

データ仮想化

データ統合のための最先端ソリューション



目次

要旨	3
序論	4
従来の統合技術	5
データ仮想化	6
データ仮想化製品における5階層	8
まとめ：データ仮想化について知っておくべき10項目	9
Denodoプラットフォーム：最先端データ仮想化プラットフォーム	10

要旨

企業や組織では、統合されたデータをより迅速に事業で活用できるようにする必要性が日に日に高まっていますが、関係者のニーズを満たすような迅速なデータ統合を実現することは容易ではありません。

従来のデータ統合技術は根本的な問題を解決していません。データの受け渡しは決まったバッチ処理であり、しかも昨今扱われるようになってきた豊富で複雑なデータ型に対応することはできません。

データ仮想化は、最先端のデータ統合手法であり、このような現状の課題に対応し、将来に向けたデータ統合の基礎を提供するものです。

本ホワイト・ペーパーでは、既存のソリューションでは満たされていない現状の課題を説明し、主要なソリューションとしてデータ仮想化を紹介します。

序論

私たちは、これまでにない種類と量のデータをこれまでにない速度で受け取り処理しています。さらに、クラウドやビッグ・データ・システムなどの新たに台頭してきている技術は、つながりのない大量のデータを生成し、状況を複雑にしています。

データはサイロ化されたソースごとに保存されており、データ・ソース間の連携はありません。最近では、データ・レイクでさえも内部は複数のサイロに分けてデータを保存しています。

ビジネス・ユーザーは、情報を素早く入手してリアルタイムで意思決定を行う必要がありますが、必要な情報が複数のソースに分散しているため迅速な意思決定を難しくしています。

同様に、クラウド・ファースト、アプリケーション・モダナイゼーション、ビッグ・データ・アナリティクスなどの最先端イニシアティブも、複数の主要なデータ・ソースからデータを統合し提供できなければ先に進めることはできません。

残念ながら、従来のデータ統合技術では、このような課題を解決するために多くの工数と時間、そしてコストがかかってしまいます。



従来の統合技術

データ統合の手法は、通常、ETL（抽出 (Extract) ・変換・加工 (Transform) ・ロード (Load)）またはそれに類似した処理を伴います。このETLは、1970年代から導入されているもので、長い年月をかけて成熟し多様化してきましたが、ETLという名称が示す通り3つの基本的な処理を行います。

1. まず、データをソースから抽出します。
2. 次に、抽出したデータを最終的に求められるフォーマットおよび構造へ変換・加工します。
3. 最後に、変換・加工したデータを運用中のデータストア、データマートやデータウェアハウスなどの最終ターゲットにロードします。

ETL処理は、データのソースや最終のターゲット・システムごとの要件に合わせてスクリプトを作成しテストする必要があり、ひとつのETLソリューションですべてのシナリオに対応できるわけではありません。

変換・加工処理を最後に行うELTと呼ばれる処理もありますが、基本的な考え方はETLと同じです。スクリプトを作成し、テストで処理を確認した後、単一または複数のデータ・ソースから大量のデータをコピーし、単一の統合システムの中で一連のバッチ処理としてデータをレプリケーションし、必要な変換・加工処理を行います。

ETL処理を行うことによるメリットは多く、そのため今日に至るまでETLの手法は採用され続けてきました。

- ETLは、データを効率的に一括処理し受け渡すことができます。
- ETLは、確立された技術であり実績あるベンダーから提供されています。
- ETLのツールは、一括またはバッチ処理でデータを受け渡すための十分な能力を備えています。
- ETLは、多くの企業や組織が社内でスキルとして持っています。

しかしながら最近では、データ処理が複雑になってきていることや、データを統合し実践的な知見を得るといった必要性がますます高まるなかで、ETLを行うことのデメリットも表面化し始めています。

- データの受け渡しが発生する手法は、工数やコストをかけて保守するレポジトリを増やすことになり、必ずしも最善の手法ではありません。
- 大手企業など大規模な組織では、何千ものETLプロセスを毎晩バッチ処理していることもあり、同期のためのスクリプトも必要に応じて簡単に変更できるわけではありません。
- ETLは、バッチ処理でデータを配信するため、エンド・ユーザーは定期的な配信を待たなくてはなりません。構成やスケジュールを調整することで、バッチ処理を高速化することも可能ですが、データを必要な時に即座に配信することはできないため、ETLは、通常夜間にデータを配信するよう設定します。
- ETLは、最近のデータ量や複雑なデータ型、例えば、分刻みで更新されるトランザクション・データや、機械センサーから送られてくる変動の多い数値データなどに対応することはできません。

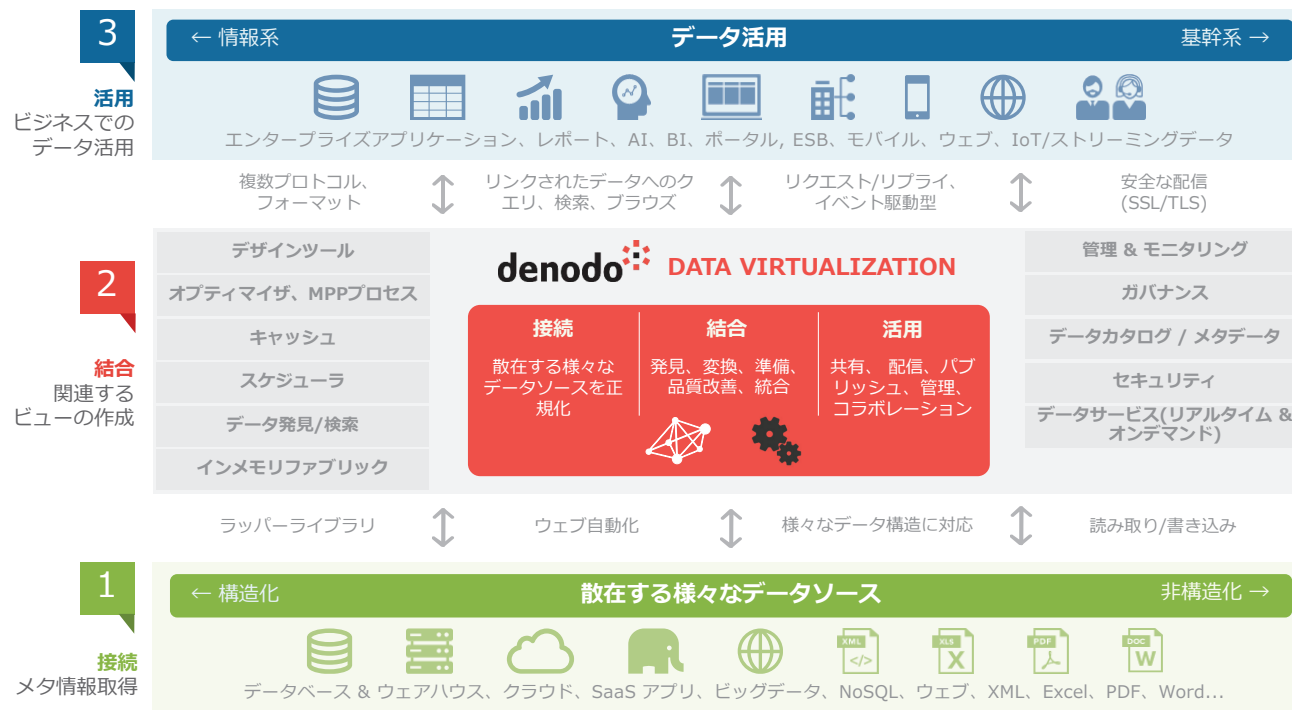
データ仮想化

データ仮想化は、これまでとは全く異なる手法を採用するデータ統合戦略であり、データを物理的に動かし新たな場所に統合するのではなく、データをそれぞれのデータ・ソースにおいたままデータを統合された形でリアルタイムに提供します。

データ仮想化の手法の中でもさらに高度なソリューションでは、データへのアクセスレイヤーをエンタープライズ環境全体にわたり設定し、企業や組織の重要なデータ・ソースにはどこからでもアクセスできるようにします。ビジネス・ユーザーは、データ仮想化レイヤーでデータを検索すれば、該当するデータ・ソースからデータを取得することができます。データ仮想化レイヤーは、データがどのようなフォーマットで何処に保存されているかなどのデータアクセスに関する複雑な要素を抽象化します。ビジネス・ユーザーは、データ仮想化レイヤーに質問を投げかけ、答えとしてデータを受け取り、複雑な処理はデータ仮想化レイヤーで処理させるよう実装することができます。

このような「セルフ・サービス」を実装する際には、ビジネス・ユーザーは、仮想化レイヤーに直接クエリをかけるのではなく、アプリケーションやウェブ・ポータルなどのユーザー指向のインターフェースにアクセスし、必要なデータを仮想化レイヤーから取得するような実装が多くみられます。このアーキテクチャでは、下記の図の通り、データ仮想化レイヤーは、すべてのデータ・ソースと、個人やアプリケーションを含めたすべてのユーザーの間に設定されることが重要な点です。

最新のリアルタイムデータ仮想化アーキテクチャ



データ仮想化で特筆すべきは、データのレプリケーションを一切行わないということです。データ仮想化レイヤー自体にデータは保存せず、必要なデータ・ソースを参照するためのメタデータを持たせます。このアーキテクチャによってデータ仮想化は軽量で実装しやすくなっていますが、メリットは他にもたくさんあります。まず、エンタープライズ環境全体のアクセス制御を個々のソース・システムではなく、データ仮想化レイヤーにまとめることができます。また、開発者は、高度に構造化されたデータから非構造化データまでさまざまなデータ・ソースにAPI経由でデータ仮想化レイヤーから一元的に連携することができるようになります。

つまり、データ仮想化はデータを統合するための最新の戦略です。従来のデータ統合ソリューションが提供する変換・加工や品質管理などの機能の多くを踏襲しながら、同時にコストを抑え、速度や俊敏性を高めたリアルタイムのデータ統合を実現します。データ仮想化は、従来のデータ統合処理とそれに付随するデータ・マートやデータ・ウェアハウスを置き換えることも、従来の機能を補完し拡張するように実装することもできます。

データ仮想化は、オリジナルのデータ・ソース、そこから派生するデータ・ソース、ETL、エンタープライズ・サービス・バス（ESB）、その他ミドルウェア、アプリケーション、デバイス間で抽象化レイヤーおよびデータ・サービス・レイヤーとして簡単に設定することができ、オンプレミスでもクラウドでも企業環境内の情報システムや事業システム間における柔軟性を実現します。

データ仮想化は、従来使用されてきたレプリケーション・ベースのデータ統合手法にはない独自のメリットを提供していることは明らかです。



- 二つ以上の独立したデータ・ソースをシームレスに連携させ、ひとつのソースのように見せ機能させます。構造化データも非構造化データも組み合わせで連携させることができます。
- データ仮想化は、クエリの最適化をリアルタイムで実施するインテリジェンス、キャッシング、インメモリ処理、あるいはソースの制約、アプリケーションのニーズ、ネットワークの要件に基いて最適化の戦略をカスタマイズするなどの付加価値機能にも対応します。
- データ・ソースへのアクセスは、標準的なデータ・ソースでもそこから派生したソースでも、統合済みでも仮想でも、任意のデータ・ソースに、API経由で、オリジナルとは異なるフォーマットやプロトコルでアクセスすることができます。アクセスは分単位で処理され、アクセス制御をかけることもできます。
- データ仮想化では、単一の仮想化レイヤーからすべてのデータへアクセスします。この仮想化レイヤーを通すことで、データの冗長、整合性、品質の問題を素早く明かにし、環境全体でガバナンスを効かせ、セキュリティ制御を実現します。

データ仮想化にもデメリットがひとつあります。ETLのように一括またはバッチでデータを受け渡すことはできません。事例によってはそのような機能が必要な場合がありますが、その場合は、上記で説明した通り、データ仮想化はETLとともに実装することができます。

データ仮想化製品を「機能」から「エンタープライズ・プラットフォーム」まで5階層に分類

データ仮想化ソリューションの広がりとともに、一部の機能が他のソリューションに取り込まれたり、アドオンやフィーチャーとして提供されるようになってきています。データ仮想化機能が、アドオンなのか、ビルトインとして組み込まれているのか、あるいは上記で説明したデータ・アクセス・レイヤーを実現できる完全なエンタープライズ向けデータ仮想化プラットフォームなのか、製品ごとの違いを見極めることが重要です。

データ仮想化製品の5階層は下記の通りです。



データ・ブレンディング機能 データ・ブレンディング機能は、BIツールの機能として提供されることが多いです。データ・ブレンディングは、複数のデータ・ソースを組み合わせるBIツールへデータをフィードしますが、アウトプット先は指定されたツールのみで、指定以外のアプリケーションからアクセスしてデータを利用することはできません。



データ・サービス・モジュール データ・サービス・モジュールは、データ統合ソリューションやデータ・ウェアハウスのベンダーから追加コストで提供されていることが多く、堅牢なデータ・モデリング機能や変換・加工機能を提供しますが、クエリ最適化、キャッシング、仮想セキュリティ・レイヤー、非構造データ・ソースへの対応および全体的な性能は弱い傾向があります。これは、データ・サービス・モジュールは、ETLまたはマスター・データ・マネジメント（MDM）ツールのプロトタイプとして設計されている場合が多いからです。



「SQL対応」製品 これは、ビッグ・データやHadoopベンダーなどを中心とした新興の分野で、ビッグ・データのインフラ技術を仮想化し、リレーショナル・データ・ソースやフラット・ファイルと組み合わせ、標準的なSQLでクエリを実施できるようにするものです。これはビッグ・データのスタックに限っては効果的です。



クラウド・データ・サービス クラウド・データ・サービスは、クラウド環境で展開され、SaaS、クラウド・アプリケーション、クラウド・データベースやマイクロソフト・エクセルなどのデスクトップやオンプレミスのツールとの連携がパッケージ化されています。しかし、階層化されたビューやクエリ実施の委任機能を備えている本来のデータ仮想化製品とは異なります。クラウド・データ・サービス製品は、クラウドのソースに対して正規化されたAPIを提供し、中規模なプロジェクトでデータを簡単に変換できるようにしていますが、ビッグデータ分析、大規模エンタープライズ・システム、メインフレーム、大規模データベース、フラットファイル、非構造化データを扱うようなプロジェクトには対応できません。



データ仮想化プラットフォーム データ仮想化プラットフォームは、エンタープライズ環境で、統合された仮想化データ・レイヤーを通して、多対多でデータを提供するためのデータ仮想化機能を実現するために設計し構築されています。データ仮想化プラットフォームは、幅広い使用目的に高速で俊敏に対応できるよう考慮し設計されており、データ・ソースやユーザーの条件を問いません。他のミドルウェア・ソリューションに代わることも連携することも可能です。

まとめ：データ仮想化について知っておくべき10項目

本ホワイト・ペーパーは、データ仮想化プラットフォームのメリットを簡単に説明するものです。このプラットフォームは、既存のデータ統合ソリューションを補完することも、完全に置き換えることもできます。まとめとし、データ統合の際に考慮すべきデータ仮想化について重要な点を10項目にまとめて提示します。



1

データ仮想化プラットフォームは、既存のデータ統合ツールよりも安価に管理することができます。

既存のツールは、データを複数回にわたり物理的にレプリケーションし、移動し、保存するために高いコストがかかります。データ仮想化は、仮想データ・レイヤーを提供するため、データのレプリケーションやストレージにかかるコストをなくすることができます。



2

データ仮想化は、リアルタイムで結果を出すため、データが配信されるまで数時間から数日待つ必要はなく、データ管理をより迅速に行うことができます。



3

データ仮想化は、既存のデータ・ウェアハウスと併せて実装し、補完することができます。



4

性能劣化は、しばしばデータ送信前の遅れによって起きますが、データ仮想化は、データ・ソースを直接参照し、リアルタイムで実践的な知見を提供するため、性能を最大化することができます。



5

セルフ・サービスBIを実現します。既存のツールは、データを複数回にわたり物理的にレプリケーションし、移動し、保存するために高いコストがかかります。データ仮想化は、仮想データ・レイヤーを提供するため、データのレプリケーションやストレージにかかるコストをなくすることができます。



6

データ仮想化は、エンタープライズ環境におけるすべての種類の情報およびメタデータへアクセスするための一元的なアクセス・ポイントを設定することで、セキュリティ管理、データ・ガバナンス、性能モニタリングを可能にし、データ・ガバナンスを実現します。



7

データ仮想化は、10年の実績をもつデータ連携技術の機能を一部集約したのですが、性能の最適化、セルフ・サービス、ディスカバリ機能を提供するなど、データ連携技術をはるかに超えています。



8

データ仮想化のプロジェクトは、通常、実装から6ヶ月以内に投資を回収しています。従来の統合手法と比較すると、データ仮想化を採用した企業は50%から80%の時間の節約を実現しており、データ仮想化は非常に大きなROIを提供します。



9

データ仮想化は、プロトタイピング機能を提供し、エンタープライズ環境規模で実装する前に戦略をテストし確認することで、従来の手法と比較して、より柔軟に実装を進めることができます。



10

データ仮想化は、データを統合し、予測分析のためにデータを整形し、リアルタイムでユーザーに提供することで、ビッグデータ活用のための環境を整えます。

Denodoプラットフォーム：データ仮想化のための最先端プラットフォーム

Denodoプラットフォームは、真にエンタープライズ向け仮想化プラットフォームとして関連製品とともにDenodo Technologiesによって開発され

DENODOプラットフォームが提供する主な機能は下記の通りです。他のどのデータ仮想化ソリューションをも凌駕するものです。



ダイナミックなデータ・カタログを提供します。検索可能で、状態を動的に認識するインターフェースからシームレスなデータアクセスを実現します。



Dynamic Query Optimizer機能を提供し、最適なクエリ戦略を選択するインテリジェンス機能で、より高速なデータ・アクセスを提供します。



インメモリ並列処理を提供し、データ・アクセスをさらに加速し比類なき速度を実現します。



全く新しく設計し直されたインターフェースを提供し、ビジネス・ユーザーおよびIT関係者のそれぞれのニーズに対応します。



ライフサイクル管理を自動化することで、ユーザーは、データ管理ではなくデータを活用して意思決定を行うためにより多くの時間を費やせるようになります。



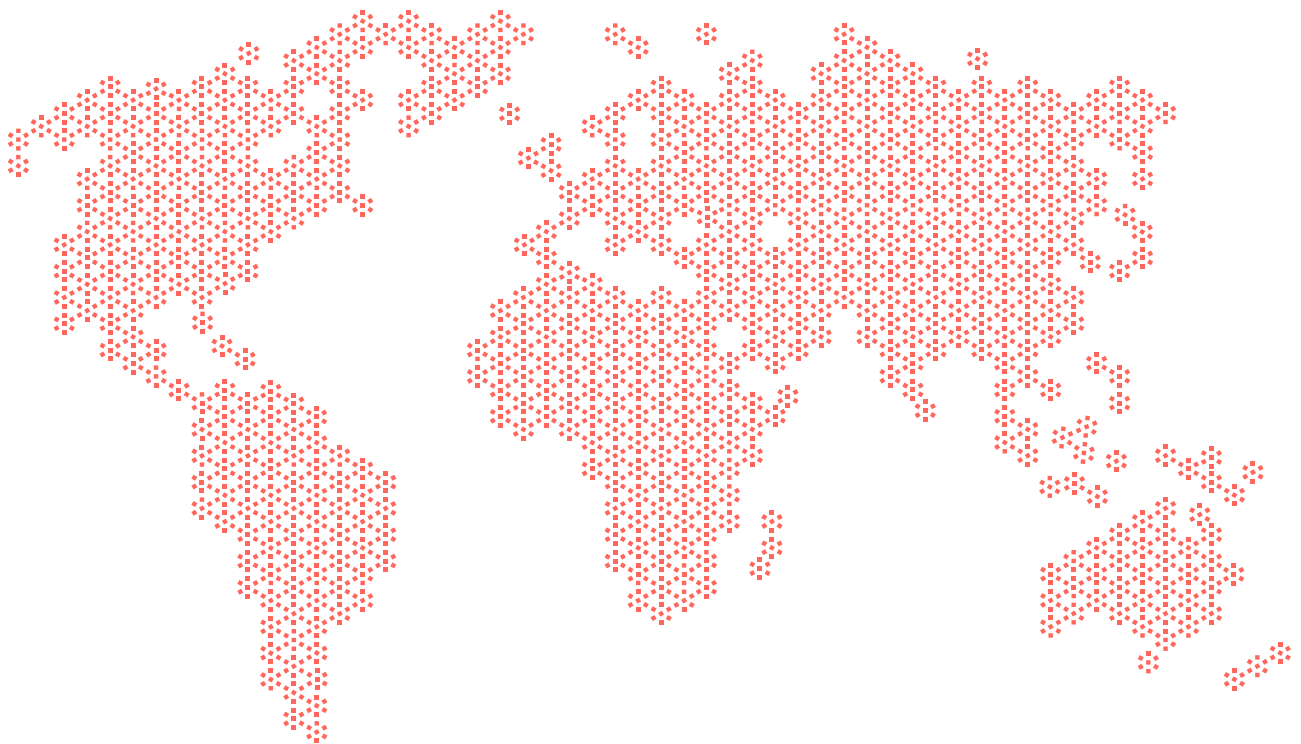
最先端のデータ・サービス・レイヤーを提供し、OAuth 2.0



セキュリティとガバナンスをシームレスに実現し、企業や組織は、単一の制御・管理点から保有するデータすべてに安全にアクセスしデータを選択できます。



AWSやMicrosoft Azureのような主要なクラウド・マーケットプレイス、さらにはDockerでも提供されています。



Denodo Technologies is the leader in data virtualization providing agile, high performance data integration, data abstraction, and real-time data services across the broadest range of enterprise, cloud, big data, and unstructured data sources at half the cost of traditional approaches. Denodo's customers across every major industry have gained significant business agility and ROI.

Visit www.denodo.com | Email info@denodo.com | Discover community.denodo.com

