

# Informatica Enterprise Data Preparation

## 利点

- アナリティクスプロジェクトに適したデータをセルフサービスで探索
- 必要なデータを迅速に準備して共有
- データの準備を簡単に運用化して再利用性を確保

## 未加工のデータを探索して、高度なアナリティクス用に準備し、再利用性を確保

デジタルトランスフォーメーションの基盤となる要素、それは間違いなく「データ」です。企業は、Apache Hadoopなどの新しいデータ処理プラットフォームを活用することで、以前であれば想像はできなかったとしても手にすることはできなかったようなインサイトを引き出しています。Apache Hadoopとデータレイクのアプローチが登場したことで、すべてのデータを蓄積するという非常に便利な環境が実現し、ユーザーはいつでもあらゆるタイプの分析にデータを使用できるようになりました。

IT部門にとって、Hadoopシステムに取り込まれる膨大なデータの量は手に余るほど圧倒的です。業務アナリストがHadoopからの質の高いデータを待っている一方で、IT担当者には手作業で時間をかけて未加工のデータを目的に合ったデータ資産にキュレーションするという負担がのしかかっています。

データのキュレーションのための、拡張性に優れ繰り返し実行が可能でインテリジェントなメカニズムがなければ、データレイクが本来提供するはずのあらゆる機会が、停滞のリスクに直面することになります。このような「データの泥沼化」の危機を解消する鍵になるのが、Informatica®のCLAIRE™エンジンです。CLAIREエンジンは、データ資産の探索やプロファイリング、関係の推論などを自動化する、メタデータ主導の人工知能テクノロジーです。

Informatica Enterprise Data Preparationにより、未加工ビッグデータの体系的な探索が可能になり、業務アナリストはセルフサービスでデータセットを信頼できる情報へと変換できるようになります。データサイエンティストや業務アナリストは、セマンティック検索とファセット検索を使用して必要なデータを速やかに見つけられるだけでなく、データリネージやデータの関係性を自動的に把握することが可能になります。また、データアナリストのチームはプロジェクトのワークスペースで簡単にコラボレーションを実現し、結果を共有できます。データセットをプロジェクトワークスペースに追加すると、機械学習アルゴリズムがバックグラウンドで機能し、チームの役に立つ可能性がある代替のデータセットを提案します。

プロジェクトのワークスペース内にあるデータセットは、Excel式の使いやすいEnterprise Data Preparationのデータ準備ツールでいつでも開くことができます。ビッグデータを信頼できる情報資産に変換し、業務上の価値を持続的に獲得するには、メタデータ主導のデータ準備アプローチが最適です。

## 主な機能

### インテリジェントな検索とインテリジェントな視覚化

インテリジェントなセマンティック検索機能と推測に基づく結果を使用して、データレイクや他のエンタープライズシステム内のデータを見つけることができます。また、動的ファセットを使用してデータ資産をシステム属性でフィルタリングしたり、Apache Zeppelinベースのチャートで視覚化したりできます。

### 包括的なデータ探索

カスタム属性、データ品質のためのプロファイリング統計、ビジネスコンテンツ向けのデータドメイン、利用状況に関する情報など、データ資産の概要を把握できるとともに、タグ付けによってデータセットに関する情報をクラウドソーシングし、メタデータをエンリッチ化してビジネスコンテキストを追加できます。また、ユーザーの認証情報に基づいてサンプルデータをプレビューすることで、データを素早く把握でき、さらに、他のテーブルやビュー、ユーザー、レポート、データドメインとの関係に基づいて、企業組織全体を通じて他の資産とデータ資産の関連性を把握できます。

Excel式の直感的なインターフェイスで、データをフィルタリング、集計、統合、結合するためのあらかじめ組み込まれた変換機能を使用して、分析用のデータをインタラクティブに準備することができます。文字列、計算、日付、論理の各演算を使用して、列レベルでデータのクレンジングとデータ変換を行います。ガイド式のインテリジェンスが、データセットをブレンドする際の結合キーの提案など、データセットの準備を支援します。また、値の分布や数値と日付の分布など、シートレベルと列レベルの記述的な統計概要を確認でき、すべてのステップがデータフローを自動生成するためのレシピに記録されます。このデータフローは、分析のインサイトを運用に活かすために、繰り返し実行するようにスケジュール設定できます。

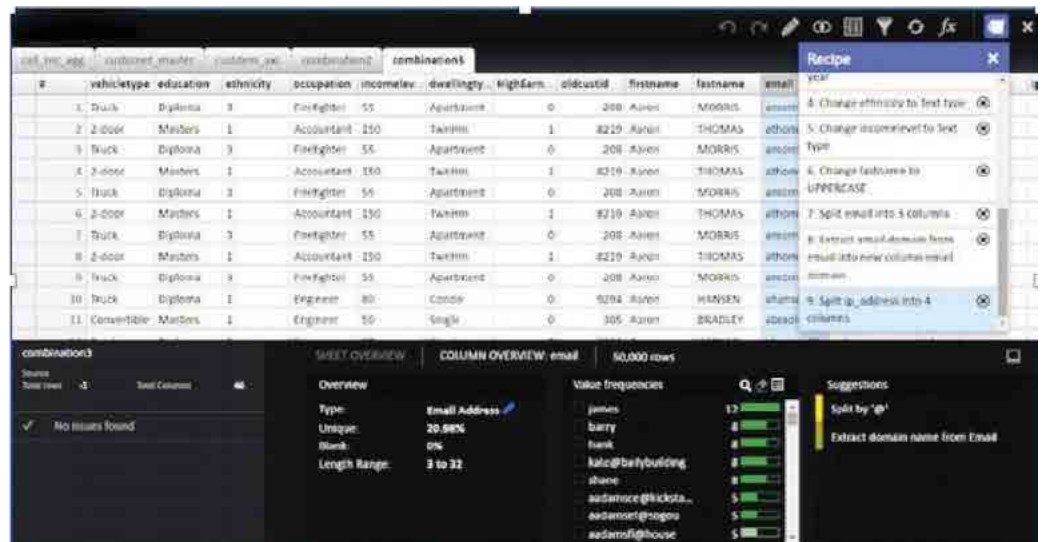


図1: インテリジェントなセマンティック検索機能と動的ファセットにより、データセットを手早く発見

### データ準備の運用化

業務アナリストは、あらかじめ組み込まれた数千ものビジネスルールを使用して、製品データの品質を体系的に強化できるだけでなく、データ準備手順をIT担当者に任せることで再利用性を確保できます。

### 社内コラボレーション

データアプリケーションをセルフサービスで管理しながら、プロジェクトのワークスペースにデータ資産を追加することで作業を整理できます。共同所有者やエディター、閲覧者など、さまざまな役割と権限のチームメンバーをプロジェクトに追加して、他のアナリストとのコラボレーションを実現できます。

### データ資産に関する提案

他のユーザーの行動と共有知識に適用される機械学習アルゴリズムを基盤にする自動提案機能によって、生産性が高まり、信頼性の高い資産の再利用が増加します。プロジェクトに追加されたデータ資産に基づいて、プロジェクトの代替追加資産が提案されます。

### ウィザードベースのデータアップロード

ウィザードベースのインターフェイスを使用して、個人の区切りファイルをデータレイクにアップロードできます。アップロードでは、Hiveテーブルが最適な形式で自動的に作成されます。アップロードしたデータには、資産を作成または上書きしたり、データを追加したりできます。

## 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、ビッグデータの特徴である「膨大な量、多様なタイプ、スピード」に合わせてデータを管理できる効率的な方法を切望しています。インテリジェントなセマンティック検索機能と動的ファセットによって結果をフィルタリングすることで、信頼できるデータ資産を簡単に発見できます。また、自動化した機械学習ベースの探索プロセスでは、関連するデータ資産が、アナリストにとって関心が高いと考えられる新しいデータ資産のインテリジェントな提案に変換されます。これによって信頼性が大幅に高まると共に、類似したプロジェクトに対して重複するデータセットが生成されることが少なくなります。

### ガバナンスに関するコラボレーション

業務アナリストは、プロジェクトワークスペースを活用してデータセットに関するコラボレーションを進めることで、ビッグデータアナリティクスプロジェクトの効率を高めることが可能になります。データセットがプロジェクトワークスペースに追加されると、アナリストはプロファイル統計、エンドツーエンドのデータセットリネージ、全関連データ資産、データドメイン、ユーザーなどを確認できます。これによって、データの品質を評価し、信頼できるデータを共有し、プロジェクトに役立つ他のデータセットを漸進的に発見できるようになります。業務アナリストも、セルフサービス環境でデータアプリケーションを管理できます。また、役割ベースのセキュリティによって、プロジェクトに追加したアナリストだけにデータへのアクセスを認めることができます。

## 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。

## 必要なデータを迅速に準備して共有

ビジネスのサイクルが短くなり続けている中、業務上の価値を獲得する競争においてデータアナリストが頼りにできる数少ない競争優位性の1つが、「スピード」です。インフォマティカによって、質の高いアナリティクスの提供に必要なデータを短時間で準備し、共有することができます。インフォマティカのセルフサービス式データ準備機能は、馴染みのある使いやすいExcelに似たインターフェイスを備えており、業務アナリストはデータを迅速に結合、フィルタリング、統合し、必要なインサイトを手に入れることができます。クラウドソーシングしたデータ資産のタグ付けおよび共有機能により、業務アナリストはデータのキュレーションプロセスを制御し、業務の生産性を向上させることができます。

## データ準備作業を再利用可能なワークフローに変換

自動化ツールやセルフサービスツールを使用できた場合でも、アナリストが新しいデータセットでデータ準備の作業を繰り返さなければならないということも珍しくありません。結果として、スケールアップや再利用性などから得られるメリットは失われてしまいます。Informatica Enterprise Data Preparationはデータ準備のステップを記録し、このステップを繰り返し実行が可能な自動化されたデータパイプラインとしてスケジューリングすることができます。これによって、データ準備は手作業のプロセスから再利用可能で持続的なシステムに変わります。

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

IN06\_0319\_03237