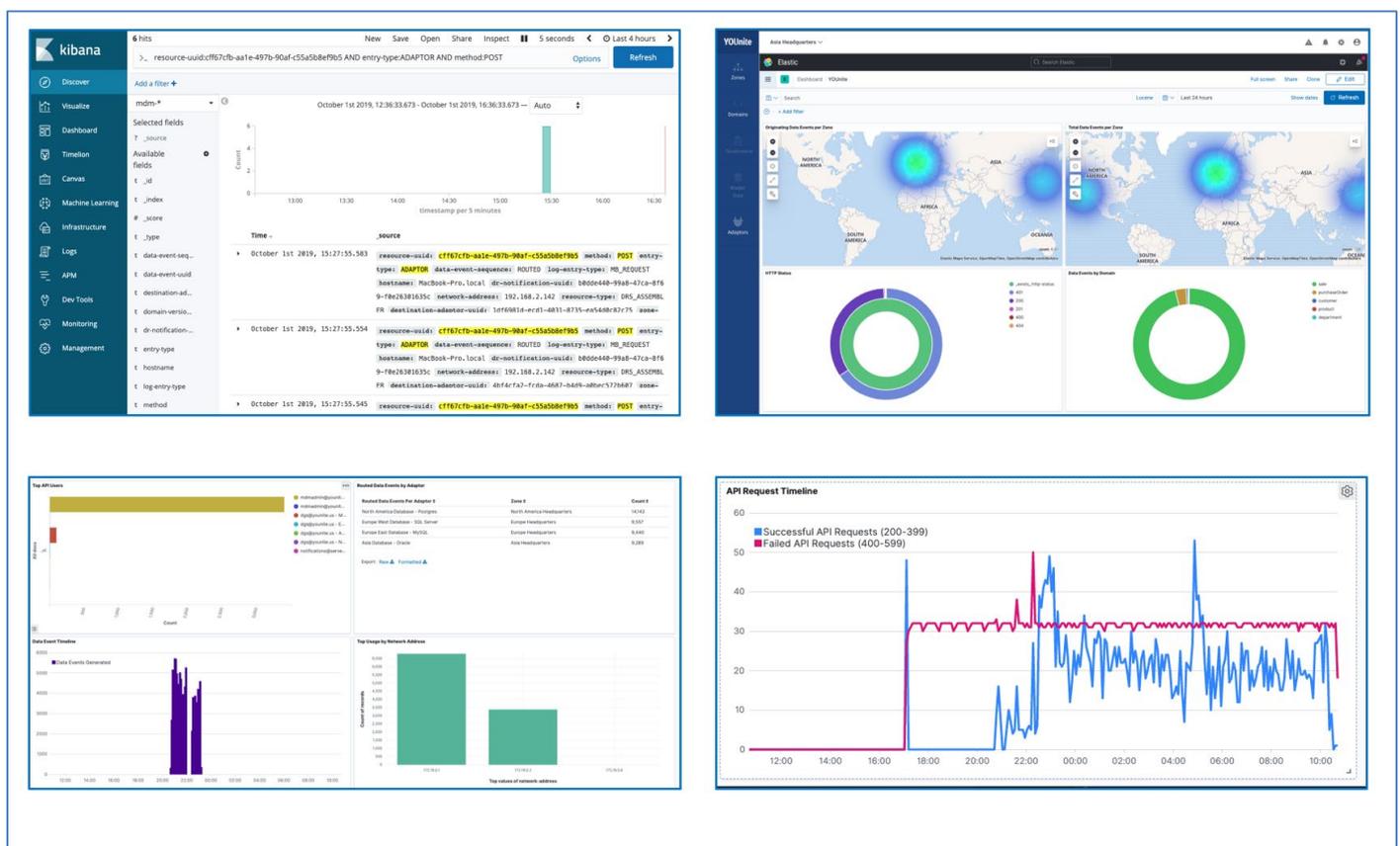


2.6 データ・リネージュ

データ・リネージュは、登録したエコシステム内で発生した全てのデータ・イベントを記録します。これは、組織のコンプライアンスにおいては必須なものです。一例として、GDPRでは72時間以内に情報の提供を求められます。

主要なステークホルダは、イベントの完全な履歴を参照することが可能です。また、データ・スチュワードは履歴からデータ品質の問題を特定することができます。

[Kibana ダッシュボード]



YOUnite Federated DUでは、イベントの履歴の可視化にElastic社のKibanaを使っています。Kibanaの強力な検索機能と、豊富な種類の表示が可能なダッシュボードで全てのイベント履歴を全体やソースごとに可視化し、分析に使うことができます。また、データ・イベントの分析や監視をダッシュボード表示から直感的に行うことが可能です。

2.7 データ・クウォリティ(Data Quality)

Data Quality とは？

多種多様なデータを扱う情報経済において、データ品質はビジネスに大きな影響を与えます。“1. DU (Data Unification) とは”で、「BI、アナリティクス、ML/DL、AIにおいてデータ・サイエンティストの全作業に占めるデータ準備作業は 80%にものぼる」と記載しました。そこには、ダーティ・データ (Dirty Data) と呼ばれる不正確、不完全、または不整合なデータを取り除くことなどの膨大な作業があります。IBM は、バッド・データが米国経済に年間 3.1 兆ドル (約 326 兆円) ものダメージを与えていると推定しています。

Data Quality は、これらのダーティ・データに Data Quality を通して、データの品質を高めるために使われます。多くの Data Quality 製品が存在し、ユーザは自分のビジネスに適した好みのものを選択すべきです。

Data Quality はなぜ必要か？

顧客データの一つをとっても、入力ミスに気付かずに、名前に異なった文字が使われていたり、同じデータが複数存在する場合があります。たとえば「斉藤」という名前の「斉」の文字は齊／齋／齊など 31 通りありますので、あるべき名前の文字と誤って入力されているケースがあります。他にもビジネスにおいて、データ品質に起因する多くの問題が発生しています。不正確なデータを使ってまともに分析などを行うことは困難であるため、これら不正なデータを排除し、データの品質を高めることにより、以下のようなメリットがあります。

正しいデータに基づいた意思決定

データ品質の向上は、組織全体の意思決定の向上につながり、意思決定の信頼性が高まります。高い品質のデータはリスクを軽減し、首尾一貫した改善が可能になります。

ベター・ターゲティング

データの品質は、より良いターゲットにつながります。高品質なデータがなければ、マーケティングの範囲を絞ることができませんが、高品質のデータがあれば、ターゲットが誰であるかをより正確に判断できます。広告キャンペーンをより正確なターゲットに対して行い、適切な人々にアピールする製品やコンテンツを開発できます。

効果的なコンテンツ&マーケティング

データ品質は、ターゲティングの改善に加え、コンテンツやマーケティング・キャンペーンの改善に役立ちます。対象者についてよく知っているほど、確実にアピールするコンテンツや広告を作成できます。たとえば、スポーツ Web サイトの運用者は Web サイトのユーザが最も興味のあるスポーツを示すデータを収集できます。ゴルフが最も人気のあるカテゴリの1つであることがわかった場合は、コンテンツ・チームにゴルフ関連の記事やビデオの作成を指示できます。

顧客との関係改善

高品質のデータは、顧客との関係の改善にも役立ちます。顧客に関するデータを収集することで、顧客の好みや興味、ニーズを知り、それらを元にしてアピールすることはビジネスの成功に不可欠です。これは、顧客との強い関係を築くことに役立ちます。

実施しやすさ

高品質のデータは、低品質のデータよりもはるかに使いやすい一面もあります。高品質のデータを使用すると、会社の効率も向上しますが、データが完全でない場合、データの修正をするためにかなりの時間が必要です。これは分析などを行う前の準備に時間がかかることを意味します。

競争優位

競合他社よりも高品質のデータがあり、そのデータを競合他社よりも効果的に使うことで競争を優位（TCA）に進めることができます。データは、高品質であればあるほど、企業が持つ最も価値のあるリソースとなります。データ品質が向上すると、競合他社よりも先にビジネス・チャンスを見つけ、売上を伸ばすことが可能です。

収益向上

高品質のデータは収益の向上につながります。より効果的なマーケティングを行い、販売数の増加に役立ちます。また、無駄な広告を減らし、キャンペーンの費用対効果を高めます。Web サイトでは、高品質のデータを使い、最も人気があり、最も収益をもたらすコンテンツをユーザに見せることができます。

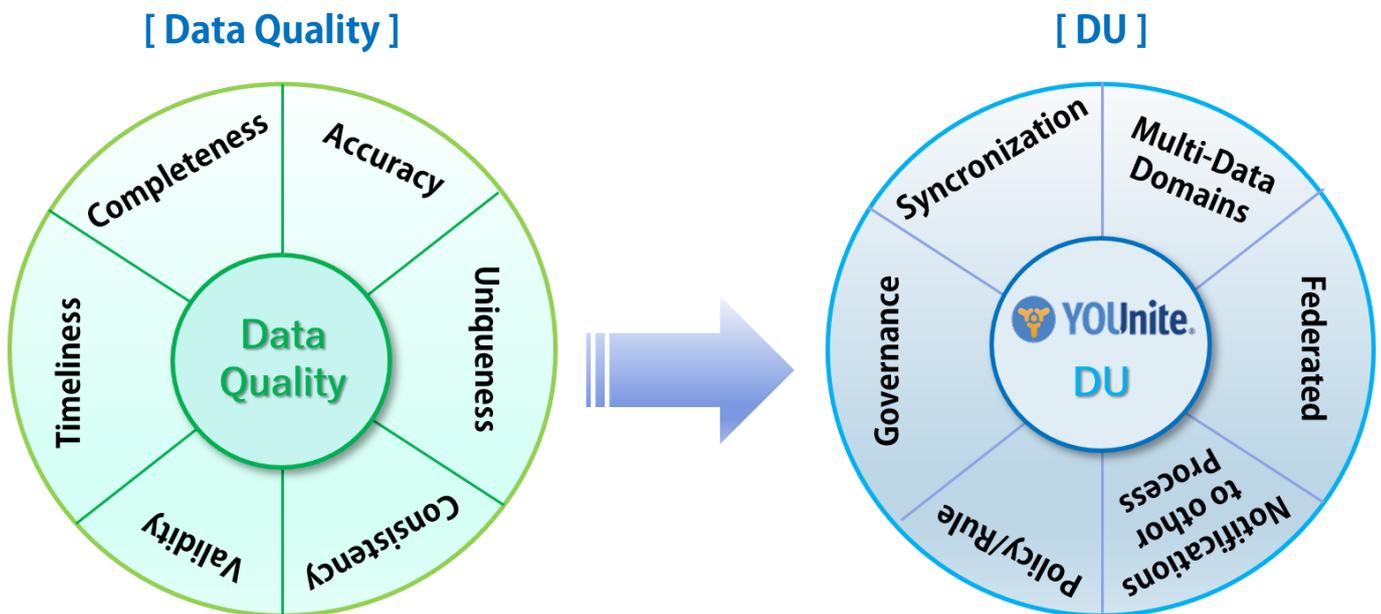
DUで生きる Data Quality の役割

Data Quality は、異なるソース・データ間の同期を行うことが目的ではなく、DU との連携が必須になります。すなわち、Data Quality は DU の一部として考えられます。

Data Quality は、以下の視点からデータ品質を高めるための役割があります。

- ・ 完全性 (Completeness)
- ・ 一貫性 (Consistency)
- ・ 正確性 (Accuracy)
- ・ 妥当性 (Validity)
- ・ 適時性 (Timeliness)
- ・ 唯一性 (Uniqueness)

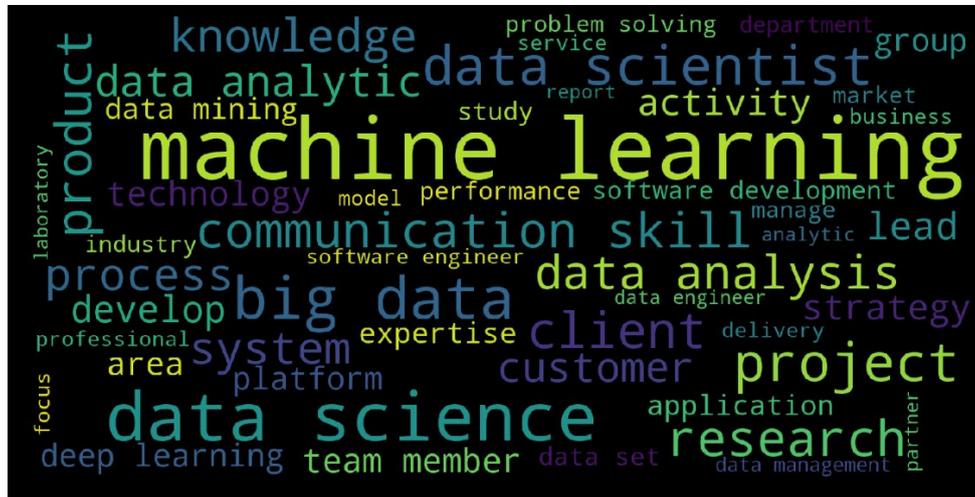
DU は、Data Quality によりダーティ・データがクリーニングされ、データ品質を高め、ガバナンスを施し、組織全体にわたり一貫して活用するデータ基盤の役割を持ちます。



機械学習と Data Quality の重要性

日々、機械学習は、ビジネスに欠かせない重要な手段になっています。機械学習では大量のデータが必要になります。順当な結果を得るためには、データ量も大事ではありますが、データの質がより重要となります。高い質のデータを扱うことで、分析の結果に深い洞察力を加えるこ

とが可能になります。現在の分析技術の進歩は目覚ましいものがありますが、データの質については機械学習に十分な水準には達していません。データの質は、ビジネスの観点から正しく使いやすいものである必要があります。



機械学習エンジニアは、最初に顧客のニーズを理解する必要があります。これは、企業が機械学習のコンサルタントと協力し、ビジネス・モデルに適合するためのガイドを作成することが必要であるという意味です。次に、機械学習エンジニアは、データ・スチュワードの支援を受けて、システムのデータ処理を行い、分類します。ここで、データ・スチュワードの協力が得られない場合、データ分類の誤りやオペレータのエラーにより、結果に誤りが発生する可能性が高くなります。機械学習エンジニアは、ほとんどの時間をデータの並べ替えなどに費やしていますが、最初に誤ったデータが提供されると、それ以降、誤ったデータでさらに悪化する事態となります。

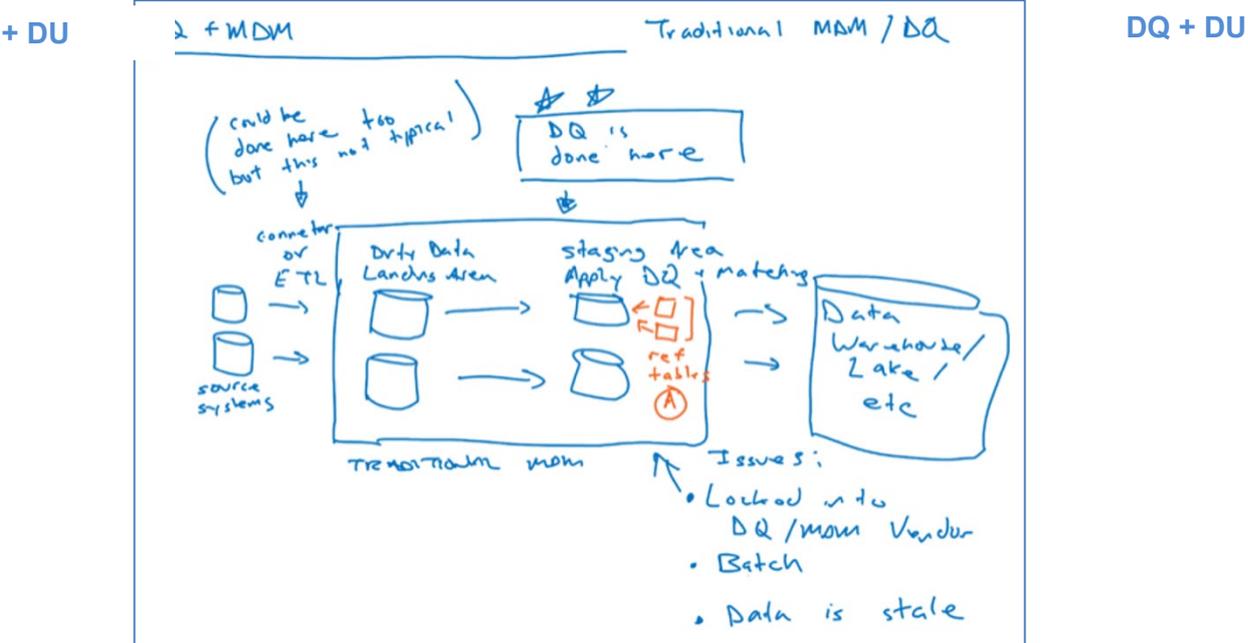
「教師あり機械学習」は、入出力ペアのサンプルに対応するアイテムにマップさせる機能のようなプロセスのことを言います。「教師なし機械学習」は、これに相反します。教師なし機械学習では、データ・ラベルがなく、アルゴリズムのパフォーマンスを測定するような方法もありません。しかし、教師なし機械学習のアルゴリズムには、人間がよく知らないデータのパターンを認識できるといったプラス面もあります。したがって、機械学習のアプローチを選択するとき、それがビジネスで使用される目的を理解することが重要となります。

機械学習では、データの質が重要です。教師なし機械学習は、ビジネスの要件を満たすために必要な質のデータが不足している場合の救い手となります。データを評価することにより、正確なビジネス洞察を提供することができます。しかし、ビジネスには万能のソリューションはないということも理解すべきです。

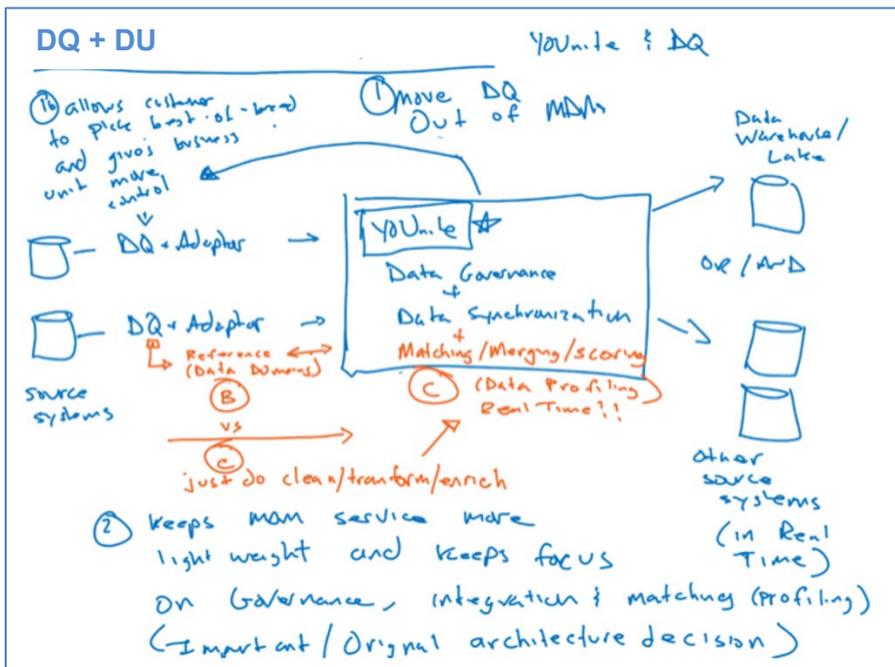
YOUnite Federate DU が取り組む Data Quality

MDM 機能を提供していると言っているベンダの中には Data Quality の機能も組み込んだと言っている企業もあります。しかしながら、マーケットにはさまざまな Data Quality 製品が多数あります。YOUnite は、このような状況下で Data Quality の製品を提供するベンダをエコシステムとして取り込み、Data Cleaning (データ・クリーニング)、Data Enrich (データ・エンリッチ) などを含めた機能をユーザの要望に沿うかたちで提供します。

[Traditional MDM / Data Quality]



[YOUnite Federated & Data Quality]



[DQ + Reference Data]

DQ + DU

DQ + Reference Data
(i.e. "Sources of Truth")

Implementing DQ? MDM reference data
(see orange notes in other drawing)

DQ + Reference Data

(A) traditional MDM see drawing (in orange)

- Newer Approach

(B) Adapter for DQ service and it would reference and throw exception when ambiguous records are encountered

(C) Let MDM do matching/margining/scoring and manage ambiguous exceptions

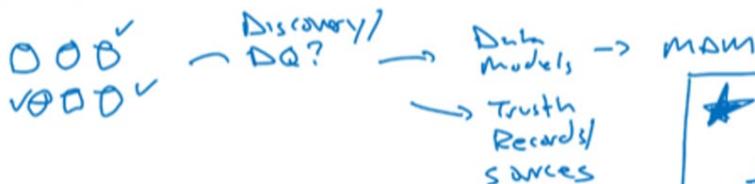
[DQ Sales Approaches]

DQ + DU

- Two Sales Approaches

- Customer must commit to full initiative ?]
- or -
- Start small

For some - DQ might be the place to start the discovery process of where the data lives and to determine how dirty it is - with a plan on how to move forward with MDM

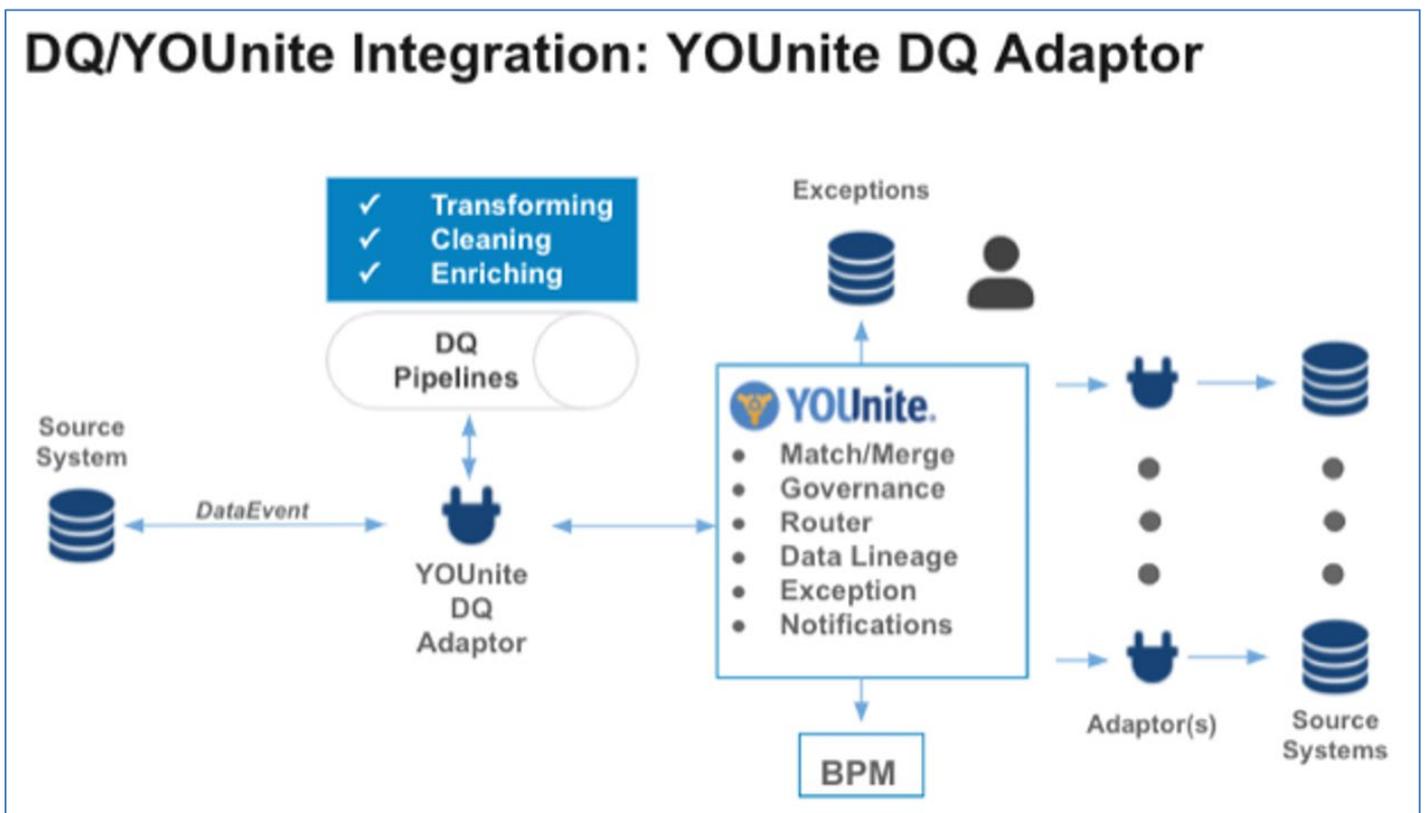


★ Great Approach to get feasibility

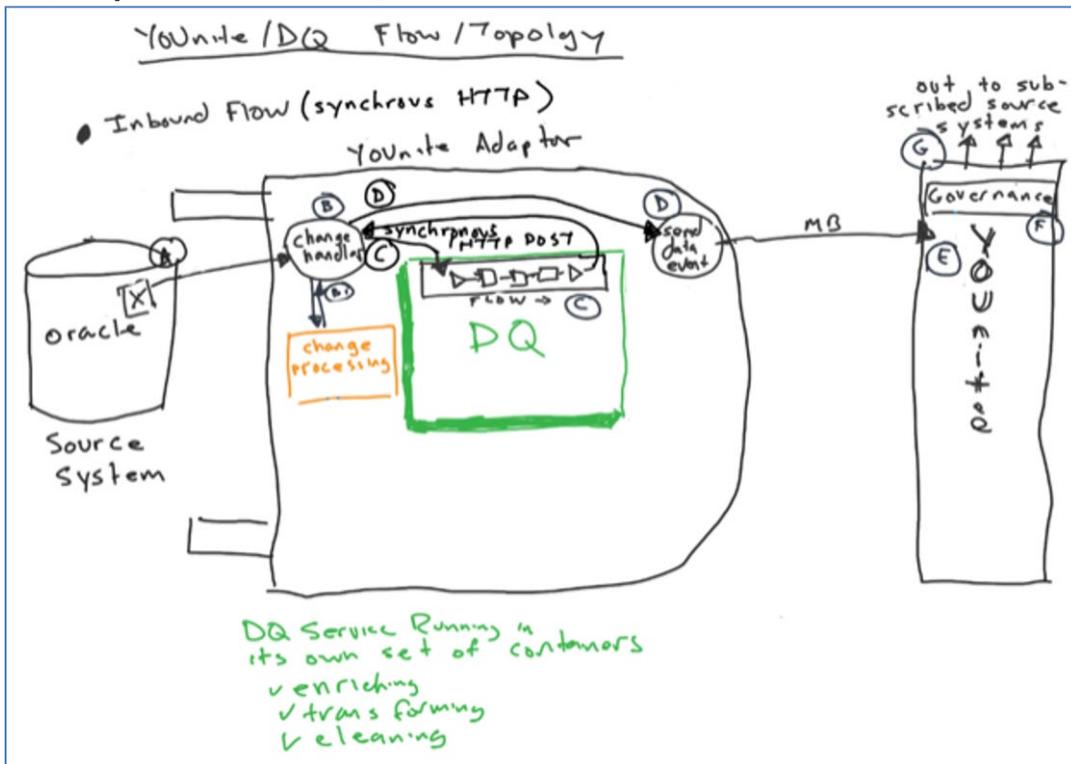
DQ Adaptor

YOUnite Federated DU は、Data Quality を”DQ Adaptor”として Adaptor に取り込み、どの Data Quality 製品を使うかについては柔軟であり、ユーザの希望する Data Quality を選択することが可能です。

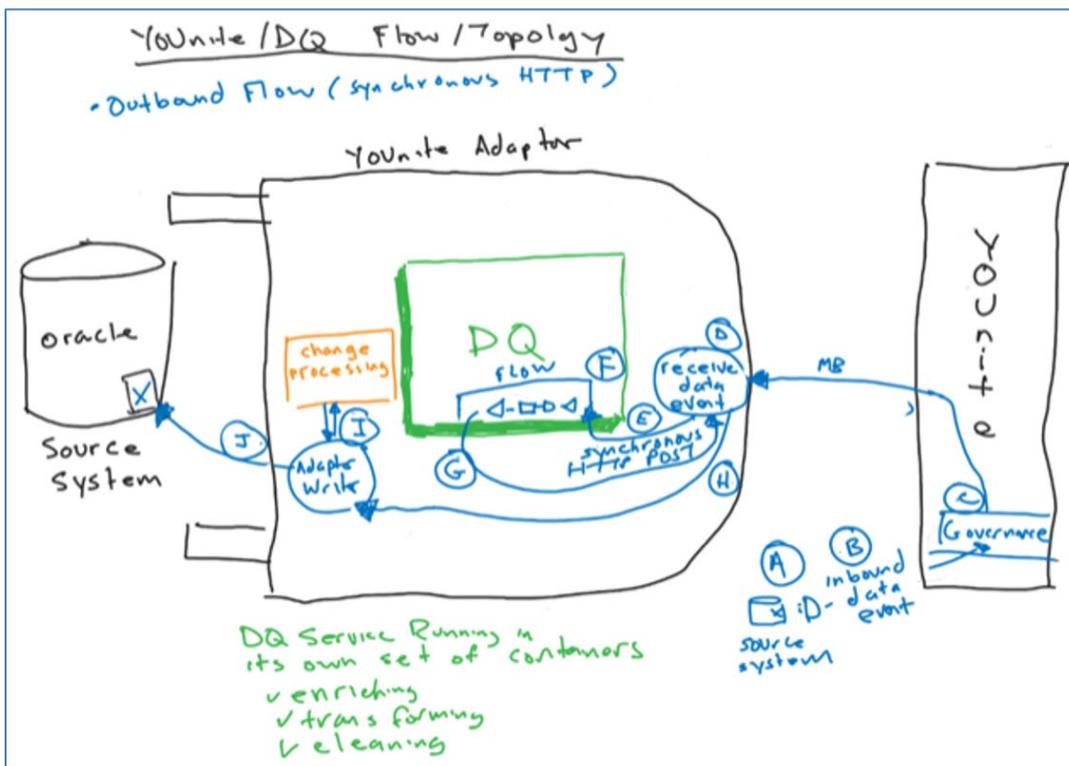
[YOUnite Data Quality Integration]



[DQ Adaptor: Inbound Flow]



[DQ Adaptor: Outbound Flow]



Matching Algorithm

Data Quality でデータ・データが改善されても、データを組織全体にわたり一貫して正しく使い、BI ツールなどでより正確なデータ分析を行うためには、似かよったデータが同一のものを表しているかどうかなどの整合性についての判断も必要になります。たとえば、引っ越し前の古い住所や、旧姓、同じ顧客の住所が異なるシステムでの表記が、○丁目△△番地×号というデータや○ー△△ー×というデータなど、これらを異なったものと判断すると、同じ顧客を複数カウントしてしまいます。営業部門では、購入した顧客を顧客データとして管理しているかもしれませんが、マーケティング部門は、コンタクトがあった人も含めて顧客データとしているかもしれません。また、カスタマ・サポート部門は、サポートが終了した人のデータは顧客データから削除している可能性もあります。これではデータの整合性が取れず、さまざまな問題が発生します。

YOUnite Federated DU では、独自のカスタム・ルールやスコア・ベースを用いた Matching Algorithm で、一部のレコードが完全に一致する場合と、そうでない場合のどちらでも、重複するレコードを識別できます。この Matching Algorithm で、似かよったデータにおけるデータの整合性を確認し、同一のものかどうかを判断します。

Matching Algorithm の詳細について YOUnite Knowledge Base に記述があります。

<https://younite.us/resources/Matching-And-Parsing-Exceptions.html>

Overview

Matching algorithms follow a SQL-like pattern with groups of comparisons wrapped in parenthesis and connected by logical AND and OR operators. YOUnite supports two similar types of matching algorithms: rules based and score based.

Rules based algorithms use a set of SQL-like rules to determine a match. If all rules evaluate to true, then the records match, otherwise they do not.

Score based algorithms also use a set of SQL-like rules to determine a match, however, matched records are then given a "score" based on another set of criteria and depending on the score the record is determined to either be a definite match, a possible match or not a match at all.

Ambiguous And Possible Matches

With both Rules Based and Score Based matching, ambiguous matches are possible.

For Rules Based matching, a match is considered ambiguous if the criteria matched to more than one Data Record.

For Score Based matching, a match is considered ambiguous if a match was given a score in the "possible" range, or if there was a "definite" match two two or more Data Records.

Ambiguous matches must be resolved through manual intervention by a user, or through a workflow. See [Matching And Parsing Exceptions](#).

3. 今後に向けて

CCC は、Gartner が「ユーザが必要とするような MDM は、現在はマーケットには存在しない」とコメントしたことから、YOUnite, Inc.との共同開発を行い、現在も継続して段階的なインプリメンテーションを進めています。

CCC での実績が認められ、YOUnite Federated DU はカリフォルニア州の次のプロジェクトである”Cradle to Career”（揺りかごからキャリア）と繋がっていきます。

「揺りかごからキャリア」”Cradle to Career”は、カリフォルニア全体のデータ・インフラストラクチャを開発し、データを広く活用することによって、より多くの学生が教育、キャリア、人生目標を達成するのに役立つ情報を提供するようにするという目的で、学校、大学、社会、サービス機関、財政援助提供者、雇用主がすでに収集しているデータを安全にセキュアに接続して活用できるものを構築するというものです。2019年にカリフォルニア州議会が、データシステム法として可決しています。

2020年は、COVID-19が世界中に蔓延し、日本においても何度も緊急事態宣言が発せられるような事態になりました。そのようなおり、各家庭に給付金を配布しようとしたときに日本はDX後進国であることがさらに露呈してしまいました。これは、縦割り行政の弊害で各省庁のシステムのデータがサイロ化され、各省庁間で共有されるべきデータが同期されていないことに大きな問題がありました。

また、地方自治体においても、自治体クラウドと自分たちが言っているものでも本来の使われ方ができておらず、データのサイロ化は解決していないのが現実です。財政力のある自治体では独自のシステムを作り上げて運用しているものもあります。さらに、地方自治体間でのデータ同期の仕組みも全くできていない状況で、且つサービスとのかかわりも全くできていない状態です。

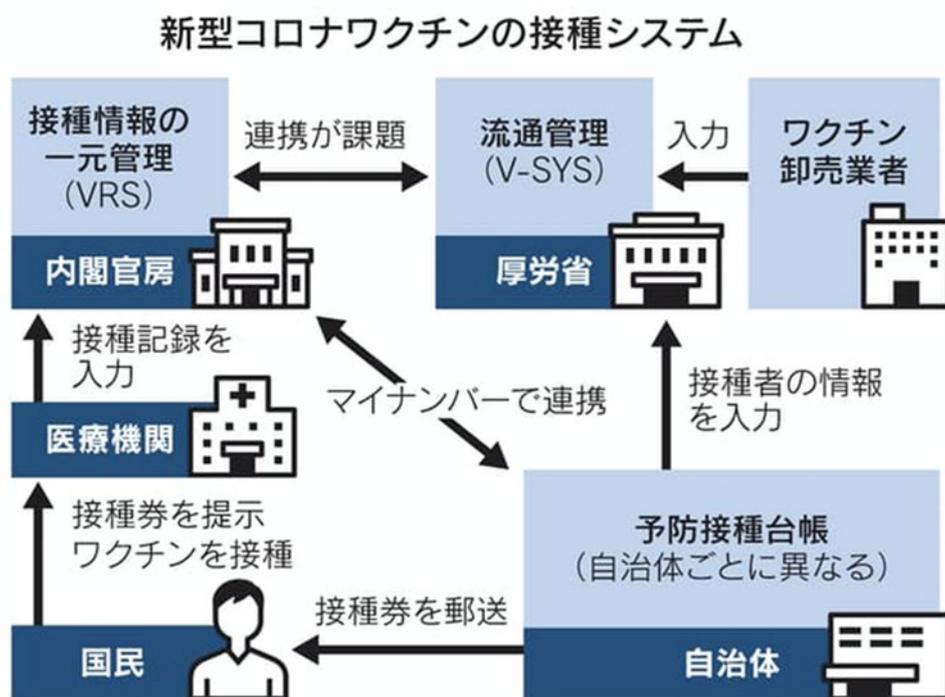
管政権になり、これらの問題を解決して、国、地方自治体のシステムを連携させる目的でデジタル庁が組織されることが決まりました。実際の活動はこれからですが、どういう手法でサイロ化しているデータを同期させるのでしょうか？ デジタル庁が音頭を取ってデータ・フォーマットを統一し、アプリケーションを含めたシステム全体を国の各省庁、市町村にいたる地方自治体のシステムにまで、それを適用するのでしょうか？

もし、そのような手法を取った場合、どれほどのコストと時間がかかってしまうのでしょうか？システムができあがるまでに何十年もの歳月を要すると財政再建もまた遠い夢となってしまいます。

経済産業省が、DXを推進しているにもかかわらず、各省庁、地方自治体ですら、DXの本来の本質であるデータが統一されていないではないですか?!

2021年5月中旬以降、高齢者への新型コロナワクチンの接種が本格化することになっています。しかしながら、デジタル化の遅れで接種を円滑にするための仕組みがなく、接種の人手不足も目立つ状況も相まってスタートからつまづいています。政府は6月末までに接種対象者の全員分のワクチンを確保できる見通しだと説明していますが、現状では早期接種に向けた計画が予定通りには機能せず、経済再開を進めることで、さらに後手に回る可能性が大いにあります。

この混乱の根本的な問題は、図のような「新型コロナワクチン接種の仕組み」でデータが必要などころで同期されていないところにあります。



出典：日本経済新聞

データが入力された時点で、その入力 that 適切なものであるかがチェックでき、必要などころにタイムリに同期できれば、混乱が最小限に抑えられるのではないのでしょうか。上図では VRS ↔ V-SYS でのデータ連携が課題とされていますが、データ・ユニフィケーションを利用すれば現行システムに変更を加えることなく問題を解決することが可能です。データ入力（データ・イベントの発生）をトリガにして他のプロセスと連動できれば、見落とし/忘れ/うっかりミスといったヒューマン・エラーを排除することもできます。

2021年5月17日から東京と大阪で行われる新型コロナワクチン大規模摂取の予約受付が開始されました。開始されると同時に二重予約が確認できないこと、架空の自治体番号や接種券番号を入力しても予約できてしまうことなどの問題点が指摘されました。

システムが個別に作成され、関係システムのデータが全く連携されていないことで起こる問題です。YOUnite が提供する Federated Data Unification でデータ基盤を構築すれば全て防げた問題です。

1. 予約サイトからデータが入力された時点で、Data Quality で入力内容を確認
2. YOUnite マッチング・アルゴリズムで重複確認
3. 同期が必要な情報を同期が必要な先(VRS、V-SYS、各自治体システムなど)へガバナンスを
チェックし同期

これらは、データ基盤をしっかり作ってさえいれば、個別に作られているシステムを変更することなく実現できます。今後も、新型コロナウイルス関連で次々にシステムが作られると思いますので、データ基盤構築は必須です。

加えて SQL インジェクションによる不正アクセスの可能性がある、というセキュリティの脆弱性の問題も発覚しました。

YOUnite フェデレイテッド・データ・ユニフィケーションは、データに誰がいつタッチしたかがリアルタイムで判りますので、SQL インジェクションが起きた瞬間に通知を受け、対策を講じることが可能です。

データ・ユニフィケーションを早急に導入すべきです!!

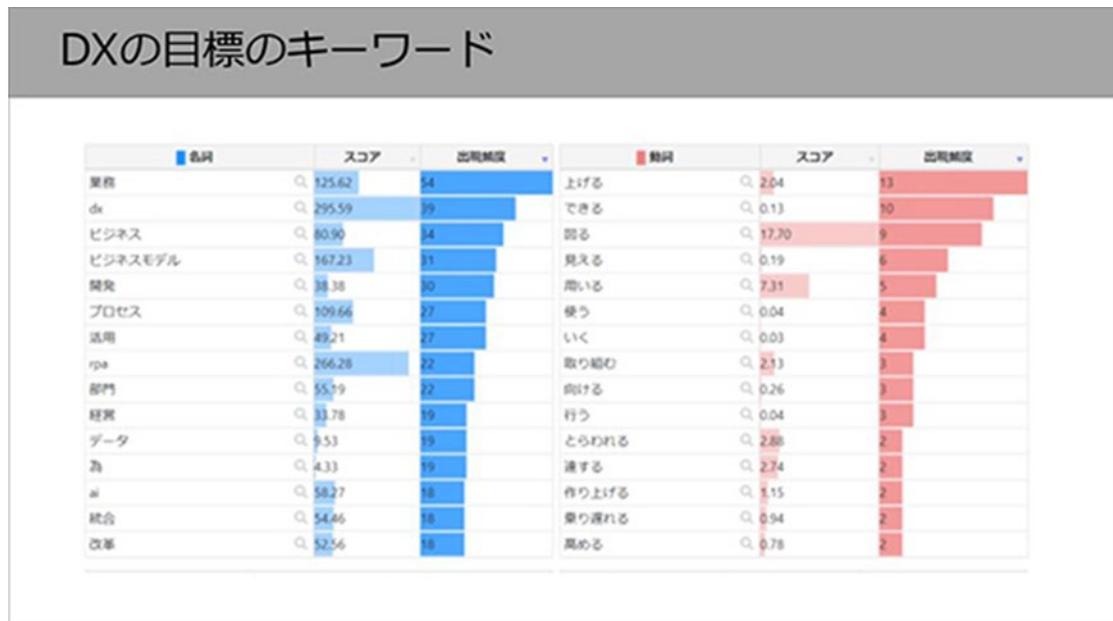
日本の各企業においても状況は全く同じです。企業によっては、システム変更の時点で一つ一つアドホックに対応していこうとしているところも多く見受けられます。このようなやり方では、国、地方自治体とおなじくコストと時間がかかってしまいます。コロナ禍で収益の低下した企業にとっては厳しく、データがサイロ化された状況が延々と続くことになってしまいます。

日本の企業が、永遠に EU GDPR に代表されるデータ保護法にも準拠できない状態が続くことになります。グローバル社会では生き残ることはできません。

デル・テクノロジーズが 2020 年 1 月に発表した「第 2 回 DX 動向調査」によると、そのやり方には、従来型の業務プロセスに主眼を置き現状の改善や改革といった“守りの DX”を目標にしているところもあれば、デジタル技術の活用でビジネス変革を目指す“攻めの DX”を目標にしているところもあります。

DX 推進のトップ 100 社の回答から抽出した DX の目標に関するキーワードでは、“攻めの DX” を目標にする取り組みも多いと解釈できます。

デル・テクノロジーズ「第 2 回 DX 動向調査」(2020 年)



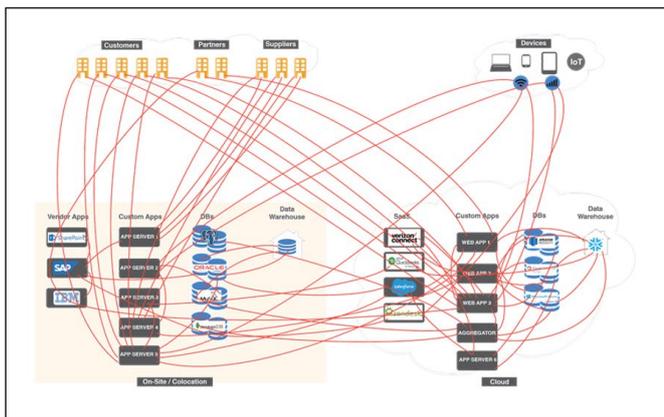
しかしながら、“攻めの DX” を目標としていながらも経済産業省が嘆いているように DX が進んでいないのが現状です。

DX を推進し、且つ成功させるには、しっかりしたデータ基盤を早急に構築することが必須です。

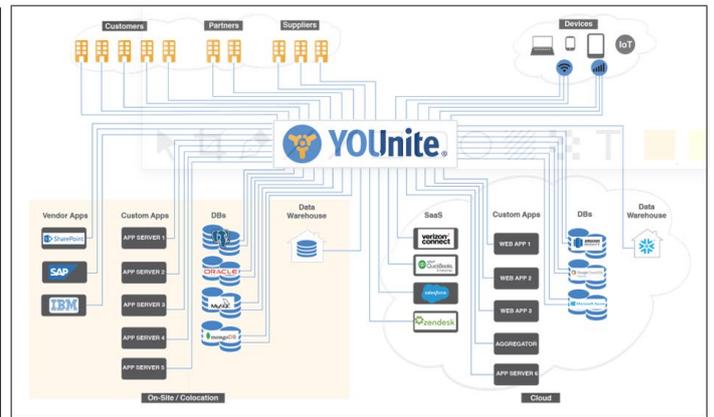
多額のコストを費やすことなく、必要なところに必要なデータがガバナンスのあるかたちでリアルタイムに統一できる、データ・ユニフィケーションを導入すべきです。これまでの MDM ライクなものには決して実現できなかったことが現実にはできるようになります。

今こそ、DU を導入すべきではないでしょうか。CCC やカリフォルニア州の次のプロジェクトである「揺りかごからキャリア」を参考にすべきです。

混沌としたシステム



YOUniteを導入すると



YOUnite は、現在の状況を理解しており、Data Quality 機能を提供するベンダをエコシステムとして取り込み、Data Cleaning（データ・クリーニング）、Data Enrich（データ・エンリッチ）などを含めた機能をユーザの要望に沿うように提供できるように DU に取り組んでいます。

以上により、データの正確性、正当性、一貫性が保たれたデータ基盤が DU で可能です。



Better Decisions with Better Data