

Informatica Enterprise Data Catalog

利点

- AI搭載のカタログ機能で社内のあらゆるタイプのデータを自動的にカタログ化および分類
- 「カタログのカタログ」により、全社規模でメタデータのSoR（記録のシステム）を構築
- 複雑なエンタープライズシステムをはじめ幅広いデータソースから詳細なメタデータを自動抽出
- Googleのようなパワフルなセマンティック検索でデータ資産を発見
- リネージ、関係ビュー、データプロファイリング統計、品質スコアカードなどの総体的なビューでデータ資産を探索および把握
- インテリジェントなキュレーションでドメインやエンティティを特定
- 管理されたクラウドソースの注釈、評価、レビューを使用してデータ資産をエンリッチ化
- ビジネス用語と技術データ資産の関連付けを自動で実行
- オープンAPIを自社環境に統合して、インテリジェントなメタデータをあらゆる場所で活用

インテリジェントなデータカタログでデータの力を最大限に引き出す

データは経済の生命線であり、データ主導型の企業は自社のデータ資産を売上や利益に変えています。データ主導のデジタルトランスフォーメーションの第一歩は、データを企業の資産として管理することです。つまり、他の重要な設備投資や運用投資と同様に、データの在庫を調べ、その価値を評価し、最大限に活用する必要があります。

データは多様化しているだけでなく、さまざまな部門やアプリケーション、データウェアハウス/データレイクに分散しているため（オンプレミス環境やクラウド環境）、どこにどのようなデータがあるのかを正確に把握することが難しくなっています。さらに、データソースの多様化に伴いデータ環境はますます複雑になっています。

Informatica® Enterprise Data Catalogは、AI搭載のデータカタログです。機械学習ベースの探索エンジンを通じて、クラウド、オンプレミス、ビッグデータなどソースに関係なくデータ資産を全社規模でスキャンおよびカタログ化することが可能です。Enterprise Data CatalogのCLAIRE®エンジンが、メタデータを活用して、データ管理タスクに関する推奨、提案、自動化をインテリジェントに実行します。これにより、IT担当者の生産性が高まるとともに、業務担当者はデータの管理と使用について全面的に協力できるようになります。

データアナリストとIT担当者は、パワフルなセマンティック検索、検索結果を絞り込む動的ファセット、データリネージ、プロファイリング統計、データ品質スコアカード、総体的な関係ビュー、データの類似性の提案、統合されたビジネス用語集を利用できます。

また、各分野の専門知識やソーシャルキュレーション、AIを組み合わせるコラボレーション機能により、ユーザー体験を改善し、データのキュレーションを自動化できます。さらに、企業データ資産の管理を簡素化および効率化して、全社規模でデータの価値を最大限に引き出せます。ユーザーは、データを素早く見つけ出し、ビジネス用語や定義、参考データなどのライフサイクルを簡単に管理できます。

Informatica Enterprise Data Catalogは、AI搭載のデータカタログです。機械学習ベースの探索エンジンを通じて、クラウド、オンプレミス、ビッグデータなどソースに関係なくデータ資産を全社規模でスキャンおよびカタログ化することが可能です。

主な機能

メタデータAPIを自社環境に統合

Informatica Enterprise Data Catalogを使用してRESTベースのAPIを自社環境に統合することで、カタログのコンテンツをどこからでも利用できるようになります。インテリジェントなメタデータ（アプリケーション、BIレポート、ダッシュボード）を業務担当者と共有できます。また、カタログのコンテンツとこれに関連付けられたエンリッチ化メタデータをエクスポートおよび共有できます。

Enterprise Data Catalog Tableau拡張機能でセルフサービスアナリティクスを管理

Informatica Enterprise Data CatalogのTableau拡張機能により、管理されたデータに基づくアジャイルなアナリティクスをセルフサービスで実行できます。TableauのネイティブユーザーインターフェイスからInformatica Enterprise Data Catalogのすべてのリソースにアクセスできます。Tableauレポートを使用する業務担当者は、Tableau内のインテリジェントな検索バーを使用して、データスチュワードの認定を受けたデータ資産を検索できます。また、ビジネスコンテキスト、技術コンテキスト、データ品質、プロファイリング統計、同僚からの推奨にもアクセスできます。さらに、レビュー、評価、Q&Aを通じてデータ資産に関するコラボレーションをTableauインターフェイス内で行うことも可能です。

インテリジェントな機能を実装したセマンティック検索

パワフルなセマンティック検索とインテリジェントな機能を使用して、最も関連性の高いデータセットを検出、発見し、分析することができます。トークンマッチングを活用した高度なキーワード検索では、カタログ内の最も関連性の高いデータ資産を見つけられます。セマンティック検索は参照先のデータドメインにも適用されるため、データ資産が発見されずに見過ごされることはありません。検索結果に基づいてインテリジェント機能を活用することで、関心のあるデータセットに検索を絞り込むことができます。

セルフサービスによるデータプロビジョニング

分析用のデータセットを見つけたら、Informatica Enterprise Data Catalog内でクリックするだけで、そのデータセットを任意のターゲットへ簡単に移動できます。Amazon Redshift、Azure Synapse Analytics（旧Azure SQL Data Warehouse）、Google BigQuery、Snowflake、TableauのようなBIツールなど、幅広いソース／ターゲットを選択できます。この機能では、Informatica Enterprise Data CatalogとInformatica Cloud Data Integrationの統合を活用します。

データリネージと影響分析

リネージビューを通じて、データソースをあらゆるレベルでインタラクティブに追跡できます（エンドポイントだけを表示する業務担当者向けのシステムレベルビューから、その他の複雑な詳細情報をすべて含んだ詳細ビューまで）。リネージビューはドリルダウンが可能で、すべてのリネージパスが展開可能なため、列レベルおよびメトリックレベルの詳細なリネージを確認できます。また、上流／下流のデータ資産に関する詳細な影響分析を実行できます。

総体的な関係探索

ナレッジグラフにより、企業データや意義のあるデータの関係性を迅速に発見して総体的に把握できます。関連性の高いデータセット、技術、業務、セマンティクス、使用状況を基準にした関係の自動探索も実行できます。総体的なデータビューには、関連するデータセット、テーブル、ビュー、データドメイン、レポート、ユーザーが表示されます。このビューは、関心のある他のデータセットのプロGRESSIVE探索にも役立ちます。

ドメインとエンティティをインテリジェントに認識して分類を自動化

すべての構造化／非構造化データ資産に関して、ドメインやエンティティ（顧客、製品、注文など）を項目、列、テーブルレベルで自動的に分類および特定します。これは、企業がデータ資産をカタログ化して管理し、価値を引き出す態勢を構築するための重要なステップです。データを分類することで、検索、検索結果の絞り込み、ビジネス用語集の提案を強化できます。インフォマティカは、電子メールやクレジットカード番号、社会保障番号、国名や都市名、URL、会社名など、あらかじめ組み込まれた60以上のデータドメインを提供します。ユーザーが作成したカスタムドメインを追加することもできます。データ資産は、データルール（ルールで定義された特定のロジックに一致するデータがある列を検出）、または列名ルール（ルールで定義された列名のロジックと一致する列を検出）を使用して分類できます。

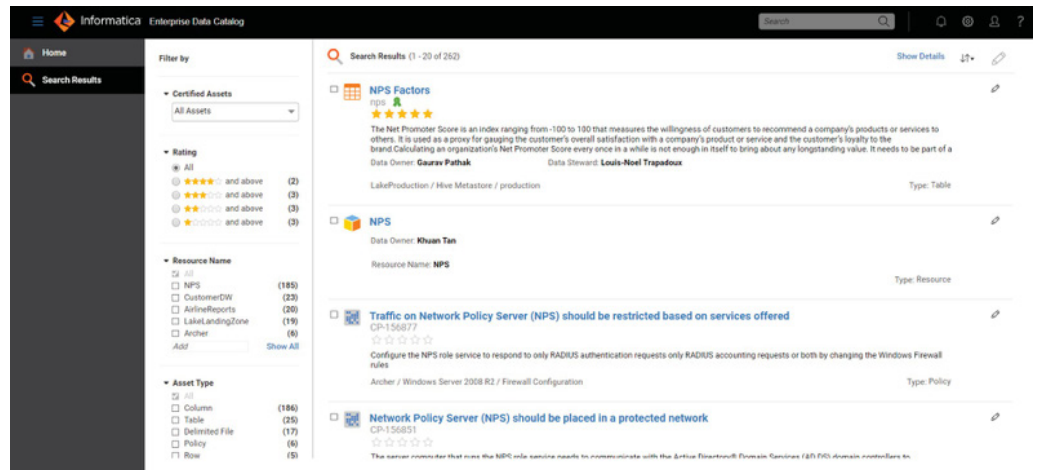


図1: スマートセマンティック検索機能と動的ファセットにより、データセットを手早く発見。認定を受けたデータセットと評価を表示

コラボレーションとソーシャルキュレーション

データアナリストとデータサイエンティストは、AI、人間の専門知識、コラボレーションの強力な組み合わせを活用して、関連性と信頼性の最も高いデータを見つけて、アナリティクスに使用できます。データ責任者と各分野の専門家は、データセットを認定できます。データの利用者がデータセットの評価やレビューを提供することで、データのソーシャルキュレーションが実現します。

ユーザーは、関心のあるデータセットをフォローし、そのデータセットに変更があった場合に通知を受けることができます。また、各分野の専門家がユーザーからの質問に回答するためのQ&Aプラットフォームもあります。さらに、データセットにカスタム属性や注釈を追加することで、業務部門とIT部門のコラボレーションや検索結果を強化できます。

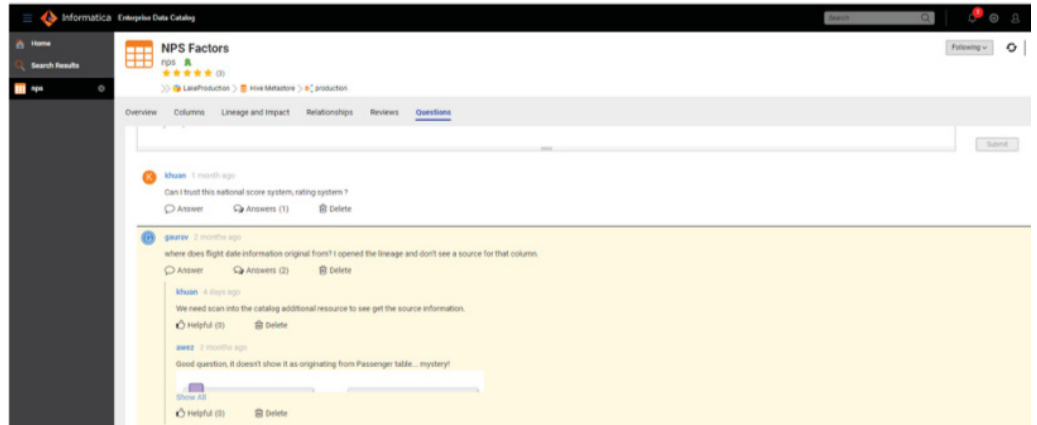


図2: コラボレーションを促進するQ&A機能

統合データ品質

技術メタデータに加えて、データプロファイリング統計、データ品質ルール、スコアカード、評価指標グループを確認し、データ資産の品質を把握してから、データを分析に使用できます。プロファイリング統計には、値の分布やパターン、データタイプやデータドメインの推測などがあります。

ビジネス用語の関連付けを自動で実行

Informatica Enterprise Data Catalogでは、用語、ポリシー、分類などのビジネスグロッサリ資産を Informatica Axon™ から簡単にインポートできます。ビジネス用語と適切な技術メタデータを関連付けることで、データに質の高いビジネスコンテキストを付加できます。Informatica Enterprise Data Catalogは、ビジネス用語と技術メタデータの関連付けを自動で実行します。これにより、データガバナンスの煩雑な手作業が不要になります。この機能により、業務部門とIT部門のデータ管理者が連携しながら、人的ワークフローの効率的な自動化を含むビジネスメタデータを管理できます。また、Informatica Business Glossaryやサードパーティツールからビジネス用語集資産をインポートすることも可能です。

インテリジェントな類似データの提案機能

高度な統計アルゴリズムと機械学習アルゴリズムが、類似するデータとデータサブセットを特定します。このパワフルな機能によって、最も関連性の高い、信頼できるデータを見つけ出すことができます。例えば、顧客の他社への乗り換えに関する分析に関心がある通信会社のアナリストが、当四半期のプリペイドカスタマーの活動を含むデータのクエリを実行するとします。この場合、Informatica Enterprise Data Catalogは、よりクリーンなデータ（データの置き換え）、前四半期の顧客の活動を含むデータ（結合可能なデータ）、データセットをエンリッチ化するための顧客情報テーブルを提案します。

ユニバーサルなメタデータ接続性と高度なスキャナー

データベース、データウェアハウス、クラウドベースのデータレイク、BIツール、Hadoopクラスター、NoSQL、複雑なエンタープライズシステム（レガシーシステムやメインフレームシステムなど）、マルチベンダーのETLツール、SQLダイレクト、各種エンタープライズアプリケーションなど、ハイブリッド環境およびマルチクラウド環境にある、社内のあらゆるタイプのデータソースからメタデータを抽出できます。

Enterprise Data Catalog Advanced Scannersでは、すべてのスクリプト、プロシージャ、プロセスを確認して、そのロジックと内部データフローを完全に理解できます。また、すべての潜在的なリネージソースと豊富な詳細情報を含む、列レベルの完全なデータリネージも取得できます。また、静的コードと動的コードをスキャンすることや、言語解析を実行して自動データリネージを取得することも可能です。

メタデータの抽出をサポートしているデータソースの例は以下のとおりです。

- データベース/データウェアハウス : Oracle、Microsoft SQL Server、SQLスクリプト、Sybase ASE、IBM Netezza、Teradata、JDBC、SAP HANA、SAP BW、SAP BW/4HANA、ストアードプロシージャ
- ビッグデータ : Cloudera Navigator、Hive (Cloudera/Hortonworks/MapR/IBM BigInsights/EMR)、HDFS、Hortonworks Atlas、Cassandra、MongoDB、Kafka
- メインフレーム : DB2 z/OS、DB2 I5/OS、COBOL、JCL
- BIおよびアナリティクス : SAP BusinessObjects、Tableau、Microsoft Power BI、Cognos、MicroStrategy、OBIEE、QlikView、Qlik Sense、Microsoft SSRSおよびSSAS、SAS
- ETL : Informatica PowerCenter、Informatica Data Engineering Integration、Informatica Intelligent Cloud Services、Informatica Data Integration Hub、Microsoft SSIS、Oracle Data Integrator、IBM InfoSphere DataStage、AWS Glue
- ビジネス用語集 : Informatica Axon Data Governance、Informatica Business Glossary
- データモデリング : Erwin Data Modeler、SAP PowerDesigner
- エンタープライズアプリケーション : Salesforce、Oracle、Workday、Informatica MDM、SAP ECC
- ファイルシステム : Microsoft SharePoint、Microsoft OneDrive、Windows/Linuxファイルシステム
- ファイル形式 : MS Excel、MS Word、MS PowerPoint、Adobe PDF、フラットファイル、CSV、区切りファイル、XML、JSON、Avro、Parquet
- クラウドプラットフォーム : AWS S3、AWS Redshift、Azure SQL DB、Azure SQL DW、Azure ADLS、Azure ADLS Gen 2、Azure Blob、Google Cloud Storage、Google BigQuery、Snowflake

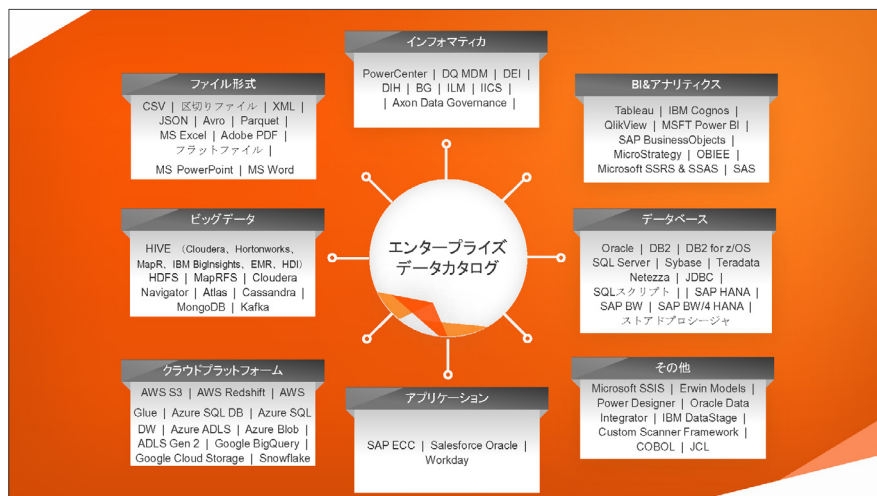


図3: Informatica Enterprise Data Catalogはユニバーサルなメタデータ接続性をサポート

リソースレベルのセキュリティ

ユーザーやグループにリソースレベルで読み取り／書き込み権限を付与できます。ユーザーは、カスタム属性の閲覧／編集、ドメインキューレーションの実行、ビジネス用語の関連付けが可能です。

エンタープライズ規模の実装

ビッグデータ規模の実装を念頭に構築されているInformatica Enterprise Data Catalogでは、数百のデータソースにわたり数千万件のデータセットをスキャンできます。また並列処理によるメタデータの取り込みと高速の分散インデックス処理をサポートしているため、カタログコンテンツを迅速に更新し、卓越した検索パフォーマンスとフォールト トレランス（耐障害性）を備えた高可用性を実現することで、24時間365日の稼働環境に対応します。さらに、Sparkベースのデータプロファイリングにより、大量のデータをプロファイリングしてデータへの理解を深めることができます。

統合システム管理

カタログリソース、メタデータ抽出スケジュール、プロファイリングの実行などのタスクを、統合されたシステム管理コンソールで管理／監視することができます。ジョブコントロールダッシュボードでは、タスク監視ビューやリソースビューを表示するウィジェットを利用できます。また、システム管理者は電子メールアラートを設定することで、カタログの問題にプロアクティブに対応することが可能になります。



図4：データの関係を示す総体的なビューでデータを把握

利点

社内のあらゆるタイプのデータをインテリジェントにカタログ化

Informatica Enterprise Data Catalogは、企業内のさまざまなタイプのデータとその関係をインテリジェントに探索します。あらかじめ組み込まれたスキャナーが、データベース、データウェアハウス、データレイク、クラウドデータストア、アプリケーション、BIツール、ETLツール、サードパーティのメタデータカタログ、NoSQLなどからメタデータを収集します。すべてのメタデータは、迅速な更新やスマート検索、高速クエリを目的として設計された拡張可能なグラフデータベースでインデックス化さ

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。

れ、カタログ化されます。生成されるデータ量が増え、組織全体に広がっていくのに伴い、類似データや重複データが増えてしまうことは避けられません。Informatica Enterprise Data Catalogは、高度な統計アルゴリズムと機械学習アルゴリズムを活用して類似データとデータのサブセットを探索することで、ユーザーが必要としている最も関連性の高い信頼できるデータを見つけ出します。

Googleのようなパワフルなセマンティック検索でデータ資産を迅速に発見

何百もの企業システムの中から必要なデータを探すという作業は、時には実りのない無駄な労力に思えることがあるかも知れません。包括的なメタデータ主導のインテリジェンスと拡張可能なインフラストラクチャを基盤にしたパワフルなセマンティック検索機能がなければ、関連性のあるデータを見つけることは決してできません。Informatica Enterprise Data Catalogでは、セマンティック検索とインテリジェントな機能によって、検索結果を詳細に絞り込むことができます。ビジネスメタデータ、技術メタデータ、運用メタデータが独自の方法で関連付けられているため、業務担当者はデータを探す際にビジネス用語を検索し、総合的な関係ビューを確認して、関連性のあるデータ資産を見つけることができます。

総合的な関係ビューとリネージでデータ資産を探索および把握

「測定できないものは管理できない」という格言がありますが、データ資産の管理についても同じことが言えます。データから価値を最大限に引き出すには、保有しているデータ、データの出所、データの変更履歴、その信頼性を把握する必要があります。こうしたニーズに応えるInformatica Enterprise Data Catalogでは、完全なエンドツーエンドのサマリーと詳細なリネージ、プロファイリング統計、データ品質スコアカード、総合的な関係ビューを通じてさらに多くの情報が提供され、ユーザーはデータを明確に把握できます。

管理されたクラウドソースの注釈とビジネスコンテキストを使用してデータ資産をエンリッチ化

Informatica Enterprise Data Catalogは、企業データ資産を項目／列レベルまで自動分類することで、データの再利用性と価値を最大化します。データの価値をさらに高めるため、Informatica Enterprise Data Catalogはクラウドソース対応のタグ、注釈、評価、レビューに加え、データの利用者や使用目的といったコンテキストも取得します。このような「集合知」を活用することでデータのエンリッチ化と取得を行い、全社規模でデータの有益性をさらに高めることが可能になります。Informatica Enterprise Data CatalogはInformatica Axonとも統合可能で、ビジネス用語、定義、ポリシーなどのビジネス用語集資産をAxonから簡単にインポートできます。これらのビジネスメタデータは技術メタデータや運用メタデータと自動的に関連付けられるので、業務アナリストやデータスチュワード（データ管理／案内人）などのユーザーは、データ資産を迅速に発見、把握し、協力してデータを活用することができます。

詳細はこちら

Informatica Enterprise Data Catalogの詳細については、次のWebサイトをご覧ください。

<https://www.informatica.com/jp/products/data-catalog/enterprise-data-catalog.html>



〒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_0620_03238

© Copyright Informatica LLC 2020. Informatica、Informatica ロゴ、CLAIRE、Axon、PowerCenter、Big Data Management、Informatica Cloud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。