データウェアハウスとデータレイクをクラウドで近代化

メリット

- データ品質を維持しながらデータ統合を実現し、ガバナンスも確保するメタデータ主導のクラウドデータマネジメントで柔軟性を強化
- 手作業のコーディングの必要性を排除するAIベースの機能でコストを削減
- コーディング不要の統合と数百のアプリケーションへのカスタマイズ不要な接続性により俊敏性を向上
- 自動化とインテリジェンスによりソフトウェア開発のライフサイクル全体の生産性を向上
- 自然言語検索機能によりセルフサービス型のデータアクセスとアナリティクスを実現

インテリジェンスと自動化を活用したクラウドレイクハウスデータマネジメント

今日、企業は開発期間の短縮やコストの削減、効率の向上、業務上の意思決定のための信頼できるインサイトの提供を目的としたデジタルトランスフォーメーションに取り組む中で、データ主導の色合いを強めています。こうした目標を達成するために、新しいクラウドデータウェアハウスおよびデータレイクを構築するか、オンプレミスのデータウェアハウスをクラウド内で統合することによって、アナリティクスイニシアチブをモダナイズ（最新化）しています。また、データウェアハウスとデータレイクを1つのデータプラットフォームに統合する最新のクラウドテクノロジーの導入も進めています。このプラットフォームは「レイクハウス」と呼ばれることもあります。

データマネジメントは、クラウドアナリティクスプロジェクトの価値を最大化するための鍵となります。クラウドの俊敏性、柔軟性、拡張性を最大限に活用するため、またオンプレミス環境で頭を悩ませてきた問題を繰り返さないようにするため、クラウドデータウェアハウス、データレイク、レイクハウスでのデータ品質とデータマネジメントには系統立てたアプローチで取り組む必要があります。

インテリジェントな自動エンタープライズクラウドデータマネジメントソリューションを導入してクラウド内でワークロードを構築して管理することで、データの透明性を高め、多様なデータソースに接続し、次第に複雑性を増すオンプレミスおよびマルチクラウド環境を管理することができます。このアプローチにより、業務アナリストからデータサイエンティスト、データエンジニアまで、社内のユーザーが迅速かつ容易にアナリティクスイニシアチブのための高品質なデータへアクセスできるようになります。その結果、イノベーションが促進され、競争上の優位性を獲得できます。

モダナイズに向けた段階的アプローチ

まずは1つの部門レベルのプロジェクト用にクラウドデータウェアハウスを作成してソリューションをパイロット運用し、クラウドオプションを評価するというケースがよくあります。これが成功し、ある程度のビジネス上の利点を得られたら、次の段階では従来のオンプレミスデータウェアハウスプロジェクトをクラウドに移行するか、データサイエンス調査用および新しいデータソースのキュレーション用の生データを格納するクラウドデータレイクを追加します。

クラウドへの移行を開始する時点では、手作業のプロセス（手作業によるコーディングなど）を選択する企業もありますが、このアプローチはすぐにITチームにとってオーバーロードとなり、結果的に保守コストの高騰を招き、拡張が困難になります。また、特定のデータマネジメントのニーズへの対処に限定したソリューションを選択する企業もあります。しかし、結局は断片化したアプローチをパッチワークのようにつなぎ合わせることになり、複雑性が高まり、不要なリスクとコストが発生してしまいます。

インフォマティカは、企業がこうした落とし穴を回避しながら、クラウドデータウェアハウスとデータレイクイニシアチブを促進することを可能にするソリューションを提供します。それが、業界最先端のクラウドデータマネジメントソリューション、Informatica Intelligent Cloud Services℠（IICS）です。IICSは、クラウドデータウェアハウスおよびデータレイクを対象とした最新の目的特化型クラウドネイティブデータマネジメントソリューションで、マイクロサービスベースかつメタデータ主導で、AIを活用しています。クラウドプラットフォームやテクノロジープロバイダー（Amazon、Microsoft、Snowflake、Databricks、Googleなど）は問いません。

主な機能

クラウドデータマネジメント

インフォマティカは、クラウドデータウェアハウスおよびデータレイクを成功に導く、データのカタログ化、取得、統合、クレンジング、ガバナンスのための業界唯一の総合的なソリューションを提供します。他社のソリューションでは複数の異なるコンポーネントが必要で、場合によっては別々のベンダーから入手する必要もあります。IICSは、クラウドデータ統合、クラウドデータ品質、メタデータマネジメントのための、AIを活用した包括的な統合クラウドネイティブ機能を備えており、これらがすべて1つのインテリジェントなプラットフォームに組み込まれています。さらに、インストールと設定が最小限で済むAIベースの自動化、自動アップグレード、高可用性を実現する統合テクノロジースタックにも対応しており、高い信頼性を保証する数多くの重要な認定（SOC、SOC2、HIPAA、ISO/IEC 27001、Cloud Security Alliance、Privacy Shield Frameworkなど）を取得しています。

データ統合

直感的かつ視覚的な開発環境であらゆるタイプのデータを迅速に取得、統合することで、クラウドデータウェアハウスおよびデータレイクにデータを供給するデータパイプラインを簡単に構築できます。IICS Cloud Data Integrationは、マルチクラウドでもオンプレミスでも、実質的にあらゆるタイプの企業データに対する、あらかじめ組み込まれたクラウドネイティブの接続機能を備えています。Cloud Data Integrationは拡張性とパフォーマンスを念頭に構築されており、プッシュダウンの最適化などの重要な最適化機能を通じてユーザーはデータを効率的に処理できます。また、自社独自のクラウド環境でデータ統合タスクを処理して、Cloud Data Integration Elastic Spark処理を活用して複雑な統合を行ったり、管理不要で自動拡張や自動調整を行う完全管理型の高度なサーバーレスランタイム環境を利用したりすることができます。

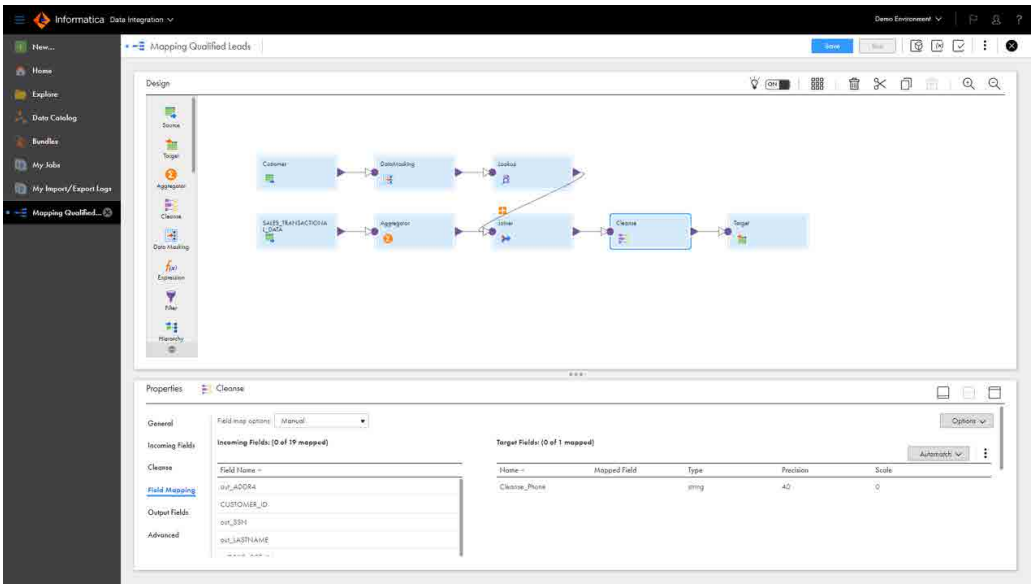

図1 : Cloud Data Integrationがクリック＆ドラッグでデータパイプラインを作成可能

ウィザードを使用すれば、データパイプラインを手早く作成できます。このウィザードには、すぐに使い始められる事前定義済みの統合、インテリジェントなデータディスカバリ、複雑なファイルの自動解析、AIベースのトランスフォーメーションに関する提案といった機能が備わっています。これによって、クラウドデータウェアハウスおよびデータレイクに取り込むデータを簡単に特定できるとともに、他のプロジェクトでデータパイプラインを再使用することもできます。IICS Cloud Mass Ingestionによって、ファイルやデータベース、変更データキャプチャ、リアルタイムデータのストリーミングなどのさまざまなソースからデータを取り込むことができます。

データ品質

クラウドデータウェアハウスには必ず信頼できるデータを格納し、データレイクで「データの泥沼化」が発生しないようにする必要があります。Informatica Cloud Data Qualityは総体的なアプローチで迅速にデータをプロファイリングできるクラウドネイティブ機能を備えているため、データをクラウドデータウェアハウスやデータレイクへ移動する前に、データの品質に関する問題を特定、修正、監視できます。Cloud Data Qualityによりデータを変換して、すべてのデータをクレンジング、標準化、エンリッチ化できます。組み込み済みの幅広いデータ品質ルールを使用することで（追加のコーディングは不要）、信頼性の高いデータ配信およびアナリティクスを実現できます。

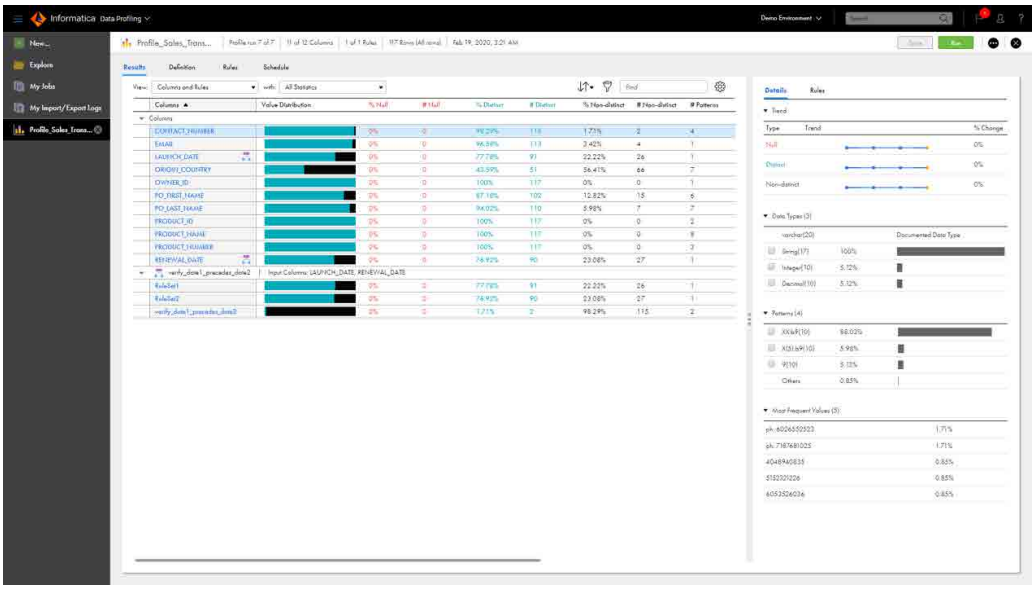

図2 : Cloud Data Qualityが自動的にプロファイルに合ったルールを特定

Cloud Data Qualityは、あらゆるプロジェクトタイプに適合するビジネスルール、標準化、マッチング、世界の住所データクレンジング、多目的のデータ品質管理などの機能を備えています。メタデータ主導の人工知能であるCLAIRE®エンジンが、類似するデータがどのように管理されているのかという情報に基づき、データ品質ルールに関するインテリジェントな提案を行います。

メタデータマネジメント

コラボレーションを促進するとともに、AIベースのデータ統合、データ品質、データカタログ化、データガバナンスのためのエンドツーエンドのプロセスを実現する、エンタープライズ規模で統合したメタデータマネジメント基盤を提供しているのはインフォマティカだけです。IICSのメタデータ主導のアプローチにより、視覚的な開発環境でデータパイプラインを構築できます。あらゆる主要なクラウドプラットフォーム（Amazon Web Services、Microsoft Azure、Google Cloud Platform）と最新の処理エンジン（Spark）上で実行でき、既存のパイプラインの再コーディングは不要です。このアプローチでは、データを自動的に特定、タグ付け、関連付けし、クラウドデータウェアハウスおよびデータレイクにプロビジョニングできます。

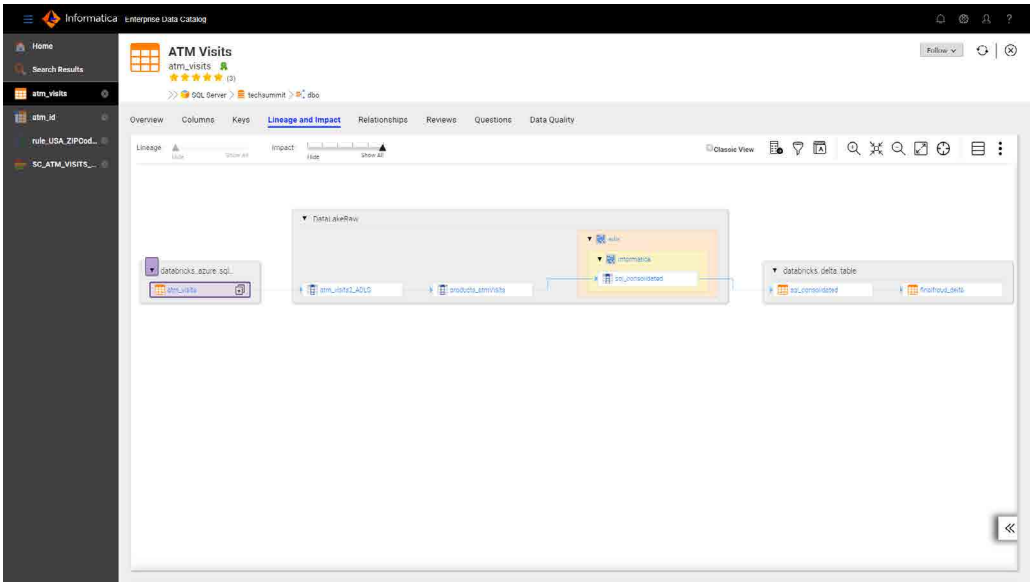

図3 : Enterprise Data Catalogがエンドツーエンドのデータリネージを実現

インフォマティカは、すべてのエンタープライズシステムからメタデータをスキャンして収集し、ビジネスのコンテキストでこれらのメタデータをキュレーションおよび強化して関係とリネージを推論することで、データ資産とその関係を明らかにして、企業データカタログを提供します。収集および分析対象のメタデータの分類は以下のとおりです。

- **技術** : データベーススキーマ、マッピングとコード、トランスフォーメーション、品質チェック
- **業務** : グロッサリ用語、ガバナンスプロセス、アプリケーションとビジネスのコンテキスト
- **運用とインフラストラクチャ** : ランタイムの統計情報、タイムスタンプ、ボリューム指標、ログ情報、システムおよび位置情報
- **使用状況** : ユーザー評価、コメント、アクセスパターン

これらの4種類のメタデータが、共通のメタデータ基盤のための土台となります。インフォマティカは、AIと機械学習を適用してテクノロジースタック全体に統合することで、これらのメタデータをアクティブにします。

主なメリット

生産性の向上とコストの削減

インフォマティカの包括的な統合クラウドデータマネジメントソリューションにより、生産性を高めプロジェクトのリスクを軽減しながら、インテリジェンス、自動化、効率向上を実現できます。高度なサーバーレスおよびSpark処理により、運用コストを削減し、リソース利用を最大限に高めることができます。これらのすべての機能を基盤に、業務上重要な意思決定の全体像を把握できます。このソリューションによって、ガバナンスは維持したままで、高品質なデータの特定、取得、統合、クレンジング、マスター、統制、セキュリティ確保を行い、すべてのユーザーに提供することができます。

- データアナリスト、データサイエンティスト、データスチュワードは、エンドツーエンドの詳細なデータリネージ、データ品質プロファイル、またクラウド／AIベースのタグ付けによってデータを迅速に把握できるため、データの分析と機械学習モデルの構築により多くの時間を向けられるようになります。
- データエンジニアは、トランスフォーメーションに関する次善策の提案、自動データパイプライン生成、企業データカタログから直接実行される自動データウェアハウスハイドレーション、IoTデータやブログデータなどの複雑なファイルの自動解析を介して、多数の実装タスクを部分的に、または場合によっては完全に自動化できます。
- データセキュリティおよびプライバシーの担当者は、簡単にデータの不正利用を検出して機密データを保護できるとともに、システムおよび地理的な境界をまたいだ機密データの自動検出と拡散により、適切なコントロールを維持していることを示すことができます。

柔軟性と拡張性の強化

インフォマティカは、データ統合、データ品質、データガバナンス、メタデータマネジメントのすべてを通じて業界最先端の卓越した機能を提供している、唯一の独立系エンタープライズクラウドデータマネジメントベンダーです。データの管理能力を大幅に制限するとともにリスクとコストを高める可能性がある、複数のポイントソリューションツールの統合を心配する必要はありません。

インフォマティカのクラウドデータマネジメントは、将来にわたり変化の速いマルチクラウド環境と進化するアナリティクススタックに適応していくことができるので、クラウドデータウェアハウスやデータレイクを柔軟に展開することができます。さまざまなオンプレミスおよびクラウドアプリケーション／データソースに接続できるため、大量のデータをシームレスに統合し、アナリティクスワークロードを高パフォーマンスで簡単に管理できます。

インフォマティカのソリューションは、あらゆるタイプのデータソースへネイティブ接続し、ユーザーがETLを取り込んでクラウドデータウェアハウスに処理をプッシュダウンできるよう支援します。このため、クラウドプラットフォームを変更する場合でも再コーディングを行わずに簡単に移行できるので、投資の保護につながります。Sparkベースのデータ統合では、ユーザーはそれぞれの要件を指定するだけで、後はCloud Data Integration Elasticが必要な容量をプロビジョニングし、最適なリソース管理と拡張性を提供します。高度なサーバーレス処理により、サーバーやクラスタを管理する必要がなくなるとともに、自動調整と自動拡張が可能になります。

インフォマティカについて

デジタルトランスフォーメーションによって我々の期待値が変化しています。より良いサービスを、素早く、便利に、低コストで利用したいという期待が高まっているのです。企業も状況に応じて変化する必要があります。そしてそのヒントは「データ」にあります。

エンタープライズ向けクラウドデータマネジメントで世界をリードするインフォマティカは、俊敏性の向上、新たな成長機会の獲得、新しいソリューションの開発を実現するための洞察を通じて、あらゆる産業や分野の企業がインテリジェントにビジネスをリードできるよう支援します。インフォマティカは、あらゆるデータを徹底的に重視し、企業の成功に必要とされる汎用性を提供します。

インフォマティカは、企業がこれからのインテリジェントな破壊的イノベーションを推進できるよう、当社が提供するあらゆるサービスを通じてデータの力を継続的に引き出すことを支援します。

俊敏性の向上

IICSは、ジョブの迅速な実装、最小限のインストールと設定、自動アップグレード、高速データオンボーディング、アジャイル開発とダウンタイムの最小化のための完全な統合テクノロジースタックを実現します。可視性を維持しながら数百のアプリケーションにカスタマイズ不要の接続性を通じて接続でき、コーディング不要の統合機能と大量のワークロードに対応した広範囲の大量取り込み機能を備えているため、高度なトレーニングの必要なしに、すぐに使い始めることができます。

クラウドの俊敏性を活かすことで、新しい機能のオンボーディングや市場展開までの期間短縮を実現するとともに、導入環境のROIを高めながら短期間で価値を引き出すことが可能になります。また、プロジェクトの迅速な実行と完了、統制したセルフサービス環境の提供、プロジェクトコストの削減、組み込みのビジネス継続性によるリスクの軽減などを通じて、より迅速にプロジェクトの価値を実現することができます。

業務アナリストとデータサイエンティストは、Googleと同様のセマンティック検索により、信頼できる豊富なデータへ簡単にアクセスできます。CLAIREは、アナリティクスの強化に役立つ新しいデータ資産について、最適な提案の提供をサポートします。これによって俊敏性が高まり、アドホックなデータサイエンスアナリティクスもIT部門の手を借りずにデータ利用者が自ら実行できるようになります。

インテリジェンスと自動化を活用したクラウドデータマネジメントにより、開発期間を短縮し、高いROIを実現することで、クラウドデータウェアハウス／データレイクプロジェクトを確実に成功に導くことができます。

今後のステップ

インフォマティカのクラウドデータマネジメントの詳細について、いつでもお気軽にお問い合わせください。

インフォマティカのクラウドデータ統合サービスの一部であるクラウドデータウェアハウスの30日間無料体験版を用意しています。今すぐご登録ください。

Informatica

〒105-6226 東京都港区愛宕2-5-1 愛宕グリーンヒルズMORIタワー26階 電話：03-6403-7600(代表) FAX：03-3433-1021

IN17_0221_03569

© Copyright Informatica LLC 2021。Informatica、Informaticaロゴ、Informatica Intelligent Cloud Services、およびCLAIREは、米国およびその他の国におけるInformatica LLCの商標または登録商標です。インフォマティカの商標の最新版は、https://www.informatica.com/jp/trademarks.htmlをご覧ください。その他すべての企業名および製品名は、各社が所有する商号または商標です。本文書に記載されている情報は、予告なく変更されることがあり、現状のまま提供され、明示または黙示を問わず一切の保証を伴いません。