

AI活用とアナリティクス開発を加速させるデータパイプラインの構築

メリット

- AIプロジェクトを成功に導き、イノベーションを促進し、ビジネスの俊敏性を強化
- 信頼できるハイブリッドデータを大量に取り込み、インサイトをより迅速に獲得
- 大規模なデータエンジニアリングパイプラインの構築と維持に要するコストを削減
- データサイエンティストが適切なデータセットを迅速に探索できる環境を確立して、モデルトレーニングを促進
- 生産性を高め、既存のリソースでより多くの成果を達成

Databricks社とインフォマティカのソリューションで、エンドツーエンドのインテリジェントなデータパイプラインを構築

先進のアナリティクスによる競争優位性の獲得を考えている企業にとって、人工知能 (AI) および機械学習 (ML) のプロジェクトは重要な優先課題です。これらのテクノロジーにはビジネスに大きな影響をもたらす潜在性がありますが、業界調査の結果からほとんどのAIプロジェクトが失敗に終わっていることが分かっています。¹なぜなのでしょう。MLのコードを記述することは、MLプロジェクトを成功に導くためのほんの一部に過ぎません。AIを通じて予測アナリティクスや不正検出アナリティクスなどの価値の高いインサイトを獲得するためには、これを支える複雑なインフラストラクチャが必要です。

適切に準備された正確なデータセットへ迅速にアクセスできなければ、データチームが正確なモデルを構築することは困難です。また、データが不正確な場合や不完全な場合、結果も不正確なものになり、AI/MLプロジェクトの全体的な信頼性が損なわれます。必要なのは、エンタープライズ規模で大量のハイブリッドデータを高速に取り込み、これらを統合して高速かつ信頼できる方法で処理するとともにデータディスカバリ機能とリネージ機能を提供できる、エンドツーエンドのソリューションです。Informatica[®]とDatabricks社は、データの迅速な探索、取り込み、準備を通じて機械学習と高度なアナリティクスの開発を高速化するための、拡張可能なデータ/機械学習ソリューションを共同で提供しています。

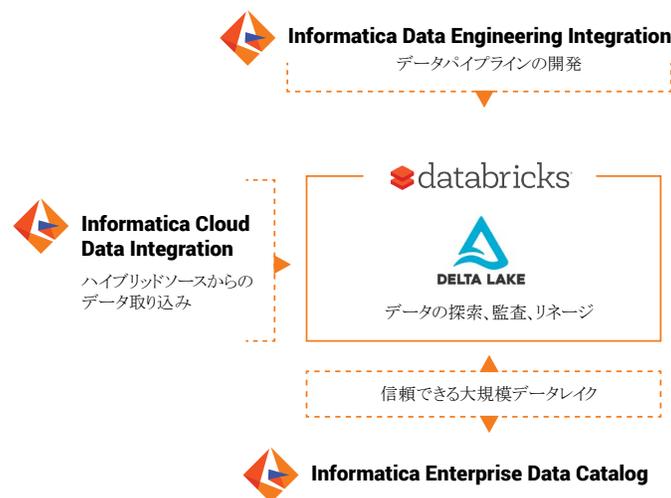


図1.インフォマティカとDatabricks社の共同ソリューションにより、AIとアナリティクスのためのデータエンジニアリングパイプラインを高速化

¹ Databricks、『2018 Trend Report: Enterprise AI Adoption』

主な機能

大量のデータを扱う大規模データパイプラインを構築

コーディング不要の視覚的な開発環境により、データパイプラインの開発を高速化。手作業でのコーディングに比べて、開発者の生産性が最大5倍向上します。Informatica Data Engineering Integrationで作成したドラッグアンドドロップ式のパイプラインをDatabricksにプッシュダウンして、最適化されているApache Spark環境で処理できます。あらかじめ組み込まれている数百のトランスフォーマー、コネクタ、パーサーに加え、動的なマッピングとテンプレートも活用できます。Data Engineering Integrationを導入すれば、機械学習向けにデータを準備して生産性を向上できるだけでなく、あらゆるコンピューティングエンジンにプッシュ可能な環境を通じて、オープンソーステクノロジーの将来的な変化にも対応できます。

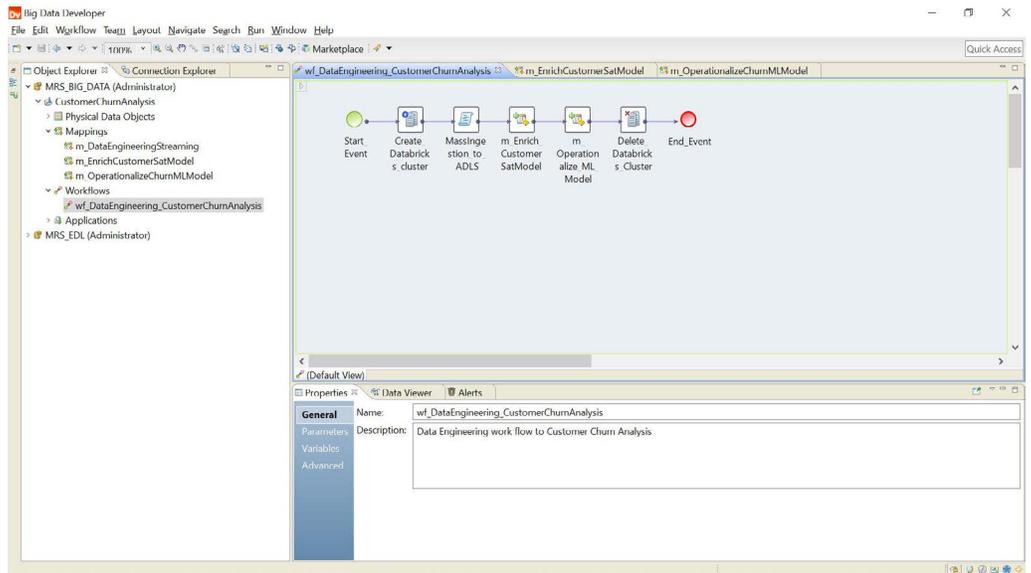


図 2. エンドツーエンドのデータパイプラインを運用化

管理されたDelta Lakeへのハイブリッドデータの高速取り込みを実現

Informatica Cloud Data IntegrationとDelta Lakeを組み合わせることで、信頼性とパフォーマンスに優れたDelta Lakeのスキーマ適用機能や圧縮機能、その他のファイル品質の向上機能を活用しながら、複数のハイブリッドソースにまたがる大量のデータをデータレイクに取り込むことができます。フローウィザードとすぐに使い始められるテンプレートを使用して迅速に統合を構築できるとともに、包括的なトランスフォーメーション機能により構造化データと非構造化データの両方をサポートできます。また、高度な例外処理機能を使用して統合プロセスのオーケストレーションを行えます。さらに、統合プロセスを自動化して、継続的統合/継続的配信システムとシームレスに統合することができます。

データサイエンティストによる社内データ資産の探索、分類、整理を実現

データスタックにおけるデータの移動を完全に可視化できるとともに、データリネージ機能によって、データがどこから来ているのか、どのように使われているのかを把握できます。Informatica Enterprise Data Catalog では、ハイブリッド/マルチクラウド環境全体にわたり信頼できるデータを容易に検索および探索することができます。また、総合的なデータ関係を探索するだけでなく、ドリルダウンしてエンドツーエンドのリネージと影響分析を把握できます。さらに、内蔵のビジネスグロッサリ、クラウドソーシングでキュレーションした業務資産、セマンティック検索などを通じてデータへの理解を深め、迅速に活用することができます。

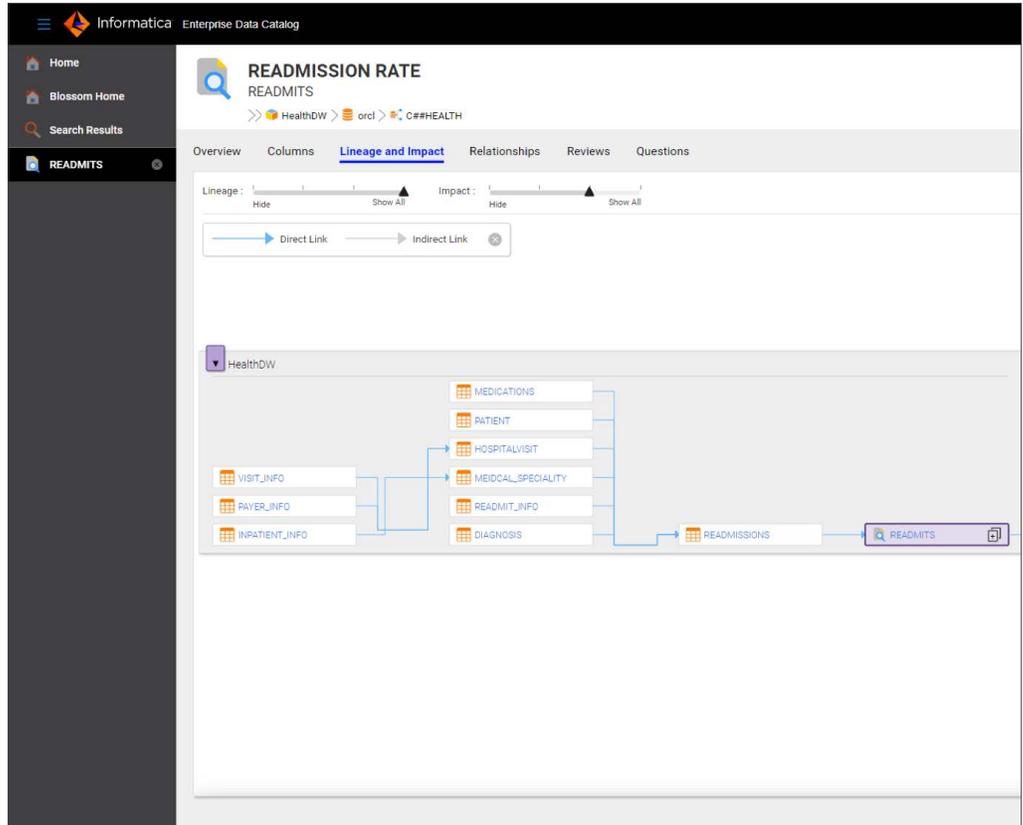


図 3. 信頼できるデータを容易に検索および探索し、機械学習モデルを構築

データレイクを最適化して、データを機械学習向けに準備

Databricks 社の管理された Delta Lake により、データレイクを最適化して、大量のデータを機械学習用に準備できます。既存のデータレイクの上位構造を提供する Delta Lake によってデータの正確性と適切な形式を確保できるだけでなく、直列化可能な分離レベルが実現するので、整合性のないデータは一切発生しません。また ACID トランザクションとスキーマ適用により、大規模環境におけるデータレイクの信頼性を確保し、質の高いデータセットを後工程のアナリティクスに提供できます。

エンドツーエンドの機械学習ライフサイクルを管理

機械学習プロセスのライフサイクルを管理する Databricks ワークスペースによって、開発から本番稼働まで、さまざまな実験を管理できます。またコラボレーションを促進する Databricks Notebook では、R や Python、Scala、SQL など複数の言語をサポートする再利用可能な AI / ML ツールを構築できます。

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ます。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管理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、企業の成功に必要な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。

利点

AI プロジェクトを成功に導き、イノベーションを促進し、ビジネスの俊敏性を強化

不正検知/予測、サプライチェーンの最適化、パーソナライズしたデジタル体験による顧客エンゲージメントの強化、患者転帰の改善など、さまざまなイニシアチブを通じてイノベーションの促進と短期間での市場展開を目指す企業が増えている中、AI への投資は右肩上がりです。インフォマティカと Databricks 社の共同ソリューションにより、企業は機械学習モデルの構築とイテレーションを高速化して、市場展開までの期間を短縮できます。

信頼できるハイブリッドデータを大量に取り込み、インサイトをより迅速に獲得

インフォマティカと Databricks 社のシームレスな統合により、データエンジニアは高い信頼性とパフォーマンスで複数のハイブリッドソースからの大量のデータをデータレイクへ迅速に取り込むことができます。

大規模なデータエンジニアリングパイプラインの構築と維持に要するコストを削減

ドラッグアンドドロップ式の使いやすいユーザーインターフェイスを通じてクラウド内の最適化された Apache Spark 環境へ処理をプッシュダウンすることで、大規模なデータパイプラインを迅速かつ低コストで構築できます。

データトレーニング用の適切なデータセットの探索機能をデータサイエンティストに提供

包括的なデータディスカバリ機能により、データサイエンティストは適切なデータセットに基づいて正確なモデルを迅速に構築できます。また、エンドツーエンドのリネージによって EU 一般データ保護規則 (GDPR) など各種規制へのコンプライアンスを確保できるとともに、データサイエンティストはモデル構築とアナリティクスに使用するデータのリネージを検証できます。

生産性を高め、既存のリソースでより多くの成果を達成

インフォマティカと Databricks 社の共同ソリューションによって、生産性を高めるとともに既存のインフラストラクチャを最大限活用することが可能になります。コーディング不要の視覚的な設計インターフェイスでは、手作業でのコーディングに比べて開発者の生産性が最大 5 倍向上します。また、AI / ML プロジェクトのインフラストラクチャリソースを最適化することで、全体的な TCO (総所有コスト) の削減を実現できます。

今後のステップ

インフォマティカと Databricks のパートナーシップについて、詳しくは informatica.com/databricks および databricks.com/informatica をご覧ください。



〒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_1019_03708