



Informatica®
10.5.7

プロファイルガイド

Informatica プロファイルガイド

10.5.7

2024 年 11 月

© 著作権 Informatica LLC 2010, 2025

本ソフトウェアおよびマニュアルは、使用および開示の制限を定めた個別の使用許諾契約のもとでのみ提供されています。本マニュアルのいかなる部分も、いかなる手段（電子的複製、写真複製、録音など）によっても、Informatica LLC の事前の承諾なしに複製または転載することは禁じられています。

Informatica および Informatica ロゴは、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。Informatica の商標の最新リストは、Web (<https://www.informatica.com/trademarks.html>) にあります。その他の企業名および製品名は、それぞれの企業の商標または登録商標です。

オプトアウトの権利の制限の下、本ソフトウェアによって、本ソフトウェアがデプロイされているコンピューティングおよびネットワーク環境に関する情報、デプロイメントのデータ使用状況およびシステム統計情報が米国の Informatica に自動的に送信されます。この送信は Informatica のプライバシーポリシーの下にサービスの一部と見なされ、Informatica は、<https://www.informatica.com/in/privacy-policy.html> のプライバシーポリシーに従って、この情報を使用または処理します。使用状況の収集は、Administrator ツールで無効にすることができます。

米政府の権利プログラム、ソフトウェア、データベース、および関連文書や技術データは、米国政府の顧客に配信され、「商用コンピュータソフトウェア」または「商業技術データ」は、該当する連邦政府の取得規制と代理店固有の補足規定に基づきます。このように、使用、複製、開示、変更、および適応は、適用される政府の契約に規定されている制限およびライセンス条項に従うものとし、政府契約の条項によって適当な範囲において、FAR 52.227-19、商用コンピュータソフトウェアライセンスの追加権利を規定します。

本ソフトウェアまたはドキュメンテーション（あるいはその両方）の一部は、第三者が保有する著作権の対象となります。必要な第三者の通知は、製品に含まれています。

特許については、<https://www.informatica.com/legal/patents.html> を参照してください。

本マニュアルの情報は、予告なしに変更されることがあります。このドキュメントで問題が見つかった場合は、infa_documentation@informatica.com までご報告ください。

Informatica 製品は、それらが提供される契約の条件に従って保証されます。Informatica は、商品性、特定目的への適合性、非侵害性の保証等を含めて、明示的または黙示的ないかなる種類の保証をせず、本マニュアルの情報を「現状のまま」提供するものとします。

発行日: 2025-02-23

目次

序文	8
Informatica のリソース	8
Informatica Network	8
Informatica ナレッジベース	8
Informatica マニュアル	8
Informatica 製品可用性マトリックス	9
Informatica Velocity	9
Informatica Marketplace	9
Informatica グローバルカスタマサポート	9
 第 I 部 : プロファイルについて	 10
第 1 章 : プロファイルについて	11
プロファイルについての概要	11
プロファイリングのプロセス	12
プロファイリングツール	12
プロファイルコンポーネント	12
 第 2 章 : カラムプロファイルの概念	 14
カラムプロファイルの概念の概要	14
カラムプロファイルオプション	15
リポジトリプロファイルのロックおよびバージョン管理されたプロファイルの管理	15
スコアカード	16
 第 3 章 : キュレーションの概念	 17
キュレーションの概念の概要	17
キュレーションタスク	17
 第 II 部 : Informatica Analyst によるプロファイリング	 19
第 4 章 : Informatica Analyst のカラムプロファイル	20
Informatica Analyst のカラムプロファイルの概要	20
カラムプロファイリングプロセス	21
プロファイルオプション	21
サンプリングのオプション	22
ドリルダウンのオプション	23
ランタイム環境	23
ネイティブ環境	23
Hadoop 環境	23
Databricks 環境	24

Informatica Analyst のオペレーティングシステムプロファイルの概要.	24
オペレーティングシステムプロファイルの選択.	25
リポジトリアセットのロックとチームベース開発の概要.	25
Informatica Analyst でのカラムプロファイルの作成.	26
カラムプロファイルの編集.	27
プロファイルの実行.	28
Spark エンジンでのプロファイルの実行.	28
Databricks クラスタでのプロファイルの実行.	28
同期オプション.	29
Informatica Analyst でのフラットファイルデータオブジェクトの同期.	29
Informatica Analyst でのリレーショナルデータオブジェクトの同期.	31
第 5 章 : Informatica Analyst のルール.	32
Informatica Analyst のルールの概要.	32
定義済みルール.	32
定義済みルールのプロセス.	33
定義済みルールの適用.	33
式ルール.	34
式ルールの作成.	34
ルール仕様を使用した式ルールの作成.	35
第 6 章 : Informatica Analyst のフィルタ.	37
Informatica Analyst のフィルタの概要.	37
フィルタの作成.	37
簡易フィルタの作成.	38
詳細フィルタの作成.	39
SQL フィルタの作成.	40
フィルタの管理.	40
第 7 章 : Informatica Analyst のカラムプロファイル結果.	42
Informatica Analyst のカラムプロファイル結果の概要.	42
サマリビュー.	43
サマリビューのプロパティ.	44
サマリビューのデフォルトフィルタ.	44
詳細ビュー.	45
[詳細ビュー] ペイン.	46
統計.	47
データプレビュー.	48
データ型.	48
異常値.	49
パターン.	50
値.	51
プロファイル実行のタイプ.	54

最新のプロファイル実行.	54
履歴プロファイル実行.	54
統合済みプロファイル実行.	54
プロファイル実行の選択.	55
複数のプロファイル結果の比較の概要.	56
複数のプロファイル結果の比較.	56
プロファイル結果の比較のサマリビュー.	57
プロファイル結果の比較の詳細ビュー.	59
カラムプロファイルのドリルダウン.	60
行データのドリルダウン.	60
ドリルダウンデータへのフィルタの適用.	60
Informatica Analyst ツールでのキュレーション.	61
データ型とデータドメインの承認.	61
データ型とデータドメインの拒否.	61
Informatica Analyst のカラムプロファイルのエクスポートファイル.	62
CSV ファイルのプロファイルエクスポート結果.	62
Microsoft Excel のプロファイルエクスポート結果.	62
Informatica Analyst からのプロファイル結果のエクスポート.	63

第 8 章 : Informatica Analyst のスコアカード. 64

Informatica Analyst のスコアカードの概要.	64
Informatica Analyst のスコアカードのプロセス.	65
Informatica Analyst でのスコアカードの作成.	66
既存のスコアカードへカラムの追加.	67
スコアカードの実行.	68
スコアカードの表示.	69
スコアカードの編集.	69
メトリック.	70
メトリックウェイト.	70
データ品質の値.	70
しきい値の定義.	71
[メトリックグループ]	71
メトリックグループの作成.	71
メトリックグループへのスコアの移動.	72
メトリックグループの編集.	72
メトリックグループの削除.	72
カラムのドリルダウン.	73
傾向グラフ.	73
スコアの傾向グラフ.	74
コストの傾向グラフ.	74
傾向グラフの表示.	75
傾向グラフのエクスポート.	76
Informatica Analyst のスコアカードエクスポートファイル.	76

Microsoft Excel のスコアカードエクスポート結果.	77
Informatica Analyst からのスコアカード結果のエクスポート.	77
スコアカード通知.	78
通知電子メールメッセージテンプレート.	78
スコアカード通知の設定.	79
スコアカード通知のグローバル設定の実行.	80
スコアカードリネージュ.	80
Informatica Analyst でのスコアカードリネージュの表示.	81

第 III 部 : Informatica Developer によるプロファイリング. 82

第 9 章 : データオブジェクトプロファイル. 83

Informatica Developer のカラムプロファイル.	83
フィルタリングのオプション.	83
サンプリングのオプション.	84
Informatica Developer での単一のデータオブジェクトプロファイルの作成.	85
Informatica Developer での複数のデータオブジェクトプロファイルの作成.	85
Informatica Developer でのフラットファイルデータオブジェクトの同期.	86
Informatica Developer でのリレーショナルデータオブジェクトの同期.	88

第 10 章 : 半構造化データソースのカラムプロファイル. 89

半構造化データソースのカラムプロファイルの概要.	89
JSON および XML データオブジェクト.	90
JSON または XML データソースからのデータオブジェクトの作成.	90
HDFS 内の半構造化データソースの複合型ファイルデータオブジェクト.	91
HDFS 内の JSON または XML データソースからの複合型ファイルデータオブジェクト.	91
HDFS 内の Avro または Parquet データソースからの複合型ファイルデータオブジェクト.	91
HDFS 接続の作成.	92
HDFS 内の JSON または XML ファイルからの複合型ファイルデータオブジェクトの作成.	92
Avro または Parquet データソースからの複合型ファイルデータオブジェクトの作成.	93
半構造化データソースでのカラムプロファイルの作成.	94

第 11 章 : Informatica Developer のルール. 96

Informatica Developer のルールの概要.	96
Informatica Developer でのルールの作成.	97
Informatica Developer でのルールの適用.	97

第 12 章 : マプレットとマッピングのプロファイリング. 98

マプレットとマッピングのプロファイリングの概要.	98
マプレットまたはマッピングオブジェクトに対するプロファイルの実行.	98
マッピングまたはマプレットオブジェクトのプロファイルの比較.	99
プロファイルからのマッピングの生成.	99

第 13 章 : Informatica Developer のカラムプロファイル結果.....	101
Informatica Developer のカラムプロファイル結果.....	101
カラム値のプロパティ.....	102
カラムパターンのプロパティ.....	102
カラム統計のプロパティ.....	102
カラムデータ型のプロパティ.....	103
Informatica Developer でのキュレーション.....	104
データ型の承認.....	104
データ型の拒否.....	104
Informatica Developer からのプロファイル結果のエクスポート.....	105
第 14 章 : Informatica Developer のスコアカード.....	106
Informatica Developer のスコアカードの概要.....	106
スコアカードの作成.....	106
スコアカードリネージュのリソースファイルのエクスポート.....	107
Informatica Developer からのスコアカードリネージュの表示.....	107
索引.....	109

序文

プロファイルを使用してデータソースのコンテンツおよび構造を分析する方法を確認するには、『*Informatica プロファイルガイド*』を使用します。値の頻度、割合、パターンなど、ソースデータのカラムの特性を判別できます。このガイドは、データアナリストおよび開発者を対象としています。

Informatica のリソース

Informatica は、Informatica Network やその他のオンラインポータルを通じてさまざまな製品リソースを提供しています。リソースを使用して Informatica 製品とソリューションを最大限に活用し、その他の Informatica ユーザーや各分野の専門家から知見を得ることができます。

Informatica Network

Informatica Network は、Informatica ナレッジベースや Informatica グローバルカスタマサポートなど、多くのリソースへの入口です。Informatica Network を利用するには、<https://network.informatica.com> にアクセスしてください。

Informatica Network メンバーは、次のオプションを利用できます。

- ナレッジベースで製品リソースを検索できます。
- 製品の提供情報を表示できます。
- サポートケースを作成して確認できます。
- 最寄りの Informatica ユーザーグループネットワークを検索して、他のユーザーと共同作業を行えます。

Informatica ナレッジベース

Informatica ナレッジベースを使用して、ハウツー記事、ベストプラクティス、よくある質問に対する回答など、製品リソースを見つけることができます。

ナレッジベースを検索するには、<https://search.informatica.com> にアクセスしてください。ナレッジベースに関する質問、コメント、ご意見の連絡先は、Informatica ナレッジベースチーム (KB_Feedback@informatica.com) です。

Informatica マニュアル

Informatica マニュアルポータルでは、最新および最近の製品リリースに関するドキュメントの膨大なライブラリを参照できます。マニュアルポータルを利用するには、<https://docs.informatica.com> にアクセスしてください。

製品マニュアルに関する質問、コメント、ご意見については、Informatica マニュアルチーム (infa_documentation@informatica.com) までご連絡ください。

Informatica 製品可用性マトリックス

製品可用性マトリックス (PAM) には、製品リリースでサポートされるオペレーティングシステム、データベースなどのデータソースおよびターゲットが示されています。Informatica PAM は、<https://network.informatica.com/community/informatica-network/product-availability-matrices> で参照できます。

Informatica Velocity

Informatica Velocity は、Informatica プロフェッショナルサービスが開発したヒントとベストプラクティスのコレクションで、多数のデータ管理プロジェクトから得た実体験に基づいています。Informatica Velocity には、世界中の組織と連携してデータ管理ソリューションを計画、開発、デプロイ、管理する Informatica コンサルタントによる集合知を表しています。

Informatica Velocity リソースには、<http://velocity.informatica.com> からアクセスしてください。Informatica Velocity についての質問、コメント、またはアイデアがある場合は、ips@informatica.com から Informatica プロフェッショナルサービスにお問い合わせください。

Informatica Marketplace

Informatica Marketplace は、お使いの Informatica 製品を拡張したり強化したりするソリューションを検索できるフォーラムです。Marketplace で、Informatica デベロッパーやパートナーからの多数のソリューションを活用すれば、生産性を向上したり、プロジェクトでの実装時間を短縮したりできます。Informatica Marketplace は、<https://marketplace.informatica.com> からアクセスしてください。

Informatica グローバルカスタマサポート

電話または Informatica Network を介してグローバルカスタマサポートに連絡できます。

各地域の Informatica グローバルカスタマサポートの電話番号は、Informatica Web サイト (<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>) を参照してください。

Informatica Network のオンラインサポートリソースを見つけるには、<https://network.informatica.com> にアクセスして eSupport オプションを選択します。

パート I: プロファイルについて

この部には、以下の章があります。

- [プロファイルについて, 11](#) ページ
- [カラムプロファイルの概念, 14](#) ページ
- [キュレーションの概念, 17](#) ページ

第 1 章

プロファイルについて

この章では、以下の項目について説明します。

- [プロファイルについての概要, 11 ページ](#)
- [プロファイリングのプロセス, 12 ページ](#)
- [プロファイリングツール, 12 ページ](#)
- [プロファイルコンポーネント, 12 ページ](#)

プロファイルについての概要

プロファイルを作成して実行すると、アプリケーション、スキーマ、または企業のデータソースの内容、品質、および構造を検索できます。データソースの内容には、値の頻度やデータ型が含まれます。データソースの構造には、キーや機能依存性が含まれます。

検出プロセスの一環として、プロファイルを作成して実行できます。プロファイルは、エンタープライズのデータソース間のデータの不整合や、データオブジェクトを危険にさらす隠れたデータの問題を検出および分析するリポジトリオブジェクトです。企業内のデータソースに対してプロファイルを実行すると、エンタープライズのデータおよびメタデータの強みと弱みについて把握することができます。

Informatica Analyst と Informatica Developer を使用して、ソースデータやメタデータを分析できます。アナリストと開発者は、これらのツールを使用して、連携しながらデータ品質の問題を特定し、データのリレーションを分析することができます。職務に基づいて、Analyst ツールまたは Developer tool のいずれかの機能を使用できます。実行できるプロファイリングのレベルは、使用するツールによって異なります。

Developer tool と Analyst ツールで次のタスクを実行できます。

- カラムプロファイリングを実行する。これには、カラム内の一意の値と NULL 値の数やデータパターンの検出が含まれます。
- ルールをカラムプロファイルに追加する。
- プロファイル結果内の推測されるデータ型をキュレーションする。
- スコアカードを使用して、データ品質を監視する。
- プロファイルからマッピングを生成する。

プロファイリングのプロセス

データ統合プロジェクトでは、多くの場合、最初にプロファイリングを行います。プロファイルを作成することで、データソースの内容、品質、および構造を分析できるようになります。プロファイリングプロセスの一環として、データソースのメタデータを検出します。

使用するプロファイルは、データ分析のタイプ（カラムプロファイルなど）によって異なります。データ品質の問題を検出し、文書化します。プロファイリングを行うには、以下のタスクを実行します。

1. データソースからデータの内容を検出して分析します。これには、データ型、値の頻度、パターンの頻度、データの統計（最小値や最大値など）が含まれます。
2. プロファイリング結果を確認します。
3. 参照データを作成します。
4. プロファイル結果をドリルダウンします。
5. データの問題を文書化します。
6. ルールを作成して実行します。
7. スコアカードを作成してデータ品質を監視します。

プロファイリングツール

複数の Informatica ツールを使用してプロファイリングプロセスを管理します。

プロファイリングプロセスの管理には、次のツールを使用できます。

Informatica Administrator

ユーザー、グループ、特権、およびロールを管理します。アナリストサービスを管理したり、Informatica Analyst のプロジェクトやオブジェクトに対する権限を管理したりすることができます。Informatica Developer のアクセス権限もこのツールで制御できます。

Informatica Developer

このツールでプロファイルを作成および実行して、1 つ以上のデータソースのメタデータを検出して分析します。プロファイルはウィザードを使用して作成します。

Informatica Analyst

Analyst ツールでは、カラムプロファイルを実行できます。プロファイルの実行後、データソース内のデータ行をドリルダウンできます。

プロファイルコンポーネント

プロファイルには複数のコンポーネントがあり、それらを使用することでデータソースの内容と構造を効果的に分析することができます。

プロファイルのコンポーネントを次に示します。

フィルタ

元のデータソースから特定の条件を満たすサブセットを作成します。その後、このサンプルデータに対してプロファイルを実行できます。

ルール

プロファイルの実行時にデータに適用される条件を定義するビジネスロジック。データを検証するには、プロファイルにルールを追加します。

タグ

ビジネス上の用途に基づいてモデルリポジトリ内のオブジェクトを定義するメタデータ。ビジネス上の用途に従ってオブジェクトをグループ化するには、タグを作成します。

コメント

プロファイルに関する説明。コメントは、プロファイルに関する情報を Analyst ツールおよび Developer ツールの他のユーザーと共有する場合に使用します。

スコアカード

カラムの有効値またはプロファイル結果内のルールの出力を視覚的に表したもの。データ品質の進捗を測定するには、スコアカードを使用します。

第 2 章

カラムプロファイルの概念

この章では、以下の項目について説明します。

- [カラムプロファイルの概念の概要, 14 ページ](#)
- [カラムプロファイルオプション, 15 ページ](#)
- [リポジトリプロファイルのロックおよびバージョン管理されたプロファイルの管理, 15 ページ](#)
- [スコアカード, 16 ページ](#)

カラムプロファイルの概念の概要

カラムプロファイルによって、値の頻度、割合、パターンなど、データソースに含まれるカラムの特性が決まります。

カラムプロファイリングでは、データに関する以下の情報が検出されます。

- カラムごとの NULL 値、重複なしの値、重複ありの値の数（数値および割合で表現される）。
- カラムごとのデータのパターン、およびそれらの値が出現する頻度。
- カラム値に関する統計。値の最大長さと最小長さや、各カラムの最初の値と最後の値など。
- 文書化されたデータ型、推測されたデータ型、および文書化されたデータ型と推測されたデータ型の間で起こりうる競合。
- パターンと値頻度の異常値。

プロファイルの作成または編集時に、次のオプションを設定できます。

- カラムプロファイルオプション。プロファイルの実行、サンプリングオプションの選択、およびドリルダウンオプションの選択を行うカラムを選択できます。
- フィルタおよびルールの追加、編集、または削除。

プロファイル結果で、プロファイルおよびプロファイルのカラムにコメントおよびタグを追加できます。ビジネス用語をカラムに割り当てられます。

モデルリポジトリでは、ユーザーがリポジトリプロファイルロックで作業を上書きしないように、プロファイルがロックされます。バージョン管理システムでは、複数バージョンのプロファイルが保存され、各バージョンにバージョン番号が割り当てられます。プロファイルをチェックアウトして、変更後にチェックインできます。プロファイルを再度チェックインする前に、プロファイルのチェックアウトアクションを取り消すことができます。

データ品質を定期的に確認するには、スコアカードを作成します。カラムの有効値を視覚的に表示できるように、プロファイルへのルールの適用前と適用後にスコアカードを作成します。

スケジューラサービスを使用して、プロファイルの実行やスコアカードの実行が特定の時刻または間隔で実行されるように、スケジュールを設定します。スケジューラサービスでは、プロファイル、スコアカード、デプロイ済みのマッピング、およびデプロイ済みのワークフローのスケジュールを管理します。スケジュールの作成、管理、および実行は、Informatica Administrator で行えます。

カラムプロファイルオプション

プロファイルを作成するときに、プロファイルウィザードを使用してフィルタ、ルール、ドリルダウンオプション、サンプリングオプション、および接続を定義できます。これらのオプションにより、プロファイルがソースデータから行を読み取る方法が決定されます。

カラムプロファイル、データドメイン検出プロファイル、またはエンタープライズ検出プロファイルの次のオプションを定義できます。

- フィルタ。フィルタを作成してプロファイルに適用できます。
- ルール。プロファイルの作成時にルールを追加できます。Analyst ツールまたは Developer tool で作成したルールは、利用できます。
- ドリルダウンオプション。データソース内の現在のデータを読み取るか、プロファイリングウェアハウスにステージングされたプロファイルデータを読み取るかを選択できます。
- サンプリングオプション。サンプリングオプションのいずれかを選択して、プロファイルを実行する行の数を決定できます。
- 接続。プロファイルは、ネイティブまたは Hadoop ランタイム環境で実行できます。Hadoop ランタイム環境では、Blaze または Spark エンジンを選択できます。

リポジトリプロファイルのロックおよびバージョン管理されたプロファイルの管理

モデルリポジトリでは、ユーザーが作業を上書きしないように、プロファイルがロックされます。プロファイルの編集を開始すると、他のユーザーがプロファイルに対する変更を保存しないように、プロファイルがロックされます。プロファイルを保存すると、ロックは解除されます。バージョン管理されたプロファイルの管理ではプロファイルのバージョンが作成され、バージョン履歴を表示できます。

Developer tool または Analyst ツールでプロファイルを編集するときに、モデルリポジトリはプロファイルをロックします。ツールが予期せずに停止した場合、ロックは保持されて、モデルリポジトリに再度接続したときにロックしたプロファイルを表示できます。プロファイルの編集を継続するか、プロファイルのロックを解除することができます。

モデルリポジトリがバージョン管理システムと統合されている場合は、プロファイルのバージョンを管理できます。例えば、プロファイルのチェックアウトとチェックイン、チェックアウトの取り消し、プロファイルの特定の履歴バージョンの表示、チェックアウトしたプロファイルの表示などが可能です。Analyst ツールでのリポジトリアセットのロックおよびバージョン管理されたアセットの管理については、『Analyst ツールガイド』を参照してください。Developer tool でのリポジトリオブジェクトのロックおよびバージョン管理されたオブジェクトの管理については、『Developer Tool ガイド』を参照してください。

スコアカード

スコアカードは、カラムの有効値またはプロファイル結果内のルールの出力を視覚的に表したものです。データ品質の進捗を測定するには、スコアカードを使用します。プロファイルからスコアカードを作成し、データ品質の進捗を時系列で監視することができます。

スコアカードには、メトリック、メトリックグループ、およびしきい値など、複数のコンポーネントがあります。プロファイルの実行後に、スコアカードにメトリックなどのソースカラムを追加し、メトリックの有効値を設定できます。スコアカードは、メトリックおよびスコアカードレベルで不良データのコストを追跡してデータ品質の値を測定するのに役立ちます。メトリックごとに不良データのコストを測定するには、メトリックにコスト単位を割り当て、固定費または変動費を設定します。スコアカードを実行すると、スコアカードの結果には、各メトリックの不良データのコストと、すべてのメトリックのコスト合計値が含まれます。

スコアカード内の関連し合うスコアを1つのまとまりとして分類するには、メトリックグループを使用します。しきい値では、レコード内のカラムで許容できる不良データの範囲を割合で指定します。正常、許容可能、または許容不可の範囲のデータに対してしきい値を設定できます。

スコアカードを実行する際、実データまたはステージングされたデータで、ドリルダウンするスコアメトリックを設定します。スコアカードを実行してスコアを表示した後、各メトリックをドリルダウンして、有効なデータレコードと無効なレコードを識別します。スコアカードリネージュは、スコアカード内のメトリックまたはメトリックグループごとにも表示できます。データ品質を効率的に追跡するには、スコアの傾向グラフとコストの傾向グラフを使用します。これらのグラフで、不良データのスコアとコストが、特定の期間にわたって変化する様子を監視します。

スコアカードのメトリックと設定情報はプロファイリングウェアハウスに保存されます。スコアカードの結果を取得してレポートを実行するようにサードパーティのアプリケーションを設定できます。スコアカードの結果は、Web アプリケーション、ポータル、またはビジネスインテリジェンスレポートなどのレポートにも表示できます。

第 3 章

キュレーションの概念

この章では、以下の項目について説明します。

- [キュレーションの概念の概要, 17 ページ](#)
- [キュレーションタスク, 17 ページ](#)

キュレーションの概念の概要

キュレーションとは、データソースで検出されたメタデータを検証し管理することで、使用とレポートに適したメタデータにするプロセスです。

キュレーションできる推測されたプロファイル結果を、次に示します。

- データ型
- データドメイン
- プライマリキー
- 外部キー

推測されたプロファイル結果をキュレーションすることで、カラム、データドメイン、およびデータベースとスキーマでのデータオブジェクト関係に関するメタデータを正確にします。そして、検出検索を使用して複数のリポジトリ全体で情報を検索した場合に、最も関連性の高いメタデータを探することができます。また、エンタープライズ検出の結果に外部キーの関係図を表示した場合にも、最も関連性の高いメタデータを探することができます。

プロファイル実行中にプロファイルによって生成された特定のメタデータの推測をキュレーションできます。例えば、カラムプロファイルの結果とデータドメイン検出の結果の中の推測されたデータ型を承認または拒否できます。また、エンタープライズ検出の結果の中の推測されたプライマリキーや外部キーも承認または却下できます。

キュレーションタスク

プロファイルの実行後にプロファイル結果をキュレーションすることができます。前回プロファイルを実行したときのキュレーションの決定を撤回することもできます。

Analyst ツールで実行できるキュレーションタスクは次のとおりです。

- 複数のカラムまたはデータドメインについて推測されたデータ型を承認または拒否します。

- [承認済み] または [拒否済み] のデータ型を [推測済み] ステータスに戻します。
- [承認済み] または [却下済み] のデータドメインを [推測済み] ステータスに戻します。
- 却下された結果の行を表示または非表示にします。
- 承認されたデータ型とデータドメインなど、メタデータの特定の設定に基づいてカラムをプロファイルの実行から除外します。

Developer ツールで実行できるキュレーションタスクは次のとおりです。

- 複数のカラムについて、推測されたデータ型を承認または拒否します。
- [承認済み] または [拒否済み] のデータ型を [推測済み] ステータスに戻します。
- [承認済み] または [却下済み] のデータドメインを [推測済み] ステータスに戻します。
- 却下された結果の行を表示または非表示にします。
- プライマリキー検出の結果の中のデータオブジェクトを承認または却下します。
- 外部キー検出の結果など、エンタープライズ検出の結果を承認または却下します。
- 承認されたデータ型とデータドメインなど、メタデータの特定の設定に基づいてカラムをプロファイルの実行から除外します。

パート II: Informatica Analyst によるプロファイリング

この部には、以下の章があります。

- [Informatica Analyst のカラムプロファイル, 20 ページ](#)
- [Informatica Analyst のルール, 32 ページ](#)
- [Informatica Analyst のフィルタ, 37 ページ](#)
- [Informatica Analyst のカラムプロファイル結果, 42 ページ](#)
- [Informatica Analyst のスコアカード, 64 ページ](#)

第 4 章

Informatica Analyst のカラムプロファイル

この章では、以下の項目について説明します。

- [Informatica Analyst のカラムプロファイルの概要, 20 ページ](#)
- [カラムプロファイリングプロセス, 21 ページ](#)
- [プロファイルオプション, 21 ページ](#)
- [ランタイム環境, 23 ページ](#)
- [Informatica Analyst のオペレーティングシステムプロファイルの概要, 24 ページ](#)
- [リポジトリアセットのロックとチームベース開発の概要, 25 ページ](#)
- [Informatica Analyst でのカラムプロファイルの作成, 26 ページ](#)
- [カラムプロファイルの編集, 27 ページ](#)
- [プロファイルの実行, 28 ページ](#)
- [Spark エンジンでのプロファイルの実行, 28 ページ](#)
- [Databricks クラスタでのプロファイルの実行, 28 ページ](#)
- [同期オプション, 29 ページ](#)

Informatica Analyst のカラムプロファイルの概要

プロファイルを作成するには、プロファイルを実行するデータオブジェクトのカラムを選択します。サンプリングおよびドリルダウンのオプションを設定することによって、プロファイリングの高速化を図ることができます。ランタイム環境を選択できます。プロファイルを作成するには、ルールおよびフィルタをプロファイルに追加できます。プロファイルの実行後に、プロファイリングの統計を調べてデータを把握できます。

最大 1000 のカラムを含む幅の広いテーブルおよびフラットファイルをプロファイリングできます。プロファイルを作成または実行するには、すべてのカラムを選択するか、プロファイルで使用する各カラムを選択できます。ドリルダウン対象としてすべてのカラムを選択し、これらのカラムの値の頻度を表示することができます。カラム名が 245 文字を超えると、区切りファイルでプロファイルを実行するカラムを選択できません。

Spark エンジンでは、半構造化データソースでプロファイルを実行できません。

Informatica Analyst の次の方法を使用してカラムプロファイルを作成できます。

- **【ライブラリ】** ワークスペースでデータオブジェクトを右クリックして、プロファイルを作成します。
- デフォルトのオプションを使用してデフォルトのカラムプロファイルを作成します。

- プロファイルの設定をカスタマイズしてカスタムプロファイルを作成します。

注: Avro、JSON、Parquet、および XML データソースに対して、プロファイルを表示および実行できます。Informatica Developer で、Avro、JSON、Parquet および XML データソースに対してカラムプロファイルを作成および編集できます。

カラムプロファイリングプロセス

カラムプロファイリングプロセスの一部として、プロファイリングにすべてのソースカラムを含めるか、特定のカラムを選択することができます。また、デフォルトのプロファイルオプションをそのまま使用したり、サンプリングオプション、ドリルダウンオプション、およびランタイム環境を設定したりすることもできます。

以下に、カラムプロファイリングプロセスの手順を示します。

1. カラムプロファイルの名前、説明、場所を選択します。
2. プロファイルを実行するインポートされたデータオブジェクトまたは外部ソースを選択します。
3. 必要に応じて、ソースデータをプレビューします。
4. プロファイルを実行するカラムを選択します。
5. デフォルトオプションを使用してプロファイルを作成するか、デフォルトのオプションを変更するかを指定します。設定できるオプションには、サンプリングオプション、ドリルダウンオプション、およびランタイム環境があります。
6. 必要に応じて、プロファイルの作成時にルールおよびフィルタを追加します。
7. プロファイルを実行します。

注: カラム名、および多言語データや Unicode データのプロファイリングについては、以下のルールとガイドラインに注意してください。

- さまざまなソースから多言語データをプロファイリングしたり、ブラウザのロケール設定に基づいてプロファイル結果を表示したりすることができます。Analyst ツールでは、ブラウザのロケールに基づいて、Datetime、Numeric、Decimal の各データ型が変更されます。
- 多言語データに基づくソート: 多言語データに基づいてソートすることができます。Analyst ツールでは、ソート順がブラウザのロケールに基づいて表示されます。
- DB2 データベースの Unicode データをプロファイリングするには、データベースで DB2CODEPAGE データベース環境変数を設定し、データ統合サービスを再起動します。

プロファイルオプション

プロファイルオプションには、データサンプリングオプションとデータドリルダウンオプションがあります。データオブジェクトのカラムプロファイルを作成または編集するときに、これらのオプションを設定できます。

プロファイルオプションを設定するには、**[検出]** ワークスペースを使用します。カラム、サンプリング、およびドリルダウンのデフォルトのオプションを使用してプロファイルを作成できます。実データとステージングされたデータの選択を切り替えるには、ドリルダウンオプションを使用します。

サンプリングのオプション

Analyst ツールでプロファイルの対象として選択される行の数は、サンプリングのオプションで決まります。サンプリングのオプションは、プロファイルの定義または実行するときに設定することができます。

以下の表に、プロファイルのサンプリングのオプションを示します。

オプション	説明
すべての行	データオブジェクト内のすべての行でプロファイルを実行します。 ネイティブ、Blaze、Spark、および Databricks ランタイム環境でサポートされます。
最初の<number>行をサンプリング	データオブジェクト内の行のうち、先頭からのサンプル行でプロファイルを実行します。行は最大で 2,147,483,647 行まで選択できます。 ネイティブおよび Blaze ランタイム環境でサポートされます。
<number>行のランダムサンプリング	データオブジェクトでランダムに選択された数の行に対してプロファイルを実行します。行は最大で 2,147,483,647 行まで選択できます。 ネイティブおよび Blaze ランタイム環境でサポートされます。
ランダムサンプリング (自動)	データオブジェクト内の行数に基づいて計算されたサンプル行でプロファイルを実行します。 ネイティブおよび Blaze ランタイム環境でサポートされます。
Limit n <number>行	データオブジェクトの行数に基づいてプロファイルを実行します。Hadoop 検証環境でプロファイルを実行するように選択すると、Spark エンジンはデータオブジェクトの複数のパーティションからサンプルを収集し、これらのサンプルを単一のノードにプッシュしてサンプルサイズを計算します。Limit n のサンプリングオプションは、Oracle、SQL Server および DB2 データベースをサポートします。Limit n のサンプリングオプションには、詳細フィルタは適用できません。 Spark ランタイム環境でサポートされます。
ランダムパーセンテージ	データオブジェクト内の行の割合に対してプロファイルを実行します。 Spark および Databricks ランタイム環境でサポートされます。
以降のプロファイル実行で、データ型とデータドメイン推測から、承認済みのデータ型とデータドメインを除外します	次のプロファイル実行で、データ型とデータドメイン推測から、承認済みのデータ型とデータドメインを除外します。

ランダムな行のサンプルでプロファイルを実行するように選択すると、ランダムサンプルアルゴリズムでは、データオブジェクト内の行をランダムに選択して、プロファイルを実行します。カラムプロファイルのランダムサンプリングオプションを選択すると、Analyst ツールは、ステージングされたデータに対してドリルダウンを実行します。これは、ドリルダウンのパフォーマンスに影響を与える可能性があります。データドメイン検索プロファイルのランダムサンプリングオプションを選択すると、Analyst ツールはライブデータのドリルダウンを実行します。

ドリルダウンのオプション

ドリルダウンのオプションは、プロファイルを定義または編集するときに設定できます。

以下の表に、プロファイルのドリルダウンのオプションを示します。

オプション	説明
Live	実データをドリルダウンして、データソースの現在のデータを読み取ります。
Staged	ステージングされたデータをドリルダウンして、プロファイリングウェアハウスでステージングされているプロファイルデータを読み取ります。
カラム選択	プロファイリング対象として選択していないカラムをドリルダウン対象として指定します。

ランタイム環境

カラムプロファイルのランタイム環境として、ネイティブ、Hadoop、または Databricks を選択できます。Hadoop ランタイム環境では、Blaze または Spark エンジンを選択できます。Databricks ランタイム環境では、Databricks Spark オプションを選択できます。Informatica Analyst では、ランタイム環境の選択後に、プロファイル定義にランタイム環境が設定されます。

ネイティブ環境

ネイティブのランタイム環境でプロファイルを実行すると、Analyst ツールがプロファイルジョブをプロファイリングサービスモジュールにサブミットします。次に、プロファイリングサービスモジュールが、プロファイルジョブを一連のマッピングに分割します。データ統合サービスは、データ統合サービスが実行する同じマシン上でこれらのマッピングを実行し、プロファイル結果をプロファイリングウェアハウスに書き込みます。デフォルトでは、すべてのプロファイルがネイティブのランタイム環境で実行されます。

ネイティブソースを使用し、ネイティブ環境でプロファイルを作成して実行できます。ネイティブのデータソースは、フラットファイル、リレーショナルソース、メインフレームソースなど、Hadoop 以外のソースのことです。また、ネイティブ環境で Hive または HDFS データソースを使用してマッピング仕様または論理データソースに対してプロファイルを実行することもできます。

Hadoop 環境

Hadoop ランタイム環境でプロファイルを実行する場合、Blaze エンジンまたは Spark エンジンを選択できます。

[Blaze] または [Spark] の選択後に、Hadoop 接続を選択できます。データ統合サービスがプロファイルのロジックを Hadoop クラスタの Blaze または Spark エンジンにプッシュして、プロファイルを実行します。

Hadoop 環境でプロファイルを実行すると、Analyst ツールがプロファイルジョブをプロファイリングサービスモジュールにサブミットします。次に、プロファイリングサービスモジュールが、プロファイルジョブを一連のマッピングに分割します。データ統合サービスは、これらのマッピングを Hadoop 接続で Hadoop 環境にプッシュします。Blaze エンジンまたは Spark エンジンはマッピングを処理し、データ統合サービスはプロファイル結果をプロファイリングウェアハウスに書き込みます。

Sqoop データソースのカラムプロファイル

Sqoop を使用するデータオブジェクトでカラムプロファイルを実行できます。Hadoop を検証環境として選択したら、Hadoop 接続で Blaze エンジンまたは Spark エンジンを選択して、カラムプロファイルを実行します。

論理データオブジェクトまたはカスタマイズデータオブジェクトのカラムプロファイルを実行すると、num-mappers 引数を設定して並行処理を実現しパフォーマンスを最適化することができます。また、Sqoop で作業単位を分割する際の基準のカラムを指定するように、split-by 引数も設定しなければなりません。

以下の構文を使用します。

```
--split-by <column_name>
```

プライマリキーの範囲の上限と下限の間で値の均等分布がない場合、データの均等分布がある別のカラムを指定して作業単位を分割するように、split-by 引数を設定できます。

split-by カラムを定義しない場合、Sqoop では次の条件に基づいて作業単位が分割されます。

- データオブジェクトに単一プライマリキーが含まれる場合、Sqoop はプライマリキーを split-by カラムとして使用します。
- データオブジェクトにコンポジットプライマリキーが含まれる場合、Sqoop のデフォルトは、split-by 引数を使用せずにコンポジットプライマリキーを処理する Sqoop の動作になります。詳細については、Sqoop のマニュアルを参照してください。
- データオブジェクトに同一のカラムを持つ 2 つのテーブルが含まれる場合、split-by カラムをテーブル修飾名で定義する必要があります。例えば、テーブル名が CUSTOMER であり、カラム名が FULL_NAME の場合、split-by カラムを次のように定義します。
--split-by CUSTOMER.FULL_NAME
- データオブジェクトにプライマリキーがない場合、m 引数の値と num-mappers 引数は、デフォルトで 1 になります。

Cloudera Connector Powered by Teradata または Hortonworks Connector for Teradata を使用し、Teradata テーブルにプライマリキーがない場合、split-by 引数は必須です。

Databricks 環境

[Databricks Spark] オプションを選択し、Databricks ランタイム環境でプロファイルを実行できます。

[Databricks Spark] オプションの選択後に、Databricks 接続を選択できます。データ統合サービスがプロファイルのロジックを Databricks クラスターの Spark エンジンにプッシュして、プロファイルを実行します。

Databricks 環境でプロファイルを実行すると、Analyst ツールがプロファイルジョブをプロファイリングサービスモジュールにサブミットします。次に、プロファイリングサービスモジュールが、プロファイルジョブを一連のマッピングに分割します。データ統合サービスは、これらのマッピングを Hadoop 接続で Spark エンジンにプッシュします。Spark エンジンはマッピングを処理し、データ統合サービスはプロファイル結果をプロファイリングウェアハウスに書き込みます。

Informatica Analyst のオペレーティングシステムプロファイルの概要

Analyst ツールでオペレーティングシステムプロファイルを選択できます。オペレーティングシステムプロファイルを選択すると、データ統合サービスによってカラムプロファイル、エンタープライズ検出プロファイル、

およびスコアカードが作成および実行されます。この処理は、オペレーティングシステムプロファイルユーザーに定義された権限に基づいて行われます。

Analyst ツールでは、デフォルトのプロファイルを使用して、プロファイルとスコアカードが実行されます。オペレーティングシステムプロファイルが 1 つしかない場合は、そのオペレーティングシステムプロファイルがデフォルトで選択されます。オペレーティングシステムプロファイルが複数ある場合は、それらの中から 1 つを選択できます。

オペレーティングシステムプロファイルの選択

Informatica Analyst でオペレーティングシステムプロファイルを選択できます。データ統合サービスでは、プロファイリングジョブを実行するために、オペレーティングシステムプロファイルユーザーの権限を使用します。

1. Informatica Analyst のヘッダー領域で、[<ユーザー名>] > [設定] をクリックします。
[設定] ダイアログボックスが表示されます。
2. オペレーティングシステムプロファイルを選択します。[保存] をクリックします。

リポジトリアセットのロックとチームベース開発の概要

モデルリポジトリでは、ユーザーが他のユーザーの作業を上書きしないように、プロファイルがロックされます。モデルリポジトリがバージョン管理システムと統合されている場合は、複数のバージョンのアセットが保存され、バージョンにバージョン番号が割り当てられます。プロファイルのチェックアウトとチェックイン、チェックアウトの取り消しなどを実行できます。チェックアウトしたプロファイルの特定のバージョンを表示できます。

Analyst ツールでプロファイルの編集を開始すると、他のユーザーがそのプロファイル編集できないようにプロファイルがロックされます。プロファイルを保存すると、ロック状態が保持されます。プロファイルを閉じると、プロファイルのロックは解除されます。

モデルリポジトリでは、バージョン管理されたアセット管理を使用してチームの他のメンバがプロファイルを上書きできないように、プロファイルが保護されます。別のユーザーによってチェックアウトされたプロファイルを編集しようとする、そのプロファイルをチェックアウトしたユーザーを示す通知を受け取ります。チェックアウト済みのプロファイルは、読み取り専用モードで開くことができるほか、別の名前でも保存することもできます。

[プロファイルのプロパティ] ダイアログボックスでプロファイルのバージョンを選択し、そのバージョンのプロファイル定義を表示できます。[アクション] メニューで [プロファイルのプロパティ] オプションにアクセスできます。リポジトリアセットのロックおよびバージョン管理されたアセットの管理の詳細については、『Analyst ツールガイド』を参照してください。

Informatica Analyst でのカラムプロファイルの作成

カスタムプロファイルまたはデフォルトプロファイルを作成できます。カスタムプロファイルを作成する際には、カラム、サンプル行、ドリルダウンオプションを設定できます。デフォルトプロファイルを作成すると、カラムプロファイルとデータドメイン検出がすべてのデータドメインの全データセットに対して実行されます。

1. **【検出】** ワークスペースで、**【プロファイル】** をクリックするか、ヘッダ領域から **【新規】** > **【プロファイル】** を選択します。

注: 【ライブラリ】 ワークスペースでデータオブジェクトを右クリックして、プロファイルを作成できます。このプロファイルでは、データオブジェクトのプロパティからプロファイル名、場所名、およびデータオブジェクトが抽出されます。デフォルトのプロファイルを作成するか、設定をカスタマイズして、カスタムプロファイルを作成できます。

新規プロファイルウィザードが表示されます。

2. **【単一ソース】** オプションがデフォルトで選択されています。**【次へ】** をクリックします。
3. **【全般プロパティの指定】** 画面で、プロファイルの名前と説明（省略可能）を入力します。**【場所】** フィールドで、プロファイルの作成先のプロジェクトまたはフォルダを選択します。**【次へ】** をクリックします。
4. **【ソースの選択】** 画面で、**【選択】** をクリックしてデータオブジェクトを選択するか、**【新規】** をクリックしてデータオブジェクトをインポートします。**【次へ】** をクリックします。

- **【データオブジェクトの選択】** ダイアログボックスで、データオブジェクトを選択します。**【OK】** をクリックします。

【プロパティ】 ペインに、選択したデータオブジェクトのプロパティが表示されます。**【データプレビュー】** ペインに、データオブジェクトのカラムが表示されます。

- **【新しいデータオブジェクト】** ダイアログボックスで、接続、スキーマ、テーブル、またはビューを選択し、プロファイルの作成、場所の選択、データオブジェクトのインポート先フォルダの作成ができます。**【OK】** をクリックします。

5. **【ソースの選択】** 画面で、プロファイルを実行するカラムを選択します。**【名前】** を選択すると、すべてのカラムを選択できます。**【次へ】** をクリックします。

デフォルトでは、すべてのカラムが選択されています。Analyst ツールによって、名前、データ型、精度、スケール、NULL 可能、および各カラムがプライマリキーに含まれているかどうかなど、カラムのプロパティがリストされます。

6. **【設定の指定】** 画面で、カラムプロファイル、データドメイン検出、またはカラムプロファイルとデータドメイン検出のうちどれを実行するかを選択します。デフォルトでは、カラムプロファイルオプションが選択されています。

- **【カラムプロファイルを実行】** を選択して、カラムプロファイルを実行します。

- **【データドメイン検出を実行】** を選択して、データドメイン検出を実行します。**【データドメイン】** ペインで、検出するデータドメインを選択して、適合条件を選択し、**【データドメイン検出のカラム選択の編集】** ダイアログボックスでデータドメイン検出対象カラムを選択します。

- **【カラムプロファイルを実行】** と **【データドメイン検出を実行】** を選択して、カラムプロファイルとデータドメイン検出を実行します。**【データドメイン】** ペインでデータドメインオプションを選択します。

注: デフォルトでは、選択したカラムはカラムプロファイルとデータドメイン検出の対象となります。**【編集】** をクリックして、データドメイン検出のカラムを選択または選択解除します。

- データドメイン検出の実行対象として、**【データ】**、**【カラム】**、**【データとカラム】** を選択します。

- サンプルングオプションを選択します。**【次に対してプロファイルを実行】** ペインで、サンプルングオプションとして **【すべての行（完全分析）】**、**【サンプリング最初の】**、**【ランダムサンプリング】**、**【ランダムサンプリング（自動）】**、**【Limit N】**、または **【ランダムパーセンテージ】** を選択できます。このサンプリングオプションはカラムプロファイルとデータドメイン検出に適用されます。

- ドリルダウンオプションを選択します。[ドリルダウン] ペインでは、[ライブ] または [ステージングされた] ドリルダウンオプションを選択したり、[オフ] を選択してドリルダウンを無効にしたりできます。必要に応じて、[カラムの選択] をクリックして、ドリルダウンするカラムを選択します。承認済みのデータ型またはデータドメインを持つカラムの場合は、データ型およびデータドメインの推測を省略できます。
 - ランタイム環境として、[ネイティブ]、[Blaze]、[Spark]、または [Databricks] を選択します。[Blaze] または [Spark] を選択した場合は、[選択] をクリックして、[Hadoop 接続の選択] ダイアログボックスで Hadoop 接続を選択します。[Databricks] を選択した場合は、[選択] をクリックして、Databricks 接続を選択します。
7. [次へ] をクリックします。
[ルールとフィルタの指定] 画面が開きます。
 8. [ルールとフィルタの指定] 画面では、次のタスクを実行できます。
 - ルールを作成、編集、または削除します。既存のルールをプロファイルに適用できます。
 - フィルタを作成、編集、または削除します。

注: このプロファイルでスコアカードを作成した場合、このプロファイル用に作成したフィルタを再利用できます。
 9. [保存して完了] をクリックしてプロファイルを作成するか、[保存して実行] をクリックしてプロファイルを作成および実行します。

カラムプロファイルの編集

プロファイル実行後にカラムプロファイルに変更を加えることができます。

1. [ライブラリ] ワークスペースでプロファイルを含むプロジェクトを選択するか、[アセット] ペインでプロファイルを選択します。
2. プロファイル名をクリックします。
サマリビューが [検出] ワークスペースに表示されます。
3. バージョン管理システムが有効になっている場合、[アクション] > [チェックアウト] をクリックしてプロファイルをチェックアウトします。
4. [アクション] > [プロファイルの編集] をクリックします。
[プロファイル] ウィザードが表示されます。
5. 変更する内容に基づいて、次のページオプションのいずれかを選択します。
 - **全般プロパティの指定。** 名前、説明、場所などの基本プロパティを変更します。
 - **ソースの選択。** プロファイルの実行対象として、別の一致するデータソースとカラムを選択します。
 - **設定の指定。** カラムプロファイルまたはカラムプロファイルとデータドメイン検出を実行するように選択します。検出対象のデータドメインを選択し、データドメイン検出、サンプリング、およびドリルダウンのオプションを変更します。
 - **ルールとフィルタの指定。** ルールとフィルタを作成、編集、または削除します。
6. [保存して完了] をクリックしてプロファイルの編集を完了するか、[保存して実行] をクリックしてプロファイルを編集および実行します。
7. バージョン管理システムが有効な場合、次のタスクを実行する必要があります。
 - [保存して完了] をクリックしてプロファイルの編集を完了します。

- [サマリ] ビューで、[チェックイン] をクリックしてプロファイルをチェックインします。
- [アクション] > [プロファイルの実行] をクリックしてプロファイルを実行します。

プロファイルの実行

プロファイルを実行してデータソースの内容および構造を分析し、ドリルダウン対象のカラムとルールを選択します。カラムとルールの実データまたはステージングされたデータをドリルダウンできます。初期プロファイルの実行後は、すべてのソースカラムに対してプロファイルを実行せずに、特定のカラムまたはルールに対してのみプロファイルを実行できます。

1. [ライブラリ] ワークスペースから、[プロジェクト] ペインで目的のプロファイルが格納されているプロジェクトまたはフォルダを選択するか、[アセット] ペインでプロファイルを選択します。

2. [アクション] > [開く] をクリックします。

サマリビューが[検出] ワークスペースに表示されます。

3. [アクション] > [プロファイルの実行] をクリックします。

Analyst ツールによってプロファイルが実行され、サマリビューにプロファイル結果が表示されます。プロファイルサマリログファイルおよびマッピングログファイルを表示して、Analyst ツールによって実行されたタスクの詳細を取得できます。

注: オペレーティングシステムプロファイルを使用してプロファイルを実行した場合、サマリログはデータ統合サービス用に構成されたログディレクトリにあり、マッピングログはオペレーティングシステムのプロファイル用に構成されたログディレクトリにあります。

Spark エンジンでのプロファイルの実行

Spark エンジンで JDBC 接続を使用してプロファイルを実行すると、プロファイルの実行が失敗します。

Spark エンジンでプロファイルを実行する前に、次の手順を実行します。

1. JDBC ウェアハウス接続を作成します。
2. データを抽出するデータベースの Data Direct JAR ファイルを取得します。
3. ファイルを次の場所にダウンロードします。<INFA_HOME>/externaljdbcjars

Databricks クラスタでのプロファイルの実行

Databricks クラスタを使用して、Spark エンジンで JDBC 接続を使用してプロファイルを実行すると、プロファイルの実行が失敗します。

Spark エンジンでプロファイルを実行する前に、次の手順を実行します。

1. JDBC ウェアハウス接続を作成します。
2. データを抽出するデータベースの Data Direct JAR ファイルを取得します。
3. ファイルを次の場所にダウンロードします。<INFA_HOME>/externaljdbcjars

同期オプション

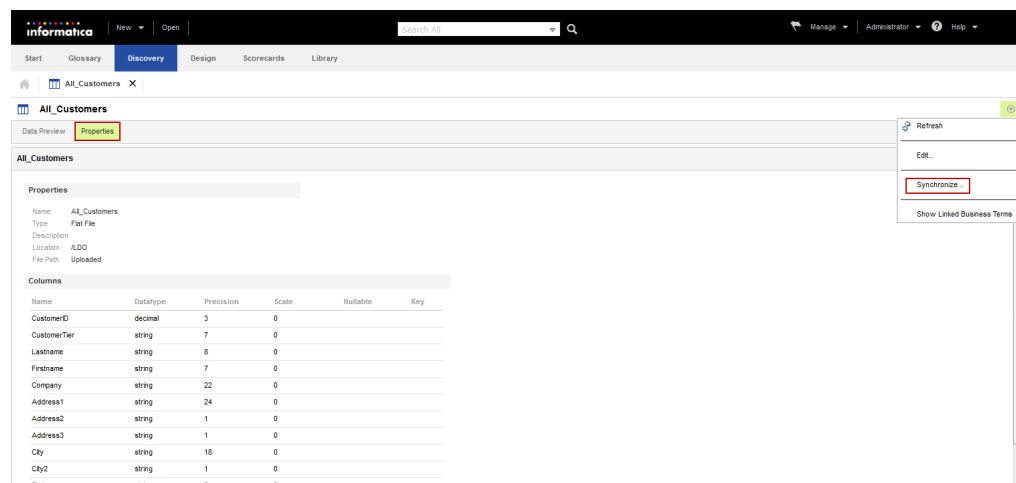
外部データソースのメタデータを変更する場合、モデルリポジトリ内のデータオブジェクトのメタデータはデフォルトでは更新されません。データオブジェクトのメタデータをデータソースのメタデータと同期させるには、[同期] オプションを使用します。[同期] オプションは、カラムプロファイル、エンタープライズ検出プロファイル、およびスコアカードに使用できます。外部データソースは、リレーショナルデータソースまたはフラットファイルデータソースにすることができます。

Informatica Analyst でのフラットファイルデータオブジェクトの同期

Analyst ツールで、外部フラットファイルデータソースに加えられた変更をそのデータオブジェクトに同期させることができます。データオブジェクトを同期させるには、**フラットファイルの同期**ウィザードを使用します。

1. **[ライブラリ]** ワークスペースを開きます。
2. **[プロジェクト]** セクションで、フラットファイルデータオブジェクトをプロジェクトから選択します。
Analyst ツールは、**[プロパティ]** タブにテーブルのプロパティを表示します。
3. **[アクション]** メニューから **[同期]** をクリックします。

次の図は、[アクション] メニューの [プロパティ] タブと [同期] オプションを示しています。



[フラットファイルの同期] ウィザードが表示されます。

4. 場所を参照するか、フラットファイルをインポートするネットワークパスを入力します。
 - 場所を参照するには、**[ファイルの選択]** をクリックして、マシンからアクセスできるディレクトリからフラットファイルを選択します。
 - ネットワークパスを入力するには、**[ネットワークパスの入力]** を選択して、ファイルパスとファイル名を設定します。

次の図は、フラットファイルの同期ウィザードを示しています。

Synchronize Flat File: Step 1 of 5

Specify a location to import the flat file from and specify how to import the flat file.

☒ Browse and Upload:

Browse...

No file selected.

☐ Enter a Network Path:

☐ Hadoop File System

Description

Upload files from a local machine. Recommended for smaller files up to 10 MB. The Analyst tool uploads a copy of the file to the node on which the Analyst Service runs. Upload the file again if you modify the file.



Back

Next

Finish

Cancel

5. **【次へ】** をクリックします。
6. 区切られたフラットファイルまたは固定幅フラットファイルのインポートを選択します。
 - 区切られたフラットファイルをインポートするには、**【区切り記号付き】** オプションを選択します。
 - 固定幅フラットファイルをインポートするには、**【固定幅】** オプションを選択します。
7. **【次へ】** をクリックします。
8. 区切りフラットファイルまたは固定幅フラットファイルのオプションを設定します。
9. **【次へ】** をクリックします。
10. 必要に応じて、カラムの属性を変更します。
11. **【次へ】** をクリックします。
12. デフォルト名をそのまま使用するか、フラットファイルの別の名前を入力します。
13. 必要に応じて、説明を入力します。
14. **【完了】** をクリックします。

操作の確認を求める同期メッセージが表示されます。
15. **【はい】** をクリックして、フラットファイルを同期します。

同期が完了したことを示すメッセージが表示されます。メタデータの変更の詳細を表示するには、**【詳細を表示】** をクリックします。
16. **【OK】** をクリックします。

Informatica Analyst でのリレーショナルデータオブジェクトの同期

外部リレーショナルデータソースに加えられた変更をそのテーブルデータオブジェクトに同期させることができます。外部データソースの変更には、ソースカラムとルールカラムの追加、変更、削除が含まれます。

1. **【ライブラリ】** ワークスペースを開きます。
2. **【プロジェクト】** セクションで、テーブルデータオブジェクトをプロジェクトから選択します。
Analyst ツールは、**【プロパティ】** タブにテーブルのプロパティを表示します。
3. **【アクション】** メニューから **【同期】** をクリックします。
操作の確認を求めるメッセージが表示されます。
4. 同期プロセスを完了するには、**【はい】** をクリックします。
同期ステータスメッセージが表示されます。
5. 同期が完了したことを示すメッセージが表示されます。
メタデータの変更の詳細を表示するには、**【詳細を表示】** をクリックします。
6. **【OK】** をクリックします。

第 5 章

Informatica Analyst のルール

この章では、以下の項目について説明します。

- [Informatica Analyst のルールの概要, 32 ページ](#)
- [定義済みルール, 32 ページ](#)
- [式ルール, 34 ページ](#)

Informatica Analyst のルールの概要

ルールは、カラムプロファイルの実行時にソースデータに適用される条件を定義するビジネスロジックです。ルールをプロファイルに追加して、データを検証できます。

ルールはさまざまな状況で使用します。例えば、1 つ以上のデータカラムをクレンジングするルールを追加したり、ソースデータでは提供されない情報を提供するルックアップルールを追加できます。また、データ品質プロジェクトやデータ統合プロジェクト用のクレンジングルールを検証するルールを追加することもできます。

カラムプロファイルを作成または編集する場合は、ルールを作成してプロファイルに追加したり、既存のルールをプロファイルに適用したりすることができます。カラムプロファイルでは、式ルールまたは定義済みのルールを使用できます。

プロファイルを実行すると、Analyst ツールによって、[概要] ビューの [ルール] 列にプロファイル結果が表示されます。特定のルールのカラム結果は、詳細ビューで確認できます。ルールの出力を 1 つ以上の仮想カラムにすることができます。仮想カラムはプロファイル結果に存在します。Analyst ツールは仮想カラムに対してプロファイルを実行します。例えば、姓と名が含まれている 1 つのカラムを FIRST_NAME と LAST_NAME の仮想カラムに分割する定義済みルールを使用するとします。Analyst ツールは、FIRST_NAME カラムと LAST_NAME カラムに対してプロファイルを実行します。

注: 他のオブジェクトタイプによって参照されているルールオブジェクトを削除しようとすると、それらのオブジェクトタイプの一覧を示すメッセージが表示されます。ルールを削除する前に、その削除の影響を判断するようにします。

定義済みルール

定義済みルールは、Developer tool で作成されたルール、または Developer tool と Analyst ツールに付属しているルールです。定義済みルールをカラムプロファイルに適用して、ソースデータを変更または検証します。

定義済みルールでは、トランスフォーメーションを使用してルールロジックを定義します。定義済みルールは複数のプロファイルで使用できます。モデルリポジトリでは、定義済みルールは入力グループ、出力グループ、およびルールロジックを定義するトランスフォーメーションが含まれるマプレットです。

定義済みルールのプロセス

プロファイルに定義済みルールを適用するには、**新しいルール**ウィザードを使用します。

定義済みルールは、次の手順に従って適用できます。

1. プロファイルを開きます。
2. 定義済みルールを選択します。
3. ルールのパラメータを確認します。
4. 入力カラムを選択します。ルールを複数のカラムに適用する必要がある場合は、複数のカラムを選択してもかまいません。
5. プロファイリングのオプションを設定します。

定義済みルールの適用

定義済みルールを適用するときは、ルールを選択し、そのルールの入力カラムと出力カラムを設定します。再利用可能なルールとして格上げされたルールを使用するか、開発者が作成したルールを使用するには、定義済みルールを適用します。

1. **【ライブラリ】** ワークスペースでプロファイルを含むプロジェクトを選択するか、**【アセット】** ペインでプロファイルを選択します。
2. **【アクション】** > **【開く】** をクリックしてプロファイルを開きます。
サマリビューが**【検出】** ワークスペースに表示されます。
3. **【アクション】** > **【プロファイルの編集】** をクリックします。
【プロファイルウィザード】 が表示されます。
4. **【ルールとフィルタの指定】** をクリックします。
5. **【ルールとフィルタの指定】** 画面の**【ルール】** パネルで、**【アクション】** > **【既存ルールの適用】** をクリックします。
【ルールの適用ウィザード】 ダイアログボックスが表示されます。
6. ルールを選択し、**【次へ】** をクリックします。
7. **【追加】** をクリックします。
【入力ポートのカラムの選択】 ダイアログボックスが表示されます。
8. フィールドと入力カラムを選択します。**【OK】** をクリックします。
入力カラムと出力カラムが**【ルールの適用ウィザード】** ダイアログボックスに表示されます。
9. **【ルールの適用ウィザード】** ダイアログボックスで、**【OK】** をクリックします。
【ルールとフィルタの指定】 画面にルールが表示されます。

式ルール

式ルールでは、式関数およびカラムを使用してルールロジックを定義します。式ルールを作成して、Analyst ツールのカラムプロファイルに追加します。

式ルールを使用して、カラムプロファイルのカラムの値を変更または検証します。1 つ以上の式ルールを作成して 1 つのプロファイルで使用できます。式関数は、ソースデータの変換に使用される SQL のような関数です。式ルールロジックを作成するときは、次のタイプの関数を使用できます。

- 文字
- 変換
- データクレンジング
- 日付
- エンコーディング
- 財務
- 数値
- 科学
- 特殊
- テスト

次のメソッドを使用して、式ルールを作成できます。

- プロファイルウィザード。列プロファイルを作成または編集するときは、プロファイルウィザードで式ルールを作成して適用できます。ルールを再利用可能なルールに昇格させ、複数のプロファイルで使用することができます。
- ルール仕様。Analyst ツールでルール仕様を構成し、カラムプロファイルでルール仕様を使用することができます。ルール仕様を設定するときは、ビジネスルールの要件を 1 つ以上のルール文に変換します。ルール文は、データセットがビジネスルールに準拠しているかどうかを判定するロジックです。ルール仕様からマップレットを生成し、Developer tool で作成するカラムプロファイルでマップレットを使用します。

式エディタを使用して式関数を追加し、カラムを関数への入力として設定して式を検証し、戻り型、精度、およびスケールを設定できます。式のルールを作成および検証したら、出力ルールカラムの精度値を編集できます。デフォルトでは、出力ルールカラムの精度値は 10 に設定されています。出力ルールカラムが設定精度値を超えている場合は、精度値が切り詰められます。

式ルールの出力は、ルールの名前をカラム名として使用する仮想カラムです。Analyst ツールが仮想カラムに対してカラムプロファイルを実行します。例えば、式ルールを使用して郵便番号を検証するとします。このルールでは、郵便番号が有効なら 1 が、無効なら 0 が返されます。ルールの出力値 1 と 0 に対して、Informatica Analyst がカラムプロファイルを実行します。

式ルールの作成

【プロファイル】 ウィザードを使用すると、式ルールを作成してプロファイルに追加できます。プロファイルのカラムの値を検証するための式ルールを作成します。

1. プロファイルを開きます。
2. サマリビューで、**【アクション】** > **【プロファイルの編集】** をクリックして、**【プロファイル】** ウィザードを開きます。
3. **【ルールとフィルタの指定】** をクリックします。
4. **【ルール】** ペインで、**【アクション】** > **【ルールの追加】** をクリックします。



【新しいルール】ダイアログボックスが表示されます。

5. 【新しいルール】ダイアログボックスで、ルールの名前と説明（省略可能）を入力します。ルールは、[関数] パネルまたは[カラム] パネルで作成できます。
 - [関数] パネルでは、関数カテゴリを選択して、右矢印ボタン（>>）をクリックします。ダイアログボックスでパラメータを指定して、[OK] をクリックします。
関数、およびカラムと値が、[式] パネルに表示されます。
 - [カラム] パネルでカラムを選択し、右矢印ボタン（>>）をクリックします。[式] パネルにカラムが表示されます。関数、式、値を追加してルールを作成します。
6. ルールを確認するには、[検証] をクリックします。
7. 必要に応じて、ルールを再利用可能なルールとして格上げするように選択し、プロジェクトとフォルダーの場所を設定します。ルールを再利用可能なルールに格上げすると、他のユーザーもこのルールを別のプロファイルで定義済みルールとして使用できるようになります。
8. [OK] をクリックします。
[ルールとフィルタの指定] 画面で、[ルール] ペインにルールが表示されます。

ルール仕様を使用した式ルールの作成

ルール仕様を使用して、Informatica Analyst で式ルールを作成できます。ルールを列プロファイルに追加して、データを検証できます。

1. ヘッダー領域で、[新規] > [ルール仕様] をクリックします。
【新しいルール仕様】ウィザードが表示されます。

2. **【新しいルール仕様】** ダイアログボックスで、ルールの名前と説明（省略可能）を入力します。
3. **【場所】** フィールドで、**【参照】** をクリックして、ルールを保存するプロジェクトまたはフォルダを選択します。
4. **【続行】** をクリックします。
ルール仕様は、**【デザイン】** ワークスペースに表示されます。
5. ルールのプロパティを入力するには、ルールで最上位の八角形の図形を選択し、**【プロパティ】** をクリックします。
6. プライマリルールセットを構成するには、ルールの **【次のレベルの四角形】** をクリックします。
7. ルールセットの入力を入力するには、**【プロパティ】** > **【入力】** をクリックします。
【入力管理】 ダイアログボックスが表示されます。
8. **【入力管理】** ダイアログボックスで、**【入力の追加】** をクリックし、入力の名前、データ型、最大長、および説明を入力します。必要に応じて、複数の入力を入力できます。
9. **【OK】** をクリックします。
入力は、**【プロパティ】** セクションに表示されます。
10. ルールロジックを定義するには、**【ルールロジック】** をクリックし、演算子、条件を入力して、**【アクション】** リストでアクションを選択します。
11. 必要に応じて、複数のルールセットを入力します。
12. ルールを検証するには、**【検証】** () アイコンをクリックします。
13. 列プロファイルにルール仕様を保存して使用するには、**【保存して終了】** をクリックします。
14. ルールを保存して作業を続行するには、**【保存して続行】** をクリックします。
15. Developer tool でルールの仕様を使用するには、**【ルールの生成】** () アイコンをクリックしてマップレットを生成します。
Analyst ツールによりモデルリポジトリにマップレットが作成されます。マップレットをルールとして検証し、Developer tool で作成したカラムプロファイルでマップレットを使用します。

第 6 章

Informatica Analyst のフィルタ

この章では、以下の項目について説明します。

- [Informatica Analyst のフィルタの概要, 37 ページ](#)
- [フィルタの作成, 37 ページ](#)
- [フィルタの管理, 40 ページ](#)

Informatica Analyst のフィルタの概要

フィルタを作成すると、フィルタ条件を満たす、元のデータソースのサブセットを作成することができます。その後、フィルタリングされたデータに対してプロファイルを実行できます。

フィルタを作成すると、プロファイル結果のうちフィルタ条件を満たすものを表示できます。サマリビューで使用可能なデフォルトフィルタを使用してプロファイル結果を表示できます。

フィルタの作成

フィルタを作成すると、フィルタ条件を満たす、元のデータソースのサブセットを作成することができます。

1. プロファイルを開きます。
2. サマリビューで、**[アクション]** > **[プロファイルの編集]** をクリックします。
[プロファイル] ウィザードが表示されます。
3. **[ルールとフィルタの指定]** をクリックします。
4. **[フィルタ]** ペインで、**[アクション]** > **[フィルタの追加]** をクリックします。
[新規フィルタ] ダイアログボックスが表示されます。
5. 簡易フィルタ、詳細フィルタ、または SQL フィルタを作成します。

注: 日付カラムに対する簡易フィルタまたは詳細フィルタでは、YYYY/MM/DD HH:MM:SS 形式で条件を指定してください。

[データプレビュー] ペインには、フィルタ条件を満たす、元のデータソースのサブセットが表示されます。

6. **[OK]** をクリックします。
[ルールとフィルタの指定] 画面で、**[フィルタ]** ペインにフィルタが表示されます。

簡易フィルタの作成

次の条件演算子を使用して簡易フィルタを作成できます: =、!=、>、<。簡易フィルタを使用すると、元のデータソースのサブセットを作成できます。

1. **【新規フィルタ】** ダイアログボックスで **【簡易】** をクリックします。

次の画像は、**【新規フィルタ】** ダイアログボックスで簡易フィルタを作成するために使用できるオプションを示しています。

New Filter

Create a filter. The filter is used to create a subset of the data rows before profiling.

Name*:

Description:

Choose the filter type*: ☒ Simple ☐ Advanced ☐ SQL

Columns	Operator	Values(s)
-Select-	-Select-	+

Filter Preview

2. 名前を入力し、必要に応じて説明を入力します。
 3. カラムを選択します。
 4. 条件演算子を選択します。
 5. 値を入力します。
 6. 必要な場合、プラス (+) アイコンをクリックしてフィルタを追加できます。
 7. **【OK】** をクリックします。
- 【ルールとフィルタの指定】** ページで、**【フィルタ】** ペインにフィルタが表示されます。

詳細フィルタの作成

AND、OR、NOT などの式を使用して詳細フィルタを作成することで、元のデータソースのサブセットを作成できます。

1. **【新規フィルタ】** ダイアログボックスで **【詳細】** をクリックします。

次の画像は、**【新規フィルタ】** ダイアログボックスの詳細フィルタオプションを示しています。

2. 詳細フィルタの名前と説明（省略可能）を入力します。
3. 詳細フィルタは、**【関数】** パネルまたは **【カラム】** パネルを使用して作成できます。
 - **【関数】** パネルでは、関数カテゴリを選択して、右矢印ボタン（>>）をクリックします。ダイアログボックスでパラメータを指定して、**【OK】** をクリックします。関数、およびカラムと値が、**【式】** パネルに表示されます。
 - **【カラム】** パネルでカラムを選択し、右矢印ボタン（>>）をクリックします。**【式】** パネルにカラムが表示されます。関数、式、値を追加して詳細フィルタを作成します。
4. 詳細フィルタを確認するには、**【検証】** をクリックします。
5. **【OK】** をクリックします。

【ルールとフィルタの指定】画面で、**【フィルタ】** ペインにフィルタが表示されます。

SQL フィルタの作成

SQL クエリを使用して SQL フィルタを作成できます。リレーショナルデータソースに対して SQL フィルタを作成できます。

1. **【新規フィルタ】** ダイアログボックスで **【SQL】** をクリックします。

次の画像は、**【新規フィルタ】** ダイアログボックスの SQL フィルタオプションを示しています。

The screenshot shows the 'New Filter' dialog box with the 'SQL' tab selected. The dialog has a title bar 'New Filter' with a close button. Below the title bar, there is a text area for 'Name*' and a larger text area for 'Description:'. Below these, there are three radio buttons for 'Choose the filter type*': 'Simple', 'Advanced', and 'SQL' (which is selected). Below the radio buttons, there is a section titled 'Type or paste a SQL statement' with a 'Validate' button. Below this section is a large text area for the SQL statement. At the bottom, there is a 'Filter Preview' section with a refresh icon. At the very bottom, there are 'Ok' and 'Cancel' buttons.

2. 必要に応じて SQL フィルタの説明を入力します。
3. テキストボックスに、SQL クエリを入力するか貼り付けます。
4. **【検証】** をクリックして、SQL クエリを検証します。
5. **【OK】** をクリックします。

【ルールとフィルタの指定】 ページで、**【フィルタ】** ペインに SQL フィルタが表示されます。

フィルタの管理

フィルタを編集および削除できます。

1. **【ライブラリ】** ワークスペースでプロファイルを含むプロジェクトを選択するか、**【アセット】** ペインでフィルタリングするプロファイルを選択します。
2. プロファイルを開きます。
3. サマリビューで、**【アクション】** > **【プロファイルの編集】** をクリックして、**【プロファイル】** ウィザードを開きます。
4. **【ルールとフィルタの指定】** をクリックします。

5. [フィルタ] ペインでフィルタを選択して、[アクション] > [フィルタの編集] をクリックします。
[フィルタの編集] ダイアログボックスが表示されます。
6. フィルタを編集して、[OK] をクリックします。
7. フィルタを削除するには、フィルタを選択して、[アクション] > [フィルタの削除] をクリックします。

第 7 章

Informatica Analyst のカラムプロファイル結果

この章では、以下の項目について説明します。

- [Informatica Analyst のカラムプロファイル結果の概要, 42 ページ](#)
- [サマリビュー, 43 ページ](#)
- [詳細ビュー, 45 ページ](#)
- [統計, 47 ページ](#)
- [プロファイル実行のタイプ, 54 ページ](#)
- [複数のプロファイル結果の比較の概要, 56 ページ](#)
- [カラムプロファイルのドリルダウン, 60 ページ](#)
- [Analyst ツールでのキュレーション, 61 ページ](#)
- [Informatica Analyst のカラムプロファイルのエクスポートファイル, 62 ページ](#)

Informatica Analyst のカラムプロファイル結果の概要

プロファイル結果を表示して、データの内容、構造、品質を把握し分析できます。サマリビューでは、プロファイルのすべてのカラムとルールを確認できます。詳細ビューでは、特定のカラムまたはルールの詳細なプロパティを確認できます。

プロファイル結果は、**[検出]** ワークスペースで確認できます。ビューヘッダーには、プロファイルのタイプ、プロファイル内のカラム数、プロファイル内のルール数、サンプリングデータ、作成日時が表示されます。

サマリビューでは、各カラムのプロパティを、値、水平棒グラフ、またはパーセンテージとして表示できます。NULL 値、重複なしの値、重複ありの値、パターン、データ型、データドメインなどのカラムプロパティを表示できます。サマリビューでは、デフォルトフィルタを使用してプロファイル結果を表示できます。

詳細ビューでは、NULL 値、重複なしの値、重複ありの値、推測されたデータ型、推測されたデータドメイン、推測されたパターン、値、ビジネス用語を表示し、各ペインでデータをプレビューできます。

最新の実行、履歴実行、統合済み実行のプロファイル結果を表示できます。2 つのプロファイル実行のプロファイル結果を比較し、サマリビューおよび詳細ビューで結果を確認できます。プロファイル統計情報の確認とデータのキュレーションができます。プロファイル統計情報には、値、データ型、異常値、カラムとルールの統計情報が含まれます。データの検出とドリルダウンを実行できます。

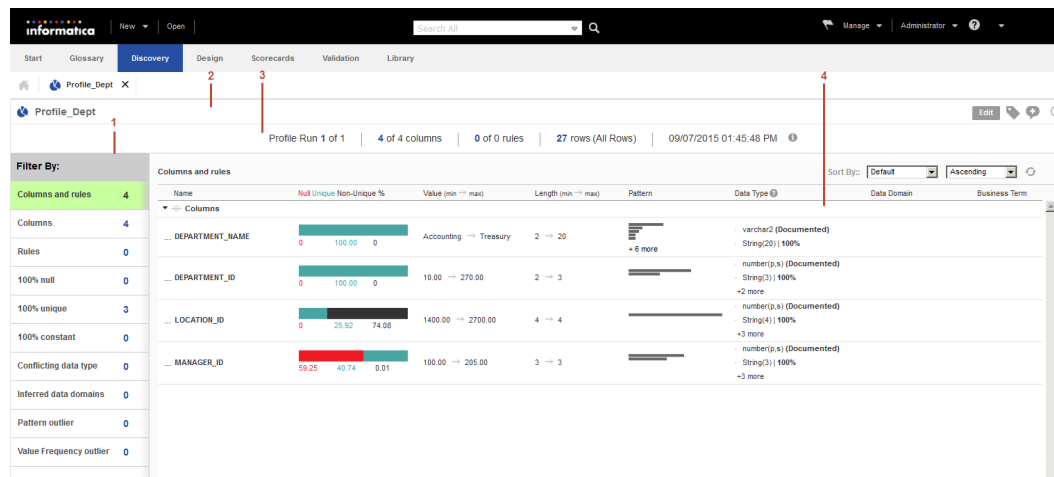
注: Avro、JSON、Parquet、および XML データソースに対して、プロファイルを表示および実行できます。最新の実行、履歴実行、統合済み実行のプロファイル結果を表示し、2 つのプロファイル実行のプロファイル結果を比較できます。

値の頻度、パターンの頻度、ドリルダウンデータ、コメント、タグ、およびビジネス用語を CSV ファイルにエクスポートできます。プロファイルのサマリ情報を Microsoft Excel ファイルにエクスポートすることで、すべてのデータをファイル内で確認して詳しく分析できます。プロファイル結果でルール情報を確認できます。表示されるプロファイル結果は、プロファイル設定およびサンプリングのオプションによって異なります。

サマリビュー

プロファイル結果のサマリは、サマリビューにグリッド形式で表示されます。サマリビューでデフォルトのフィルタを使用して、特定の統計を表示できます。例えば、[ルール] を選択すると、サマリビューにプロファイルのすべてのルールが表示されます。

次の図は、サマリビューのグラフィカルビューの例を示しています。



1. デフォルトフィルタ。サマリビューでは、デフォルトフィルタを使用してプロファイル結果を表示できます。
2. プロファイルヘッダー。ヘッダーでプロファイル名を表示できます。[編集] ボタンを使用してプロファイルの編集、タグおよびコメントアイコンを使用してタグやコメントの追加と編集をして、[アクション] メニューからオプションを選択します。
3. サマリビューのヘッダー。サマリビューのヘッダーにはプロファイル固有の情報を表示できます。プロファイルのプロファイル実行番号、プロファイル実行の総数、カラムおよびルールの数、および行数を表示できます。
4. サマリビュー。プロファイルのすべてのカラムとルールのプロパティを表示できます。

サマリビューでは、プロファイルの実行と編集、パターンまたは値頻度の異常値の検出、スコアカードへのカラムの追加、プロファイル実行の選択、2 つのプロファイル実行の比較、Microsoft Excel スプレッドシートへのプロファイル結果またはデータドメイン検出結果のエクスポート、複数カラムの推測結果の検証、コメントとタグの追加と削除、またはプロファイルプロパティの表示ができます。

サマリビューのプロパティ

サマリビューには、プロファイルのすべてのカラムとルールのプロパティが表示されます。サマリビューには、プロパティを視覚的に表したものが表示されます。各サマリプロパティをクリックするとプロパティの値をソートできます。

次の表に、サマリプロパティのプロファイル結果を示します。

プロパティ	説明
名前	プロファイルのカラムまたはルールの名前が表示されます。
NULL、重複なし、重複ありの割合	カラムまたはルール出力について、NULL 値、重複なしの値、重複ありの値が割合で表示されます。値を水平棒グラフで表示できます。
パターン	カラムの複数のパターンが水平棒グラフで表示されます。マウスポインタを棒グラフの上に移動すると、パターンの特性やカラムに出現する類似パターンの数をパーセンテージとして確認できます。
値	カラムまたはルール出力の最小値と最大値が表示されます。
長さ	カラムまたはルール出力の値の最小長と最大長が表示されます。
データ型	<p>カラムまたはルールの文書化されたデータ型が表示されます。マウスをフィールド上に移動すると、推測されたデータ型が表示されます。Analyst ツールは次のデータ型を推測できます。</p> <ul style="list-style-type: none">- String- Varchar- Decimal- Integer- Date <p>推測されたデータ型に基づいて一致率を表示することもできます。</p> <p>注: Analyst ツールでは、精度が 38 より大きい数値カラムの値からデータ型を導出することはできません。また、精度が 255 より大きい文字列カラムの値からデータ型を導出することはできません。1800 年より前の年の値を含む日付カラムに対してカラムプロファイルを作成している場合、推測されたデータ型が固定長文字列として表示される可能性があります。その場合は、InferDateTimeConfig.xml の中で year-minimum パラメータのデフォルト値を必要に応じて変更してください。</p>
データドメイン	カラムに関連付けられているデータドメインの名前と、一致率と一致する行数が一緒に表示されます。
ビジネス用語	カラムに割り当てられたビジネス用語が表示されます。

サマリビューのデフォルトフィルタ

サマリビューでは、デフォルトフィルタを使用してプロファイル結果を表示できます。

サマリビューには、すべてのソースカラム、仮想カラム、およびルールカラムのプロファイル結果がデフォルトで表示されます。[フィルタ基準] ペインには、デフォルトフィルタを適用できるカラム数が表示されます。

サマリビューでは、次のデフォルトフィルタオプションを使用して、プロファイル結果を表示できます。

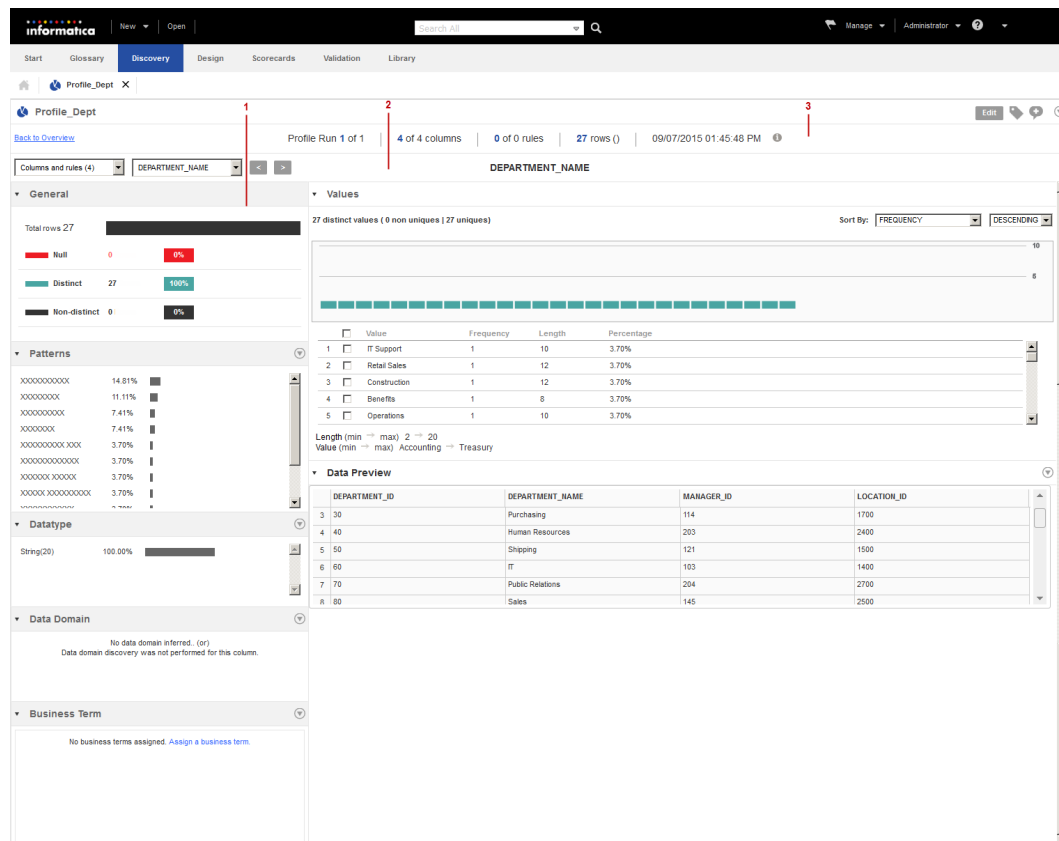
デフォルトフィルタオプション	説明
カラムとルール	ソースカラムとルールカラムのプロファイル結果を表示します。ソースカラムとルールカラムを展開および縮小して結果を表示できます。
カラム	ソースカラムのプロファイル結果を表示します。
ルール	ルールカラムのプロファイル結果を表示します。
100% NULL	100% NULL 値を持つカラムのプロファイル結果を表示します。
100%重複なし	100%重複のない値を持つカラムのプロファイル結果を表示します。
100%定数	すべてのレコードで同じ値を持つカラムのプロファイル結果を表示します。例えば、Country カラムのプロファイル結果が一律に"USA"である場合、100%定数フィルタには、Country カラムのプロファイル結果が含まれます。
競合しているデータ型	文書化されたデータ型と推測されたデータ型が一致しないカラムのプロファイル結果を表示します。例えば、CustomerTier カラムの文書化されたデータ型が Integer (2)、推測されたデータ型が String の場合、CustomerTier カラムはこのフィルタによって表示されます。
推測されたデータドメイン	推測されたデータドメインと設定済みデータドメインが同じカラムのプロファイル結果を表示します。
パターン異常値	パターン異常値を持つカラムのプロファイル結果を表示します。
値/頻度の異常値	値または頻度の異常値を持つカラムのプロファイル結果を表示します。

詳細ビュー

詳細ビューにカラム結果が表示されます。カラムプロパティの詳細を表示することができます。

カラムの詳細ビューは、サマリビューでカラムをクリックすると表示されます。

次の画像は、詳細ビューに表示されたカラムプロファイルのグラフィカルビューの例を示しています。



1. ペイン。全般プロパティ、カラム内の値、データプレビュー、推測されたパターン、推測されたデータ型、推測されたデータドメイン、およびビジネス用語を各ペインに表示できます。
2. カラムの詳細ヘッダ。ドロップダウンリストでカラムを選択するか、ナビゲーションボタンを使用して、カラム結果を表示できます。
3. サマリビューのヘッダー。サマリビューのヘッダーにはプロファイル固有の情報を表示できます。プロファイル実行と、プロファイル実行内のカラム数、ルール数、および行数のほか、プロファイル実行の日時を表示できます。

詳細ビューでは、プロファイルの実行または編集、スコアカードへのカラムの追加、プロファイル実行の選択、2つのプロファイル実行の比較、Microsoft Excel スプレッドシートへのプロファイル結果のエクスポートができます。また、値の頻度、パターンの頻度、データ型、選択した値のドリルダウンデータ、または選択したパターンのドリルダウンデータを csv ファイルにエクスポートしたり、カラムにコメントやタグを追加または削除したり、プロファイルプロパティを表示したりできます。

カラムのプロパティに対してさらなるアクションを実行するには、各ペインの [アクション] メニューを使用します。ペインは縮小または展開できます。

[詳細ビュー] ペイン

詳細ビューの各ペインには、重複なしの値/重複ありの値/NULL 値の数と割合、パターン、推測されたデータ型、推測されたデータドメイン、値、データプレビュー、およびリンクされたビジネス用語などのカラムプロパティが表示されます。

カラムまたはルールをクリックできる場合は、そのカラムまたはビューの詳細ビューが開きます。

次の表に、詳細ビューの各ペインを示します。

ペイン	説明
全般	NULL 値、重複なしの値、重複ありの値を含む行の数が異なる色で表示されます。値をパーセンテージで表示できます。連続するプロファイル実行の各実行における全般値の増減をスパークラインとして表示できます。スパークラインには、最新の連続した 5 回のプロファイル実行について、NULL 値、重複なしの値、または重複ありの値の数の変動が折れ線グラフとして表示されます。スパークラインの各プロファイル実行にポインタを移動すると、値の数と割合を参照できます。カラムにはタグとコメントを追加できます。
パターン	カラム値のパターンが表示されます。カラムにパターンの出現する頻度が、水平棒グラフとパーセンテージで表示されます。パターンをドリルダウンしたり、参照テーブルにパターンを追加したり、選択したパターンを持つデータドメインを作成したりできます。
データ型	カラムの推測されたデータ型が表示されます。カラムのデータ型の頻度は、水平棒グラフとパーセンテージとして表示されます。データ型にドリルダウンすることができ、推測されたデータ型のうち選択したものを承認、拒否、またはリセットすることができます。 【拒否済みを表示】 オプションには、拒否された推測データ型が表示されます。
データドメイン	カラムの推測データドメインが表示されます。適合する行、適合しない行、または NULL 値が含まれる行を対象に、データドメインでドリルダウンできます。データドメイン値を、承認、拒否、またはリセットできます。 【拒否済みを表示】 オプションには、拒否されたデータドメインが表示されます。データドメイン値を確認できます。
ビジネス用語	カラムの割り当て済みビジネス用語が表示されます。カラムに対してビジネス用語の割り当てまたは割り当て解除を実行できます。
値	カラムのすべての値が、頻度、長さ、パーセンテージと共にグラフィカルに表示されます。それぞれの値にドリルダウンすることができます。値の参照テーブルへの追加、値の頻度ルールを作成、データドメインの作成を実行できます。
データプレビュー	選択したパターン、データ型、データドメイン、または値について、ドリルダウンデータが表示されます。

統計

統計情報（値、パターン、データ型、データドメイン、プロファイルのカラムとルールの異常値）を表示できます。

プロファイル統計情報はサマリビューで、カラム統計情報はサマリビューと詳細ビューで表示できます。最新のプロファイル実行、履歴プロファイル実行、統合済みプロファイル実行の統計情報を表示できます。2 つのプロファイル実行のプロファイル結果を比較し、プロファイルとカラムの統計情報をサマリビューおよび詳細ビューで確認できます。

データプレビュー

[データプレビュー] ペインでは、選択したパターン、データ型、データドメイン、または値に対するドリルダウンデータを表示できます。

[データプレビュー] ペインは詳細ビューで表示することが可能です。サマリビューでカラムをクリックすると、詳細ビューが表示され、[データプレビュー] ペインは折りたたまれます（デフォルト設定）。カラムデータを表示するには、[アクション] > [プレビューの表示] をクリックします。

以下の表に、[データプレビュー] ペインの [アクション] メニューに含まれるオプションを示します。

オプション	説明
フィルタに追加	プロファイル結果のサブセットでデータの不整合を分析できるように、ドリルダウンデータをフィルタするためのドリルダウンフィルタを作成します。
フィルタの保存	ドリルダウンフィルタを保存します。
プレビューの表示	ソース行を表示します。
データのエクスポート	ドリルダウン結果を CSV ファイルまたは Microsoft Excel ファイルにエクスポートします。

データ型

データ型には、プロファイル結果の各カラムの推測されたデータ型がすべて含まれています。

データ型は、サマリビューと詳細ビューで確認できます。サマリビューでは、文書化されたデータ型と推測されたデータ型を表示できます。[競合しているデータ型] フィルタは、文書化されたデータ型と推測されたデータ型が一致しないカラムを表示します。詳細ビューでは、カラムの推測されたデータ型を表示できます。カラムのデータ型の頻度は、水平棒グラフとパーセンテージとして表示されます。選択した推測データ型をドリルダウン、承認、拒否、またはリセットできます。[拒否済みを表示] オプションには、拒否された推測データ型が表示されます。

次の表に、データ型のプロパティを示します。

プロパティ	説明
データ型	プロファイルのカラムの文書化されたデータ型と推測されたデータ型の一覧を表示します。
頻度	カラムに特定のデータ型が出現する回数（数値）を表示します。
割合	カラムに特定のデータ型が出現するパーセンテージを表示します。

プロパティ	説明
ドリルダウン	<p>カラムのデータ型に基づいて特定のソース行にドリルダウンします。</p> <p>注: 推測されたデータ型を複数選択した場合、ドリルダウンアクションは実行できません。</p>
ステータス	<p>データ型のステータスを示します。ステータスは推測済み、承認済み、または拒否済みです。</p> <p>推測済み</p> <p>カラムのデータ型が、Analyst ツールによって推測されたものであることを示します。</p> <p>承認済み</p> <p>カラムのデータ型が承認済みであることを示します。データ型を承認すると、そのデータ型がモデルリポジトリにコミットされます。</p> <p>拒否済み</p> <p>カラムのデータ型が拒否されたことを示します。</p>

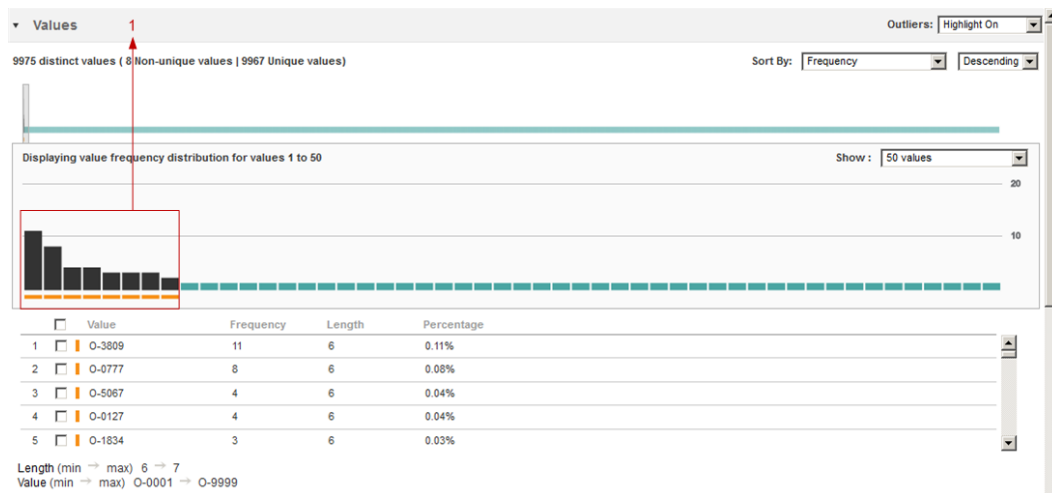
異常値

異常値とは、プロファイル結果の特定のカラムで、期待される値の範囲に収まらないパターン、値、頻度のことです。

データ統合サービスのプロファイリングプラグインは、カラムの大多数の値の範囲内に収まらない値を識別するアルゴリズムを実行します。特定のカラムに期待される大多数の値の範囲内に収まらないすべてのパターン、値、または頻度は異常値となります。

デフォルトでは、Analyst ツールはプロファイル結果の異常値を特定しません。サマリビューで、異常値を実行し、異常値結果を表示できます。パターンの異常値フィルタを使用すると、カラムのパターンに基づいて異常値が表示されます。値頻度の異常値フィルタを使用すると、カラムの値または頻度に基づいて異常値が表示されます。異常値の検出はバックグラウンドで実行されるため、サマリビューで他のアクションを実行できます。

詳細ビューで、リストから **【ハイライトをオン】** オプションを選択すると、[値] ペインに異常値を表示できます。異常値は、オレンジ色の下線が付いた垂直バーとして表示されます。異常値のみを表示するには、リストから **【フィルタ】** オプションを選択する必要があります。



1. 異常値。異常値は、オレンジ色の下線が付いた垂直バーとして表示されます。

異常値の実行

異常値を実行すると、カラムの期待される値の範囲に収まらないパターン、値、または頻度を特定できます。

1. サマリビューで、**【アクション】** > **【異常値の検出】** をクリックします。
【フィルタ基準】 ペイン内のパターン異常値と値頻度の異常値が「N/A」から、検出された異常値の数に変わります。
2. **【フィルタ基準】** ペインで、**【パターン異常値】** をクリックします。
サマリビューに、パターン異常値が出現するカラムが表示されます。
3. **【フィルタ基準】** ペインで、**【値頻度の異常値】** をクリックします。
サマリビューに、値または頻度の異常値が出現するカラムが表示されます。
4. 詳細ビューで、**【異常値】** ドロップダウンリストから **【ハイライトをオン】** を選択します。
【値】 ペインで、異常値は、オレンジ色の下線が付いた垂直バーとして表示されます。
5. **【異常値】** ドロップダウンリストで **【フィルタ】** をクリックすると、異常値のみが表示されます。

パターン

カラム値のパターンおよびパターンが出現する頻度を、サマリビューおよび詳細ビューで確認できます。

サマリビューでは、カラムの複数のパターンを水平棒グラフで表示できます。マウスポインタを棒グラフの上に移動すると、パターンの特性やカラムに出現する類似パターンの数をパーセンテージとして確認できます。詳細ビューでは、カラムにパターンが出現する頻度を、水平棒グラフとパーセンテージで表示できます。パターンをドリルダウンして参照テーブルにパターンを追加したり、選択したパターンを持つデータドメインを作成できます。

プロファイリングウェアハウスには、一意の最高頻度値が 16,000 個まで格納されます。これには、プロファイル結果の NULL 値もデフォルトで含まれます。プロファイル結果に NULL 値が 1 つ以上存在する場合、Analyst ツールでは NULL 値をパターンとして表示することができます。

注: Analyst ツールでは、精度が 38 より大きい数値カラムのパターンを導出することはできません。また、精度が 255 より大きい文字列カラムのパターンを導出することはできません。

以下の表に、カラムパターンのプロパティを示します。

プロパティ	説明
パターン	プロファイルのカラムのパターンが表示されます。
頻度	カラムでパターンが出現する回数（数値）が表示されます。
割合	カラムに特定のパターンが出現するパーセンテージが表示されます。

以下の表に、パターン文字と、各パターン文字が何を表わすかを示します。

Character	説明
「B」または「b」 または「」	空白（ブランクスペース）を表します。
「C」または「c」	任意の文字を表します。
「L」または「l」	任意の小文字の英字を表します。
「T」または「t」	タブを表します。
「U」または「u」	任意の大文字の英字を表します。
9	任意の数字を表します。Informatica Analyst は、最高 3 文字を「9」形式で別々に表示します。4 文字以上の場合、括弧内に値として示されます。例えば、「9 (8)」というフォーマットは 8 桁の数値を表します。
「X」または「x」	任意の英字を表します。Informatica Analyst は、最高 3 文字を「X」形式で別々に表示します。4 文字以上の場合、括弧内に値として示されます。例えば、「X (6)」というフォーマットは、値「Boston」に相当する場合があります。 注: パターン文字 X は、大文字と小文字を区別しません。このため、ソースデータ内の大文字または小文字に相当する可能性があります。
「P」または「p」	「(」(開き括弧)を表します。
「Q」または「q」	「)」(閉じ括弧)を表します。

注: カラムパターンには特殊文字を含めることができます。例えば、~、[,、=、-、?、=、{、*、-、>、<、\$などの特殊文字を使用できます。

値

カラムの値とその値がカラムに出現する頻度を表示できます。

サマリビューで、特定のカラムの最小値と最大値を確認できます。詳細ビューでは、カラムの値プロパティを確認できます。

サマリビューの値

サマリビューでは、最新のプロファイル実行、履歴プロファイル実行、統合済みプロファイル実行について、すべてのカラムとルール の最小値と最大値を確認できます。

例

小売店データベースの Employee テーブルに Employee という名前のカラムがあり、100～250 の従業員 ID、および名前（Bob や Robert など）が格納されています。Employee テーブルに対してカラムプロファイルを実行すると、サマリビューに従業員 ID の値カラムとして、100 --> Robert が表示されます。

詳細ビューの値

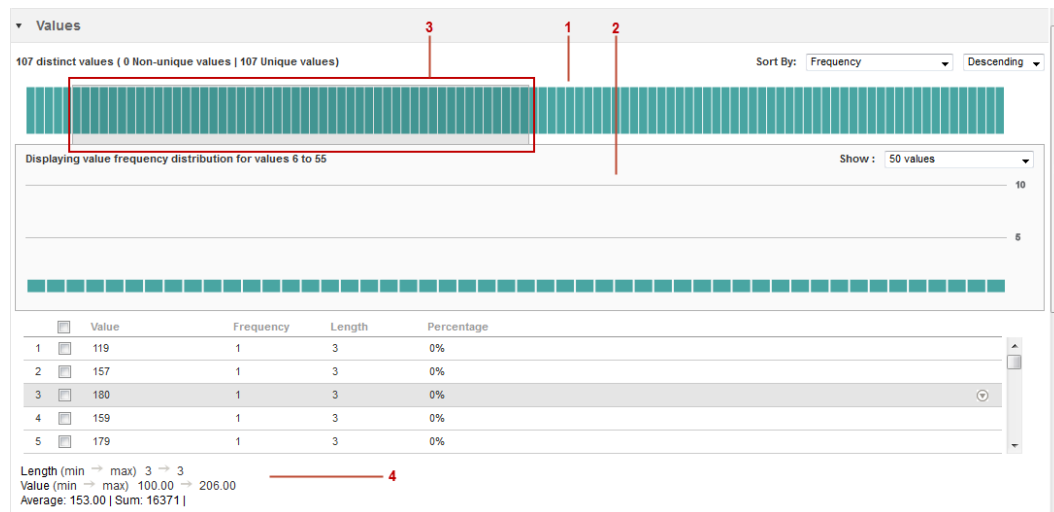
詳細ビューのカラム値には、カラムの値とその値がカラムに出現する頻度が表示されます。

【値】 ペインに、カラム値が視覚的に表示されます。各値の頻度、長さ、パーセンテージを確認できます。値は、値または頻度に基づいてソートできます。データのドリルダウン、値の参照テーブルへの追加、値の頻度ルール作成、データドメインの作成を実行できます。NULL 値は赤い縦棒、値の頻度は黒い縦棒、異常値はオレンジ色に強調表示された縦棒として表示されます。異常値の強調表示、異常値の無効化、またはカラム内の異常値だけを表示するための結果のフィルタを行えます。

【値】 ペインは、グラフィカルなレイアウトと値セクションで構成されます。

グラフィカルレイアウトは 2 つのパネルに分かれています。

次の画像は、詳細ビューの【値】 ペインを示しています。



1. 上部パネル。値を垂直棒グラフとして表示できます。頻度と値を基準にして値をソートできます。昇順または降順で値をソートできます。異常値は、オレンジ色で強調表示された縦棒として表示できます。
2. 下部パネル。下部パネルのスライダに含まれる値を表示できます。ここでは、各値が縦棒で表されています。値のドリルダウン、値の参照テーブルへの追加、値の頻度ルール作成、値に対するデータドメインの作成を実行できます。一度に 50 個、75 個、または 100 個の値を表示するように選択できます。
3. スライダ。上部パネルの値の上でスライダをスライドさせることができます。下部パネルには、スライダ内の値が表示されます。
4. 値のプロパティ。値のプロパティセクションには、値とプロパティが表示されます。

次の表に、グラフィカルレイアウトの各パネルを示します。

パネル	説明
上部パネル	すべての値が垂直棒グラフで表示されます。上部パネルには、最大 16,000 個まで値を表示できます。スライダを使用して一連の値を表示することができます。
下部パネル	上部パネルで選択したバッチの各値が表示されます。デフォルトでは、Analyst ツールによって、50 個の値が表示されます。1 度に 75 個または 100 個の値を表示するよう選択できます。

次の表に、値セクションに表示されるカラム値のプロパティを示します。

プロパティ	説明
値	上部パネルで選択したバッチの値のリストが表示されます。 注: カラム値の CLOB、BLOB、Raw、およびバイナリの各データ型は、Analyst ツールによって除外されます。
頻度	カラムで値が出現する回数（数値）が表示されます。
長さ	カラム値の長さが表示されます。
割合	カラムに特定の値が出現するパーセンテージが表示されます。

次の表で、選択したカラムの統計について説明します。

統計	説明
長さ（最小～最大）	カラムの最短値と最長値の長さが表示されます。
値（最小～最大）	カラムの最小値と最大値が表示されます。
平均	カラムの値の平均が表示されます。
合計	カラム内のすべての値の合計が表示されます。

詳細ビューの値によるプロファイル結果の比較

詳細ビューの「値」ペインには、プロファイル結果を比較できるように、値のプロパティとして、重複しない値の数、値の最小値、最大値、最小長および最大長、平均、標準偏差、および合計が表示されます。

カラムの詳細ビューには、プロファイル結果を比較できるように、値のプロパティ、値、値の頻度と水平棒グラフが表示されます。

次の表に、2 つのプロファイル実行の結果を比較するとき、詳細ビューに表示されるカラム値のプロパティを示します。:

プロパティ	説明
重複しない値の数	カラムの重複しない値の数が表示されます。
最小値	カラムの最小値が表示されます。
最大値	カラムの最大値が表示されます。

プロパティ	説明
長さ（最小～最大）	カラムの最短値と最長値の長さが表示されます。
平均	カラムの値の平均が表示されます。
標準偏差	カラムのすべての値について、カラム値の標準偏差（カラム値のばらつき）が表示されます。
合計	カラム内のすべての値の合計が表示されます。

プロファイル実行のタイプ

最新のプロファイル実行、履歴プロファイル実行、統合済みプロファイル実行のプロファイル結果を確認できます。プロファイル実行結果は、サマリビューで確認できます。

最新のプロファイル実行

サマリビューで、特定のプロファイルの最新のプロファイル実行の結果を確認できます。

次の操作を行うと、最新のプロファイル実行のプロファイル結果をサマリビューに表示できます。

- プロファイルを作成、保存して、実行する。
- 以前実行したプロファイルを【ライブラリ】ワークスペースから開く。
- 統合済みプロファイル実行のサマリビューまたは詳細ビューで【**最後のプロファイルの実行**に戻る】リンクをクリックする。
- 履歴プロファイル実行のサマリビューまたは詳細ビューで【**最後のプロファイルの実行**に戻る】リンクをクリックする。
- 【**プロファイル実行の選択**】ダイアログボックスで最新のプロファイル実行を選択し、【OK】をクリックする。

履歴プロファイル実行

サマリビューで、以前のプロファイル実行の結果を確認できます。

プロファイリングウェアハウスには、プロファイルのすべてのプロファイル実行の結果が保存されています。以前実行したプロファイル実行の結果を表示するには、【プロファイル実行の選択】ダイアログボックスでプロファイル実行を選択します。

統合済みプロファイル実行

サマリビューで、プロファイルの各カラムについて最新のプロファイル結果を確認できます。

統合済みプロファイル実行では、プロファイルの各カラムについて最新の結果を確認できます。【**プロファイル実行の選択**】ダイアログボックスで統合済みプロファイル実行を選択すると、プロファイリングウェアハウスによって特定のプロファイルのすべてのプロファイル実行から最新のカラム結果が取得されます。サマリビューで結果を確認できます。サマリビューのヘッダーには差分プロファイル実行が表示されます。

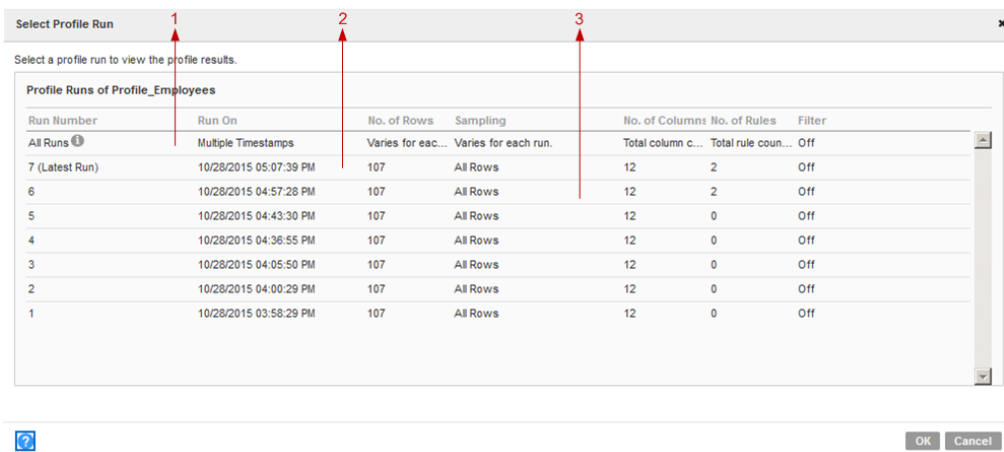
例

データアナリストは、プロファイルの各カラムの最新の結果を確認できます。例えば、プロファイル実行 A にはカラム 1、2、および 3 を選択し、プロファイル実行 B にはカラム 3、4、および 5 を選択します。このとき、すべてのカラムの最新の結果を表示するには、[プロファイル実行の選択] ダイアログボックスで統合済みプロファイル実行を選択します。サマリビューには、実行 A のカラム 1 と 2 の結果、および実行 B のカラム 3、4、5 の結果が表示されます。

プロファイル実行の選択

プロファイル結果を表示する際には、履歴プロファイル実行、最新のプロファイル実行、または統合済みのプロファイル実行を選択できます。プロファイル結果はサマリビューで、カラム結果は詳細ビューで確認できます。

1. **【ライブラリ】** ワークスペースでプロファイルを含むプロジェクトまたはフォルダを選択するか、**【アセット】** ペインでプロファイルを選択します。
2. **【アクション】** > **【開く】** をクリックしてプロファイルを開きます。
サマリビューが **【検出】** ワークスペースに表示されます。
3. サマリビューで、**【アクション】** > **【プロファイル実行の選択】** をクリックします。
【プロファイル実行の選択】 ダイアログボックスが表示されます。
次の画像は、**【プロファイル実行の選択】** ダイアログボックスを示しています。



1. 統合済みプロファイル実行このプロファイル実行を選択すると、サマリビューで、各カラムの最新のプロファイル結果を確認できます。
2. 最新のプロファイル実行このプロファイル実行を選択すると、サマリビューで、特定のプロファイルの最新のプロファイル結果を確認できます。
3. 履歴プロファイル実行このプロファイル実行を選択すると、サマリビューで、以前のプロファイルの履歴プロファイル結果を確認できます。
4. **【プロファイル実行の選択】** ダイアログボックスで、いずれかのプロファイル実行を選択してそのプロファイル結果を表示します。
 - 最新のプロファイル実行のプロファイル結果を表示するには、最新のプロファイル実行を選択して、**【OK】** をクリックします。
 - 履歴プロファイル実行のプロファイル結果を表示するには、最新以外のプロファイル実行を選択して、**【OK】** をクリックします。
 - 統合済みプロファイル実行のプロファイル結果を表示するには、**【すべての実行】** を選択し、**【OK】** をクリックします。サマリビューに、各カラムの最新のプロファイル結果が表示されます。

Analyst ツールによってプロファイルが実行され、サマリビューにプロファイル結果が表示されます。

5. サマリビューでカラムをクリックすると、カラム結果が表示されます。
詳細ビューが表示されます。

複数のプロファイル結果の比較の概要

2つのプロファイル実行のプロファイル結果を比較できます。サマリビューで比較結果を、詳細ビューでカラム結果を確認できます。

サマリビューでは、両方のプロファイル実行のすべてのカラムの比較結果を確認できます。

複数のプロファイル結果の比較

2つのプロファイル実行を比較した場合は、サマリビューで、プロファイル結果の比較を表示できます。

1. サマリビューで、[アクション] > [プロファイル実行の比較] をクリックします。

次の画像は、[プロファイル実行の比較] ダイアログボックスを示しています。

Compare Profile Runs

Select the profile runs that you want to compare.

Select a profile run

Run Number	Run On	No. of Rows	Pr. Sampling	No. of Columns	No. of Rules	Pr. Filter
3 (Latest Run)	09/09/2015 02:47:10 PM	27	No Sampling	4	0	Off
2	09/09/2015 02:45:24 PM	27	No Sampling	4	0	Off
1	09/07/2015 01:45:48 PM	27	No Sampling	4	0	Off

Select another profile run

Run Number	Run On	No. of Rows	Pr. Sampling	No. of Columns	No. of Rules	Pr. Filter
3 (Latest Run)	09/09/2015 02:47:10 PM	27	No Sampling	4	0	Off
2	09/09/2015 02:45:24 PM	27	No Sampling	4	0	Off
1	09/07/2015 01:45:48 PM	27	No Sampling	4	0	Off

OK Cancel

1. 実行 A。プロファイル実行を実行 A として選択します。
2. 実行 B。プロファイル実行を実行 B として選択します。
[プロファイル実行の比較] ダイアログボックスが表示されます。
2. [実行 A] ペインでプロファイルを選択し、[実行 B] ペインで別のプロファイルを選択します。
3. [OK] をクリックします。
サマリビューにプロファイル結果の統合ビューが表示されます。

プロファイル結果の比較のサマリビュー

2つのプロファイル実行の比較時に、サマリビューに結果をグリッド形式で表示できます。サマリビューでデフォルトのフィルタを使用して、特定の統計を表示できます。

次の画像は、サマリビューに表示された、2つのプロファイル実行のプロファイルの比較結果を示しています。

No	Name	Null	Unique	Non Unique	Pattern	Data Type	Data Domain
1	DEPARTMENT_ID	—	—	—	—	—	—
2	DEPARTMENT_NAME	—	—	—	—	—	▲ 1
3	LOCATION_ID	—	—	—	—	—	—
4	MANAGER_ID (Removed)	—	—	—	—	—	—

1. デフォルトフィルタ。サマリビューでは、デフォルトフィルタを使用してプロファイル比較結果を表示できます。
2. プロファイルヘッダー。ヘッダーでプロファイル名を表示できます。
3. サマリビューのヘッダー。サマリビューのヘッダーにはプロファイル固有の情報を表示できます。プロファイル実行の比較、プロファイル実行間の行の増減、プロファイルの行数、およびプロファイル実行の日時を表示できます。
4. サマリビュー。2つのプロファイル実行のカラム間の比較を表示できます。

プロファイル結果の比較のためのサマリビューのプロパティ

プロファイル結果の比較のためのサマリビューのプロパティとして、重複なしの値、重複ありの値、NULL 値、パターン、推測されたデータ型、推測されたデータドメイン、リンク済みビジネス用語のそれぞれの数とパーセンテージがあります。サマリビューには、プロパティを視覚的に表したものが表示されます。各サマリのプロパティをクリックすると、そのプロパティの値に基づいてソートできます。

サマリビューでは、データ統合サービスによって、すべてのカラムおよびルールに昇順で数値が割り当てられます。

注: 数値カウントに付いている上矢印は、2つのプロファイル実行を比較したとき一方のプロパティの値が増加したことを示します。数値カウントに付いている下矢印は、2つのプロファイル実行を比較したとき一方のプロパティの値が減少したことを示します。

次の表に、プロファイル結果の比較のためのサマリプロパティを示します。

プロパティ	説明
×	カラムまたはルールが表示されます。
名前	プロファイルのカラムまたはルールの名前が表示されます。

プロパティ	説明
Null	NULL 値の増減が表示されます。
重複なし	重複なしの値の増減が表示されます。
重複あり	重複ありの値の増減が表示されます。
パターン	複数のプロファイル実行間のパターンのバリエーションが表示されます。
データ型	2つのプロファイル実行のカラムまたはルールについて、推測されたデータ型のバリエーションが表示されます。
データドメイン	2つのプロファイル実行のカラムまたはルールに関連付けられた推測データドメインのバリエーションが表示されます。

サマリビューのプロファイル結果の比較で使用するデフォルトフィルタ

サマリビューでは、デフォルトフィルタを使用してプロファイル結果を表示できます。

サマリビューでソースカラムと仮想カラムを表示できます。ルールの出力は、サマリビューに仮想カラムとして表示されます。ルールの出力ポートを変更し、プロファイル実行と履歴実行を比較すると、履歴ルールの出力カラムは【**削除されたルール**】フィルタに表示され、新規ルールの出力カラムは【**追加されたルール**】フィルタに表示されます。1つの出力ルールのルールロジックを変更した場合、またはプロファイル実行で複数のルール出力の入力を変更して履歴実行と比較した場合、【**追加されたルール**】および【**削除されたルール**】のフィルタ出力は変更されません。フィルタはカラムの名前変更のみをフィルタに対する有効な入力とみなすため、フィルタ出力は変更されません。

次のデフォルトのフィルタオプションを使用し、特定の条件を満たすプロファイル結果を表示できます。

デフォルトフィルタオプション	説明
すべてのカラムとルール	ソースカラム、仮想カラム、およびルールカラムのプロファイル結果を表示します。ソースカラムとルールカラムを展開および縮小して結果を表示できます。
すべてのカラム	ソースカラムと仮想カラムのプロファイル結果を表示します。
共通カラム	両方のプロファイル実行結果で利用できるカラムを表示します。
追加されたカラム	最新のプロファイル実行で使用可能なカラムを表示します。例えば、実行5と実行3を比較する場合、【 追加されたカラム 】には、実行3ではなく実行5で使用可能なカラムが表示されます。
削除されたカラム	履歴プロファイル実行で使用可能なカラムを表示します。例えば、実行5と実行3を比較する場合、【 削除されたカラム 】には、実行5ではなく実行3で使用可能なカラムが表示されます。
すべてのルール	すべてのルールカラムのプロファイル結果を表示します。

デフォルトフィルタオプション	説明
追加されたルール	最新のプロファイル実行で使用可能なルールを表示します。例えば、実行 5 と実行 3 を比較する場合、[追加されたルール] には、実行 3 ではなく実行 5 で使用可能なルールが表示されます。
削除されたルール	履歴プロファイル実行で使用できるルールを表示します。例えば、実行 5 と実行 3 を比較する場合、[削除されたルール] には、実行 5 ではなく実行 3 で使用可能なルールが表示されます。

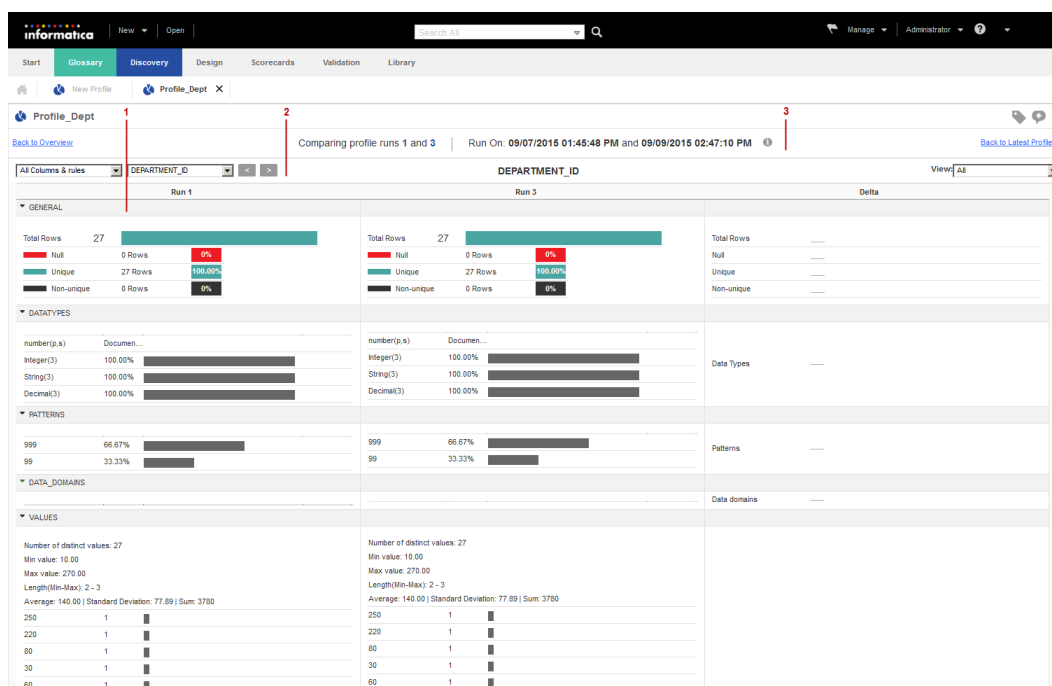
サマリビューには、すべてのソースカラムおよび仮想カラムのプロファイル結果がデフォルトで表示されます。

プロファイル結果の比較の詳細ビュー

カラム結果は、詳細ビューにグリッド形式で表示されます。カラムの詳細には、重複なしの値、重複ありの値、NULL 値、パターン、データ型、データドメイン、ビジネス用語、値、データプレビューなどの全般的な情報が含まれます。

カラムの詳細ビューは、カラム名をクリックすると表示されます。実行 A と実行 B のカラム結果は別のカラムとして表示できます。データの比較は差分カラムに表示されます。

次の画像は、詳細ビューに表示された特定のカラムのプロファイル結果の比較を示しています。



1. ペイン。2つのプロファイル実行のカラムのプロファイル結果と統計、および2つのプロファイル実行のカラムの差分情報を各ペインに表示できます。
2. プロファイルヘッダー。ドロップダウンリストでカラムを選択するか、ナビゲーションボタンを使用して、カラム結果を表示できます。カラム名を表示できます。また、[表示] ドロップダウンリストのオプションを使用して特定の結果を表示することもできます。
3. サマリビューのヘッダー。サマリビューのヘッダーにはプロファイル固有の情報を表示できます。比較しているプロファイル実行と、プロファイル実行の日時を表示できます。

プロファイル結果の比較のための [詳細ビュー] ペイン

詳細ビューには、2つのプロファイル実行について、特定のカラムの詳細なプロファイル結果と比較結果が表示されます。

詳細ビューでは、実行 A と実行 B のカラム結果が表示され、差分カラムにデータの比較が表示されます。他のカラム結果を表示するには、[フィルタ] ドロップダウンリストからフィルタを選択するか、[カラム] ドロップダウンリストからカラムを選択します。

カラムプロファイルのドリルダウン

カラムプロファイルのドリルダウンオプションを使用すると、カラム値に基づいてデータソース内の特定の行にドリルダウンできます。ドリルダウンの対象として、データソース内の現在のデータを読み取るか、プロファイルウェアハウスにステージングされたプロファイルデータを読み取るかを選択できます。ステージングされたプロファイルデータの特定の行にドリルダウンする場合は、Analyst ツールによって、一致するカラムの値に絞り込むためのドリルダウンフィルタが作成されます。ドリルダウンの実行後、ドリルダウンフィルタの編集、再呼び出し、リセット、および保存ができます。

プロファイリング対象として選択していないカラムも、ドリルダウン対象のカラムとして選択できます。ドリルダウンの対象として、データソース内の現在のデータを読み取るか、プロファイリングウェアハウスにステージングされたプロファイルデータを読み取るかを選択できます。カラム値のドリルダウンを実行した後に、選択した値またはパターンのドリルダウンデータを、任意の場所の CSV ファイルにエクスポートすることができます。Informatica Analyst にはドリルダウンデータの最初の 200 個しか値が表示されませんが、CSV ファイルにはすべての値がエクスポートされます。

行データのドリルダウン

プロファイルを実行した後に、カラム値、データ型、またはパターンに一致する特定の行にドリルダウンすることができます。

1. プロファイルを実行します。
サマリビューにプロファイル結果が表示されます。
2. サマリビューでカラム名をクリックします。
詳細ビューにカラム結果が表示されます。
3. 詳細ビューで [値] ペインの値を右クリックし、[ドリルダウン] を選択します。
[データプレビュー] ペインにドリルダウンデータが表示されます。

ドリルダウンデータへのフィルタの適用

プロファイル結果のサブセットのデータの不整合を分析できるように、ドリルダウンデータを繰り返しフィルタすることができます。

1. [値] タブでカラム値を選択します。
2. 右クリックして [ドリルダウン] を選択します。
[データプレビュー] ペインにドリルダウン結果が表示されます。
3. フィルタ条件を追加するには、[データプレビュー] ペインでカラム値を右クリックして、[フィルタに追加] を選択します。
フィルタ条件を示す [ドリルダウンフィルタ] ダイアログボックスが表示されます。

4. 必要なフィルタ条件を追加して、[OK] をクリックします。
推測されるデータ型にドリルダウンフィルタを適用できません。
5. フィルタを保存するには、[アクション] > [フィルタの保存] をクリックします。
6. ドリルダウンフィルタをクリアするには、[アクション] > [更新] をクリックします。
7. ドリルダウンデータを Microsoft Excel スプレッドシートにエクスポートするには、[アクション] > [データのエクスポート] をクリックします。

Analyst ツールでのキュレーション

キュレーションとは、データソースで検出されたメタデータを検証し管理することで、使用とレポートに適したメタデータにするプロセスです。Analyst ツールでメタデータをキュレーションする場合、プロファイル結果の中の推測されたデータ型またはデータドメインを承認、拒否、リセットすることができます。

1つのカラムに対して1つのデータ型およびデータドメインを承認できます。カラムの拒否されたデータ型またはデータドメインは、非表示にできます。推測されたデータ型またはデータドメインを承認または拒否した後も、データ型またはデータドメインをリセットすると、ステータスを推測済みに戻すことができます。

データ型とデータドメインの承認

プロファイル結果には、データソースの各カラムについて、推測されたデータ型とデータドメインが含まれています。Analyst ツールで、各カラムについて、1つのデータ型と1つのデータドメインを選択および承認できます。

1. プロファイルを実行します。
サマリビューにプロファイル結果が表示されます。
2. サマリビューでカラム名をクリックします。
詳細ビューにカラム結果が表示されます。
3. 詳細ビューの【データ型】ペインでデータ型を選択するか、【データドメイン】ペインでデータドメインを選択します。
4. [アクション] > [承認] をクリックします。
5. データ型またはデータドメインのステータスを推測済みに戻すには、データ型またはデータドメインを選択して、[アクション] > [リセット] をクリックします。

データ型とデータドメインの拒否

詳細ビューで、データ型またはデータドメインを拒否できます。拒否されたデータ型およびデータドメインの表示/非表示を切り替えることができます。

1. プロファイルを実行します。
サマリビューにプロファイル結果が表示されます。
2. サマリビューでカラム名をクリックします。
詳細ビューにカラム結果が表示されます。
3. 詳細ビューの【データ型】ペインでデータ型を選択するか、【データドメイン】ペインでデータドメインを選択します。

4. **【アクション】** > **【拒否】** をクリックします。
Analyst ツールによって、拒否されたデータ型がデータ型リストから削除されます。
5. 拒否されたデータドメインを表示するには、**【アクション】** > **【拒否済みを表示】** をクリックします。

Informatica Analyst のカラムプロファイルのエクスポートファイル

プロファイル結果の一部と、完全な結果サマリのどちらを選択するかに基づいて、カラムプロファイル結果を CSV ファイルまたは Microsoft Excel ファイルにエクスポートすることができます。

選択した値およびパターンについて、値の頻度、パターンの頻度、データ型、またはドリルダウンデータを CSV ファイルにエクスポートすることができます。すべてのカラムのプロファイリング結果のサマ리를 Microsoft Excel ファイルにエクスポートすることができます。データ統合サービスの特権「**結果のドリルダウンおよびエクスポート**」を使用して、プロファイル結果のエクスポートを誰に許可するかをユーザーまたはグループ単位で決定します。

CSV ファイルのプロファイルエクスポート結果

値の頻度、パターンの頻度、データ型、ドリルダウンデータをエクスポートして、1 つのファイルでデータを確認することができます。Analyst ツールは CSV ファイル内の情報を保存します。

推測されるカラムパターンをエクスポートすると、Analyst ツールは異なる形式のカラムパターンをエクスポートします。例えば、推測されるカラムパターン X (5) をエクスポートすると、形式 XXXXX のカラムパターンが CSV ファイルに表示されます。

Microsoft Excel のプロファイルエクスポート結果

完全なプロファイル結果のサマ리를エクスポートすると、Analyst ツールは Microsoft Excel ファイルの複数のワークシートに情報を保存します。Analyst ツールにより、ファイルは「xlsx」形式で保存されます。

次の表に、エクスポートファイルの各ワークシートに表示される情報を示します。

タブ	説明
カラムプロファイル	プロファイルの実行後にサマリビューからエクスポートされるサマリ情報。これには、カラム名、ルール名、重複なしの値の数、NULL 値の数、推測されたデータ型、プロファイルの最終実行日時などが含まれます。
値	カラムとルールの値、および各カラムの値が出現する頻度。
パターン	プロファイルを実行したカラムとルールの値のパターン、およびそのパターンが出現する頻度。
データ型	カラムに対するすべてのデータ型、各データ型の頻度、パーセンテージ値、および [推測済み]、[承認済み]、[拒否済み] などのデータ型のステータス。

タブ	説明
統計	それぞれのカラムとルールに関する統計。これには、平均、長さ、上位の値、下位の値、標準偏差などが含まれます。
プロパティ	【プロパティ】ビューの情報。これには、プロファイルの名前、タイプ、サンプリングポリシー、行数が含まれます。

Informatica Analyst からのプロファイル結果のエクスポート

プロファイルの結果を「.csv」または「.xlsx」ファイルにエクスポートして、ファイルでデータを表示することができます。

1. **【ライブラリ】** ワークスペースで、プロファイルを含むプロジェクトまたはフォルダを選択します。
2. プロファイルをクリックして開きます。
サマリビューにプロファイル結果が表示されます。
3. サマリビューで、**【アクション】 > 【データのエクスポート】** をクリックします。
【ファイルへのデータのエクスポート】 ダイアログボックスが表示されます。
4. **【ファイルへのデータのエクスポート】** ダイアログボックスで、ファイル名を入力します。必要に応じて、デフォルトのファイル名を使用します。
5. **【すべて（サマリ、値、パターン、統計、プロパティ）】** または **【データドメイン検出結果】** を選択し、**【コードページ】** を選択します。 **【OK】** をクリックします。
データが Microsoft Excel スプレッドシートにエクスポートされます。
6. サマリビューでカラムをクリックします。
詳細ビューにカラム結果が表示されます。
7. 詳細ビューで、**【アクション】 > 【データのエクスポート】** をクリックします。
【ファイルへのデータのエクスポート】 ダイアログボックスが表示されます。
8. **【ファイルへのデータのエクスポート】** ダイアログボックスで、ファイル名を入力します。必要に応じて、デフォルトのファイル名を使用します。
9. 次のいずれかのオプションを選択します。
 - **【すべて】**（サマリ、値、パターン、統計、プロパティ）
 - **【選択されたカラムの値の頻度】**
 - **【選択されたカラムのパターンの頻度】**
 - **選択されたカラムのデータ型**
 - **選択された値でデータをドリルダウン**
 - **選択されたパターンでデータをドリルダウン**
 - **選択したデータ型のデータをドリルダウン**
10. ファイル形式を入力します。形式は、**【すべて】** オプションの場合は **【Excel】**、その他のオプションの場合は **【CSV】** です。フィールド名をファイルの先頭行としてエクスポートするように選択できます。
11. ファイルのコードページを選択します。
12. **【OK】** をクリックします。
データはファイルにエクスポートされます。

第 8 章

Informatica Analyst のスコアカード

この章では、以下の項目について説明します。

- [Informatica Analyst のスコアカードの概要, 64 ページ](#)
- [Informatica Analyst のスコアカードのプロセス, 65 ページ](#)
- [Informatica Analyst でのスコアカードの作成, 66 ページ](#)
- [既存のスコアカードへカラムの追加, 67 ページ](#)
- [スコアカードの実行, 68 ページ](#)
- [スコアカードの表示, 69 ページ](#)
- [スコアカードの編集, 69 ページ](#)
- [メトリック, 70 ページ](#)
- [\[メトリックグループ\], 71 ページ](#)
- [カラムのドリルダウン, 73 ページ](#)
- [傾向グラフ, 73 ページ](#)
- [Informatica Analyst のスコアカードエクスポートファイル, 76 ページ](#)
- [スコアカード通知, 78 ページ](#)
- [スコアカードリネージュ, 80 ページ](#)

Informatica Analyst のスコアカードの概要

スコアカードは、プロファイル内のカラムの有効値を視覚的に表したものです。スコアカードを作成して、実データまたはステージングされたデータをドリルダウンすることができます。

データ品質の進捗を測定するには、スコアカードを使用します。例えば、データ品質ルールを適用する前に、スコアカードを作成してデータ品質を測定することができます。データ品質ルールを適用した後に、別のスコアカードを作成して、データ品質に対するルールの影響を比較することができます。

スコアカードには、カラムの値の頻度がスコアとして表示されます。スコアにはカラムの有効値の割合が反映されます。プロファイルの実行後に、プロファイルからスコアカードにカラムをメトリックとして追加できます。関連し合うメトリックをグループ化する単一のエンティティとして、メトリックグループを作成できます。レコード内のカラムで許容できる不良データの範囲を指定するしきい値を定義し、メトリックごとにメトリックウェイトを割り当てます。スコアカードを実行すると、Analyst ツールは、各メトリックグループの加重平均を生成します。データ品質を詳細に評価するために、各メトリックに固定費または変動費を割り当てること

もできます。スコアカードを実行すると、Analyst ツールでメトリックごとに不良データのコストの合計を計算して合計コストを表示します。

スコアカードを作成または編集すると、ソースデータに基づくスコアカードフィルタを作成できます。スコアカードフィルタにより、フィルタ条件に基づくメトリックスコアを再計算できます。有効なデータレコードと無効なレコードを識別するには、各メトリックをドリルダウンします。傾向グラフを使用して、メトリックスコアおよびメトリックの不良データのコストが、特定の期間にわたって変化する様子を追跡できます。スコアカードでプロファイルフィルタを再利用できます。

Analyst ツールでバージョン管理システムが有効になっていると、複数のバージョンのスコアカードを作成し、スコアカードのバージョン履歴を表示できます。デフォルトでは、スコアカードは作成後にチェックアウトされます。他のユーザーがスコアカードを編集できるようにするには、スコアカードをチェックインする必要があります。

【スコアカード】 ワークスペースでスコアカードダッシュボードを表示できます。スコアカードダッシュボードでは、スコアカードのあるデータオブジェクト、プロジェクトのスコアカード、過去 6 か月間のスコアカード実行傾向、1 か月間に実行されたすべてのスコアカードに対する、正常、許容可能、および許容不可のメトリックの集計を表示できます。

Informatica Analyst で、スコアカードの電子メール通知を設定および管理できます。電子メールサービスを使用して、電子メール通知を管理します。電子メールサービスは、Informatica Administrator で設定できるシステムサービスです。

Informatica Analyst のスコアカードのプロセス

Developer tool および Analyst ツールでは、スコアカードを作成、編集できます。Analyst ツールではスコアカードを実行できます。スコアカードは、データオブジェクトの現在のデータに対して実行するか、プロファイルリングウェアハウスにステージングされているデータに対して実行することができます。

【スコアカード】 ワークスペースでスコアカードを表示できます。スコアカードを実行したら、**【スコアカード】** パネルでスコアを表示できます。スコアカード内のスコアからデータオブジェクトを選択し、そのデータオブジェクトに移動できます。データオブジェクトは別のタブで開きます。

スコアカードに関する以下の作業を実行できます。

1. Developer tool または Analyst ツールでスコアカードを作成し、プロファイルからカラムを追加します。
2. Analyst ツールでスコアカードを開きます。
3. プロファイルの実行後に、プロファイルカラムをメトリックとしてスコアカードに追加します。
4. 必要に応じて、ソースデータに基づくスコアカードフィルタを作成します。
5. オプションで、各メトリックの無効なデータのコストを設定します。
6. スコアカードを実行してカラムのスコアを生成します。
7. スコアカードを表示して、レコード内の各カラムのスコアを確認します。
8. スコアのカラムをドリルダウンします。
9. スコアカードを編集します。
10. スコアカードの各メトリックに対してしきい値を設定します。
11. グループを作成して、スコアカードにおいて関連し合うメトリックを追加または移動します。
12. 必要に応じて、グループを編集または削除します。
13. 各スコアのスコア傾向グラフを表示して、スコアの変化を時系列で監視します。
14. 必要に応じて、各メトリックのコストの傾向グラフを表示して、データ品質の値を監視します。

15. メトリックまたはメトリックグループごとにスコアカードリネージュを表示します。
16. 読み取りアクセス権限のあるスコアカードに関する統合情報を表示します。

Informatica Analyst でのスコアカードの作成

スコアカードを作成し、プロファイルからスコアカードにカラムを追加します。スコアカードにカラムを追加する前に、プロファイルを実行する必要があります。

1. **【ライブラリ】** ワークスペースで、プロファイルを含むプロジェクトまたはフォルダを選択します。
2. プロファイルをクリックして開きます。
プロファイル結果が **【検出】** ワークスペースのサマリビューに表示されます。
3. **【アクション】** > **【スコアカードに追加】** をクリックします。
【スコアカードに追加】 ウィザードが表示されます。
4. **【スコアカードに追加】** 画面で、新しいスコアカードを作成するように選択したり、既存のスコアカードを編集してカラムを定義済みのスコアカードに追加したりできます。デフォルトでは、**【新しいスコアカード】** オプションが選択されています。**【次へ】** をクリックします。
5. **【手順 2/8】** 画面で、スコアカードの名前を入力します。必要に応じて、スコアカードの説明を入力できます。スコアカードを保存するプロジェクトおよびフォルダを選択します。**【次へ】** をクリックします。
デフォルトでは、プロファイルで定義されているカラムとルールが選択されます。プロファイルに含まれていないカラムは追加できません。
6. **【手順 3/8】** 画面で、スコアカードにメトリックとして追加するカラムとルールを選択します。必要に応じて、左側のカラムヘッダのチェックボックスをクリックしてすべてのカラムを選択します。必要に応じて、**【カラム名】** を選択してカラム名をソートします。**【次へ】** をクリックします。
7. **【手順 4/8】** 画面で、メトリックにフィルタを追加できます。
プロファイルに対して作成したフィルタをメトリックに適用したり、新しいフィルタを作成したりできます。**【メトリックフィルタ】** ペインでメトリックを選択し、**【フィルタの管理】** アイコンをクリックして、**【フィルタの編集: カラム名ダイアログボックス】**を開きます。**【フィルタの編集: カラム名ダイアログボックス】**で、以下のいずれかのタスクを実行するように選択できます。
 - プロファイル用に作成したフィルタを選択します。**【次へ】** をクリックします。
 - 既存のフィルタを選択します。**【編集】** アイコンをクリックし、**【フィルタの編集】** ダイアログボックスでフィルタを編集します。**【次へ】** をクリックします。
 - プラス (+) アイコンをクリックし、**【新規フィルタ】** ダイアログボックスでフィルタを作成します。**【次へ】** をクリックします。必要に応じて、選択したフィルタをスコアカードのすべてのメトリックに適用することもできます。
フィルタが **【メトリックフィルタ】** ペインに表示されます。
8. **【手順 4/8】** 画面で **【次へ】** をクリックします。
9. **【手順 5/8】** 画面で、**【メトリック】** ペインの各メトリックを選択し、次のタスクを実行します。
 - 有効な値を設定します。**【スコアで使用: 値】** ペインで、**【利用可能な値】** ペインの 1 つ以上の値を選択し、右矢印ボタンをクリックしてそれらの値を **【有効な値】** ペインに移動します。メトリックの有効な値の総数が **【利用可能な値】** ペインの上部に表示されます。
 - メトリックのしきい値を設定します。**【メトリックのしきい値】** ペインで、**【正常】**、**【許容可能】**、および **【許容不可】** スコアに対してしきい値を設定します。

- 無効なデータのコストを設定します。メトリックのコストに定数値を割り当てるには、**【固定費】**を選択します。数値カラムを変動費としてメトリックに添付するには、**【変動費】**を選択し、**【カラムの選択】**をクリックして数値カラムを選択します。必要に応じて、**【コスト単位の変更】**をクリックしてコストの単位を変更します。メトリックに対して無効なデータのコストを設定しない場合は、**【なし】**を選択します。
10. **【次へ】**をクリックします。
 11. **【手順 6/8】** 画面で、メトリックを追加できるメトリックグループを選択したり、新しいメトリックグループを作成したりできます。新しいメトリックグループを作成するには、グループアイコンをクリックします。**【次へ】**をクリックします。
 12. **【手順 7/8】** 画面で、グループ内のメトリックのウェイトと、グループのしきい値を指定します。
 13. **【手順 8/8】** 画面で、スコアカードを実行するためのランタイム環境として **【ネイティブ】**、**【Hadoop】**、または **【Databricks】** を選択します。Hadoop ランタイム環境では **【Blaze】** または **【Spark】** エンジンを選択できます。**【Databricks】** を選択した場合は、**【参照】** をクリックして、プロファイルを実行する Databricks 接続を選択します。**【Blaze】** または **【Spark】** エンジンを選択した場合は、**【参照】** をクリックして、プロファイルを実行する Hadoop 接続を選択します。
 14. **【保存】** をクリックしてスコアカードを保存するか、**【保存して実行】** をクリックしてスコアカードを保存してから実行します。
スコアカードが **【スコアカード】** ワークスペースに表示されます。

既存のスコアカードへカラムの追加

プロファイルの実行後に、既存のスコアカードにカラムを追加できます。

1. プロファイルをクリックして開きます。
サマリビューにプロファイル結果が表示されます。
2. カラムを選択します。**【アクション】** > **【スコアカードに追加】** をクリックします。
【スコアカードに追加】 ウィザードが表示されます。
注: スコアカードにカラムを追加するときは、以下のルールとガイドラインに従います。
 - カラム名とスコアカード名が同じ場合、スコアカードにカラムを追加することはできません。
 - カラム名を変更しても、同じカラムをスコアカードに 2 回追加することはできません。
3. **【既存のスコアカード】** を選択して定義済みのスコアカードにカラムを追加します。**【次へ】** をクリックします。
4. **【手順 2/7】** 画面で、カラムを追加するスコアカードを選択します。**【次へ】** をクリックします。
スコアカードに関連付けられている既存のメトリックとメトリックグループを確認できます。
5. **【手順 3/7】** 画面で、スコアカードにメトリックとして追加するカラムとルールを選択します。必要に応じて、左側のカラムヘッダのチェックボックスをクリックしてすべてのカラムを選択します。**【カラム名】** をクリックして、カラム名をソートします。**【次へ】** をクリックします。
6. **【手順 4/7】** 画面で、メトリック用のフィルタを作成できます。プロファイルに対して作成したフィルタをメトリックに適用することもできます。
7. **【手順 5/7】** 画面で、次のタスクを実行できます。
 - **【メトリック】** ペインで、他のペインの各メトリックを選択し、メトリック値を設定します。

- **【スコアで使用: 値】** ペインで、**【利用可能な値】** ペインの複数の値を選択し、右矢印ボタンをクリックして値を**【有効な値】** ペインに移動します。
メトリックの有効な値の総数が**【利用可能な値】** ペインの上部に表示されます。
 - **【メトリックのしきい値:]** ペインで、**【正常】**、**【許容可能】**、および**【許容不可】** スコアに対してしきい値を設定できます。
 - **【無効なデータのコスト】** では以下の操作が可能です。
 - 各メトリックを選択して、メトリックの無効なデータのコストを設定する。
 - **【固定費】** オプションを選択して、メトリックのコストに定数値を割り当てる。コストの単位を変更するには**【コスト単位の変更】** をクリックします。
 - **【変動費】** オプションを選択して、数値カラムを変動費としてメトリックに添付する。数値カラムを選択するには**【カラムの選択】** をクリックします。
8. **【次へ】** をクリックします。
 9. **【手順 6/7】** 画面で、次のタスクを実行できます。
 - メトリックを追加するメトリックグループを選択します。
 - **【デフォルト - メトリック】** ペインで、デフォルトのメトリックウェイト「0」をダブルクリックして値を変更できます。
 - **【メトリックのしきい値:]** ペインで、**【正常】**、**【許容可能】**、および**【許容不可】** スコアに対してしきい値を設定できます。
 10. **【次へ】** をクリックします。
 11. **【手順 7/7】** 画面で、ランタイム環境を選択します。
 12. **【保存】** をクリックしてスコアカードを保存するか、**【保存して実行】** をクリックしてスコアカードを保存してから実行します。

スコアカードの実行

スコアカードを実行してカラムのスコアを生成します。

1. **【アセット】** パネルで、実行するスコアカードを選択します。
2. スコアカードをクリックして開きます。
スコアカードが**【スコアカード】** ワークスペースに表示されます。
3. **【アクション】** > **【スコアカードの実行】** をクリックします。
4. **【メトリック】** ペインからスコアを選択し、**【カラム】** ペインからドリルダウンするカラムを選択します。
5. **【ドリルダウン】** オプションで、実データをドリルダウンするかステージングされたデータをドリルダウンするかを選択します。
最適なパフォーマンスを実現するには、実データをドリルダウンします。
6. **【実行】** をクリックします。

スコアカードの表示

スコアカードを実行して、各メトリックのスコアを表示します。スコアカードには、スコアが円グラフおよび棒グラフとして表示されます。有効なデータまたは無効なデータを表示します。また、データオブジェクトの名前、スコアカードの情報、メトリックグループスコア、およびメトリックウェイトを表示することもできます。

1. スコアカードを実行してスコアを表示します。
2. 表示するスコアが含まれているメトリックを選択します。
3. **【アクション】** > **【ドリルダウン】** をクリックして、カラムで有効なデータの行および無効なデータの行を表示します。

Analyst ツールにより、無効なデータの行が **【ドリルダウン】** セクションにデフォルト表示されます。

スコアカードの編集

スコアカードのメトリックの有効な値を編集します。スコアカードを編集する前に、そのスコアカードを実行する必要があります。

1. **【ライブラリ】** ワークスペースの **【アセット】** ペインで、編集するスコアカードをクリックします。
スコアカードが **【スコアカード】** ワークスペースに表示されます。
2. バージョン管理システムを有効にしている場合は、**【アクション】** > **【チェックアウト】** をクリックします。
3. **【アクション】** > **【編集】** > **【全般】** をクリックします。
【スコアカードの編集】 ダイアログボックスが表示されます。
4. **【全般】** タブで、必要に応じてスコアカードの名前と説明を編集できます。
5. **【メトリック】** タブをクリックします。
6. **【メトリック】** ペインでスコアを選択し、**【スコアで使用: 値】** ペイン内のすべての値のリストから有効な値を設定します。
7. **【メトリックしきい値】** ペインで、必要に応じてスコアしきい値を変更できます。
8. 各メトリックの無効なデータのコストを確認し、必要に応じて変更を加えます。
9. **【スコアカードフィルタ】** タブをクリックします。
10. フィルタを追加、編集、または削除できます。
11. **【メトリックグループ】** タブをクリックします。
12. メトリックグループを作成、編集、または削除できます。
また、**【メトリックグループ】** タブで、メトリックウェイトおよびメトリックしきい値を編集することもできます。
13. **【通知】** タブをクリックします。
14. 必要に応じて、スコアカード通知設定を変更できます。
メトリックおよびメトリックグループにグローバル設定およびカスタム設定を行うことができます。
15. ランタイム環境として、**【ネイティブ】**、**【Blaze】**、**【Spark】**、または **【Databricks】** を選択します。ランタイム環境に **【Blaze】** または **【Spark】** を選択した場合、**【参照】** をクリックして Hadoop 接続を選択します。**【Databricks】** を選択した場合は、**【選択】** をクリックして、Databricks 接続を選択します。

16. **【保存】** をクリックしてスコアカードを保存するか、**【保存して実行】** をクリックして変更を保存してからスコアカードを実行します。
17. **【チェックイン】** をクリックします。

メトリック

メトリックは、スコアカードに含まれるルール出力またはデータソースのカラムです。スコアカードを作成する際に、各メトリックにウェイトを割り当てることができます。スコアカード内の関連し合うメトリックを1つのまとまりとして分類するには、メトリックグループを作成します。

メトリックウェイト

スコアカードを作成する際に、各メトリックにウェイトを割り当てることができます。ウェイトのデフォルト値は0です。

スコアカードを実行すると、Analyst ツールは、ユーザーが各メトリックに割り当てたメトリックスコアに基づいて、各メトリックグループの加重平均を計算します。

例えば、メトリック M1 にウェイト W1、メトリック M2 にウェイト W2 を割り当てるとします。Analyst ツールは、次の式を使用して加重平均を計算します。

$$(M1 \times W1 + M2 \times W2) / (W1 + W2)$$

データ品質の値

ソースデータのデータ品質の測定値は、組織のデータアセット管理において重要な情報です。スコアカードに示されたメトリックの無効なデータのコストは、組織でソースデータのデータ品質を監視する際に、値を導出するのに役立ちます。データアナリストとして、値（通貨単位やカスタム単位など）をメトリックやメトリックグループに関連付けることができます。このような場合、スコアカードを実行して、ソースデータ内の無効なデータの合計コストを表示できます。

ビジネスニーズに合わせて、メトリックのコスト単位を定義できます。スコアカードの作成時または編集時に、各メトリックの変動費または固定費を設定することもできます。

固定費

固定費とは、スコアカード内のメトリックに割り当てた定数値です。事前定義されたコスト単位を選択するか、ビジネスニーズに合わせてカスタムのコスト単位を作成できます。

変動費

変動費は、データソースの数値カラムの値に基づいてメトリックに割り当てた値です。データ統合サービスでは、コストに割り当てたカラムまたは仮想カラムに基づいてメトリックの変動費を計算します。

例

ローン担当者として、顧客がローン返済額を提示できるように、顧客に支払記入帳を用意する必要があります。スコアカードを使用して顧客アドレスの精度を測定し、確実に支払記入帳を配送します。アドレス精度メトリックの月次支払額カラムに変動費を設定できます。スコアカードを実行して、顧客が月次額を期限内に支払わなかった場合にローン部門が失う合計コストを計算します。

しきい値の定義

スコアカードの各スコアに対してしきい値を設定できます。しきい値では、レコード内のカラムで許容できる不良データの範囲を割合で指定します。正常、許容可能、または許容不可の範囲のデータに対してしきい値を設定できます。スコアカードにカラムを追加するとき、またはスコアカードを編集するときに、各カラムに対してしきい値を定義できます。

スコアカードのカラムに対してしきい値を定義する前に、次の前提条件タスクのいずれかを完了します。

- **【スコアカードに追加】** ダイアログボックスで、プロファイルを開いて、プロファイルからのカラムをのスコアカードに追加します。
 - 必要に応じて、**【ライブラリ】** ワークスペース内のスコアカードをクリックし、**【アクション】** > **【編集】** を選択して、**【スコアカードの編集】** ダイアログボックスでスコアカードを編集します。
1. **【スコアカードに追加】** ダイアログボックスまたは **【スコアカードの編集】** ダイアログボックスの **【メトリック】** ペインで、各メトリックを選択します。
 2. **【メトリックのしきい値】** ペインで、許容不可範囲の上限および許容可能範囲の下限を示すしきい値を入力します。
しきい値は小数点以下2桁まで設定できます。
 3. **【次へ】** または **【保存】** をクリックします。

【メトリックグループ】

スコアカード内の関連し合うスコアを1つのまとまりとして分類するには、メトリックグループを作成します。デフォルトでは、すべてのスコアがデフォルトメトリックグループに分類されます。

メトリックグループを作成したら、デフォルトメトリックグループから別のメトリックグループにスコアを移動できます。デフォルトメトリックグループをはじめとするメトリックグループは、名前と説明を変更して、編集することができます。不要になったメトリックグループは削除できますが、デフォルトメトリックグループは削除できません。

メトリックグループの作成

メトリックグループを作成して、そのグループにスコアカード内の関連し合うスコアを追加します。

1. **【ライブラリ】** ワークスペースの **【アセット】** ペインで、編集するスコアカードをクリックします。
スコアカードが **【スコアカード】** ワークスペースに表示されます。
2. **【アクション】** > **【編集】** をクリックします。
【スコアカードの編集】 ウィンドウが表示されます。
3. **【メトリックグループ】** タブをクリックします。
デフォルトグループが **【メトリックグループ】** パネルに表示され、デフォルトグループのスコアが **【メトリック】** パネルに表示されます。
4. **【新規グループ】** アイコンをクリックし、メトリックグループを作成します。
【メトリックグループ】 ダイアログボックスが表示されます。
5. 名前を入力し、必要に応じて説明を入力します。
6. **【OK】** をクリックします。
7. **【保存】** をクリックして、変更内容をスコアカードに保存します。

メトリックグループへのスコアの移動

メトリックグループを作成した後で、関連し合うスコアをそのメトリックグループに移動できます。

1. **【ライブラリ】** ワークスペースの **【アセット】** ペインで、編集するスコアカードをクリックします。
スコアカードが **【スコアカード】** ワークスペースに表示されます。
2. **【アクション】** > **【編集】** をクリックします。
【スコアカードの編集】 ウィンドウが表示されます。
3. **【メトリックグループ】** タブをクリックします。
デフォルトグループが **【メトリックグループ】** パネルに表示され、デフォルトグループのスコアが **【メトリック】** パネルに表示されます。
4. **【メトリック】** パネルでメトリックを選択し、**【メトリックの移動】** アイコンをクリックします。
【メトリックの移動】 ダイアログボックスが表示されます。
注: 複数のスコアを選択するには、Shift キーを押しながら選択します。
5. スコアの移動先のメトリックグループを選択します。
6. **【OK】** をクリックします。

メトリックグループの編集

メトリックグループを編集して、名前と説明を変更します。デフォルトメトリックグループの名前も変更できます。

1. **【ライブラリ】** ワークスペースの **【アセット】** ペインで、編集するスコアカードをクリックします。
スコアカードが **【スコアカード】** ワークスペースに表示されます。
2. **【アクション】** > **【編集】** をクリックします。
【スコアカードの編集】 ウィンドウが表示されます。
3. **【メトリックグループ】** タブをクリックします。
デフォルトメトリックグループが **【メトリックグループ】** パネルに表示され、デフォルトグループのメトリックが **【メトリック】** パネルに表示されます。
4. **【メトリックグループ】** パネルでメトリックグループを選択し、**【グループの編集】** アイコンをクリックします。
【編集】 ダイアログボックスが表示されます。
5. 名前を入力し、必要に応じて説明を入力します。
6. **【OK】** をクリックします。

メトリックグループの削除

無効になったメトリックグループを削除できます。メトリックグループを削除するときに、メトリックグループ内のスコアをデフォルトメトリックグループに移動するように選択できます。デフォルトメトリックグループは削除できません。

1. **【ライブラリ】** ワークスペースの **【アセット】** ペインで、編集するスコアカードをクリックします。
スコアカードが **【スコアカード】** ワークスペースに表示されます。
2. **【アクション】** > **【編集】** をクリックします。
【スコアカードの編集】 ウィンドウが表示されます。

3. **【メトリックグループ】** タブをクリックします。
デフォルトメトリックグループが**【メトリックグループ】** パネルに表示され、デフォルトグループのメトリックが**【メトリック】** パネルに表示されます。
4. **【メトリックグループ】** パネルでメトリックグループを選択し、**【グループの削除】** アイコンをクリックします。
【グループの削除】 ダイアログボックスが表示されます。
5. メトリックグループを削除する前に、メトリックグループ内のメトリックを削除するオプション、またはメトリックをデフォルトメトリックグループに移動するオプションを選択します。
6. **【OK】** をクリックします。

カラムのドリルダウン

スコアに対応するカラムをドリルダウンして、有効なデータ行または無効なデータ行を表示したときに示されるカラムを選択します。ドリルダウン対象として選択したカラムは**ドリルダウン**パネルに表示されます。

1. スコアを表示するためにスコアカードを実行します。
2. 表示するスコアが含まれているカラムを選択します。
3. **【アクション】** > **【ドリルダウン】** をクリックして、カラムの有効なデータ行または無効なデータ行を表示します。
4. **【アクション】** > **【カラムのドリルダウン】** をクリックします。
選択したスコアに対応するカラムが**【ドリルダウン】** パネルに表示されます。デフォルトでは、Analyst ツールにはカラムの有効なデータ行が表示されます。無効なデータ行を表示する場合は、**【無効】** をクリックします。

傾向グラフ

傾向グラフを使用して、メトリックスコアおよびメトリックの無効なデータのコストが、特定期間にわたって変化する様子を監視します。

傾向グラフには、縦軸にスコアまたはコストの値を配置し、横軸にすべてのスコアカード実行を配置した、スコアグラフとコストグラフの両方が含まれます。デフォルトでは、傾向グラフには、最新の 10 のスコアカード実行で生成されたデータが表示されます。傾向グラフには、メトリックの合計行数と無効な行数を表示できます。傾向グラフには、スコアおよびコストの傾向が一定であるか、最後のスコアカード実行に基づいて上昇または下降しているのかも表示されます。

Analyst ツールでは、日付ごとの履歴スコアカード実行データと最新の有効なスコアの値を使用してスコアが計算されます。また、グラフの最新のしきい値設定を使用してスコアポイントの色が表示されます。スコアに対して、**【正常】**、**【許容可能】**、および**【許容不可】**のしきい値を表示できます。しきい値は、スコアカードでのスコアの値の編集後、スコアカードを実行するたびに更新されます。スコアカードをエクスポートすると、エクスポートファイル内のスコア情報とコスト情報を含め、Analyst ツールに傾向グラフの情報が含まれます。

スコアの傾向グラフ

スコアの傾向グラフは、複数のプロファイル実行にわたって、メトリックのスコアの変化の様子を視覚的に表したものです。スコアの傾向グラフでは、メトリックのスコアの値が縦軸に配置され、すべてのスコアカード実行が横軸に配置されます。

次の画像は、スコアの傾向グラフの例を示しています。



例

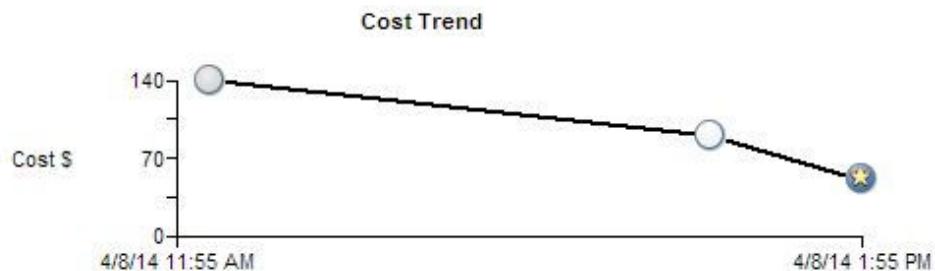
データアナリストはデータ品質を監視して、マッピングおよびその他のプロセスの変更によってデータ品質のスコアが上昇しているかどうかを分析できます。データ品質の変化を測定した後は、組織がデータ品質の変化を分析および使用できるようレポートできます。例えば、複数のスコアカード実行の終了で、社会保障番号カラム内の有効な値の割合が 84 から 90 に動いたとします。素早く分析するには、このデータ品質の変化を視覚的なグラフとしてレポートします。

コストの傾向グラフ

コストの傾向グラフは、複数のプロファイル実行にわたって、メトリックの無効なデータの変化の様子を視覚的に表したものです。コストの傾向グラフでは、組織内のデータ品質の影響を測定することができます。コストの傾向グラフでは、コストの値が縦軸に配置され、すべてのスコアカードの実行が横軸に配置されます。また、メトリックの無効なデータと有効な値の合計コストが、コストの傾向グラフの下グリッドに表示されます。

コスト傾向グラフは、high-value レコードでの無効なデータの影響を追跡するのに役立ちます。固定費を使用して無効なデータを計算した場合、high-value レコードでの無効なデータの影響を見落とす恐れがあります。傾向グラフに、複数のスコアカードの実行にわたるスコアの改善とコスト全体の減少が示されることが原因で、この問題が発生します。ただし、high-value レコードには、スコアカードに現れる少数のデータ品質の問題が存在する可能性があります。

次の画像は、コストの傾向グラフの例を示しています。



例

ある金融機関には、多額の銀行預金額と投資額（\$1,000 万など）を持つ残高の大きい顧客が複数存在します。残高の小さい顧客も多数存在します。スコア傾向グラフに、一定期間にわたるスコアの改善が示されることがあります。ただし、残高の大きい少数の顧客口座の住所または性別の間違いが、組織の最も価値の高い顧客との関係に影響を及ぼす恐れがあります。預金残高カラムを、無効なデータを計算するための変動費のカラムと

して設定できます。カラムが原因で無効なデータのコストが高くなっている場合、合計値にリスクがあるとみなして、即座に対処できます。

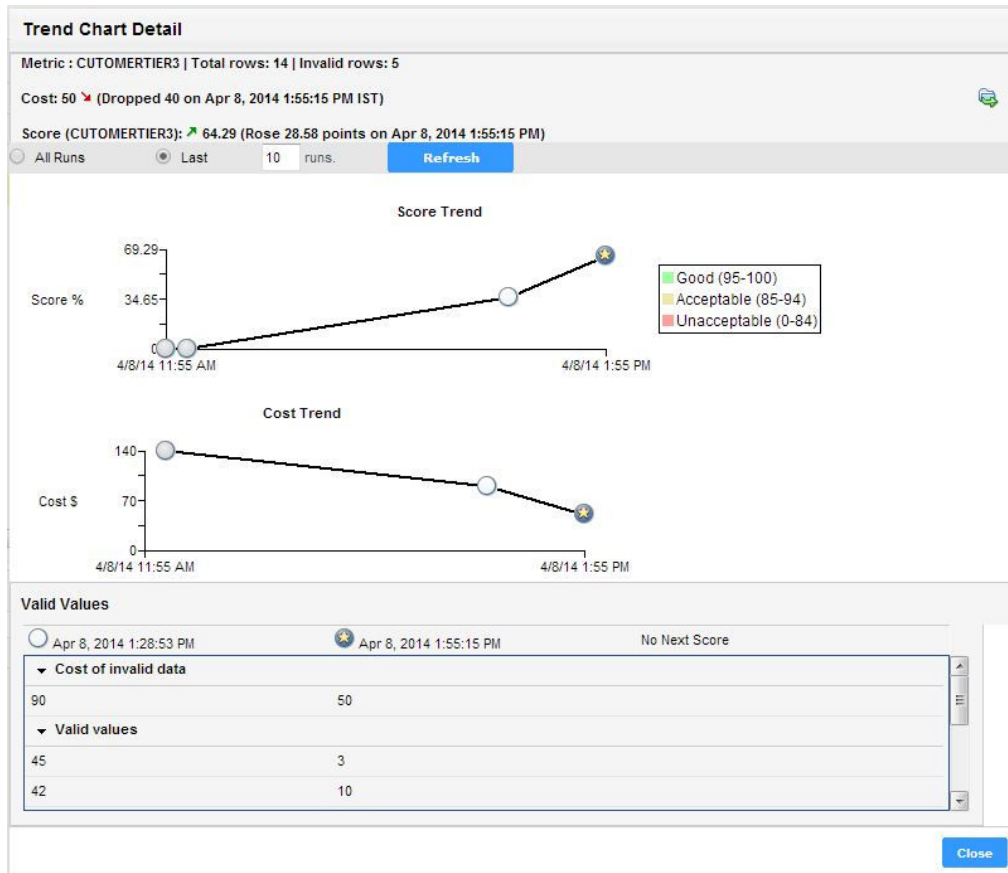
傾向グラフの表示

メトリックごとに傾向グラフを表示して、スコアまたは無効なデータのコストの変化を時系列で監視できます。

1. **【ライブラリ】** ワークスペースでスコアカードを含むプロジェクトまたはフォルダを選択します。
2. スコアカードをクリックして開きます。
スコアカードが**【スコアカード】** ワークスペースに表示されます。
3. **【スコアカード】** ビューでメトリックを選択します。
4. **【アクション】** > **【傾向グラフの表示】** をクリックします。

【傾向グラフの詳細】 ダイアログボックスが表示されます。

次の画像は、**【傾向グラフの詳細】** ダイアログボックスを示しています。



スコアとコストの値の変化を時系列で確認できます。ダイアログボックスの上部には、合計行数と無効な行の数が表示されます。Analyst ツールでは、日付ごとの履歴スコアカード実行データと最新の有効なスコアの値を使用してスコアが計算されます。スコアの傾向グラフとコストの傾向グラフの下には、メトリックの有効な値と、無効なデータのコストが表示されます。

傾向グラフのエクスポート

スコアの傾向グラフおよびコストの傾向グラフを.xlsx ファイルにエクスポートして、ファイル内のデータを表示できます。

1. スコアカードを開きます。
2. メトリックを選択して、**[アクション]** > **[傾向グラフの表示]** をクリックします。
[傾向グラフの詳細] ダイアログボックスが表示されます。
3. **[データのエクスポート]** アイコンをクリックします。
[ファイルへのデータのエクスポート] ダイアログボックスが表示されます。
4. **[ファイル名]** フィールドにファイル名を入力します。必要に応じて、デフォルトのファイル名を使用できます。
デフォルトのファイル形式は Microsoft Excel です。
5. **[無効な行]** フィールドに、エクスポートする無効な行の数を入力します。フィールドには最大 100,000 までの値を入力できます。
6. **[コードページ]** フィールドで、ファイルのコードページを選択できます。
7. **[OK]** をクリックします。

Informatica Analyst のスコアカードエクスポートファイル

スコアカードの結果は Microsoft Excel ファイルにエクスポートできます。Analyst ツールにより、ファイルは XLSX 形式で保存されます。スコアカードをエクスポートすると、Microsoft Excel ファイルにはスコアカードのサマリ、傾向グラフ、無効な行、およびスコアカードのプロパティが複数のワークシートに表示されます。

スコアカードをエクスポートするときに、次のオプションを設定できます。

[データ]

データをエクスポートする際は次のオプションのいずれかを選択します。

- **[すべて]**。スコアカードのサマリ、傾向グラフ、無効な行、およびスコアカードのプロパティを Microsoft Excel ファイルにエクスポートします。
- **[サマリビュー]**。スコアカードのサマリ、傾向グラフ、およびスコアカードのプロパティを Microsoft Excel ファイルにエクスポートします。

[無効な行]

エクスポートする無効な行の数を入力します。フィールドには最大 100,000 までの値を入力できます。デフォルトでは、フィールドに 100 が表示されます。Analyst ツールは、各メトリックに対して最大で 100 の無効な行をワークシートにエクスポートします。

1つのメトリックに対して100を超える無効な行をエクスポートすることを選択した場合は、データ統合サービスは次の手順を実行して残りのデータをエクスポートします。

1. スコアカード用のフォルダを<INFA_HOME>/tomcat/bin/rejectの場所に作成します。データ統合サービスは、スコアカード名とファイル作成時刻に基づいて、<scorecard_name>_HH_MM_SS形式でフォルダに名前を付けます。
例えば、スコアカード名がSD1でファイル作成時刻が2:23:15の場合、データ統合サービスはフォルダにSD1_02_23_15という名前を付けます。
2. スコアカードの各メトリックに対してサブフォルダを作成します。
例えば、スコアカードSD1にM1、M2、およびM3という名前のメトリックがある場合、データ統合サービスはM1、M2、およびM3という名前の3つのサブフォルダを作成します。
3. メトリックサブフォルダにMicrosoft Excel ファイルを作成します。ファイルには、メトリック名と増分数に基づいて、<metric_name>_<IncrementalNumberStartingWith0>形式で名前が付けられます。最後に作成されたファイルは<metric_name>_Remaining形式を使用します。各Excelファイルには、最大10Kの無効な行を含めることができます。
例えば、メトリックM1に30Kの無効な行がある場合、データ統合サービスはM1_0、M1_1、およびM1_Remainingという名前の3つのMicrosoft Excel ファイルを作成し、各ファイルに10Kの無効な行を保存します。
4. データ統合サービスは、スコアカード内の他のすべてのメトリックに対して手順3を繰り返します。

コードページ

コードページを選択します。

Microsoft Excel のスコアカードエクスポート結果

スコアカードの結果をエクスポートすると、Analyst ツールによって Microsoft Excel ファイルの複数のワークシートにその情報が保存されます。スコアカードのサマリ、傾向グラフ、無効な行、およびスコアカードのプロパティは、ファイル内のワークシートとして表示されます。Analyst ツールにより、ファイルは「xlsx」形式で保存されます。

次の表に、エクスポートファイルの各ワークシートに表示される情報を示します。

タブ	説明
スコアカードサマリ	エクスポートされたスコアカード結果のサマリ。この情報には、スコアカード名、各カラムの行の合計数、有効ではない行数、スコア、およびメトリックウェイトが含まれます。
傾向グラフ	スコアの傾向グラフです。
無効な行	各カラムの無効な行の詳細です。Analyst ツールは、最大で100行をワークシートにエクスポートします。 [ファイルへのデータのエクスポート] ダイアログボックスで [データ] > [すべて] を選択すると、[無効な行] ワークシートが表示されます。
プロパティ	名前、タイプ、説明、場所などのスコアカードのプロパティです。

Informatica Analyst からのスコアカード結果のエクスポート

スコアカード結果を「.xlsx」ファイルにエクスポートして、データをファイルで表示することができます。

1. スコアカードを開きます。
2. [アクション] > [データのエクスポート] をクリックします。

【ファイルへのデータのエクスポート】ダイアログボックスが表示されます。

3. ファイル名を入力します。必要に応じて、デフォルトのファイル名を使用します。
デフォルトのファイル形式は Microsoft Excel です。
4. ファイルのコードページを選択します。
5. **[OK]** をクリックします。

スコアカード通知

特定のメトリックのスコア、メトリックグループのスコア、またはメトリックコストがしきい値を超えた場合は Analyst ツールで電子メールを送信するように、スコアカード通知設定を構成します。メトリックのスコアまたはメトリックグループのスコアは、しきい値を超えるか特定のスコア範囲（許容不可、許容可能、正常など）内にとどまります。メトリックコストの値は、設定したコストしきい値の上限または下限を超えて移動することがあります。

個々のメトリックスコア、メトリックグループ、メトリックコストに電子メール通知を設定できます。スコアにグローバル設定を使用すると、特定のメトリックのスコアがしきい値を超えて、スコア範囲である [正常] から [許容可能]、および [許容可能] から [不良] に変化した場合に、Analyst ツールが通知電子メールを送信します。また、スコアカードを再度実行してもスコアが [許容不可] のスコア範囲にとどまっている場合は、スコアカードの実行ごとに通知電子メールが送信されます。メトリックコストにグローバル設定を使用すると、選択したメトリックでの無効なデータのコストがしきい値の上限と下限を超えたときに、Analyst ツールが通知電子メールを送信します。

スコアがスコア範囲である [許容不可] から [許容可能]、および [許容可能] から [正常] に変化した場合にスコアカードユーザーが電子メール通知を受信するよう、通知設定をカスタマイズすることもできます。スコアカードの実行ごとにメトリックスコアまたはメトリックコストが特定の範囲内にとどまっている場合は、電子メール通知を送信するよう選択できます。通知設定で各メトリックの無効なデータの現在のコストを表示し、この値に基づいてコストしきい値を設定できます。

スコアカードを設定して電子メール通知を送信する前に、管理者は Administrator ツールで電子メールサービスを設定する必要があります。

通知電子メールメッセージテンプレート

Analyst ツールがスコアカード通知の一部として受信者に送信する電子メールメッセージのメッセージテキストおよび構造を設定できます。電子メールテンプレートには、オプションの導入部テキストセクション、読み取り専用のメッセージ本文セクション、およびオプションの結びのテキストセクションがあります。

次の表で、電子メールテンプレートのタグについて説明します。

タグ	説明
ScorecardName	スコアカードの名前。
ObjectURL	スコアカードへのハイパーリンク。ユーザー名とパスワードを入力する必要があります。
MetricGroupName	メトリックが属しているメトリックグループの名前。
CurrentWeightedAverage	現在のスコアカード実行に含まれるメトリックグループの加重平均値。

タグ	説明
CurrentRange	現在のスコアカード実行に含まれるメトリックグループのスコア範囲（許容不可、許容可能、および正常など）。
PreviousWeightedAverage	前回のスコアカード実行に含まれるメトリックグループの加重平均値。
PreviousRange	前回のスコアカード実行に含まれるメトリックグループのスコア範囲（許容不可、許容可能、および正常など）。
MetricName	メトリックの名前。
MetricGroupName	メトリックグループの名前。
CurrentScore	最終スコアカード実行に基づくスコア。
CurrentRange	最終スコアカード実行に基づく現在のスコアが残っているスコア範囲。
PreviousScore	前回のスコアカード実行に基づくスコア。
PreviousRange	前回のスコアカード実行に基づくスコア範囲。
CurrentCost	最終スコアカード実行に基づく、メトリック内の無効なデータのコスト。
PreviousCost	前回のスコアカード実行に基づく、メトリック内の無効なデータのコスト。
ColumnName	メトリックが割り当てられているソースカラムの名前。
ColumnType	ソースカラムのタイプ。
RuleName	ルールの名前。
RuleType	ルールのタイプ。
DataObjectName	ソースデータオブジェクトの名前。

スコアカード通知の設定

スコアカード通知は、メトリックレベルとメトリックグループレベルのいずれに対しても設定できます。グローバル通知設定は、個別に通知設定を持っていないメトリックやメトリックグループに適用されます。

1. Analyst ツールでスコアカードを実行します。
2. **【アクション】** > **【編集】** をクリックします。
3. **【通知】** タブをクリックします。
4. **【通知を有効にする】** を選択して、スコアカード通知の設定を開始します。
5. メトリックまたはメトリックグループを選択します。
6. **【通知】** チェックボックスをクリックし、メトリックまたはメトリックグループのグローバル設定を有効にします。
7. **【カスタム設定の使用】** を選択し、メトリックまたはメトリックグループの設定を変更します。

スコアが【許容不可】、【許容可能】、および【正常】の範囲内にあり、しきい値を超えた場合に通知電子メールを送信するよう選択できます。メトリックコストがしきい値の上限または下限を超えた場合にも、通知電子メールを送信できます。

8. スコアカード通知のグローバル設定を編集するには、【**グローバル設定の編集**】アイコンをクリックします。

【**グローバル設定の編集**】ダイアログボックスが表示されます。このダイアログボックスで、電子メールテンプレートなどの設定を編集できます。

スコアカード通知のグローバル設定の実行

グローバルスコアカード通知設定を選択したときに、スコアが【許容不可】の範囲にある場合は、Analyst ツールがターゲットユーザに電子メールを送信します。通知設定を構成して、メトリックスコアまたはメトリックコストがしきい値を超えたときに電子メールを送信するようにもできます。電子メールアドレスとスコアカードのメッセージテキストを含む電子メールテンプレートを設定できます。

1. Analyst ツールでスコアカードを実行します。
2. 【**アクション**】 > 【**編集**】 > 【**通知**】 をクリックして、【**スコアカードの編集**】ダイアログボックスを開きます。
3. 【**通知を有効にする**】を選択して、スコアカード通知の設定を開始します。
4. 【**グローバル設定の編集**】アイコンをクリックします。

【**グローバル設定の編集**】ダイアログボックスが表示されます。このダイアログボックスで、電子メールテンプレートなどの設定を編集できます。

5. 【**許容可能スコアへの変化**】および【**許容不可スコアへの変化**】チェックボックスを使用して、メトリックスコアの電子メール通知を送信するタイミングを選択します。
 6. 【**コストの変化**】チェックボックスを使用して、メトリックコストの電子メール通知を送信するタイミングを選択します。
 7. 【**宛先**】フィールドに、受信者の電子メール ID を入力します。複数の電子メール ID を区切るには、セミコロンを使用します。
- 送信者のデフォルトの電子メール ID は、ドメインの SMTP プロパティで設定された【**送信側電子メールアドレス**】です。
8. 電子メールの件名にテキストを入力します。
 9. 【**ボディ**】フィールドで、電子メールメッセージの導入部のテキストおよび結びのテキストを追加します。
 10. グローバル設定を適用するには、**すべてのメトリックおよびメトリックグループに設定を適用する**を選択します。

11. 【**OK**】をクリックします。

スコアカードリネージュ

スコアカードリネージュは、データの発生元、パス、そしてメトリックまたはメトリックグループに対するデータフローを示します。スコアカードリネージュを使うと、メトリックまたはメトリックグループでスコアが

許容範囲を超えてばらつくことの根本原因を分析することができます。Analyst ツールでスコアカードリネージュを表示します。

スコアカードリネージュを表示するには、次のタスクを完了します。

1. Informatica Administrator で Metadata Manager Service を Analyst Service に関連付けます。
2. プロジェクトを選択して、その中のスコアカードオブジェクトを、Developer tool 中の [メタデータマネージャのリソースファイルのエクスポート] オプションまたは `infacmd oie exportResources` コマンドを使用して XML ファイルにエクスポートします。
3. Metadata Manager で、エクスポートした XML ファイルを使用してリソースを作成してロードします。
注: 作成して Metadata Manager にロードするリソースファイルに名前を付ける際は、<MRS 名>_<プロジェクト名>という形式にします。リソースファイルの作成とロードの方法に関する詳細は、『*Informatica PowerCenter Metadata Manager ユーザーガイド*』を参照してください。
4. Analyst ツールでスコアカードを開いてメトリックまたはメトリックグループを選択します。
5. スコアカードリネージュを表示します。

Informatica Analyst でのスコアカードリネージュの表示

メトリックまたはメトリックグループのスコアカードリネージュを表示できます。Analyst ツールでスコアカードリネージュダイアグラムを表示できるようにするには、スコアカードリネージュと Metadata Manager のメタデータをロードする必要があります。

1. **【ライブラリ】** ワークスペースの **【アセット】** ペインで、表示するスコアカードをクリックします。
スコアカードが **【スコアカード】** ワークスペースに表示されます。
2. **【スコアカード】** ビューで、メトリックまたはメトリックグループを選択します。
3. 右クリックして **【リネージュを表示する】** を選択します。
スコアカードリネージュダイアグラムが別のウィンドウで表示されます。

重要: スコアカードオブジェクトのエクスポートされた XML 形式で Metadata Manager のリソースを作成またはロードしない場合、リソースはカタログで利用できないというエラーメッセージが出ることがあります。スコアカードリネージュへの XML ファイルのエクスポートの詳細については、[「スコアカードリネージュのリソースファイルのエクスポート」 \(ページ 107\)](#)を参照してください。

パート III: Informatica Developer によるプロファイリング

この部には、以下の章があります。

- [データオブジェクトプロファイル, 83 ページ](#)
- [半構造化データソースのカラムプロファイル, 89 ページ](#)
- [Informatica Developer のルール, 96 ページ](#)
- [マプレットとマッピングのプロファイリング, 98 ページ](#)
- [Informatica Developer のカラムプロファイル結果, 101 ページ](#)
- [Informatica Developer のスコアカード, 106 ページ](#)

第 9 章

データオブジェクトプロファイル

この章では、以下の項目について説明します。

- [Informatica Developer のカラムプロファイル, 83 ページ](#)
- [Informatica Developer での単一のデータオブジェクトプロファイルの作成, 85 ページ](#)
- [Informatica Developer での複数のデータオブジェクトプロファイルの作成, 85 ページ](#)
- [Informatica Developer でのフラットファイルデータオブジェクトの同期, 86 ページ](#)
- [Informatica Developer でのリレーショナルデータオブジェクトの同期, 88 ページ](#)

Informatica Developer のカラムプロファイル

データソースのカラムの特性（値の割合やパターンなど）を分析するには、カラムプロファイルを使用します。フィルタを追加することで、プロファイルの実行時に条件を満たす行のみ読み取るようにすることができます。プロファイルでは、フィルタ条件を満たさない行は処理されません。

プロファイルのカラムに対して実行して検出できる情報は、次のタイプです。

- カラムの値が出現する回数。
- カラムの各値が出現する頻度（割合または行数で表される）。
- カラム内の値の文字パターン。
- 統計値（カラムの値の最大長と最小長、および最初の値と最後の値など）。
- 推測されるデータ型、頻度、データドメイン検出の適合条件、およびデータ型の推測ステータス。

カラムプロファイルは、マッピングまたはマップレット内のデータオブジェクトか、モデルリポジトリ内のオブジェクトについて定義できます。リポジトリ内のオブジェクトは、単一のデータオブジェクトプロファイル、複数のデータオブジェクトプロファイル、またはエンタープライズ検出プロファイルに含めることができます。

カラムプロファイルのサンプリングオプション、ドリルダウンオプション、およびランタイム環境を選択できます。カラムプロファイルにルールやフィルタを追加できます。

フィルタリングのオプション

詳細フィルタまたは SQL フィルタを追加し、カラムプロファイルの実行時にカラムプロファイルで使用される行を特定できます。プロファイルでは、フィルタ条件を満たさない行は処理されません。

サンプリングのオプション

サンプリングオプションは、Developer tool がプロファイルを実行する行数を決定します。サンプリングのオプションは、プロファイルを定義または実行するときに設定することができます。

以下の表に、プロファイルのサンプリングのオプションを示します。

プロパティ	説明
すべての行	データオブジェクト内のすべての行でプロファイルを実行します。 ネイティブ、Blaze、Spark、および Databricks ランタイム環境でサポートされます。
最初の<number>行をサンプリング	データオブジェクト内の行のうち、先頭からのサンプル行でプロファイルを実行します。行は最大で 2,147,483,647 行まで選択できます。 ネイティブおよび Blaze ランタイム環境でサポートされます。
<number>行のランダムサンプリング	データオブジェクトでランダムに選択された数の行に対してプロファイルを実行します。行は最大で 2,147,483,647 行まで選択できます。 ネイティブおよび Blaze ランタイム環境でサポートされます。
ランダムサンプリング（自動）	データオブジェクト内の行数に基づいて計算されたサンプル行でプロファイルを実行します。 ネイティブおよび Blaze ランタイム環境でサポートされます。
Limit n <number>行	データオブジェクトの行数に基づいてプロファイルを実行します。 Hadoop 検証環境でプロファイルを実行するように選択すると、Spark エンジンではデータオブジェクトの複数のパーティションからサンプルを収集し、これらのサンプルを単一のノードにプッシュしてサンプルサイズを計算します。Limit n のサンプリングオプションは、Oracle、SQL Server および DB2 データベースをサポートします。Limit n のサンプリングオプションには、詳細フィルタは適用できません。 Spark ランタイム環境でサポートされます。
ランダムパーセンテージ	データオブジェクト内の行の割合に対してプロファイルを実行します。 Spark および Databricks ランタイム環境でサポートされます。
以降のプロファイル実行で、データ型とデータドメイン推測から、承認済みのデータ型とデータドメインを除外します	次のプロファイル実行で、データ型とデータドメイン推測から、承認済みのデータ型とデータドメインを除外します。

ランダムな行のサンプルでプロファイルを実行するように選択すると、ランダムサンプルアルゴリズムでは、データオブジェクト内の行をランダムに選択して、プロファイルを実行します。カラムプロファイルのランダムサンプリングオプションを選択すると、Developer tool は、ステージングされたデータに対してドリルダウンを実行します。これは、ドリルダウンのパフォーマンスに影響を与える可能性があります。データドメイン検出プロファイルのランダムサンプリングオプションを選択すると、Developer tool はライブデータのドリルダウンを実行します。

Informatica Developer での単一のデータオブジェクトプロファイルの作成

データオブジェクトの 1 つ以上のカラムに対して単一のデータオブジェクトプロファイルを作成し、プロファイルオブジェクトをモデルリポジトリに格納することができます。

1. **【オブジェクトエクスプローラ】** ビューで、プロファイリングするデータオブジェクトを選択します。
2. **【ファイル】 > 【新規】 > 【プロファイル】** をクリックしてプロファイルウィザードを開きます。
3. **【プロファイル】** を選択して **【次へ】** をクリックします。
4. プロファイルの名前を入力し、プロジェクトの場所を確認します。必要に応じて、新しい場所を参照します。
5. 必要に応じて、プロファイルの説明を入力します。
6. 選択したデータオブジェクトの名前が **【データオブジェクト】** セクション内に表示されることを確認します。
7. **【次へ】** をクリックします。
8. 実行するプロファイルの操作を P 設定します。以下の操作を設定できます。
 - カラムプロファイリング
 - プライマリキー検出
 - 機能依存性検出
 - データドメイン検出

注: プロファイルの操作を有効にするには、その操作に対して **["プロファイルの実行"アクションの一部として有効化]** を選択します。カラムプロファイリングはデフォルトで有効になっています。

9. プロファイルのオプションを確認します。

すべてのプロファイルタイプのカラム選択を編集できます。カラムプロファイルのフィルタおよびサンプリングのオプションを確認します。プライマリキー、機能依存性、およびデータドメイン検出の推測オプションを確認できます。また、データドメイン検出のデータドメイン選択も確認できます。
10. ドリルダウンのオプションを確認し、必要に応じて編集します。デフォルトでは、**【行のドリルダウンを有効にする】** オプションが選択されています。カラムプロファイルでは、ドリルダウンのオプションを編集できます。このオプションでは、ドリルダウン操作の読み取り元をデータソースにするか、またはステージングされたデータにするか、および以前のプロファイルの実行による結果データを格納するかも決定します。
11. **【実行の設定】** セクションで、検証環境を選択します。検証環境として、**【ネイティブ】**、**【Hadoop】**、または **【Databricks】** を選択します。ランタイム環境として **【ネイティブ】**、**【Blaze】**、または **【Spark】** を選択できます。**【Blaze】** または **【Spark】** を選択すると、Hadoop 接続を選択できます。**【Databricks】** を選択した場合は、Databricks 接続を選択できます。
12. **【完了】** をクリックします。

Informatica Developer での複数のデータオブジェクトプロファイルの作成

複数のデータオブジェクトに対して複数のデータオブジェクトプロファイルを実行する場合、Developer tool はデフォルトのカラムプロファイリングオプションを使用して、1 つ以上のデータオブジェクトのカラムプロ

ファイルを生成します。必要に応じて、複数のデータオブジェクトに対してプロファイルを実行するためのエンタープライズ検出プロファイルを作成することもできます。

1. **【オブジェクトエクスプローラ】**ビューで、プロファイリングするデータオブジェクトを選択します。
2. **【ファイル】** > **【新規】** > **【プロファイル】** をクリックして **【新しいプロファイル】** ウィザードを開きます。
3. **【新規】** ウィザードで **【複数のプロファイル】** オプションを選択して、**【次へ】** をクリックします。
4. **【複数のプロファイル】** ウィンドウで、プロファイルを作成する場所を選択します。プロファイリングしたオブジェクトと同じ場所に各プロファイルを作成するか、またはプロファイルの共通の場所を指定することができます。
5. 選択したデータオブジェクトの名前が **【データオブジェクト】** セクション内に表示されることを確認します。

必要に応じて、**【追加】** をクリックして別のデータオブジェクトを追加します。

6. 必要に応じて、プロファイリングする行数を指定し、ウィザードの完了時にプロファイルを実行するかどうかを選択します。
7. **【次へ】** をクリックします。
8. **【検証環境】** セクションで、**【ネイティブ】** を選択します。

注: 複数のデータオブジェクトプロファイルを実行するには、必ず **【ネイティブ】** オプションを選択します。Hadoop ランタイム環境の Blaze または Spark エンジンで複数のデータオブジェクトを実行するには、エンタープライズデータ検出プロファイルを選択します。Databricks ランタイム環境の Databricks Spark で複数のデータオブジェクトを実行するには、エンタープライズデータ検出プロファイルを選択します。

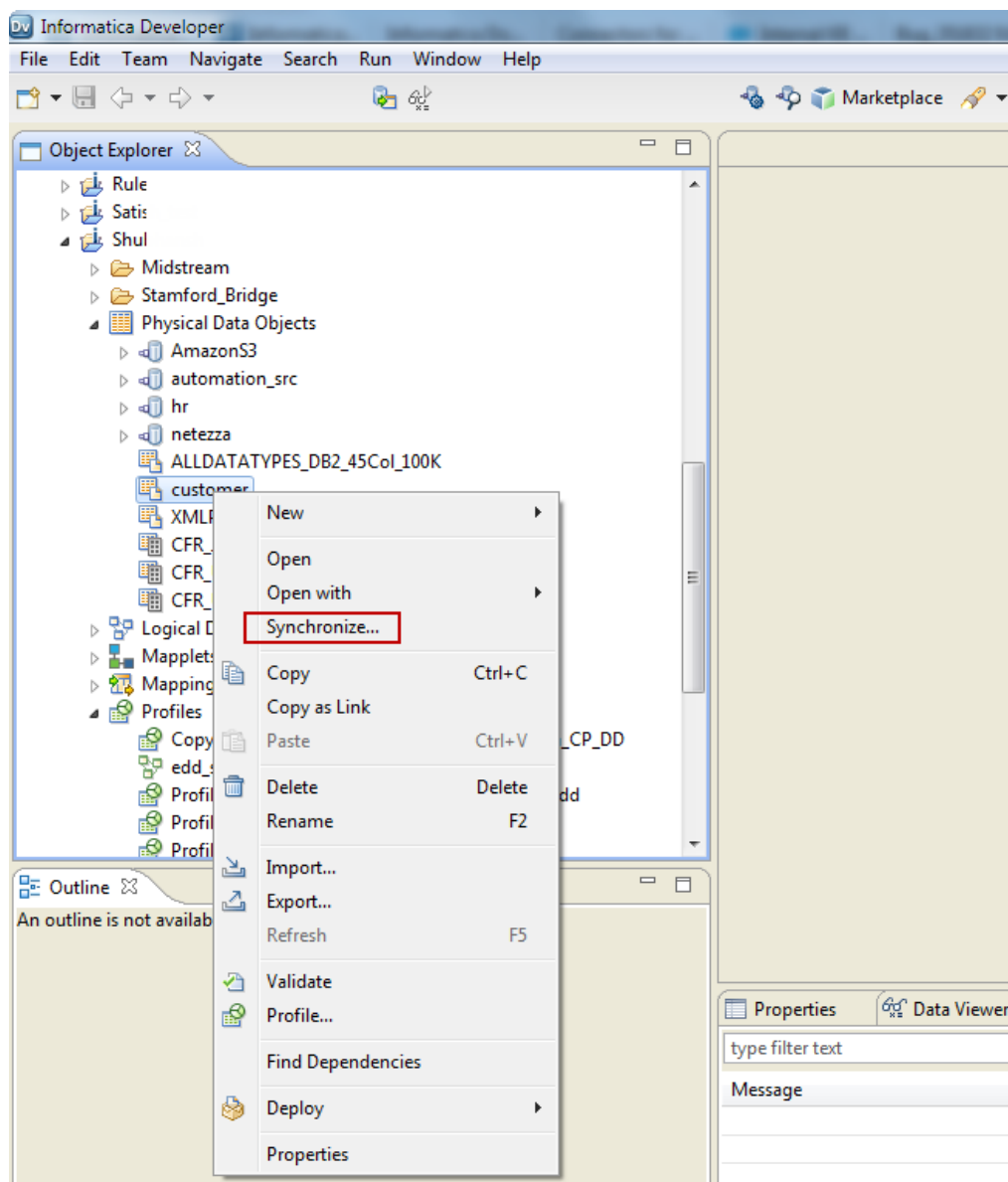
9. **【完了】** をクリックします。
10. 必要に応じて、プロファイル名に追加するプレフィックスおよびサフィックス文字列を入力します。
11. **【OK】** をクリックします。

Informatica Developer でのフラットファイルデータオブジェクトの同期

Informatica Developer で、外部フラットファイルデータソースに加えられた変更をそのデータオブジェクトに同期させることができます。データオブジェクトを同期させるには、**【フラットファイルの同期】** ウィザードを使用します。

1. **【オブジェクトエクスプローラ】** ビューで、フラットファイルデータオブジェクトを選択します。
2. 右クリックし、**【同期】** を選択します。

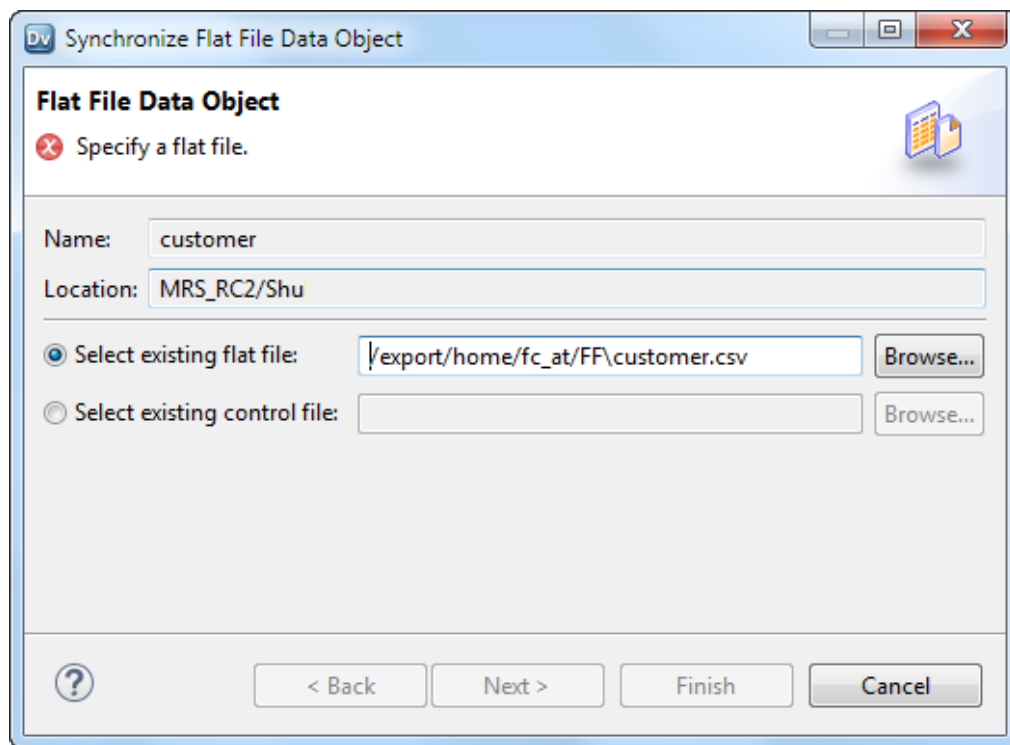
次の図は、データオブジェクトの同期オプションを示しています。



【フラットファイルデータオブジェクトの同期】ウィザードが表示されます。

3. 【フラットファイルデータオブジェクトの同期】ウィザードで、[既存のフラットファイルを選択] フィールドのフラットファイルパスを確認します。

次の図は、フラットファイルデータオブジェクトの同期ウィザードを示しています。



4. **【次へ】** をクリックします。
5. 必要に応じ、コードページ、形式、区切り形式プロパティ、およびカラムプロパティを選択します。
6. **【完了】** をクリックし、**【OK】** をクリックします。

Informatica Developer でのリレーショナルデータオブジェクトの同期

Informatica Developer で、リレーショナルデータソースの外部データソース変更をそのデータオブジェクトに同期させることができます。外部データソース変更には、カラムの追加、変更、削除や、ルールの変更などがあります。

1. **【Object Explorer】** ビューで、リレーショナルデータオブジェクトを選択します。
2. 右クリックし、**【同期】** を選択します。
操作の確認を求めるメッセージが表示されます。
3. 同期プロセスを実行するには、**【OK】** をクリックします。
同期プロセスのステータスメッセージが表示されます。
4. 「**同期が完了しました**」というメッセージが表示された時点で、**【OK】** をクリックします。
このメッセージには、データオブジェクトに加えられたメタデータ変更の概要が示されます。

第 10 章

半構造化データソースのカラムプロフィール

この章では、以下の項目について説明します。

- [半構造化データソースのカラムプロフィールの概要, 89 ページ](#)
- [JSON および XML データオブジェクト, 90 ページ](#)
- [HDFS 内の半構造化データソースの複合型ファイルデータオブジェクト, 91 ページ](#)
- [HDFS 接続の作成, 92 ページ](#)
- [HDFS 内の JSON または XML ファイルからの複合型ファイルデータオブジェクトの作成, 92 ページ](#)
- [Avro または Parquet データソースからの複合型ファイルデータオブジェクトの作成, 93 ページ](#)
- [半構造化データソースでのカラムプロフィールの作成, 94 ページ](#)

半構造化データソースのカラムプロフィールの概要

Avro、JSON、Parquet、および XML データソースからデータオブジェクトを作成し、次に、そのデータオブジェクトのカラムプロフィールを作成できます。

Avro、JSON、Parquet、および XML の各形式は、半構造化データソースです。カラムプロフィールの作成に半構造化データソースを使用するために、次のタスクを実行できます。

1. 半構造化データソースに対して、物理データオブジェクトを作成する。
2. 物理データオブジェクトでカラムプロフィールを作成して実行する。

JSON または XML データソースの場合は、フラットファイルデータオブジェクトを作成できます。Hadoop 分散ファイルシステム（HDFS）内の Avro、JSON、Parquet、および XML データソースの場合は、複合型ファイルデータオブジェクトを作成できます。

JSON および XML データオブジェクト

JSON または XML データソースから、フラットファイルデータオブジェクトまたは複合型ファイルデータオブジェクトを作成できます。データオブジェクトにカラムプロファイルを作成して実行できます。

JSON または XML データソースのパスが含まれるテキストファイルを作成して、そのファイルをデータソースとして使用して、フラットファイルデータオブジェクトを作成します。このテキストファイルに、複数の JSON または複数の XML データソースへのファイルパスを追加することもできます。

複合ファイルリーダーを使用すると、JSON または XML データソースから複合型ファイルデータオブジェクトを作成できます。複合ファイルリーダーは、データプロセッサトランスフォーメーションに入力を渡します。このトランスフォーメーションにより、ファイルが解析され、ソースデータがフラットなカンマ区切り値のレコードに変換されます。

注: Developer tool は、UTF-8 エンコードを使用した JSON データソースには対応していません。

JSON または XML データソースからのデータオブジェクトの作成

JSON または XML データソースから、フラットファイルデータオブジェクトまたは複合型ファイルデータオブジェクトを作成できます。

1. Developer tool の **【オブジェクトエクスプローラ】** ビューで、データオブジェクトおよびカラムプロファイルを作成するプロジェクトを選択します。
2. **【ファイル】 > 【新規】 > 【データオブジェクト】** をクリックします。
【新規】 ダイアログボックスが表示されます。
3. フラットファイルデータオブジェクトまたは複合型ファイルデータオブジェクトの作成を選択できます。
 - フラットファイルデータオブジェクトを作成するには、次のタスクを実行します。
 1. **【物理データオブジェクト】 > 【フラットファイルデータオブジェクト】** を選択し、**【次へ】** をクリックします。
【新しいフラットファイルデータオブジェクト】 ダイアログボックスが表示されます。
 2. **【既存のフラットファイルから作成】** を選択し、**【参照】** をクリックしてテキストファイルを選択します。**【次へ】** をクリックします。
 3. コードページが **【MS Windows Latin 1 (ANSI)、Latin 1 のスーパーセット】** であり、形式が **【区切り】** であることを確認します。**【次へ】** をクリックします。
 4. 区切り文字が **【カンマ】** に設定されていることを確認します。**【完了】** をクリックします。
 - 複合型ファイルデータオブジェクトを作成するには、次のタスクを実行します。
 1. **【物理データオブジェクト】 > 【複合型ファイルデータオブジェクト】** を選択し、**【次へ】** をクリックします。
【新しい複合型ファイルデータオブジェクト】 ダイアログボックスが表示されます。
 2. データオブジェクトの名前を入力します。アクセスタイプとして **【ファイル】** を選択します。
 3. **【参照】** をクリックして JSON または XML ファイルを選択します。 **【完了】** をクリックします。
Developer サーバーが Linux にある場合、データソースのファイルパスをそのサーバー上の場所に更新する必要があります。ファイルパスを更新するには、複合型ファイルデータオブジェクトを選択して **【データオブジェクト操作】** タブの **【読み取り】** をクリックし、**【データオブジェクト操作の詳細】** ペインの **【詳細】** タブでファイルパスを追加します。

データオブジェクトがプロジェクトフォルダに表示されます。

HDFS 内の半構造化データソースの複合型ファイルデータオブジェクト

HDFS を使用する Avro、JSON、Parquet、または XML ファイルに対して、カラムプロファイルを作成して実行できます。HDFS 内の JSON または XML ファイルを読み取るには、JSON または XML 形式の入力をデータプロセッサトランスフォーメーションに渡すための複合型ファイルリーダーを使用します。

HDFS 内の JSON または XML データソースからの複合型ファイルデータオブジェクト

JSON または XML ファイルから、複合型ファイルデータオブジェクトを作成できます。このデータオブジェクトに対して、カラムプロファイルを作成して実行できます。

HDFS 内の JSON または XML ファイルに対してデータオブジェクトを作成する前に、HDFS への接続を作成します。

次のいずれかの方法を使用して、HDFS 内の JSON または XML ファイルからデータオブジェクトを作成できます。

- 1 つの JSON または XML ファイルに対して、複合型ファイルデータオブジェクトを作成する。
- 複数の JSON または複数の XML ファイルが格納されたフォルダに対して、複合型ファイルデータオブジェクトを作成する。

データオブジェクトを作成したら、そのデータオブジェクトに対してカラムプロファイルを作成および実行できます。

HDFS 内の Avro または Parquet データソースからの複合型ファイルデータオブジェクト

HDFS 内の Avro または Parquet データソースから、複合型ファイルデータオブジェクトを作成できます。このデータオブジェクトを使用して、カラムプロファイルを作成して実行できます。

複合型ファイルデータオブジェクトの作成は、1 つの Avro または Parquet ファイルから、または複数の Avro ファイルや複数の Parquet ファイルが含まれるフォルダに対して行うことができます。Avro および Parquet データソースからの複合型ファイルデータオブジェクトは、ファイルまたは接続のアクセスタイプで、Binary、Avro、または Parquet のリソース形式で作成できます。Avro および Parquet データソースから複合型ファイルデータオブジェクトを作成する前に、HDFS 接続を作成しておく必要があります。

注: リソース形式として **[Avro]** または **[Parquet]** を選択できるのは、フラットな構造の Avro および Parquet データソースの場合のみです。

HDFS 内の Avro および Parquet ファイルからデータオブジェクトを作成する場合は、次のオプションのいずれかを選択できます。

- アクセスタイプをファイルにし、リソース形式を Binary として選択する。
- アクセスタイプをファイルにし、リソース形式を Avro または Parquet として選択する。
- アクセスタイプを接続にし、リソース形式を Avro または Parquet として選択する。

HDFS 接続の作成

Informatica Developer で HDFS 接続を設定して、HDFS 内の Avro、JSON、Parquet、および XML データソースに対してカラムプロファイルを作成します。HDFS 接続の作成後に、複合型ファイルデータオブジェクトを作成できます。

1. **【ウィンドウ】** > **【設定】** をクリックします。
2. **【Informatica】** > **【接続】** を選択します。
3. ドメインを展開します。
4. 接続タイプを **【ファイルシステム】** > **【Hadoop ファイルシステム】** で選択し、**【追加】** をクリックします。
5. 接続名を入力します。
6. 必要に応じて、接続の説明を入力します。
7. **【次へ】** をクリックします。
8. 接続のプロパティを設定します。
9. **【テスト接続】** をクリックし、HDFS への接続を確認します。
10. **【完了】** をクリックします。

HDFS 内の JSON または XML ファイルからの複合型ファイルデータオブジェクトの作成

HDFS を使用した JSON または XML ソースファイルの複合型ファイルデータオブジェクトを作成し、そのデータオブジェクトに対するカラムプロファイルを作成できます。

1. Developer tool の **【オブジェクトエクスプローラ】** ビューで、物理データオブジェクトおよびカラムプロファイルを作成するプロジェクトを選択します。
2. **【ファイル】** > **【新規】** > **【データオブジェクト】** をクリックします。
【新規】 ダイアログボックスが表示されます。
3. **【物理データオブジェクト】** > **【複合型ファイルデータオブジェクト】** を選択し、**【次へ】** をクリックします。
【新しい複合型ファイルデータオブジェクト】 ダイアログボックスが表示されます。
4. データオブジェクトの名前を入力します。アクセスタイプとして **【接続】** を選択します。
5. 1 つの JSON または XML ファイルから、または複数の JSON ファイルまたは複数の XML ファイルが含まれるフォルダに対して、データオブジェクトを作成できます。
 - JSON または XML ファイルから複合型ファイルデータオブジェクトを作成するには、次の手順を実行します。
 1. **【参照】** をクリックして接続を選択します。
 2. **【リソースの追加】** ダイアログボックスで、**【追加】** をクリックして JSON ファイルまたは XML ファイルを選択します。
 3. **【完了】** をクリックします。
データオブジェクトがプロジェクトフォルダに表示されます。

- 複数の JSON ファイルまたは複数の XML ファイルが含まれるフォルダに対して、複合型ファイルデータオブジェクトを作成するには、次の手順を実行します。
 1. **【参照】** をクリックして接続を選択します。
 2. **【リソースの追加】** ダイアログボックスで、**【追加】** をクリックしてフォルダ内の JSON または XML ファイルを選択します。
 3. **【完了】** をクリックします。
データオブジェクトがプロジェクトフォルダに表示されます。
 4. プロジェクトフォルダ内のデータオブジェクトを選択して、**【詳細】** > **【ランタイム: 読み取り】** > **【ソースファイルのディレクトリ】** をクリックします。
 5. ファイルパスのファイル名を削除し、フォルダ名はそのまま保持します。

Avro または Parquet データソースからの複合型ファイルデータオブジェクトの作成

Avro または Parquet データソースから **【ファイル】** または **【接続】** のアクセスタイプで複合型ファイルデータオブジェクトを作成できます。このデータオブジェクトに対して、カラムプロファイルを作成できます。

1. **【オブジェクトエクスプローラ】** ビューで、プロジェクトを選択します。
2. **【ファイル】** > **【新規】** > **【データオブジェクト】** をクリックします。
【新規】 ダイアログボックスが表示されます。
3. **【物理データオブジェクト】** > **【複合型ファイルデータオブジェクト】** を選択し、**【次へ】** をクリックします。
【新しい複合型ファイルデータオブジェクト】 ダイアログボックスが表示されます。
4. データオブジェクトの名前を入力します。
5. アクセスタイプを **【接続】** または **【ファイル】** から選択できます。
 - アクセスタイプに **【接続】** を選択した場合は、次の手順を実行します。
 1. **【参照】** をクリックして HDFS 接続を選択します。
 2. **【接続の選択】** ダイアログボックスで、データソースを選択して **【OK】** をクリックします。
 3. **【新しい複合型ファイルデータオブジェクト】** ダイアログボックスで、**【完了】** をクリックします。
データオブジェクトがプロジェクトフォルダに表示されます。
 - アクセスタイプに **【ファイル】** を選択し、リソース形式に **【Binary】** を選択した場合は、次の手順を実行します。
 1. **【参照】** をクリックして、ローカルマシン上の Avro または Parquet ファイルを選択します。
 2. **【新しい複合型ファイルデータオブジェクト】** ダイアログボックスで、**【完了】** をクリックします。
データオブジェクトがプロジェクトフォルダに表示されます。
 3. プロジェクトフォルダ内のデータオブジェクトを選択して、**【データオブジェクト操作】** ビューをクリックします。
 4. **【データオブジェクト操作】** ビューで、**【読み取り】** > **【詳細】** タブをクリックします。
 5. **【詳細】** タブの **【ファイルパス】** フィールドに、Linux または Windows マシンにあるデータソースのファイルパスを入力します。
 6. ファイル形式に **【カスタム入力】** を指定します。

7. Avro データソース用の【入力形式】フィールドに **com.informatica.avro.AvroToXML** と入力し、Parquet データソース用の【入力形式】フィールドに **com.informatica.parquet.ParquetToXML** と入力します。入力形式を追加すると、データプロセッサトランスフォーメーションによって実行時に Avro または Parquet 形式のデータソースが処理され、XML 形式のデータソースに変換されます。
- アクセスタイプに【ファイル】を選択し、リソース形式に【Avro】または【Parquet】を選択した場合は、次の手順を実行します。
 1. 【参照】をクリックして、ローカルマシンの Avro または Parquet ファイルを選択します。
 2. 【新しい複合型ファイルデータオブジェクト】ダイアログボックスで、【完了】をクリックします。データオブジェクトがプロジェクトフォルダに表示されます。
 3. データオブジェクトを作成したら、【データオブジェクト操作】 > 【読み取り】 > 【詳細】タブに移動して、【ファイルパス】フィールドのファイルパスが、Linux または Windows マシンのデータソースに対応するかどうかを確認します。

注: リソース形式として【Avro】または【Parquet】を選択できるのは、フラットな構造の Avro および Parquet データソースの場合のみです。

複数の Avro ファイルまたは複数の Parquet ファイルが含まれるフォルダを 1 つ選択して、データオブジェクトを作成できます。データオブジェクトを作成したら、【データオブジェクト操作】 > 【読み取り】 > 【詳細】タブに移動して、【ファイルパス】フィールドのファイルパスが、Linux または Windows マシンのデータソースが含まれるフォルダを指すかどうかを確認します。

半構造化データソースでのカラムプロファイルの作成

Avro、JSON、Parquet、または XML データソースからフラットファイルデータオブジェクトまたは複合型ファイルデータオブジェクトを作成後、そのデータオブジェクトに対してカラムプロファイルを作成して実行できます。

1. 【オブジェクトエクスプローラ】ビューで、Avro、JSON、Parquet または XML ファイルのデータオブジェクトを選択します。
2. 【ファイル】 > 【新規】 > 【プロファイル】をクリックします。
【新規】ダイアログボックスが表示されます。
3. 【プロファイル】を選択します。【次へ】をクリックします。
【新規プロファイル】ダイアログボックスが表示されます。
4. 【新規プロファイル】ダイアログボックスで、プロファイルの名前と説明（省略可能）を入力します。
5. 【拡張ファイル形式の処理】オプションを選択します。【次へ】をクリックします。

次の画像は、【新しいプロファイル】ウィザードで【拡張ファイル形式の処理】オプションが選択された状態を示しています。

1. [拡張ファイル形式の処理]。このオプションを選択し、半構造化データソースを処理します。

注: リソース形式を **[Avro]** または **[Parquet]** として選択した場合、Avro および Parquet の各データソースに対しては **[拡張ファイル形式の処理]** オプションは表示されません。

6. **[単一のデータオブジェクトプロファイル]** ページで、必要に応じて、**[カラムの選択]** および **[データドメイン検出]** の下にあるカラムとオプションを選択します。**[完了]** をクリックします。

注: Linux マシンに Developer tool がインストールされていて、JSON または XML 物理データオブジェクトが、テキストファイルを含むフラットファイルデータオブジェクトの場合は、次のタスクを実行します。

1. **[概要]** タブの **[精度]** 値を更新して、サーバーのデータソースのファイルパスの文字数が含まれるようにします。
2. フラットファイルデータオブジェクトでプロファイルを作成後、サーバー内の場所へのデータソースファイルパスを更新します。ファイルパスを更新するには、**[詳細]** タブの **[ランタイム: 読み取り]** > **[ソースファイルディレクトリ]** をクリックし、ファイルパスを追加します。
7. プロファイルを右クリックし、**[プロファイルの実行]** を選択します。
プロファイル結果が表示されます。

第 11 章

Informatica Developer のルール

この章では、以下の項目について説明します。

- [Informatica Developer のルールの概要, 96 ページ](#)
- [Informatica Developer でのルールの作成, 97 ページ](#)
- [Informatica Developer でのルールの適用, 97 ページ](#)

Informatica Developer のルールの概要

ルールは、カラムプロファイルの実行時にソースデータに適用される条件を定義するビジネスロジックです。ルールをプロファイルに追加して、データを検証できます。ルール、定義済みルール、または再利用可能なルールとして検証されるマップレットをカラムプロファイルで使用できます。

次の方法を使用して、列プロファイルでルールを使用できます。

- Developer tool で、マップレットを作成し、ルールとして検証します。ルールは、Analyst ツールに再利用可能なルールとして表示されます。ルールをカラムプロファイルに適用するには、Analyst ツールと Developer tool を使用します。
- カラムプロファイルでは、定義済みのルールを使用できます。Informatica は、Developer tool と Analyst ツールを使用して定義済みのルールを提供します。
- Analyst ツールで、ルール仕様を作成し、マップレットを生成します。ルール仕様は、Analyst ツールのカラムプロファイルに適用できます。Developer tool で、マップレットをルールとして検証します。このルールは、カラムプロファイルで使用できる再利用可能なルールとして表示されます。

注: Developer tool では、カラムプロファイルにルール仕様を追加、編集、または削除することはできません。

ルールは、以下の要件を満たす必要があります。

- 入力トランスフォーメーションと出力トランスフォーメーションを含める必要がある。ルールでデータソースを使用することはできません。
- 式トランスフォーメーション、ルックアップトランスフォーメーション、およびパッシブなデータ品質トランスフォーメーションを含めることができる。他のタイプのトランスフォーメーションを含めることはできません。例えば、一致トランスフォーメーションはアクティブなトランスフォーメーションなので、ルールに含めることはできません。
- 入力グループ間のカーディナリティが指定されていない。

Informatica Developer でのルールの作成

Developer ツールでルールを作成するには、マプレットをルールとして検証する必要があります。

Developer ツールでマプレットを作成します。

1. マプレットエディタを右クリックします。
2. **【検証】** > **【ルール】** を選択します。

Informatica Developer でのルールの適用

保存したカラムプロファイルにルールを追加できます。結合分析用に設定したプロファイルにルールを追加することはできません。

1. **【Object Explorer】** ビューを参照して必要なプロファイルを見つけます。
2. プロファイルを右クリックし、**【開く】** を選択します。
プロファイルがエディタで開かれます。
3. **【定義】** タブをクリックし、**【ルール】** を選択します。
4. **【追加】** をクリックします。
【ルールの適用】 ダイアログボックスが開きます。
5. **【参照】** をクリックして、適用するルールを見つけます。
リポジトリプロジェクトからルールを選択して、**【OK】** をクリックします。
6. **【入力値】** の下の **【値】** カラムをクリックし、ルールの入力ポートを選択します。
7. 必要に応じて、**【出力行】** の下の **【値】** カラムをクリックし、ルールの出力ポートの名前を編集します。
ルールが **【定義】** タブに表示されます。

第 12 章

マプレットとマッピングのプロファイリング

この章では、以下の項目について説明します。

- [マプレットとマッピングのプロファイリングの概要, 98 ページ](#)
- [マプレットまたはマッピングオブジェクトに対するプロファイルの実行, 98 ページ](#)
- [マッピングまたはマプレットオブジェクトのプロファイルの比較, 99 ページ](#)
- [プロファイルからのマッピングの生成, 99 ページ](#)

マプレットとマッピングのプロファイリングの概要

マプレットまたはマッピングのオブジェクトに対するカラムプロファイルを定義できます。マッピングまたはマプレットの設計を確認するときにプロファイル結果を保存しない場合は、マプレットまたはマッピングオブジェクトに対してプロファイルを実行します。また、プロファイルからマッピングを生成することもできます。

マプレットまたはマッピングオブジェクトに対するプロファイルの実行

マプレットまたはマッピングオブジェクトに対してプロファイルを実行すると、プロファイルがすべてのデータカラムに対して実行され、データオブジェクトのステージングされたデータに対するドリルダウン操作が可能になります。複数の出力ポートを使用して、マプレットまたはマッピングオブジェクトに対してプロファイルを実行できます。

プロファイルでは、選択したオブジェクトの出力ポートへのマッピングによってソースデータがトレースされます。マッピングを実行した場合にこれらのポートに表示されるデータが分析されます。

1. マプレットまたはマッピングを開きます。
2. マプレットまたはマッピングが有効であることを確認します。
3. データオブジェクトまたはトランスフォーメーションを右クリックし、[今すぐプロファイル] を選択します。

トランスフォーメーションに複数の出力グループがある場合は、**【出力グループの選択】** ダイアログボックスが表示されます。

4. トランスフォーメーションに複数の出力グループがある場合は、必要な出力グループを選択します。
5. **【OK】** をクリックします。

プロファイルの**【結果】** タブにプロファイル結果が表示されます。プロファイルサマリログファイルおよびマッピングログファイルを表示して、Developer ツールによって実行されたタスクの詳細を取得できます。

注: オペレーティングシステムプロファイルを使用してプロファイルを実行した場合、サマリログはデータ統合サービス用に構成されたログディレクトリにあり、マッピングログはオペレーティングシステムのプロファイル用に構成されたログディレクトリにあります。

マッピングまたはマップレットオブジェクトのプロファイルの比較

マップレットまたはマッピングの2つのオブジェクトを分析し、これらのオブジェクトのカラムプロファイルの結果を比較するプロファイルを作成できます。

単一のマッピングまたはマップレットオブジェクトのカラムプロファイルと同様に、プロファイル比較はすべてのデータカラムに対して実行され、データオブジェクトにステージングされているデータに対してドリルダウン操作を実行できるようにします。ソーステーブルからターゲットテーブルにデータを移動したあと、プロファイルを比較してデータの移行を確認できます。また、時間の経過と共に変化するデータソースのプロファイルを比較することもできます。

1. マップレットまたはマッピングを開きます。
2. マップレットまたはマッピングが有効であることを確認します。
3. エディタで **Ctrl** キーを押しながら2つのオブジェクトをクリックします。
4. いずれかのオブジェクトを右クリックし、**【プロファイルの比較】** を選択します。
5. 必要に応じて、一方のオブジェクトから他方のオブジェクトまでのカラムを照合するようにプロファイル比較を設定します。
6. 必要に応じて、一方のオブジェクトのカラムをクリックして他方のオブジェクトのカラムにドラッグし、カラムを照合します。
7. 必要に応じて、プロファイルですべてのカラムを分析するか一致したカラムのみを分析するかを選択します。
8. **【OK】** をクリックします。

プロファイルからのマッピングの生成

プロファイルからマッピングオブジェクトを作成できます。作成したマッピングオブジェクトを使用し、有効なマッピングを作成します。作成するマッピングには、プロファイリングしたオブジェクトに基づくデータソースが含まれます。また、プロファイルルールロジックに基づくトランスフォーメーションを含めることもできます。マッピングの作成後に、オブジェクトを追加してマッピングを完成させます。

1. **Object Explorer** ビューで、マッピングを作成するプロファイルを見つけます。
2. プロファイル名を右クリックし、**【マッピングの生成】** を選択します。

【マッピングの生成】ダイアログボックスが表示されます。

3. マッピング名を入力します。必要に応じて、マッピングの説明を入力します。

4. マッピングのフォルダの場所を確認します。

デフォルトでは、マッピングはプロファイルと同じプロジェクトの **Mappings** フォルダに作成されます。
マッピングの場所として別の場所を選択するには、【参照】をクリックします。

5. Developer ツールでマッピングを作成するために使用されるプロファイル定義を確認します。別のプロファイルを使用するには、【プロファイルの選択】をクリックします。

6. 【完了】をクリックします。

マッピングが **Object Explorer** に表示されます。

オブジェクトをマッピングに追加してマッピングを完成させます。

第 13 章

Informatica Developer のカラム プロファイル結果

この章では、以下の項目について説明します。

- [Informatica Developer のカラムプロファイル結果, 101 ページ](#)
- [カラム値のプロパティ, 102 ページ](#)
- [カラムパターンのプロパティ, 102 ページ](#)
- [カラム統計のプロパティ, 102 ページ](#)
- [カラムデータ型のプロパティ, 103 ページ](#)
- [Informatica Developer でのキュレーション, 104 ページ](#)
- [Informatica Developer からのプロファイル結果のエクスポート, 105 ページ](#)

Informatica Developer のカラムプロファイル結果

カラムプロファイル分析は、データ品質に関して、値の頻度、データのパターン、および統計に焦点を当てた情報を提供します。

以下の表に、それぞれのタイプの分析のプロファイル結果を示します。

プロファイルタイプ	プロファイル結果
カラムプロファイル	<ul style="list-style-type: none">- 一意の値と NULL 値の割合と数に関する統計- 推測されたデータ型- データソースでデータに対して宣言されているデータ型- 最大値と最小値- 最後にプロファイルを実行した日時- カラム内の一意の各データ要素の割合と数に関する統計- カラム内の一意の各文字パターンの割合と数に関する統計
プライマリキープロファイル	<ul style="list-style-type: none">- 推測されるプライマリキー- キー違反
機能依存性プロファイル	<ul style="list-style-type: none">- 推測される機能依存性- 機能依存性違反

カラム値のプロパティ

カラム値のプロパティは、プロファイリングしたカラムの値、および各値が各カラムに出現する頻度を示します。頻度は数、割合、および棒グラフで表されます。

カラム値のプロパティを表示するには、**【表示】** リストから **【値】** を選択します。カラム値をダブルクリックし、その値を含む行にドリルダウンします。

以下の表に、カラム値のプロパティを示します。

プロパティ	説明
値	プロファイルのカラムのすべての値を示す一覧。
頻度	カラムの値が出現する回数。
割合	カラムの値が出現する回数（カラム内のすべての値に対する割合で表される）。
チャート	割合の棒グラフ。

カラムパターンのプロパティ

カラムパターンのプロパティは、プロファイリングしたカラムにおけるデータのパターンおよびパターンが各カラムに出現する頻度を示します。パターンは数、割合、および棒グラフで表されます。

パターン情報を表示するには、**【表示】** リストから **【パターン】** を選択します。パターンをダブルクリックし、そのパターンを含む行にドリルダウンします。

以下の表に、カラム値のパターンのプロパティを示します。

プロパティ	説明
パターン	選択されたカラムのパターン。
頻度	カラムのパターンが出現する回数。
割合	カラムのパターンが出現する回数（カラム内のすべての値に対する割合で表される）。
チャート	割合の棒グラフ。

カラム統計のプロパティ

カラム統計には、値の最大長と最小長、および最初の値と最後の値などのプロパティが含まれます。

統計情報を表示するには、**【表示】** リストから **【統計】** を選択します。

以下の表に、カラム統計のプロパティを示します。

プロパティ	説明
最大長	カラムの最長値の長さ。
最小の長さ	カラムの最短値の長さ。
下	カラム内の最後の 5 つの値。
上	カラム内の最初の 5 つの値。
合計	数値データ型を持つカラムのすべての値の合計。

注: Integer タイプのカラムの場合、プロファイルには平均と標準偏差の統計も表示されます。

カラムデータ型のプロパティ

カラムデータ型には、プロファイル結果の各カラムの推測されたデータ型がすべて含まれています。

データ型情報を表示するには、**[表示]** リストから **[データ型]** を選択します。特定のデータ型をダブルクリックすると、そのデータ型を含む行にドリルダウンします。

次の表に、カラムデータ型のプロパティを示します。

プロパティ	説明
データ型	プロファイルのカラムのすべての推測されたデータ型の一覧。
頻度	カラムに特定のデータ型が出現する回数（数値で表現される）。
一致率（%）	カラムに特定のデータ型が出現するパーセンテージ。
ステータス	データ型のステータスを示します。ステータスは推測済み、承認済み、または拒否済みです。 推測済み カラムのデータ型が、Developer tool によって推測されたものであることを示します。 承認済み カラムのデータ型が承認済みであることを示します。データ型を承認すると、そのデータ型がモデルリポジトリにコミットされます。 却下済み カラムのデータ型が拒否されたことを示します。

Informatica Developerでのキュレーション

キュレーションとは、データソースで検出されたメタデータを検証し管理することで、使用とレポートに適したメタデータにするプロセスです。Informatica Developerでメタデータをキュレーションする場合、プロファイル結果の中の推測されたデータ型またはデータドメインを承認、拒否、リセットすることができます。

1つのカラムに対して1つのデータ型またはデータドメインを承認できます。カラムの拒否されたデータ型またはデータドメインは、非表示にできます。推測されたデータ型またはデータドメインを承認または拒否した後でも、データ型またはデータドメインをリセットすると、ステータスを推測済みに戻すことができます。

データ型の承認

プロファイルの結果には、データソースのカラムごとに推測データ型、頻度、一致率、および推測ステータスが含まれています。各カラムに対して1つのデータ型を選択して承認することができます。

1. **【オブジェクトエクスプローラ】** ビューでプロファイルを選択して、それを開きます。
2. **【結果】** タブが表示されていることを確認します。
3. **【カラムプロファイリング】** ビューで、カラムを選択して値の頻度、パターン、データ型、統計を右側のパネルに表示します。
4. **【詳細】** パネルで、**【データ型】** を**【表示】** リストから選択します。
カラムの推測されたデータ型が表示されます。
5. 承認する必要があるカラムを右クリックして、**【承認】** をクリックします。
データ型のステータスが**【承認済み】** に変わります。
6. データ型のステータスを推測済みに戻すには、データ型を右クリックしてから**【リセット】** をクリックします。

データ型の拒否

Informatica Developerでは、デフォルトで、プロファイル結果に推測されたデータ型が表示されます。推測または承認されたデータ型を拒否することもできます。拒否されたデータドメインを表示するか非表示にするかを選択できます。

1. **【Object Explorer】** ビューで、プロファイルを選択します。
2. プロファイルをダブルクリックして開きます。
タブにプロファイルが開きます。
3. **【カラムプロファイリング】** ビューで、行を選択します。
4. 推測されたカラムデータ型を拒否するには、右側のパネルで**【データ型】** ビューを選択します。拒否する必要がある推測されたデータ型を選択し、その行を右クリックして**【拒否】** を選択します。
Informatica Developerでは、拒否されたデータ型がデータ型リストで灰色に表示されます。
5. 拒否されたデータドメインを非表示にするには、その行を右クリックして**【拒否済みを非表示】** を選択します。
6. 拒否されたデータ型を表示するには、行のうち1つを右クリックして**【拒否済みを表示】** を選択します。

Informatica Developer からのプロファイル結果のエクスポート

カラムプロファイル結果を.csv ファイルまたは Microsoft Excel ファイルにエクスポートできます。プロファイル結果を Microsoft Excel ファイルにエクスポートすると、Developer tool は情報を.xlsx ファイルに保存します。

1. **[Object Explorer]** ビューで、プロファイルを開きます。
2. 必要に応じて、プロファイルを実行してプロファイル結果を更新します。
3. **[結果]** ビューを選択します。
4. カラムを選択します。
5. **[詳細]** で、**[値]**、**[パターン]**、または **[データ型]** を選択し、**[エクスポート]** アイコンをクリックします。
[データをファイルにエクスポート] ダイアログボックスが開きます。
6. デフォルトのファイル名を受け入れるか変更します。
7. エクスポートするデータのタイプを選択します。**[選択したカラムの値]**、**[選択したカラムのパターン]**、**[選択したカラムのデータ型]**、または **[すべて (概要、値、パターン、データ型、統計、プロパティ)]** を選択できます。
8. **[参照]** をクリックして場所を選択し、ファイルをコンピュータ上にローカルに保存します。
9. フィールド名を最初の行としてエクスポートしない場合は、**[フィールド名を最初の行としてエクスポート]** チェックボックスをオフにします。
10. **[OK]** をクリックします。

第 14 章

Informatica Developer のスコアカード

この章では、以下の項目について説明します。

- [Informatica Developer のスコアカードの概要, 106 ページ](#)
- [スコアカードの作成, 106 ページ](#)
- [スコアカードリネージュのリソースファイルのエクスポート, 107 ページ](#)
- [Informatica Developer からのスコアカードリネージュの表示, 107 ページ](#)

Informatica Developer のスコアカードの概要

スコアカードは、プロファイルでの品質解析結果を視覚的に表したものです。スコアカードは Developer ツールで表示できます。Developer ツールでスコアカードを作成した後、Analyst ツールに接続してそのスコアカードを編集用に開くことができます。データオブジェクトの現在のデータに対して、またはプロファイリングウェアハウスにステージングされているデータに対してスコアカードを実行します。

Analyst ツールでスコアカードを編集または実行したり、メトリックまたはメトリックグループのスコアカードリネージュを表示したりできます。

スコアカードの作成

スコアカードを作成し、プロファイルからスコアカードにカラムを追加します。スコアカードにカラムを追加する前に、プロファイルを実行する必要があります。

1. **[Object Explorer]** ビューで、スコアカードを作成するプロジェクトまたはフォルダーを選択します。
2. **[ファイル]** > **[新規]** > **[スコアカード]** をクリックします。
[新しいスコアカード] ダイアログボックスが表示されます。
3. **[追加]** をクリックします。
[プロファイルの選択] ダイアログボックスが表示されます。追加するカラムが含まれているプロファイルを選択します。
4. **[OK]** をクリックし、**[次へ]** をクリックします。
5. スコアカードに追加するカラムを選択します。

デフォルトでは、プロファイルで定義されているカラムとルールが選択されます。プロファイルに含まれていないカラムは追加できません。

6. **【完了】** をクリックします。
スコアカードが作成されます。
7. 必要に応じて、**【Informatica Analyst で開く】** をクリックして Analyst ツールに接続し、Analyst ツールでスコアカードを開きます。

スコアカードリネージュのリソースファイルのエクспорт

スコアカードを含むプロジェクトや、依存オブジェクトを Metadata Manager のリソースファイルとしてエクспортできます。XML 形式でエクспортされたリソースファイルを使用して Metadata Manager のスコアカードリネージュのリソースを作成およびロードします。

1. **エクспортウィザード**を開くには、**【ファイル】 > 【エクспорт】** の順にクリックします。
2. **【Informatica】 > 【Metadata Manager のリソースファイル】** を選択します。
3. **【次へ】** をクリックします。
4. **【参照】** をクリックしてエクспортするスコアカードオブジェクトとリネージュを含むプロジェクトを選択します。
5. **【次へ】** をクリックします。
6. エクспортするスコアカードオブジェクトを選択します。
7. エクспортファイルの名前とファイルの場所を入力します。
8. 選択したオブジェクトとともに**エクспортウィザード**でエクспортされる依存オブジェクトを表示するには、**【次へ】** をクリックします。
エクспортウィザードに依存オブジェクトが表示されます。
9. **【完了】** をクリックします。
オブジェクトが XML ファイルにエクспортされます。

Informatica Developer からのスコアカードリネージュの表示

Developer ツールからメトリックまたはメトリックグループのスコアカードリネージュを表示するには、Analyst ツールを起動します。

1. **【Object Explorer】** ビューで、スコアカードが含まれるプロジェクトまたはフォルダを選択します。
2. スコアカードをダブルクリックして開きます。
タブにスコアカードが表示されます。
3. **【Informatica Analyst で開く】** をクリックします。
ブラウザウィンドウ内に Analyst ツールが開きます。

4. Analyst ツールの【スコアカード】ビューで、メトリックまたはメトリックグループを選択します。
5. 右クリックして【リネージュを表示する】を選択します。
スコアカードリネージュダイアグラムがダイアログボックスに表示されます。

索引

A

Analyst ツールでのカラムプロファイル結果
 インタフェース [44](#), [57](#), [58](#)
 カラムの詳細 [45](#), [59](#)
 サマリ [43](#)

I

Informatica Analyst
 カラムプロファイル結果 [42](#), [54](#)
 カラムプロファイルの概要 [20](#), [56](#)
 ルール [32](#)
 ロックとバージョン管理 [25](#)
Informatica Developer
 ルール [96](#)

S

Sqoop 設定
 プロファイリング [24](#)

い

異常値
 検出 [50](#)

え

export
 XML へのスコアカードリネージュ [107](#)

か

カラムプロファイル
 Informatica Developer [83](#)
 オプション [15](#)
 オペレーティングシステムプロファイル [25](#)
 ドリルダウン [60](#)
 プロセス [21](#)
 概要 [14](#)
カラムプロファイルの作成
 プロファイル [26](#)
カラムプロファイル結果
 Informatica Developer [101](#)

き

キュレーション
 Informatica Analyst [61](#)

キュレーション (続く)
 Informatica Developer [104](#)
 概念 [17](#)
 タスク [17](#)

け

傾向グラフ
 Informatica Analyst からのエクスポート [76](#)
 コスト [74](#)
 スコア [74](#)
 表示 [75](#)

し

式ルールの作成
 ルール [34](#)

す

スコアカード
 Informatica Analyst [64](#)
 Informatica Analyst でのプロセス [65](#)
 Informatica Developer [106](#)
 概要 [16](#)
 グローバル通知設定の実行 [80](#)
 傾向グラフ [73](#)
 固定費 [70](#)
 しきい値の定義 [71](#)
 実行中 [68](#)
 スコアカードへのカラムの追加 [67](#)
 スコアの移動 [72](#)
 通知 [78](#)
 通知の設定 [79](#)
 ドリルダウン [73](#)
 表示 [69](#)
 変動費 [70](#)
 無効なデータのコスト [70](#)
 メトリック [70](#)
 メトリックウェイト [70](#)
 メトリックグループの削除 [72](#)
 メトリックグループの作成 [71](#)
 メトリックグループの編集 [72](#)
 編集 [69](#)
スコアカードの結果
 Excel へのエクスポート [77](#)
 Informatica Analyst からのエクスポート [77](#)
 エクスポート [76](#)
スコアカードリネージュ
 Informatica Analyst での表示 [81](#)
 Informatica Developer からの表示 [107](#)

て

- 定義済みルール
 - プロセス [33](#)
- データオブジェクトプロファイル
 - 単一のプロファイルの作成 [85](#)
 - 複数のプロファイルの作成 [86](#)
- テーブルデータオブジェクト
 - 同期 [31](#)

ふ

- フィルタ
 - 概要 [37](#)
- フラットファイルデータオブジェクト
 - 同期 [29](#)
- プロファイリング
 - ツール [12](#)
 - プロセス [12](#)
 - ロックとバージョン管理 [15](#)
- プロファイル
 - Avro または Parquet 形式 [91](#)
 - XML 形式と JSON 形式 [90](#), [91](#)
 - 概要 [11](#)
 - カラムプロファイルの編集 [27](#)
 - カラムプロファイルの作成 [26](#)
 - コンポーネント [12](#)
 - 実行中 [28](#), [55](#), [56](#)
 - フィルタの作成 [37](#)
 - フィルタの編集 [40](#)
- プロファイル結果
 - Developer tool でのデータドメインの拒否 [104](#)
 - Excel [62](#)
 - Informatica Analyst からのエクスポート [63](#)
 - Informatica Developer でのエクスポート [105](#)
 - Informatica Developer でのデータ型の承認 [104](#)
 - エクスポート [62](#)

- プロファイル結果 (続く)
 - カラム値 [51](#)
 - カラムデータ型 [48](#), [103](#)
 - カラムパターン [50](#)
 - サマリ [57](#), [60](#)
 - サマリビュー [44](#)
 - 詳細ビュー [46](#)
 - データ型の拒否 [61](#)
 - データ型の承認 [61](#)
 - ドリルダウン [60](#)

ま

- マッピングオブジェクト
 - プロファイルの実行 [98](#)
- マプレットとマッピングのプロファイリング
 - 概要 [98](#)

ら

- ランタイム環境
 - Analyst ツール [23](#)
 - Hadoop [23](#), [24](#)

る

- ルール
 - Informatica Developer での作成 [97](#)
 - Informatica Developer での適用 [97](#)
 - 式 [34](#)
 - 式ルールの作成 [34](#)
 - 定義済み [32](#)
 - 定義済みルールの適用 [33](#)
 - ルール仕様を使用した式ルールの作成 [35](#)