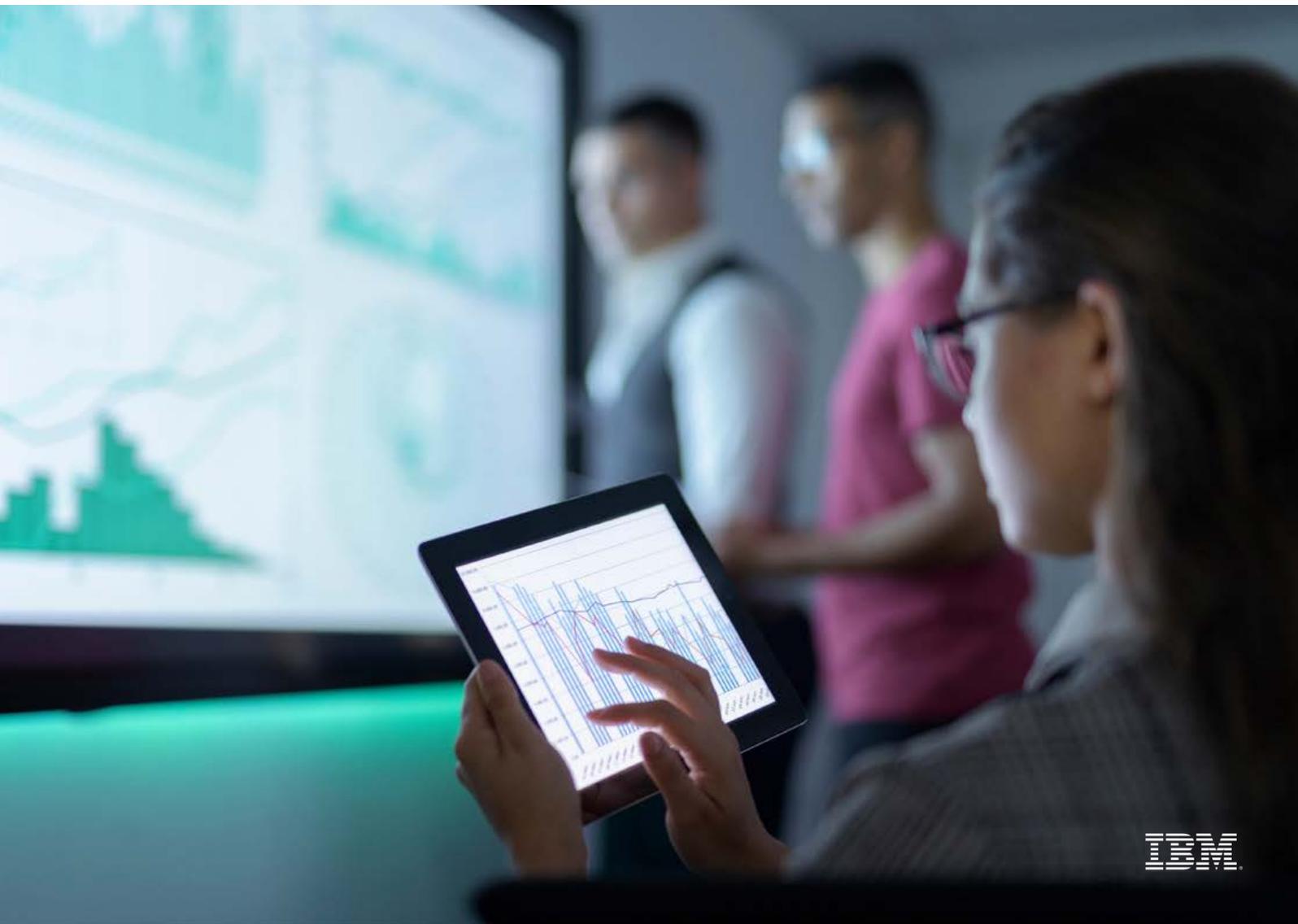


データ・サイロを解消する：
多数のシステムをまとめて照会
IBM Cloud Pak for Data に
おけるデータ仮想化



ハイライト

- 複数のデータベースやビッグデータ・リポジトリを、個別でも一括でも照会できます
- アクセス制御とガバナンスを一元管理できます
- 世界中に分散した多数のデータベースでも、1つのデータベースのように見せることができます
- 拡張が容易で高性能なプラットフォームにより、データ・アナリティクスを簡素化できます

背景

データはいたるところにあり、今日の世界におけるほとんどのビジネスはデータ・ドリブンで進められています。ビジネスでは、規模も多様性も増大の一途をたどるデータ・ソースからデータを収集、分析しながら、事業を運営しており、そうしたデータ・ソースはおそらく数百万にも達します。そのようなデータを一元的に収集、管理、保管、処理、分析する上での複雑さやコスト、時間、およびエラーのリスクは、指数関数的に増大しています。同時に、すべてのデータの源であるデータベースやリポジトリは、豊富な処理機能やデータ・ストレージ機能が備わり、より高性能になってきています。

データ仮想化の概要

IBM® Cloud™ Pak for Data (旧称 IBM Cloud Private for Data)におけるデータ仮想化は、こうしたすべてのデータ・ソースを、コンステレーションと呼ばれる、データ・ソースやデータベースの単一の自己バランス型のコレクションへと結びつける、独自の新技术です。図1をご覧ください。データを中央の保管場所にコピーおよび保管の上、そこでアナリティクスや照会を実行するようなことは、もはや必要ありません。アナリティクス・アプリケーションが送信した照会を、当該データ・ソースが存在するサーバー上で処理されます。照会結果はコンステレーション内で統合され、元のアプリケーションに返されます。データは、コピーされることなく、元のデータ・ソースのみに存在します。

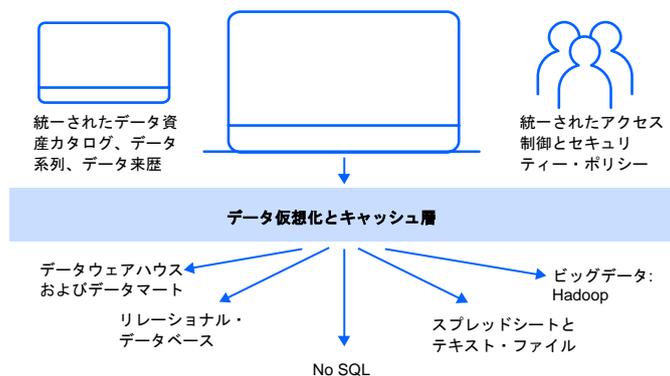


図1: Cloud Pak for Data におけるデータ仮想化

データ仮想化の働き

アプリケーションは、あたかも単一の IBM Db2®データベースに接続しているかのように、IBM Data Virtualization に接続します。接続したアプリケーションは、まるで単一のデータ・ソース・データベースに対して照会を実行しているかのように、システムに対して照会を送信できます。扱われたすべてのデータ・ソースのうち照会に適合するデータを持つものが協働してワークロードを分類、計算します。

価値ある機能

IBM Data Virtualization には、企業が自社のデータをより効果的に処理できるようにするための重要な機能が数多くあります。

協調的なコンピューティング

あらゆるデータ・ソースの処理能力を利用し、それぞれのデータ・ソースに物理的に保管されているデータにアクセスすることで、データ移動やコピーにかかる待ち時間が発生しません。さらに、すべてのリポジトリ・データにリアルタイムでアクセスできる上、ガバナンスやデータの誤りといった問題が事実上排除されます。データ・ストレージの抽出/変換/ロード(ETL)や複製の必要もないため、処理時間が短縮されます。この処理により、意思決定アプリケーションやアナリストに、既存の方式に比べてより迅速で信頼性の高い、リアルタイムの洞察がもたらされます。また、既存方式に対する高度な補完性も維持しているため、履歴、アーカイブ、法令順守といった目的で一部のデータのコピーや移動が引き続き必要な場合でも、既存方式と容易に共存できます。

スキーマ・フォールディング

分散データ・システムにおける一般的なシナリオでは、多数のデータベースが共通スキーマ下でデータを保管します。例えば、テナント・グループごと、あるいは地域ごとの販売データやトランザクション・データが保管されているデータベースが複数存在している場合があります。IBM Data Virtualization は、システム間で共通のスキーマを自動的に検出し、これをデータ仮想化で単一のスキーマとして見えるようにします。これが、スキーマ・フォールディングと呼ばれる処理です。例えば、20個のデータベースそれぞれに存在する SALES 表を、1つの SALES 表として表示でき、構造化照会言語(SQL)を通じて1つの仮想テーブルとして照会できます。

シンプルなビュー結合ツール

図2に示すように、明快なインライン・ツールにより、さまざまな種類の、多くの場合地理的にも分散しているデータベース群を跨いで、テーブル・ビューを定義できます。

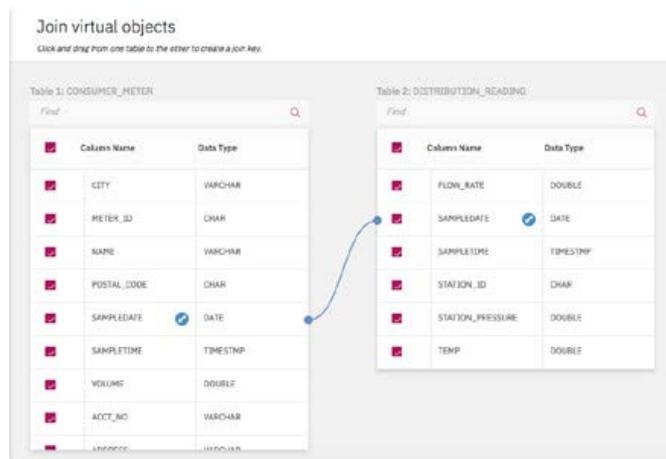


図2: 直感的なインターフェースによりテーブル・ビューを簡単に結合

セキュリティ

コンストレインション内の通信もアプリケーションに戻される通信もすべて、セキュリティ性の高い、堅牢かつ高性能な IBM テクノロジー、Secure Sockets Layer (SSL)、および Transport Layer Security (TLS) 暗号化によって、標準プロトコルを使用して暗号化されます。

対応しているプラットフォーム

IBM Data Virtualization は、アプリケーションからは、単一の Db2 データベース・インスタンスのように見えます。そのため、一般に広く使用されている Db2 接続のクライアントやアプリケーションを IBM Data Virtualization に接続して、変更せずに機能させることができます。これは、照会の対象となるデータ・ソースのコレクションに、次のような多様な種類のデータ・ソースが混在している場合でも可能です。

- PostgreSQL
- Oracle
- Netezza®
- Microsoft SQL Server

IBM Data Virtualization テクノロジーによって、あらゆる SQL 方言との変換が行われます。したがって、ご使用のアプリケーションでは、SQL、プロシージャ型言語/SQL (PL/SQL)、および SQL PL を、あたかもそれらが Db2 データベースに対して直接機能するかのように自由にコーディングできます。ターゲット・データ・システムがその構文をサポートしているかどうかを判断する必要はありません。いかなる変更もアップグレードも必要なく IBM Data Virtualization に接続できる、一般的なツールとしては、次のようなものがあります。

- IBM Cognos® Business Intelligence (BI) ソフトウェア
 - Tableau
 - MicroStrategy
 - Looker
 - Plotly
 - R
 - Jupyter
- アプリケーションの接続先となるデータ仮想化サービス・ノードは、Cloud Pak for Data 内のマイクロサービスです。

Apache Hive	IBM Informix® データベース・サーバー
Cloudera Impala	MariaDB
Db2 ソフトウェア	MySQL
IBM Db2 Big SQL	Netezza
IBM Db2 Event Store	Oracle
DerbyDB	PostgreSQL
Excel およびコンマ区切り値 (CSV) ファイル	SQL Server
Hortonworks Data Platform (HDP) と Apache Hive	

表 1: サポートされているデータ・ソース

現在サポートされているデータ・ソースです。今後のリリースで、新たなデータ・ソースが追加される予定です。

最小ハードウェア要件

Cloud Pak for Data におけるデータ仮想化には、以下の構成が必要です。

- 16 (v) コアのプロセッサ
- 少なくとも 64GB の物理ランダム・アクセス・メモリー (RAM)
- 200GB のディスク・スペース (推奨)

IBM Data Virtualization の一般的なシナリオ

IBM Data Virtualization は、データやアナリティクスの結果に時間的制約のある、高度に分散されたデータ・セットを対象とするアナリティクスに適しています。また、そうした特定のデータ・セットに対してアナリティクスが一回限りの操作になる可能性がある場合にも効果的です。

さらに、一部のデータ・ソースからのバッチ・コピーの待ち時間が、アナリティクス結果に求められるビジネス上のニーズを上回っているようなシナリオにも適用できます。

多くの組織では、データを複製して新しいデータ・リポジトリを作成することで、アナリティクスに求められる業務部門 (LOB) のニーズを満たしています。このプロセスの場合、物理資産を構成し、リポジトリに対してデータをロードしたり変換したりするための新規 ETL を作成し維持することが必要になります。LOB でやっとならざるを得なくなった時には、既にそのデータは古くなっている、ということがよくあります。

多くの IT 組織にとって、既存のアプローチでは限界に近づきつつあります。データ・ソースの数や多様性が増し、アナリティクスに対するニーズも高まる中で、このアプローチはもはやスケラブルとはいえません。IBM Data Virtualization なら、IT 組織の生産性を向上させ、LOB から企業全体のデータにアクセスできるスケラブルなアプローチを提供できます。

多くの場合、データ (例えば個人情報など) のコピーや移動にはポリシー上の問題や法的な問題が絡みます。こうした制約事項は、ターゲット層の分析結果を必要とするビジネスの妨げとなりかねません。IBM Data Virtualization なら、保護データをデータ・ソースに残したまま、ターゲット層の照会結果だけを返すことができるため、こうした問題を解消できます。

今日、データ・サイエンティストがアナリティクスによって仮定を検証するには、事前に、データレイクを作成し、関心のあるデータ・ソースからデータをコピーし、そのデータを統合しなければなりません。IBM Data Virtualization なら、データレイクの必要性はありません。データ・サイエンティストは、IBM Watson® Studio のようなツールを直接データ・ソースにつなぐだけで、仮定の検証に必要なデータを統合できます。

鍵となる分析プロジェクトに俊敏性をもたらす

IBM Data Virtualization を利用すれば、実用的な統一データを、必要な時に、必要な方法で、また分析上のニーズを満たすスピードで簡単に獲得できます。この技術は、統合のスピードならびにパフォーマンスの向上や、変化の絶えないビジネス・ニーズへの適応を支援する意思決定の改善につながります。

Cloud Pak for Data で IBM Data Virtualization を利用することで、次のような、広範囲の主要なイニシアチブに対応できます。

- SoE(Systems of Engagement)の構築をより迅速かつ簡単に行うためのモダナイゼーション
- ビジネスの即時的なニーズを満たすリアルタイム・アナリティクス
- 組織データにアクセスするためのコストと複雑さを軽減するための最適化

IBM Data Virtualization が可能にするセルフサービス BI。 仮想的で再利用可能なデータ資産によって、ビジネスで使いやすいデータ表現が提供されるため、ユーザーは、物理データ層やデータがどこに保管されているのかといった複雑なことについての知識がなくても、データと対話できるようになります。また、複数の BI ツールやレポート・ツールで、データ仮想化層からデータを取得できるようにもなります。

IBM Data Virtualization が提供する、統一された 360 度のビュー。 仮想化されたデータ資産によって、完全なデータ・ビューがリアルタイムで提供されます。仮想データ層が、組織データに対するユーザーの理解力や活用力を高める、統一されたビジネス情報の統合ビューとして機能します。

IBM Data Virtualization が提供する、俊敏性の高いサービス指向アーキテクチャー(SOA)によるデータ・サービス。 データ仮想化層は、SOA アプリケーションにデータサービス層を提供します。基礎となるソースに触れる必要がなく、データ・アクセス・ロジックをカプセル化するオートディスカバリーおよびメタデータ・マッピングにより、仮想資産の作成を高速化します。

また、データ仮想化により、複数のビジネス・サービスで中央の場所からデータを取得できるようになるとともに、ビジネス・サービスと物理データ・ソース間の疎結合が提供されます。

IBM Data Virtualization による情報制御面の向上。 一元管理型のアクセス制御を通じてデータ品質が向上し、堅牢なセキュリティ・インフラストラクチャーが提供されるとともに、物理的なデータ・コピーの削減によりリスクが減少します。組織のデータ・ストアや、各種データ・ソース内のデータ間の関係がメタデータ・リポジトリに登録されるため、透過性と可視性が生まれます。

まとめ: 意思決定の方法を変革し、意思決定を迅速化
Cloud Pak for Data におけるデータ仮想化は、以下を模索している企業にとって理想的です。

- 収益性、成長、リスク軽減
- 俊敏性と生産性の劇的向上
- 既存の IT 投資の最適化

サーバーやストレージに対する既存の投資を有効に活用しながら、データの不要な複製を削減し、複製やインフラストラクチャー管理にかかる関連コストを軽減します。管理の簡素化と、一連の SQL アプリケーション・プログラミング・インターフェース(API)により、ビジネスでリアルタイム・アナリティクスによる利点を活用できるようになります。

IBM Data Virtualization について詳しくは、以下のサイトをご覧ください:

<https://www.ibm.com/jp-ja/analytics/data-virtualization>

© Copyright IBM Corporation 2019

東京都中央区
日本橋箱崎町 19 番 21 号
日本アイ・ビー・エム株式会社

Produced in Japan
January 2019

IBM、IBM ロゴ、ibm.com、Cognos、Db2、IBM Cloud、IBM Watson、および Informix は、世界の多くの国で登録された International Business Machines Corporation の商標です。他の製品名およびサービス名等は、それぞれ IBM または各社の商標である場合があります。現時点での IBM の商標リストについては、<http://www.ibm.com/legal/copytrade.shtml> をご覧ください。

Netezza は、IBM International Group B.V., an IBM Company の登録商標です。

Microsoft、Excel および SQL Server は、Microsoft Corporation の米国およびその他の国における商標です。

本書の情報は最初の発行日の時点で得られるものであり、予告なしに変更される場合があります。すべての製品が、IBM が営業を行っているすべての国において利用可能なわけではありません。

IBM 以外の製品とプログラムの操作またはサービスの評価および検証は、お客様の責任で行っていただきます。本書に掲載されている情報は特定物として現存するままの状態を提供され、第三者の権利の侵害の保証、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任なしで提供されています。IBM 製品は、IBM 所定の契約書の条項に基づき保証されます。

適切なセキュリティの実施について: IT システム・セキュリティには、企業内外からの不正アクセスの防止、検出、および対応によって、システムや情報を保護することが求められます。不正アクセスにより、情報の改ざん、破壊もしくは悪用を招くおそれがあり、またはシステムの損傷や、他のシステムへの攻撃を含む悪用につながるおそれがあります。完全に安全と見なすことができる IT システムまたは IT 製品は存在せず、また単一の製品、サービス、またはセキュリティ対策が、不正使用や不正アクセスを防止する上で、完全に有効となることもありません。IBM のシステムおよび製品は、法律に沿った包括的なセキュリティの取り組みの一部となるように設計されており、これらには必ず追加の運用手順が伴います。また、最高の効果を得るために、他のシステム、製品、またはサービスを必要とする場合があります。IBM は、何者かの悪意のある行為または違法行為によって、システム、製品、またはサービスのいずれも影響を受けないこと、またはお客様の企業がそれらの行為によって影響を受けないことを保証するものではありません。

IBM の将来の方向性および指針に関する記述は、予告なく変更または撤回される場合があります。これらは目標および目的を提示するものすぎません。