

セルフサービスアナリティクスのためのデータ探索と準備

メリット

- 使いやすいインターフェイスであらゆるデータを探索してセルフサービスアナリティクス用に準備
- エンドツーエンドのデータ準備パイプラインを通じて、信頼できる高品質なデータを提供
- データガバナンスとデータ保護を通じて、データ準備のコラボレーションを促進
- データパイプラインを大規模に運用化

Informatica Enterprise Data Preparationが、エンドツーエンドのデータ探索、変換、エンリッチ化、保護を実現

業務担当者には重要な業務データへのオンデマンドアクセスが必要

今日のデータ主導型の経済環境では、ビッグデータの分析から高品質なインサイト（知見）を迅速に引き出すことが、企業の最終的な成功を左右します。業務担当者は、顧客や製品に関する豊富なインサイトを獲得して、価格設定を最適化し、売上を増やし、コストを削減しなければなりません。データサイエンティストは、より正確な予測モデルを開発して、業務担当者の予測やトレンド分析を支援しなければなりません。そのためには、より多くのデータが必要です。

セルフサービスのデータ準備には時間がかかる

業務担当者やデータサイエンティストが業務上の最優先課題へ対応するためには、豊富な業務データにタイムリーにアクセスする必要があります。しかし、ユーザーがデータを使用する前段階で、データの探索と準備に80%もの時間が浪費されてしまうこともあります。必要なのは、コンプライアンスを確保し、個人データや機密データを保護しながら迅速にデータにアクセスできる、使いやすいセルフサービスのアナリティクスソリューションです。セルフサービスのアナリティクスがなければ、業務担当者はデータの整理を手作業で行い、分析よりもデータの探索と取得に多くの時間を費やすこととなります。

セルフサービスアナリティクスにはガバナンスと法規制へのコンプライアンスが不可欠

データを民主化してセルフサービスアナリティクスを実現するためには、より多くのデータを、より多くのユーザーに提供しなければなりません。データの信頼性を確保し、規則に従った使用を徹底するためには、品質、アクセス、保護に関するポリシーとルールが必要です。データエンジニアやデータアナリストがガバナンスに則ってデータパイプラインを構築しなければ、品質の低下やプライバシー規制へのコンプライアンス違反、顧客からの信頼性の低下など、さまざまなリスクが生じることとなります。

Informatica Enterprise Data Preparationソリューションを導入することで、データエンジニアやデータアナリストはセルフサービス機能を通じてデータを迅速に準備、保護、提供できるようになり、データからより多くの価値を引き出すことが可能になります。また、このソリューションはAIと機械学習に基づく自動化、インテリジェントな検索、提案も行います。



図1: Enterprise Data Preparationパイプラインでは、データカタログを使用して検索を行い、データ準備機能で変換を実行し、データ品質機能によりビジネスルールを適用するとともに、データ保護機能を使用してマスキングを行い、拡張可能なSparkエンジンによって運用化を実現します。

主なメリット

あらゆるデータを発見、アクセス

業務アナリストは豊富なデータへ手軽にアクセスして、Google式のセマンティック検索と動的ファセットを用いてデータを検索し、データ資産をフィルタリングおよび集計できます。CLAIRE™エンジンに基づくメタデータ主導のAIが、アナリティクスの強化に役立つ新規データ資産を提案することで、データディスカバリとデータ変換の各プロセスを支援します。これにより、データの信頼性を大幅に高め、機械学習モデルの精度に悪影響を及ぼす重複データを削減できます。

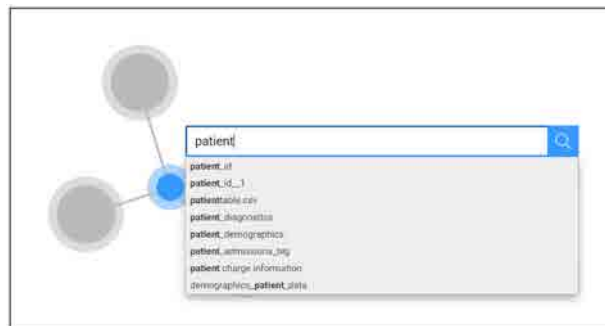


図2: Google式のセマンティック検索を使用したデータディスカバリ。

エンドツーエンドのデータ準備パイプラインを提供

インフォマティカのセルフサービス式のデータ準備環境には、カタログ、準備、データ品質、データ保護、運用化の機能があらかじめ統合されています。また、業務担当者はExcel式のインターフェイスでデータの結合とフィルタリングを実行できるだけでなく、アナリティクスインサイト/機械学習モデルへの統合も迅速に実行できます。データセットに関するクラウドソースの情報を通じてビジネスコンテキストと関連性を把握できるので、データの再利用性も高まります。

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。

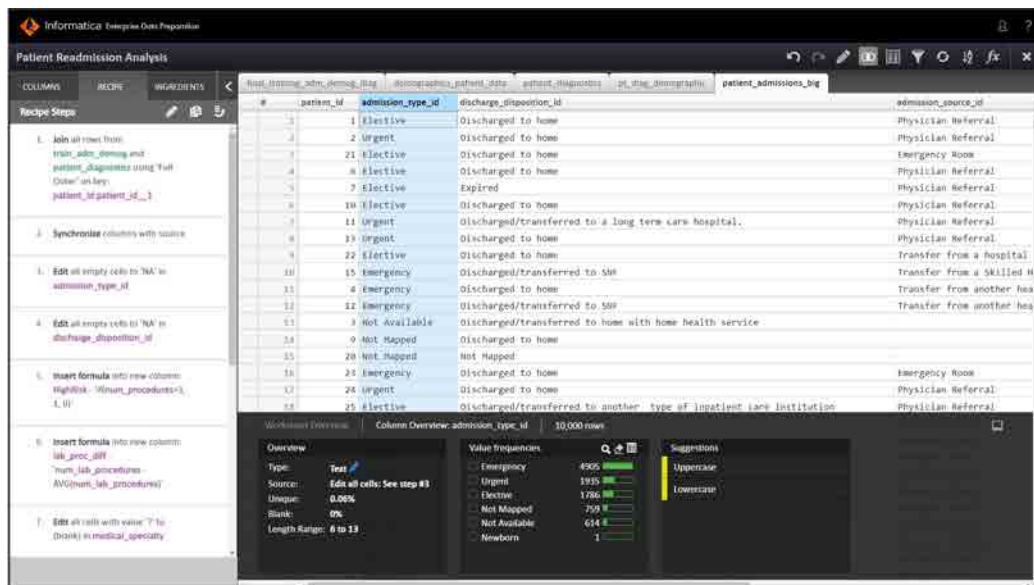


図3: Excel式のインターフェイスを使用したデータ準備。再利用と自動化のために変換のステップを記録。

データガバナンスを通じてコラボレーションを促進

データ準備パイプラインにはデータエンジニアやデータアナリスト、業務アナリスト、データサイエンティストなど多数の担当者が関与し、それぞれの継続的なコラボレーションが必要になります。データ準備のプロセスを記録するレシピ的なアプローチにより、注釈や共有可能なマッピングなどの拡張機能を通じて、ユーザー間のコラボレーションを促進できます。すべてのユーザーがプロジェクトワークスペースを使用してコラボレーションを行い、データセット、データリネージ、プロファイル統計、関係、変換を共有できます。また、ガバナンスポリシーとルールにより、許可されたユーザーにのみデータアクセス権を付与して、コンプライアンスに則った倫理的な使用を徹底できます。

データ準備を大規模に運用化

高度なSparkサポートに基づいてデータパイプライン処理を運用化することで、優れたパフォーマンスと拡張性を獲得し、大規模データパイプラインのライフサイクルを管理できます。また、データの取り込み、準備、配信のスケジュールを設定することで、パイプラインを自動化することも可能です。

人工知能／機械学習プロジェクトを促進

機械学習では、大量のデータへ継続的にアクセスして、モデルの精度を強化および改善し続ける必要があります。データサイエンティストは、Enterprise Data Preparationで作成したデータセットと外部の機械学習プラットフォームを活用して、機械学習モデルを開発、実行できます。外部プラットフォームには、例えばDatabricks NotebooksやDataRobot、AWS SageMaker、またPythonやRなどのプログラミング言語も含まれます。

今後のステップ

詳細は、[Informatica Enterprise Data Preparation](#)のページをご覧ください。



〒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_0619_03710