



Informatica® Data Quality
10.4.0

入門ガイド

Informatica Data Quality 入門ガイド

10.4.0

2019 年 12 月

© 著作権 Informatica LLC 2011, 2020

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。

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。

Informatica、PowerCenter、PowerExchange、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。

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。

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: 2020-02-07

目次

序文	7
Informatica のリソース	7
Informatica Network	7
Informatica ナレッジベース	7
Informatica マニュアル	7
Informatica 製品可用性マトリックス	8
Informatica Velocity	8
Informatica Marketplace	8
Informatica グローバルカスタマサポート	8
第 1 章 : スタートガイド概要	9
Informatica ドメインの概要	9
機能の可用性	11
Informatica Analyst について	12
Informatica Developer の概要	12
Informatica Developer のようこそページ	13
チートシート	13
データ品質およびプロファイリング	14
チュートリアルの内容	14
チュートリアルの構造	15
Informatica Analyst のチュートリアル	15
Informatica Developer tool	16
チュートリアルの前提条件	17
第 I 部 : Informatica Analyst の基本操作	18
第 2 章 : レッスン 1. Informatica Analyst の設定	19
Informatica Analyst の設定の概要	19
タスク 1. Informatica Analyst へのログイン	20
タスク 2. プロジェクトの作成	20
タスク 3. フォルダーの作成	21
Informatica Analyst の設定のまとめ	21
第 3 章 : レッスン 2. データオブジェクトの作成	22
データオブジェクトの作成の概要	22
タスク 1. フラットファイルデータオブジェクトの作成	23
タスク 2. データオブジェクトのプロパティの表示	23
データオブジェクトの作成のまとめ	24

第4章: レッスン 3. デフォルトプロファイルの作成	25
デフォルトプロファイルの作成の概要	25
タスク 1. デフォルトプロファイルの作成と実行	26
タスク 2. サマリビューでのプロファイル結果の表示	26
デフォルトプロファイルの作成のサマリ	27
第5章: レッスン 4. カスタムプロファイルの作成	28
カスタムプロファイルの作成の概要	28
タスク 1. カスタムプロファイルの作成	29
タスク 2. プロファイルの実行	30
タスク 3. プロファイル結果でのドリルダウン	30
カスタムプロファイルの作成のまとめ	31
第6章: レッスン 5. 式ルールの作成	32
式ルールの作成の概要	32
タスク 1. 式ルールの作成とプロファイルの実行	33
タスク 2. 式ルールの出力の表示	33
タスク 3. 式ルールの編集	34
式ルールの作成のまとめ	34
第7章: レッスン 6. スコアカードの作成と実行	35
スコアカードの作成と実行の概要	35
タスク 1. プロファイル結果からのスコアカードの作成	36
タスク 2. スコアカードの実行	37
タスク 3. スコアカードの表示	37
タスク 4. スコアカードの編集	38
タスク 5. しきい値の設定	39
タスク 6. スコアの傾向グラフの表示	39
スコアカードの作成と実行のまとめ	39
第8章: レッスン 7. プロファイルカラムからの参照テーブルの作成	41
プロファイルカラムからの参照テーブルの作成の概要	41
タスク 1. プロファイルカラムからの参照テーブルの作成	42
タスク 2. 参照テーブルの編集	43
プロファイルカラムからの参照テーブルの作成のまとめ	43
第9章: レッスン 8. 参照テーブルの作成	44
参照テーブルの作成の概要	44
タスク 1. 参照テーブルの作成	45
参照テーブルの作成のまとめ	45

第 II 部 : Informatica Developer の基本操作	46
第 10 章 : レッスン 1. Informatica Developer の設定	47
Informatica Developer の設定の概要.....	47
タスク 1. Informatica Developer の起動.....	48
タスク 2. ドメインの追加.....	48
タスク 3. モデルリポジトリの追加.....	49
タスク 4. プロジェクトの作成.....	49
タスク 5. フォルダの作成.....	49
タスク 6. デフォルトのデータ統合サービスの選択.....	50
Informatica Developer の設定のまとめ.....	50
第 11 章 : レッスン 2: 物理データオブジェクトのインポート	51
物理データオブジェクトのインポートの概要.....	51
タスク 1. Boston_Customers フラットファイルデータオブジェクトのインポート.....	52
タスク 2. LA_Customers フラットファイルデータオブジェクトのインポート.....	57
タスク 3. All_Customers フラットファイルデータオブジェクトのインポート.....	58
物理データオブジェクトのインポートのまとめ.....	58
第 12 章 : レッスン 3. データソースでのプロファイルの実行	59
データのプロファイリングの概要.....	59
タスク 1. 2 つのデータソースでの結合分析の実行.....	60
タスク 2. 結合分析結果の表示.....	61
タスク 3. データソースでのプロファイルの実行.....	61
タスク 4. カラムプロファイル結果の表示.....	62
データのプロファイリングのまとめ.....	62
第 13 章 : レッスン 4. データの解析	63
データの解析の概要.....	63
タスク 1. ターゲットデータオブジェクトの作成.....	64
手順 1.LA_Customers_tgt データオブジェクトの作成.....	64
手順 2.読み取りオプションおよび書き込みオプションの設定.....	65
手順 3.データオブジェクトへのカラムの追加.....	65
タスク 2. データを解析するためのマッピングの作成.....	66
手順 1.マッピングの作成.....	66
手順 2.データオブジェクトのマッピングへの追加.....	66
手順 3.パーサートランスフォーメーションのマッピングへの追加.....	67
手順 4.パーサートランスフォーメーションの設定.....	67
タスク 3. パーサートランスフォーメーションでのプロファイルの実行.....	68
タスク 4. マッピングの実行.....	68
タスク 5. マッピング出力の表示.....	68
データの解析のまとめ.....	69

第 14 章 : レッスン 5. データの標準化 70

データの標準化の概要.....	70
タスク 1. ターゲットデータオブジェクトの作成.....	71
手順 1.All_Customers_Stdz_tgt データオブジェクトの作成.....	71
手順 2.読み取りオプションおよび書き込みオプションの設定.....	72
タスク 2. データを標準化するためのマッピングの作成.....	72
手順 1.マッピングの作成.....	73
手順 2.データオブジェクトのマッピングへの追加.....	73
手順 3.標準化トランスフォーメーションのマッピングへの追加.....	73
手順 4.標準化トランスフォーメーションの設定.....	74
タスク 3. マッピングの実行.....	75
タスク 4. マッピング出力の表示.....	75
データの標準化のまとめ.....	75

第 15 章 : レッスン 6. 住所データの検証..... 76

アドレスデータの検証の概要.....	76
タスク 1. ターゲットデータオブジェクトの作成.....	77
手順 1.All_Customers_av_tgt データオブジェクトの作成.....	78
手順 2.読み取りオプションおよび書き込みオプションの設定.....	78
手順 3.ポートのデータオブジェクトへの追加.....	79
タスク 2. アドレスを検証するためのマッピングの作成.....	79
手順 1.マッピングの作成.....	79
手順 2.データオブジェクトのマッピングへの追加.....	80
手順 3.アドレスバリデータトランスフォーメーションのマッピングへの追加.....	80
タスク 3. アドレスバリ Data Transformation の設定.....	80
手順 1. アドレス検証のデフォルトの国の設定.....	81
手順 2.アドレスバリデータトランスフォーメーションの入力ポートの設定.....	81
手順 3.アドレスバリデータトランスフォーメーションの出力ポートの設定.....	82
手順 4.未使用のデータソースポートのデータターゲットへの接続.....	83
タスク 4. マッピングの実行.....	83
タスク 5. マッピング出力の表示.....	84
アドレスデータの検証のまとめ.....	86

付録 A : FAQ（よくある質問） 87

Informatica Analyst の FAQ（よくある質問）.....	87
Informatica Developer の FAQ（よくある質問）.....	87

索引..... 89

序文

Data Quality 入門ガイドを読むと、Data Quality の主な機能を確認し、Informatica Developer および Informatica Analyst でデータ品質タスクを実行する方法を学習できます。

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 からグローバルサポートセンターに連絡できます。

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 で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、eSupport オプションを選択します。

第 1 章

スタートガイド概要

この章では、以下の項目について説明します。

- [Informatica ドメインの概要, 9 ページ](#)
- [Informatica Analyst について, 12 ページ](#)
- [Informatica Developer の概要, 12 ページ](#)
- [チュートリアルの内容, 14 ページ](#)
- [チュートリアルの構造, 15 ページ](#)

Informatica ドメインの概要

Informatica には、サービス指向のアーキテクチャが備わっています。これにより、サービスを拡張し、複数のマシンでリソースを共有することができます。Informatica ドメインは、サービスの管理用の主要な単位です。

Informatica をインストールしたら、Informatica Administrator にログインできます。ドメインを管理したり、他のアプリケーションクライアントにアクセスする前に必須のアプリケーションサービスを設定したりするには、Administrator ツールを使用します。

Informatica ドメインには以下のコンポーネントが含まれます。

- アプリケーションクライアント。基本的な Informatica の機能にアクセスするために使用するクライアントのグループです。アプリケーションクライアントは、サービスマネージャやアプリケーションサービスに対する要求を作成します。
- アプリケーションサービス。サーバーベースの機能を表すサービスのグループです。Informatica ドメインには、アプリケーションサービスのサブセットを含めることができます。アプリケーションクライアントに必要なアプリケーションサービスを作成および設定します。

アプリケーションサービスには、ドメイン内に 1 つのインスタンスを持つことができるシステムサービスが含まれます。システムサービスは、ドメインを作成すると自動的に作成されます。各システムサービスは、そのサービスが提供する機能を使用するように設定し、有効化できます。

- プロファイルウェアハウス。データ統合サービスがプロファイル結果の格納に使用するリレーショナルデータベース。
- 参照データウェアハウス。モデルリポジトリ内の参照テーブルオブジェクト向けの参照データ値を格納するリレーショナルデータベース。
- リポジトリ。アプリケーションクライアントからのユーザー要求の処理に必要なオブジェクトおよびプロセスに関するメタデータを格納するリレーショナルデータベースのグループです。

- サービスマネージャ。すべてのドメイン操作を管理するドメインに組み込まれているサービスです。サービスマネージャは、アプリケーションサービスを実行し、認証、承認、ロギングなどのドメイン機能を実行します。
- ワークフローデータベース。ワークフロー向けのランタイムメタデータを格納するリレーショナルデータベース。

以下の表に、クライアントで必要なアプリケーションクライアント（Administrator ツールを除く）、アプリケーションサービス、およびリポジトリを一覧表示します。

アプリケーションクライアント	アプリケーションサービス	リポジトリ
Informatica Analyst	<ul style="list-style-type: none"> - アナリストサービス - コンテンツ管理サービス - データ統合サービス - モデルリポジトリサービス - 検索サービス 	モデルリポジトリ
Informatica Developer	<ul style="list-style-type: none"> - アナリストサービス - コンテンツ管理サービス - データ統合サービス - モデルリポジトリサービス 	モデルリポジトリ
Metadata Manager	<ul style="list-style-type: none"> - Metadata Manager サービス - PowerCenter 統合サービス - PowerCenter リポジトリサービス 	<ul style="list-style-type: none"> - Metadata Manager リポジトリ - PowerCenter リポジトリ
PowerCenter(R)クライアント	<ul style="list-style-type: none"> - PowerCenter 統合サービス - PowerCenter リポジトリサービス 	PowerCenter リポジトリ
Web サービス Hub コンソール	<ul style="list-style-type: none"> - PowerCenter 統合サービス - PowerCenter リポジトリサービス - Web サービス Hub 	PowerCenter リポジトリ

以下のアプリケーションサービスは、Informatica アプリケーションクライアントからアクセスされません。

- PowerExchange(R)リスナサービス。PowerExchange リスナでのバルクデータ移動および変更データキャプチャを管理します。PowerCenter 統合サービスは、リスナサービスを介して PowerExchange リスナに接続します。
- PowerExchange ロガーサービス。PowerExchange ロgger（Linux、UNIX、Windows 用）を管理して、変更データをキャプチャし、PowerExchange ロggerのログファイルにそのデータを書き込みます。変更データは、DB2 リカバリログ、Oracle REDO ログ、Microsoft SQL Server 分散データベース、あるいは i5/OS または z/OS システムのデータソースから取得できます。
- SAP BW サービス。SAP BI の RFC 要求をリスンし、PowerCenter 統合サービスでワークフローを実行して SAP BI から抽出したり、SAP BI にロードしたりするように要求します。

機能の可用性

Informatica 製品では、共通のアプリケーションセットが使用されます。使用できる製品機能は、製品ライセンスによって異なります。

次の表に、ライセンスオプションと各オプションで使用可能なアプリケーション機能を示します。

ライセンスオプション	Informatica Developer の機能	Informatica Analyst の機能
Data Quality	<ul style="list-style-type: none"> - マッピングの作成と実行 - マップレットとルールの作成と実行 - プロファイルを作成して実行します。プロファイルには、エンタープライズ検出、プライマリキーおよび外部キー検出、機能依存関係検出があります。 - 推測されたプロファイル結果をキュレーションする - スコアカードの作成と実行 - 参照テーブルの管理 - データソース内での例外レコードの特定 - PowerCenter へオブジェクトをエクスポートします。 	<ul style="list-style-type: none"> - エンタープライズ検出を含むプロファイリング - 検出検索を使用して、データとメタデータがプロファイルリポジトリのどこにあるかを探す - スコアカードの作成と実行 - 推測されたプロファイル結果をキュレーションする - プロファイリングルールの作成と実行 - 参照テーブルの管理 - マップレットへのルール仕様の作成、およびコンパイル - 例外レコードの確認および編集
Data Services	<ul style="list-style-type: none"> - 論理データオブジェクトモデルの作成 - Data Services トランスフォーマーを使用したマッピングの作成と実行 - SQL データサービスの作成 - Web サービスの作成 - PowerCenter へオブジェクトをエクスポートします。 	<ul style="list-style-type: none"> - 参照テーブルの管理
Data Services とプロファイリングオプション	<ul style="list-style-type: none"> - 論理データオブジェクトモデルの作成 - Data Services トランスフォーマーを使用したマッピングの作成と実行 - SQL データサービスの作成 - Web サービスの作成 - PowerCenter へオブジェクトをエクスポートします。 - Data Services トランスフォーマーを使用したルールの作成と実行 - プロファイリングの実行。 	<ul style="list-style-type: none"> - 参照テーブルの管理

注: Informatica 製品を Data Engineering ライセンスで使用する場合、一連の機能が異なる可能性があります。例えば、Data Engineering アプリケーションは PowerCenter と統合しません。例外レコード管理は、Data Engineering Quality では実行できません。

Informatica Analyst について

Informatica Analyst は、アナリストがエンタープライズ内のデータを分析、クレンジング、標準化、プロファイル、およびスコアリングするために使用できる Web ベースのアプリケーションクライアントです。

ライセンスによっては、ビジネスアナリストと開発者は、Analyst ツールをデータドリブンコラボレーションに使用します。カラムプロファイリングとルールプロファイリング、スコアカードの作成、および不良レコードと重複レコードの管理を実行できます。データ品質ソリューションでは開発者に対して、参照データの管理と提供も行うことができます。

Informatica Developer の概要

Developer tool は、データ統合、データ品質、データプロファイリング、データサービス、およびデータエンジニアリングに関連したソリューションを設計および実装するためのアプリケーションです。

Developer tool を使用して、メタデータのインポート、接続の作成、およびデータオブジェクトの作成を実行できます。また、Developer tool を使用して、プロファイル、マッピング、およびワークフローを作成して実行することもできます。

Informatica Developer のビュー

Developer tool のワークベンチには、エディタとビューがあります。エディタでマッピングなどのオブジェクトを編集します。Developer tool では、エディタ内で選択されているオブジェクトに応じたビューが表示されます。

追加のビューの選択、ビューの非表示、Developer tool のワークベンチ内の別の場所へのビューの移動などの操作が可能です。

表示するビューを選択するには、**[ウィンドウ] > [ビューの表示]** をクリックします。

Developer ツールでは、デフォルトで次のビューが表示されます。

[接続エクスプローラ] ビュー

リレーショナルデータベースへの接続が表示されます。

[データビューア] ビュー

ソースデータ、プロファイル結果を表示し、トランスフォーメーションの出力をプレビューします。

[オブジェクトエクスプローラ] ビュー

ドメインと、ドメイン内の設計時オブジェクトおよびランタイムオブジェクトを表示します。設計時オブジェクトは、モデルリポジトリ内のプロジェクトおよびフォルダに保存されます。ランタイムオブジェクトは、データ統合サービス上にランタイムアプリケーションの一部として保存されます。

[アウトライン] ビュー

[オブジェクトエクスプローラ] ビューで選択したオブジェクトに応じたオブジェクトが表示されます。

[進捗状況] ビュー

マッピング実行などの Developer tool での操作の進捗状況が表示されます。

[プロパティ] ビュー

エディタで選択されたオブジェクトのプロパティが表示されます。

次のビューは **[ビューの表示]** メニューを使用して表示することもできます。

【警告】 ビュー

接続ステータス警告が表示されます。

【チェックアウトされたオブジェクト】 ビュー

チェックアウトしたすべてのオブジェクトが表示されます。

【通知】 ビュー

ヒューマンタスクのすべての作業が完了したときにユーザーまたはグループに通知するオプションが表示されます。

【オブジェクトの依存関係】 ビュー

オブジェクトの表示、編集、または削除を行うときにオブジェクトの依存関係が表示されます。

【検索】 ビュー

検索結果が表示されます。また、検索オプションダイアログボックスを起動することもできます。

【タグ】 ビュー

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タグが表示されます。

Informatica Developer のようこそページ

Developer tool を初めて開くと、ようこそページが表示されます。ようこそページを使用して、Developer tool の設定方法を確認し、作業を始めてください。

ようこそページには次のオプションが表示されます。

- 概要: データ品質およびデータサービスソリューションの概要を表示するには、[概要] ボタンをクリックします。
- 最初の手順: Developer tool の設定および Informatica Data Quality Service と Informatica Data Services のレッスンへのアクセスに関する知識を深めるには、[最初の手順] ボタンをクリックします。
- チュートリアル: データ品質およびデータサービスソリューションのチュートリアルレッスンを表示するには、[チュートリアル] ボタンをクリックします。
- Web リソース: Informatica ナレッジベースへのリンクにアクセスするには、[Web リソース] ボタンをクリックします。Informatica How-To ライブラリにアクセスできます。Informatica How-To ライブラリには、Informatica Data Quality、Informatica Data Services、および他の Informatica 製品に関する記事が掲載されています。
- ワークベンチ: Developer tool での作業を開始するには、[ワークベンチ] ボタンをクリックします。

閉じた後にようこそページにアクセスするには、[ヘルプ] > [ようこそ] をクリックします。

チートシート

Developer ツールには、オンラインヘルプの一部としてチートシートが含まれています。チートシートは段階を追った手引きで、Developer ツールで 1 つまたは複数のタスクを完了するために役立ちます。

チートシートを完了して、タスクを完了して結果を確認します。例えば、チートシートを完了してリレーショナルデータオブジェクトをインポートおよびプレビューすると、リレーショナルデータベーステーブルはインポートされ、Developer ツールでのデータはプレビューされています。

チートシートにアクセスするには、[ヘルプ] > [チートシート] をクリックします。

データ品質およびプロファイリング

Developer tool のデータ品質機能を使用して、データの内容および構造を分析します。ビジネスニーズに合わせてデータを強化できます。

Developer tool を使用して、次の目標を実現するプロセスを実行します。

- データのプロファイリング。プロファイリングにより、データの内容と構造が明らかになります。プロファイリングは、どのようなデータプロジェクトにおいても重要です。プロファイリングを実施すると、データの長所と短所が明らかになり、プロジェクト計画の定義に役立ちます。
- スコアカードを作成して、データ品質を確認します。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。
- データ値の標準化。データを標準化し、プロファイルを実行したときに見つかるエラーや不整合を除去します。句読点、形式、およびスペルの差異を標準化できます。例えば、市区町村、州、および郵便番号の値に整合性を持たせることができます。
- レコードの解析。データレコードを解析してレコード構造を改善し、データから追加の情報を取り出します。自由形式のデータを含む単一のフィールドを、さまざまなタイプの情報を含む複数のフィールドに分割できます。レコードに情報を追加することもできます。例えば、顧客レコードに個人顧客または企業顧客のフラグを設定できます。
- 郵便アドレスの検証。アドレス検証では、郵便アドレスデータの精度と配信性を評価および強化します。アドレス検証では、アドレスレコードと国内の郵便運送業者の参照データを照合することで、アドレスの誤りを修正し、部分的なアドレスを完全なアドレスにします。アドレス検証では、郵便情報を追加して、郵便の配達に要する時間とコストを削減することもできます。
- 重複レコードの検索。重複レコード分析では、レコードセットのレコードを相互に比較し、選択したデータカラムの類似した値または一致する値を見つけます。フィールド値間の良好な一致を示す類似性レベルを設定します。照合計算で各カラムに割り当てられる相対的なウェイトを設定することもできます。例えば、名前よりも姓を優先することができます。
- データ品質ルールの作成と実行。Informatica には、事前作成されたルールが用意されています。このルールは、そのまま実行することも、プロジェクト目標に合わせて編集することもできます。Developer tool でルールを作成できます。
- Informatica ユーザーとの共同作業。モデルリポジトリに追加したルールと参照データテーブルは、Developer ツールおよび Analyst ツールで使用できます。ユーザーはプロジェクトを共同で実施できます。また、プロジェクトのさまざまな段階でさまざまなユーザーがオブジェクトの所有権を取得できます。
- PowerCenter へのマッピングのエクスポート。PowerCenter にマッピングをエクスポートして、物理データの統合のためにメタデータを再利用したり、Web サービスを作成したりできます。

チュートリアルの内容

HypoStores Corporation は、本社が Boston で、いくつかの州に店舗を持つ全国的な小売組織です。全国の店舗の業務データを本社のデータストアと定期的に統合しています。最近、Los Angeles に店舗を開店しました。

本社には、すべてのコンポジットおよび BI アプリケーションに共通のデータサービス層を提供する、管理者、開発者、およびアーキテクトから成る中央の ICC チームがあります。BI アプリケーションには、請求やマーケティングに使用されるマスター顧客データファイルが含まれる CRM システムが組み込まれています。

HypoStores Corporation は、以下のタスクを実行し、ロサンゼルスストアの業務データをボストン本社にあるデータと統合する必要があります。

- データ品質の問題がないか、ボストンとロサンゼルのデータを調査する。

- ロサンゼルスからのデータから情報を解析する。
- ボストンとロサンゼルスからのデータにまたがるアドレス情報を標準化する。
- CRM 用のデータに含まれている郵便アドレス情報の精度を検証する。

チュートリアル

スタートガイドには、レッスンとタスクから構成されるチュートリアルが含まれています。

レッスン

各レッスンでは、タスクを理解してレッスンを完了するために役立つ概念を紹介します。レッスンでは、内容全体のビジネス要件が示されています。レッスンの目標では、ビジネス要件を満たすために完了するタスクの概略を説明します。レッスンごとに、完了までの推定時間が示されています。レッスン内のタスクを完了すると、レッスンのサマリを確認できます。

ツール内の環境が設定されていない場合は、各チュートリアルの最初のレッスンでそれを行う方法が示されています。

タスク

タスクでは、段階を追った説明が示されています。示された順序ですべてのタスクを完了すると、レッスンは完了です。

Informatica Analyst のチュートリアル

このチュートリアルでは、アナリストが Analyst ツールへのログイン、プロジェクトとフォルダーの作成、プロファイルとルール作成、データのスコアリング、および参照テーブルの作成を行います。

実行可能なレッスンは、Informatica Data Quality、Informatica Data Services 製品のどれを所有しているかによって異なります。

次の表に、所有している製品によって異なる、実行可能なレッスンを示します。

レッスン	説明	製品
レッスン 1。 Informatica Analyst の設定	Analyst ツールにログインし、チュートリアルのレッスン用のプロジェクトとフォルダーを作成します。	Data Quality Data Services
レッスン 2。 データオブジェクトの作成	フラットファイルをデータオブジェクトとしてインポートし、データをプレビューします。	Data Quality
レッスン 3。 クイックプロファイルの作成	クイックプロファイルを作成して、データ品質を迅速に把握します。	Data Quality
レッスン 4。 カスタムプロファイルの作成	カスタムプロファイルを作成して、カラム、サンプリングオプション、およびドリルダウンオプションを設定します。	Data Quality
レッスン 5。 式ルールの作成	式ルールを作成して、カラム値を変更およびプロファイルします。	Data Quality
レッスン 6。 スコアカードの作成と実行	スコアカードを作成および実行して、長期間にわたってデータ品質の進捗を測定します。	Data Quality

レッスン	説明	製品
レッスン 7。プロファイル結果からの参照テーブルの作成	ソースデータの標準化に使用できる参照テーブルを作成します。	Data Quality Data Services
レッスン 8。参照テーブルの作成	参照テーブルを作成して、ソースデータと有効な標準値の間にリレーションシップを確立します。	Data Quality Data Services

Informatica Developer tool

このチュートリアルでは、Developer tool を使用していくつかのデータ品質操作を実行します。

Informatica Data Quality ユーザーは、Developer tool を使用してデータ品質を高めるプロセスを設計および実行します。また、Informatica Data Quality ユーザーは、Developer tool を使用してデータの内容と構造を分析するプロファイルを作成および実行します。

データ品質チュートリアルでは、次のレッスンを完了します。

レッスン 1. Informatica Developer の設定

ドメイン内でモデルリポジトリサービスによって管理されるモデルリポジトリへの接続を作成します。チュートリアルのレッスンの作業を格納するプロジェクトとフォルダを作成します。ドメインに複数のデータ統合サービスが含まれている場合は、サービスを選択します。

レッスン 2. 物理データオブジェクトのインポート

物理データオブジェクトに関連付けられた顧客のデータファイルのデータ品質プロセスを定義します。

レッスン 3. データのプロファイリング

プロファイリングにより、データの内容と構造が明らかになります。

プロファイリングには、結合分析（2 つのデータカラム間で有効な結合が可能であるかどうかを判断する分析の形式）が含まれます。

レッスン 4. データの解析

解析により、データレコードが拡充され、レコード構造が改良されます。データ内の有用な情報を検出することや、現在のデータから新しい情報を派生させることもできます。

レッスン 5. データの標準化

標準化により、プロファイリング中に検出されたデータのエラーや不整合が取り除かれます。

レッスン 6. 住所データの検証

住所の検証により、郵便の宛先の正確さや配達可能であるかどうかが評価され、住所の誤りや住所の抜けが修正されます。

チュートリアル的前提条件

チュートリアルのレッスンを開始する前に、少なくとも 1 つのノード設定で Informatica ドメインが実行されている必要があります。

インストーラには、レッスンを完了するために使用するチュートリアルファイルが含まれています。すべてのファイルは、クライアントのインストールディレクトリとサーバーのインストールディレクトリの両方に置かれています。

- チュートリアルファイルは、Developer ツールのインストールパスの次の場所に置かれています。

<Informatica Installation Directory>\clients\DeveloperClient\Tutorials

- チュートリアルファイルは、サービスのインストールパスの次の場所に置かれています。

<Informatica Installation Directory>\server\Tutorials

チュートリアルのレッスンでは次のファイルが必要です。

- All_Customers.csv
- Boston_Customers.csv
- LA_customers.csv

パート I: Informatica Analyst の 基本操作

この部には、以下の章があります。

- [レッスン 1. Informatica Analyst の設定, 19 ページ](#)
- [レッスン 2. データオブジェクトの作成, 22 ページ](#)
- [レッスン 3. デフォルトプロファイルの作成, 25 ページ](#)
- [レッスン 4. カスタムプロファイルの作成, 28 ページ](#)
- [レッスン 5. 式ルールの作成, 32 ページ](#)
- [レッスン 6. スコアカードの作成と実行, 35 ページ](#)
- [レッスン 7. プロファイルカラムからの参照テーブルの作成, 41 ページ](#)
- [レッスン 8. 参照テーブルの作成, 44 ページ](#)

第 2 章

レッスン 1. Informatica Analyst の設定

この章では、以下の項目について説明します。

- [Informatica Analyst の設定の概要, 19 ページ](#)
- [タスク 1. Informatica Analyst へのログイン, 20 ページ](#)
- [タスク 2. プロジェクトの作成, 20 ページ](#)
- [タスク 3. フォルダーの作成, 21 ページ](#)
- [Informatica Analyst の設定のまとめ, 21 ページ](#)

Informatica Analyst の設定の概要

このチュートリアルレッスンを開始する前に、Analyst ツールを設定する必要があります。Analyst ツールを設定するには、Analyst ツールにログインし、作業内容を格納するプロジェクトおよびフォルダーを作成します。

Informatica ドメインは、Informatica 環境を定義するノードとサービスの集合です。ドメイン内のサービスとして、アナリストサービスとモデルリポジトリサービスがあります。アナリストサービスは Analyst ツールを実行するサービスで、モデルリポジトリサービスはモデルリポジトリを管理するサービスです。Analyst ツールでの作業で作成したアセットは、モデルリポジトリに格納されます。

Analyst ツールでアセットを作成する前に、プロジェクトを作成する必要があります。プロジェクトには、Analyst ツールのアセットが含まれます。プロジェクトには、同じビジネス要件の一部であるデータオブジェクトなどの関連するアセットを格納するフォルダーも含まれます。

目標

このレッスンでは、以下の作業を完了します。

- Analyst ツールにログインする。
- Analyst ツールで作成したアセットを格納するプロジェクトを作成する。
- 関連するアセットを格納するプロジェクト内にフォルダーを作成する。

要件

このレッスンを開始する前に、次の要件を確認してください。

- 管理者が Administrator ツールでモデルリポジトリサービスとアナリストサービスの設定を完了している。

- Analyst ツールのホスト名とポート番号がわかっている。
- アナリストサービスにアクセスするためのユーザー名とパスワードがわかっている。この情報は管理者から提供されます。

時間

このレッスンの完了には、5～10 分見ておいてください。

タスク 1. Informatica Analyst へのログイン

Analyst ツールにログインし、チュートリアルを開始します。

1. Microsoft Internet Explorer または Google Chrome ブラウザを起動します。
2. [アドレス] フィールドに、次に示す Informatica Analyst の URL を入力します。
`http[s]://<fully qualified host name>:<port number>/analyst`
3. ドメインで LDAP またはネイティブ認証を使用する場合、ログインページ用のユーザー名とパスワードを入力します。
4. [ネイティブ] または特定のセキュリティドメインの名前を選択します。
Informatica ドメインで LDAP または Kerberos 認証を使用している場合は、[セキュリティドメイン] フィールドが表示されます。ユーザーアカウントが所属するセキュリティドメインがわからない場合は、Informatica ドメイン管理者に問い合わせてください。
5. [ログイン] をクリックします。
Analyst ツールにより、[開始] ワークスペースが開きます。

タスク 2. プロジェクトの作成

このタスクでは、Analyst ツールで作成したアセットを含むプロジェクトを作成します。プロジェクトのフォルダーを含むようにチュートリアルプロジェクトを作成します。

1. [管理] ヘッダーの [プロジェクト] をクリックします。
[プロジェクト] ワークスペースが表示されます。
2. [アクション] メニューから、[新規] > [プロジェクト] を選択します。
[新規プロジェクト] ウィンドウが表示されます。
3. プロジェクト名として「Tutorial_自分の名前」を入力します。
4. [OK] をクリックします。

タスク 3. フォルダの作成

このタスクでは、関連するアセットを格納するフォルダを作成します。フォルダは、プロジェクトまたは別のフォルダ内に作成することができます。Data Quality プロジェクトに関連するアセットを格納する Customers という名前のフォルダを作成します。

1. **【プロジェクト】** パネルで、チュートリアルプロジェクトを選択します。
2. **【アクション】** メニューから、**【新規】** > **【フォルダ】** をクリックします。
【新しいフォルダ】 ウィンドウが表示されます。
3. フォルダ名に「Customers」を入力します。
4. **【OK】** をクリックします。
フォルダにはチュートリアルプロジェクトの下に表示されます。

Informatica Analyst の設定のまとめ

このレッスンでは、Analyst ツールではプロジェクトおよびフォルダにアセットが格納されることを学習しました。プロジェクトとフォルダはモデルリポジトリに格納されます。Analyst ツールは、アナリストサービスで実行されます。モデルリポジトリサービスでは、モデルリポジトリが管理されます。アナリストサービスとモデルリポジトリサービスは、Informatica ドメイン内のアプリケーションサービスです。

Analyst ツールにログインし、プロジェクトとフォルダを作成しました。

これで、Analyst ツールを使用してこのチュートリアルその他のレッスンを実行できます。

第 3 章

レッスン 2. データオブジェクトの作成

この章では、以下の項目について説明します。

- [データオブジェクトの作成の概要, 22 ページ](#)
- [タスク 1. フラットファイルデータオブジェクトの作成, 23 ページ](#)
- [タスク 2. データオブジェクトのプロパティの表示, 23 ページ](#)
- [データオブジェクトの作成のまとめ, 24 ページ](#)

データオブジェクトの作成の概要

Analyst ツールで、データオブジェクトはフラットファイルまたはリレーショナルデータベーステーブルに基づいたデータの表現です。フラットファイルまたはテーブルオブジェクトを作成し、そのフラットファイルまたはリレーショナルデータベーステーブル内のデータに対してプロファイルを実行します。Analyst ツール内でフラットファイルデータオブジェクトを作成すると、Analyst ツールを実行するマシン上のフラットファイルキャッシュにファイルをアップロードすることや、フラットファイルが格納されるネットワーク上の場所を指定することができます。

内容

HypoStores は、Los Angeles の顧客データをフラットファイルに保存しています。HypoStores は、そのデータをプロファイルおよび分析し、データ品質タスクを実行する必要があります。

目標

このレッスンでは、以下の作業を完了します。

1. フラットファイルをフラットファイルキャッシュの場所にアップロードし、データオブジェクトを作成します。
2. フラットファイルデータオブジェクトのデータをプレビューします。

要件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスン 1 を完了している。
- LA_Customers.csv フラットファイルがある。このファイルは、<Installation Root Directory>\<Release Version>\clients\DeveloperClient\Tutorials フォルダにあります。

時間

このタスクの完了には、5～10 分見ておいてください。

タスク 1。フラットファイルデータオブジェクトの作成

このタスクでは、LA_Customers ファイルからフラットファイルデータオブジェクトを作成します。

1. Analyst ツールで **【新規】** > **【フラットファイルデータオブジェクト】** をクリックします。
フラットファイルの追加ウィザードが表示されます。
2. **【参照およびアップロード】** を選択し、**【参照】** をクリックします。
3. LA_Customers.csv の場所を参照し、**【開く】** をクリックします。
4. **【次へ】** をクリックします。
インポートの種類の選択パネルには、**【区切り記号付き】** および **【固定幅】** オプションが表示されます。
【区切り記号付き】 オプションを選択します。デフォルトのオプションは **【区切り記号付き】** です。
5. **【次へ】** をクリックします。
6. **【データで使用されている区切り文字とテキスト修飾子を指定する】** で、テキスト修飾子として **【二重引用符】** を選択します。
7. **【インポートする行を指定する】** で、**【先頭行からインポート】** を選択してブランク以外の先頭行のカラム名をインポートします。
【プレビュー】 パネルが更新されて、先頭行のカラムの見出しが表示されます。
8. **【次へ】** をクリックします。
カラム属性パネルに、カラムごとのデータ型、精度、スケール、および形式が表示されます。
9. **【次へ】** をクリックします。
【名前】 フィールドに LA_Customers が表示されます。
10. **【チュートリアル】** プロジェクトおよび **【Customers】** フォルダを選択します。
11. **【完了】** をクリックします。
Customers フォルダのフォルダー内容にデータオブジェクトが表示されます。

タスク 2。データオブジェクトのプロパティの表示

このタスクでは、LA_Customers データオブジェクトのプロパティを表示できます。

1. **【開く】** をクリックして **【ライブラリ】** ワークスペースを開きます。
2. **【ライブラリ】** ワークスペースで、**【アセット】** パネルの **【データオブジェクト】** をクリックします。
【データオブジェクト】 パネルにデータオブジェクトのリストが表示されます。
3. **LA_Customers** フラットファイルをクリックします。
【データプレビュー】 パネルとともに、LA_Customers データオブジェクトから取得したデータが表示されます。

4. **【データプレビュー】** パネルで、LA_Customers データオブジェクトの構造および内容を確認します。
Analyst ツールによって、フラットファイルデータオブジェクトの最初の 100 行が表示されます。
5. **【プロパティ】** をクリックします。
【プロパティ】 パネルには、データオブジェクトの名前、タイプ、説明、および場所が表示されます。データオブジェクトのカラム名とカラムプロパティも表示できます。

データオブジェクトの作成のまとめ

このレッスンでは、データオブジェクトがフラットファイルまたはリレーショナルデータベースソースに基づいたデータの表現であることを学習しました。フラットファイルデータオブジェクトを作成し、その中のデータをプレビューできることを学習しました。

フラットファイルをアップロードしてフラットファイルデータオブジェクトを作成し、データオブジェクトのデータをプレビューし、データオブジェクトのプロパティを表示しました。

データオブジェクトの作成後、レッスン 3 でデータオブジェクトのデフォルトプロファイルを作成し、レッスン 4 でデータオブジェクトのカスタムプロファイルを作成します。

第 4 章

レッスン 3. デフォルトプロファイルの作成

この章では、以下の項目について説明します。

- [デフォルトプロファイルの作成の概要, 25 ページ](#)
- [タスク 1. デフォルトプロファイルの作成と実行, 26 ページ](#)
- [タスク 2. サマリビューでのプロファイル結果の表示, 26 ページ](#)
- [デフォルトプロファイルの作成のサマリ, 27 ページ](#)

デフォルトプロファイルの作成の概要

プロファイルは、データの内容と構造に基づいたデータ品質の分析です。デフォルトプロファイルは、デフォルトのオプションで作成するプロファイルです。プロファイルにカラムとオプションを設定せずにプロファイル結果を取得するには、デフォルトプロファイルを使用します。

データ品質プロジェクトの開始時に、デフォルトプロファイルを作成および実行してデータの品質を分析します。デフォルトプロファイルオブジェクトの作成時には、分析するデータオブジェクトとデータオブジェクトカラムを選択します。デフォルトプロファイルでは、プロファイルカラムとオプションの設定がスキップされます。Analyst ツールは、フラットファイルデータオブジェクトの実フラットファイルでプロファイリングを実行します。

内容

HypoStores は、新たに確保した Los Angeles オフィスからデータウェアハウスにデータを組み込む必要があります。データはデータウェアハウスに組み込む前にクレンジングする必要があります。あなたは、データの品質を評価し、そのデータのクレンジングを担当する開発者に情報を渡すアナリストです。プロファイル結果を迅速に表示し、データの品質についての基本的な考えを得る必要があります。

目標

このレッスンでは、以下の作業を完了します。

1. LA_Customers フラットファイルデータオブジェクトのデフォルトプロファイルを作成および実行します。
2. プロファイル結果を表示します。

前提条件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスンの1から2までを完了している。

時間

このレッスンの完了には、5～10分見ておいてください。

タスク 1. デフォルトプロファイルの作成と実行

このタスクでは、データオブジェクト内のすべてのカラムのデフォルトプロファイルを作成し、デフォルトのサンプリングオプションとドリルダウンオプションを使用します。

1. **【ライブラリ】** ワークスペースの **【アセット】** パネルでデータオブジェクトを選択します。
2. データオブジェクトを右クリックして、**【プロファイルの作成】** を選択します。
新規プロファイルウィザードが表示されます。
3. **【全般プロパティの指定】** 画面では、デフォルトで、名前、説明、場所が入力されています。
4. **【次へ】** をクリックします。
5. **【ソースの選択】** 画面では、デフォルトでデータオブジェクトが表示されます。**【カラムを選択】** ペインでカラムを表示できます。
6. **【次へ】** をクリックします。
7. **【設定の指定】** 画面では、次のオプションがデフォルトで選択されています。
 - **カラムプロファイルの実行**
 - **【次に対してプロファイルを実行】** ペインの **【すべての行】**
 - **【ドリルダウン】** ペインの **【ライブ】**
 - **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。**
 - **【ランタイム環境】** ペインの **【ネイティブ】** オプション。
8. **【次へ】** をクリックします。
9. **【ルールとフィルタの指定】** 画面で、**【保存して実行】** をクリックしてプロファイルを作成し、実行します。

Analyst ツールでプロファイルを作成して実行します。サマリビューにプロファイル結果が表示されます。

タスク 2. サマリビューでのプロファイル結果の表示

このタスクでは、LA_Customers プロファイルのサマリビューを使用して、プロファイル結果の簡単な概要を表示します。

1. **【ライブラリ】** > **【アセット】** > **【プロファイル】** ペインで、**LA_Customers** プロファイルをクリックします。
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、**【フィルタ基準】** ペインの **【カラム】** をクリックして、カラムのプロファイル結果を表示します。
3. パーセントで値を表示するには、水平棒グラフ上にマウスのポインタを移動します。
4. **【NULL 重複なし 重複あり %】** セクションに、カラムの Null 値、重複しない値、重複する値をパーセントで表示できます。
5. **【パターン】** セクションでは、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。
6. **【長さ】** セクションでは、カラム内の値の最小長および最大長を表示できます。
7. **【値】** セクションでは、カラム内の最小値および最大値を表示できます。
8. **【データ型】** セクションでは、マウスポインタを値の上に移動すると、カラムのすべての推測されたデータ型と文書化されたデータ型を表示できます。
9. **【データドメイン】** セクションでは、マウスポインタを値の上に移動すると、カラムのすべての推測されるデータドメインを表示できます。
10. 異常値データを表示するには、**【アクション】** > **【異常値の検出】** をクリックして、プロファイル結果で異常値を検出します。
11. プロファイル結果に含まれる異常値を表示するには、**【パターン異常値】** または **【値頻度の異常値】** フィルタをクリックします。
12. 詳細ビューにカラムのプロファイル結果を表示するには、カラム名をクリックします。

デフォルトプロファイルの作成のサマリ

このレッスンでは、デフォルトプロファイルを使用すると、プロファイルにカラムと行のサンプリングオプションを設定せずにプロファイル結果が表示されることを学習しました。データ品質プロジェクトの開始時に、デフォルトプロファイルを作成および実行してデータの品質を分析することも学習しました。また、Analyst ツールにより、フラットファイルデータオブジェクトの実フラットファイルでプロファイリングが実行されることを学習しました。

デフォルトプロファイルを作成し、プロファイル結果を分析しました。NULL 値やデータ型など、プロファイル内のカラムに関する詳細な情報を取得しました。また、カラム値とパターンを使用して、データ品質の問題を特定しました。

クイックプロファイルの結果を分析すると、次のタスクを完了できます。

- カスタムプロファイルを作成して、プロファイルからカラムを除外し、関心のあるカラムのみを組み込みます。
- 式ルールを作成して仮想カラムを作成し、それらをプロファイルします。
- カラムの有効な値を含める参照テーブルを作成します。

第 5 章

レッスン 4. カスタムプロファイルの作成

この章では、以下の項目について説明します。

- [カスタムプロファイルの作成の概要, 28 ページ](#)
- [タスク 1. カスタムプロファイルの作成, 29 ページ](#)
- [タスク 2. プロファイルの実行, 30 ページ](#)
- [タスク 3. プロファイル結果でのドリルダウン, 30 ページ](#)
- [カスタムプロファイルの作成のまとめ, 31 ページ](#)

カスタムプロファイルの作成の概要

プロファイルは、データの内容と構造に基づいたデータ品質の分析です。カスタムプロファイルは、プロファイリングを迅速に行うために、カラム、サンプリングオプション、およびドリルダウンオプションの設定時に作成するプロファイルです。ソースのサンプル行を選択するには、サンプリングオプションを設定します。ソースデータまたはステージングされたデータでデータ行にドリルダウンするには、ドリルダウンオプションを設定します。プロファイルは、Hadoop またはネイティブの環境で実行することを選択できます。

データ品質プロジェクトの開始時に、プロファイルを作成および実行してデータの品質を分析します。プロファイルオブジェクトを作成するときは、プロファイルを実行するデータオブジェクトとデータオブジェクトカラムを選択します。

内容

HypoStores は、新たに確保した Los Angeles オフィスからデータウェアハウスにデータを組み込む必要があります。HypoStores は、LA の顧客データファイル内の顧客層データの品質にアクセスする必要があります。あなたは、データの品質を評価し、そのデータのクレンジングを担当する開発者に情報を渡すアナリストです。

目標

このレッスンでは、以下の作業を完了します。

1. フラットファイルデータオブジェクトのカスタムプロファイルを作成し、NULL 値のカラムを除外します。
2. プロファイルを実行して、CustomerTier カラムの内容と構造を分析します。
3. プロファイル結果の行にドリルダウンします。

前提条件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスン 1、2、および 3 を完了している。

時間

このレッスンの完了には、5～10 分見ておいてください。

タスク 1. カスタムプロファイルの作成

このタスクでは、カスタムプロファイルを作成します。カスタムプロファイルの作成時には、プロファイルを実行するデータオブジェクトとカラムを選択します。サンプリングオプションとドリルダウンオプションも設定します。

1. **【新規】 > 【プロファイル】** をクリックします。
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、次のオプションを設定します。
 - **【名前】** フィールドに、「**Profile_LA_Customers**」と入力します。
 - **【場所】** フィールドで、**Customers** フォルダを選択します。
4. **【次へ】** をクリックします。
5. **【ソースの選択】** 画面で、**【選択】** 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が表示されます。
6. **【データオブジェクトの選択】** ダイアログボックスで、**【LA_Customers】** を選択します。**【OK】** をクリックします。
7. **【ソースの選択】** 画面で、Address2、Address3、および City2 の各カラムをクリアします。
8. **【次へ】** をクリックします。
9. **【設定の指定】** 画面で、次のオプションを設定します。
 - **【カラムプロファイルの実行】**。
 - **【次に対してプロファイルを実行】** ペインで、**【ランダムサンプリング】** オプションを選択します。
 - **【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】** オプションを選択します。
 - **【ドリルダウン】** ペインで **【ステージングされた】** オプションを選択します。
10. **【次へ】** をクリックします。
11. **【ルールとフィルタの指定】** 画面で **【保存して完了】** をクリックして、プロファイルを作成、実行します。
Analyst ツールによってプロファイルが作成され、**【検出】** ワークスペースにプロファイルが表示されます。プロファイルを実行して結果を表示する必要があります。

タスク 2. プロファイルの実行

このタスクでは、プロファイルを実行してデータオブジェクトについてプロファイリングを実行し、プロファイル結果を表示します。Analyst ツールは、フラットファイルデータオブジェクトのステージングされたフラットファイルでプロファイリングを実行します。

1. **【検出】** ワークスペースが表示されていることを確認します。
ワークスペースに **Profile_LA_Customers_Custom** プロファイルが表示されています。
2. ワークスペースで、**Profile_LA_Customers_Custom** をクリックします。
3. プロファイル画面が表示されます。この画面でプロファイルの編集またはプロファイルの実行を選択できます。**【実行】** をクリックします。
4. サマリビューにプロファイル結果が表示されます。

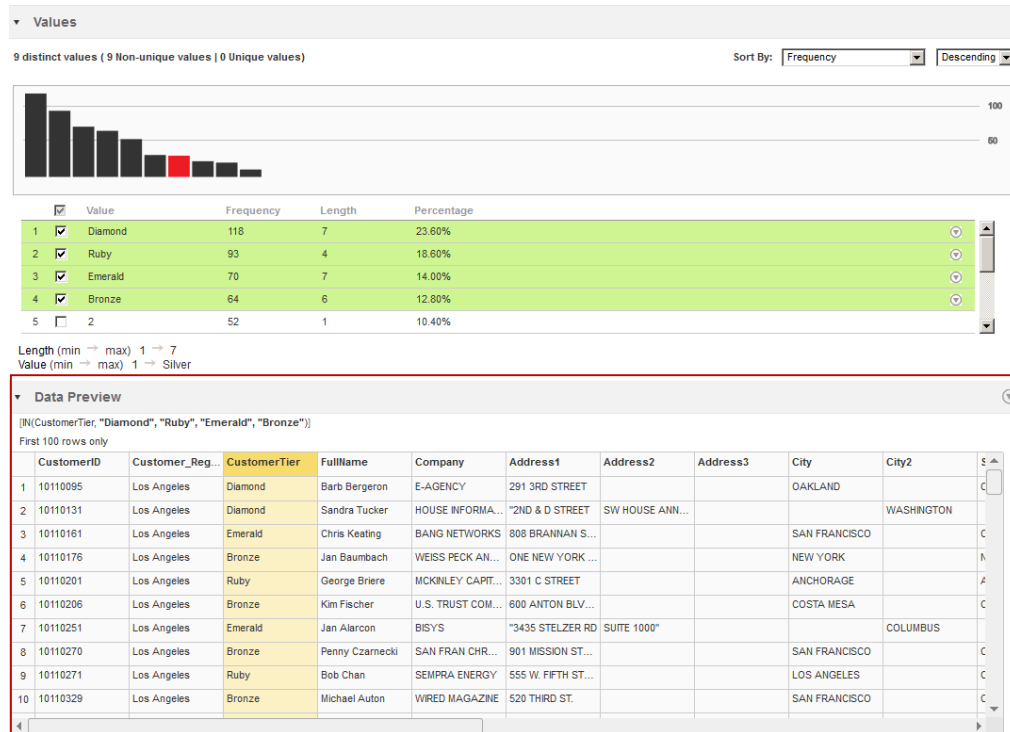
タスク 3. プロファイル結果でのドリルダウン

このタスクでは、CustomerTier カラムの値でドリルダウンして、プロファイルのデータオブジェクト内のソース行を確認します。

1. **Profile_LA_Customers** プロファイルで、プロファイル結果のサマリビューを開いていることを確認します。
2. **【CustomerTier】** カラムをクリックします。
カラムのプロファイル結果が詳細ビューに表示されます。

3. 詳細ビューで、Diamond、Ruby、Emerald、および Bronze の値を選択します。【値】 ペインで値を右クリックし、【ドリルダウン】を選択します。

値が Diamond、Ruby、Emerald、Bronze のカラムの行が【データプレビュー】 ペインに表示されます。次の画像は、Diamond、Ruby、Emerald、または Bronze の値でドリルダウンするときに【データプレビュー】 ペインに表示されるドリルダウン結果を示しています。



【データプレビュー】 ペインには、選択したカラムの最初の 100 行が表示されます。【データプレビュー】 ペインのタイトルには、ソースカラムで使用されたロジックが表示されます。

カスタムプロファイルの作成のまとめ

このレッスンでは、プロファイルされるカラムを設定できることと、サンプリングオプションとドリルダウンオプションを設定できることを学習しました。ドリルダウンしてカラム値の基礎となる行を確認できることと、カラム値を表示するときに含まれるカラムを設定できることを学習しました。

CustomerTier カラムが含まれるカスタムプロファイルを作成し、結果内の CustomerTier カラムの基礎となる行にドリルダウンしました。

このカスタムプロファイルオブジェクトを使用して、レッスン 5 で式ルールを作成します。

第 6 章

レッスン 5. 式ルールの作成

この章では、以下の項目について説明します。

- [式ルールの作成の概要, 32 ページ](#)
- [タスク 1. 式ルールの作成とプロファイルの実行, 33 ページ](#)
- [タスク 2. 式ルールの出力の表示, 33 ページ](#)
- [タスク 3. 式ルールの編集, 34 ページ](#)
- [式ルールの作成のまとめ, 34 ページ](#)

式ルールの作成の概要

式ルールは、式関数とソースカラムを使用してルールの論理を定義します。式ルールを作成して、Analyst ツールのプロファイルに追加することができます。式ルールは、1 つまたは複数のプロファイルと関連付けることができます。

式ルールの出力は、プロファイル内の仮想カラムです。プロファイルを実行すると、Analyst ツールによって仮想カラムがプロファイルされます。

式ルールを使用して、ソースカラムを検証したり、ソースカラムの値に基づいて追加のソースカラムを作成したりすることができます。

内容

HypoStores は、新たに確保した Los Angeles オフィスからデータウェアハウスにデータを組み込む必要があります。HypoStores は、顧客名を分析して、名と姓に分離する必要があります。HypoStores は、式ルールを使用して名と姓が含まれるカラムを別々の仮想カラムに解析し、それらのカラムをプロファイルする必要があります。HypoStores は、これらのルールの出力を分析する必要がある他のアナリストがそれらのルールを使用できるようにする必要があります。

目標

このレッスンでは、以下の作業を完了します。

1. 式ルールを作成して、FullName カラムを名カラムと姓カラムに分離します。氏名から名を分離するルールを作成します。名から姓を分離するもう 1 つのルールを作成します。これらのルールは Profile_LA_Customers プロファイルに対して作成します。
2. プロファイルを実行し、プロファイル内のルールの出力を表示します。
3. ルールを編集して、他の Analyst ツールユーザーが使用できるようにします。

前提条件

このレッスンを開始する前に、次の要件を確認してください。

- レッスン 1～4 を完了している。

時間

このレッスンの所要時間は 10～15 分です。

タスク 1. 式ルールを作成とプロファイルの実行

このタスクでは、2 つの正規表現を作成して、FullName カラムを FirstName および LastName という名前の 2 つの仮想カラムに解析します。ルール名は FirstName と LastName です。

1. **【ライブラリ】** ワークスペースで、**Profile_LA_Customers** プロファイルをクリックします。
サマリビューにプロファイル結果が表示されます。
2. **【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】** 画面をクリックします。
4. **【ルール】** ペインで、**【アクション】** > **【ルールの作成】** をクリックします。
5. **【名前】** フィールドに、「**FirstName**」と入力します。
6. **【式】** セクションで次の式を入力して、**【名前】** カラムから名前を分離します。
`SUBSTR(FullName,1,INSTR(FullName,' ', -1,1) - 1)`
7. **【検証】** をクリックします。
8. **【OK】** をクリックします。
9. 手順 4～8 を繰り返して、**LastName** という名前のルールを作成します。次の式を入力して、**【名前】** カラムから姓を分離します。
`SUBSTR(FullName,INSTR(FullName,' ', -1,1),LENGTH(FullName))`
10. **【保存して実行】** をクリックし、プロファイルを保存して実行します。

タスク 2. 式ルールの出力の表示

このタスクでは、プロファイルの実行後に名と姓を分離する式ルールの出力を表示します。

1. サマリビューで **【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
2. プロファイルウィザードで、**【ソースの選択】** をクリックします。
3. **【ソースの選択】** 画面で、ツールバーの **【名前】** の横にあるチェックボックスを選択して、すべてのカラムをクリアします。
【カラム】 セクションでは、少なくとも 1 つのカラムを選択する必要があるため、デフォルトでカラムの 1 つが選択されています。
4. **【FullName】** カラム、**【FirstName】** ルール、および **【LastName】** ルールを選択します。
5. 他のカラムが選択されていれば、それをクリアします。

6. **【保存して実行】** をクリックします。
サマリビューにプロファイル結果が表示されます。
7. **【FirstName】** ルールをクリックすると、ルールのプロファイル結果が詳細ビューに表示されます。
8. **【値】** ペインで値を選択します。その値を右クリックして、**【ドリルダウン】** をクリックします。
【FullName】 カラムの値、および **【FirstName】** ルールと **【LastName】** ルールが、他のカラムの値と一緒に **【データプレビュー】** ペインに表示されます。Analyst ツールにより、**【FullName】** カラムが名と姓に分離されています。

タスク 3. 式ルールの編集

このタスクでは、すべての Analyst ツールユーザーに対して式ルールを再利用可能および使用可能にします。

1. **Profile_LA_Customers** プロファイルのサマリビューで、**【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。
2. **【ルールとフィルタの指定】** 画面をクリックします。
3. **【ルールとフィルタの指定】** 画面で、**【FirstName】** ルールを選択して、**【アクション】** > **【ルールの編集】** をクリックします。
【ルールの編集】 ダイアログボックスが表示されます。
4. **【このルールを再利用可能なルールとして保存しますか?】** オプションを選択してから、ルールを保存する場所を選択します。
5. **【OK】** をクリックします。
6. **【LastName】** ルールを選択し、手順 3~5 を繰り返します。
7. **【保存して完了】** をクリックしてプロファイルを保存します。

Analyst ツールのすべてのユーザーは、FirstName ルールと LastName ルールを使用して、姓名を含むカラムを個別のカラムに分割できます。

式ルールの作成のまとめ

このレッスンでは、式ルールが式関数とソースカラムを使用してルールの論理を定義することを学習しました。式ルールの出力はプロファイル内の仮想カラムであることを学習しました。プロファイルを実行すると、Analyst ツールによって仮想カラムが組み込まれます。

2つの式ルールを作成してプロファイルに追加し、そのプロファイルを実行しました。ルールの出力を表示し、それらのルールをすべての Analyst ツールユーザーが使用できるようにしました。

第 7 章

レッスン 6. スコアカードの作成と実行

この章では、以下の項目について説明します。

- [スコアカードの作成と実行の概要, 35 ページ](#)
- [タスク 1. プロファイル結果からのスコアカードの作成, 36 ページ](#)
- [タスク 2. スコアカードの実行, 37 ページ](#)
- [タスク 3. スコアカードの表示, 37 ページ](#)
- [タスク 4. スコアカードの編集, 38 ページ](#)
- [タスク 5. しきい値の設定, 39 ページ](#)
- [タスク 6. スコアの傾向グラフの表示, 39 ページ](#)
- [スコアカードの作成と実行のまとめ, 39 ページ](#)

スコアカードの作成と実行の概要

スコアカードは、グラフィカル形式の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です。長期間にわたってデータ品質の進捗を測定して監視するには、スコアカードを使用します。

スコアカードを作成するには、プロファイル内のカラムをメトリックとしてスコアカードに追加し、メトリックにウェイトを割り当て、スコアしきい値を設定します。ソースデータに基づいて、フィルタをスコアカードに追加できます。スコアカードを実行するには、メトリックの有効値を選択します。スコアカードを実行して、メトリックのスコアを確認します。

スコアカードには、プロファイル内のカラムの値の頻度がスコアとして表示されます。スコアにはメトリックの有効値の割合が反映されます。

内容

HypoStores は、新たに確保した Los Angeles オフィスからデータウェアハウスにデータを組み込む必要があります。組織は、データをマージする前に、さまざまな顧客層と状態のデータのデータ品質が分析されていることを確認する必要があります。あなたは、データ品質分析の実行の進捗を監視するアナリストです。顧客層プロファイルカラムと状態プロファイルカラムからスコアカードを作成し、データ品質のしきい値を設定し、スコアの傾向グラフを表示して時間の経過とともにスコアがどのように改善するかを確認する必要があります。

目標

このレッスンでは、以下のタスクを完了します。

1. Profile_LA_Customers_Custom プロファイルの結果からスコアカードを作成して、CustomerTier カラムと State カラムのスコアを表示します。
2. スコアカードを実行して、CustomerTier カラムと State カラムのスコアを生成します。
3. スコアカードを表示して、各カラムのスコアを確認します。
4. スコアカードを編集して、スコアのさまざまな有効値を指定します。
5. スコアのしきい値を設定し、スコアカードを実行します。
6. スコアの傾向グラフを表示して、時間の経過とともにスコアがどのように改善するかを確認します。

前提条件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスンの 1 から 5 までを完了している。

時間

このレッスンのタスクの所要時間は 15 分です。

タスク 1. プロファイル結果からのスコアカードの作成

このタスクでは、Profile_LA_Customers_Custom プロファイルからスコアカードを作成して、CustomerTier カラムと State カラムの値をスコアリングします。

1. **【ライブラリ】** ワークスペースで、**Profile_LA_Customers** プロファイルをクリックします。
プロファイル結果のサマリビューが表示されます。
2. サマリビューで **【CustomerTier】** カラムを選択し、そのカラムを右クリックして、**【次に追加】 > 【スコアカード】** を選択します。
スコアカードに追加ウィザードが表示されます。
3. **スコアカードに追加ウィザード**では、デフォルトで **【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。
4. **【手順 2/8】** 画面で、スコアカード名を「**sc_LA_Customer**」と入力し、スコアカードの場所の **Customers** フォルダに移動します。
5. **【次へ】** をクリックします。
6. **【手順 3/8】** 画面で **【CustomerTier】** カラムと **【State】** カラムを選択し、スコアカードに追加します。
7. **【次へ】** をクリックします。
8. **【手順 4/8】** 画面では、メトリックのフィルタを作成、編集、または削除できます。このチュートリアルでは、スコアカードフィルタを作成しません。**【次へ】** をクリックします。
9. **【手順 5/8】** 画面の **【メトリック】** ペインで、**CustomerTier** メトリックを選択します。
10. **【スコアで使用: 値】** ペインですべての値を選択し、**【すべて追加】** ボタンをクリックして、選択した値を **【有効な値】** セクションに移動します。
Shift キーを使用して複数の値を選択します。

11. **【メトリック】** ペインで、**State** メトリックを選択し、**【スコアで使用: 値】** セクションで 2 文字の州コードを持つ値を選択します。
12. **【追加】** ボタンをクリックして、値を **【有効な値】** セクションに移動します。
セクションの最上部に有効値の総数および有効値の割合が表示されます。
13. **【メトリック】** セクションの各メトリックで、**【メトリックしきい値】** セクションのスコアしきい値のデフォルト設定をそのまま使用します。
14. **【次へ】** をクリックします。
15. **【手順 6/8】** 画面で、必要に応じて、メトリックを追加するメトリックグループを選択できます。デフォルトでは、Analyst ツールはメトリックを **【デフォルト】** メトリックグループに追加します。
16. **【次へ】** をクリックします。
17. **【手順 7/8】** 画面の **【デフォルト - メトリック】** ペインで、**CustomerTier** メトリックの **【ウェイト】** カラムをダブルクリックします。
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。
18. **CustomerTier** および **State** メトリックのウェイトを入力します。
19. **【次へ】** をクリックします。
20. **【手順 8/8】** 画面では、デフォルトで **【ネイティブ】** オプションが選択されています。**【保存】** をクリックして、スコアカードを作成します。
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。

タスク 2. スコアカードの実行

このタスクは、sc_LA_Customer スコアカードを実行して、CustomerTier カラムと State カラムのスコアを生成します。

1. **【ライブラリ】** ワークスペースで、**【アセット】** > **【スコアカード】** をクリックします。
2. **【スコアカード】** ペインでスコアカードを選択します。
3. **【アクション】** > **【開く】** をクリックします。
スコアカードが表示されます。
4. **【アクション】** > **【スコアカードの実行】** をクリックします。
【スコアカードの実行】 ダイアログボックスが表示されます。
5. ダイアログボックスの設定を確認し、**【実行】** をクリックします。
【スコアカード】 ワークスペースに、CustomerTier カラムと State カラムのスコアが表示されます。

タスク 3. スコアカードの表示

このタスクでは、sc_LA_Customer スコアカードを表示して、CustomerTier カラムと State カラムのスコアを確認します。

1. 表示する State スコアが含まれる **State** 行を選択します。

[**sc_LA_Customer - メトリック**] セクションでは、次のスコアカードのプロパティを表示できます。

- スコアカード名。
 - スコアカード内の行の総数。
 - 無効な行数。
 - スコアと水平棒グラフ。
 - スコアの傾向。スコアの傾向をクリックすると、**【傾向グラフの詳細】** 画面にその傾向を視覚的に表示することができます。
 - メトリックのウェイト。
 - 無効なデータのコスト。
 - コスト傾向。
 - データオブジェクト。データオブジェクトのデータプレビューを **【検出】** ワークスペースで表示するには、データオブジェクトをクリックします。
 - カラムまたはルール名。
 - ソースのタイプ。
 - ドリルダウンアイコン。
2. **【状態】** 行でドリルダウンアイコンをクリックします。
【状態】 カラムで無効なスコアは、**【ドリルダウン】** ペインの **【無効な行】** セクションに表示されます。
 3. **【有効な行】** を選択し、State カラムに対して有効なスコアを表示します。
 4. **【CustomerTier】** 行でドリルダウンアイコンをクリックします。
【CustomerTier】 カラムのスコアは、すべて有効です。

タスク 4. スコアカードの編集

このタスクでは、sc_LA_Customer スコアカードを編集して、Ruby の値を CustomerTier スコアに対して有効ではないと指定します。

1. **【スコアカード】** ワークスペースを開いており、sc_LA_Customer スコアカードが開かれていることを確認します。
2. **【アクション】** > **【編集】** > **【メトリック】** 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。
3. **【メトリック】** セクションで **CustomerTier** を選択します。
4. **【スコアで使用: 値】** セクションで、**Ruby** を **【有効な値】** セクションから **【使用可能な値】** セクションに移動します。
【メトリックしきい値】 セクションのデフォルト設定をそのまま使用します。
5. **【保存して実行】** をクリックし、スコアカードの変更内容を保存して、スコアカードを実行します。
6. CustomerTier スコアをもう一度表示します。
CustomerTier のスコアが 81.4 パーセントに変更されています。

タスク 5. しきい値の設定

このタスクでは、sc_LA_Customer スコアカード内の State スコアのしきい値を設定して、State カラムのデータの許容範囲を決定します。CA などの 2 文字のコードの値は許容され、Calif などの 2 文字を超えるコードは許容されません。

1. **【スコアカード】** ワークスペースを開いており、sc_LA_Customer スコアカードが開かれていることを確認します。
2. **【アクション】** > **【編集】** > **【メトリック】** 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。
3. **【メトリック】** セクションで、**State** を選択します。
4. **【メトリックしきい値】** セクションの **【正常】** スコアと **【許可不可】** スコアに次のように入力します。**【正常】** スコアに 90~100%、**【許可不可】** スコアに 0~50%、**【許可】** スコアに 51%~89%の範囲を入力します。
しきい値は許容可能な範囲と正常な範囲の下方境界を表します。
5. **【保存して実行】** をクリックし、スコアカードの変更内容を保存して、スコアカードを実行します。
【スコアカード】 パネルで、スコアの割合の変更およびバーとして表示される State スコアのスコアを表示します。

タスク 6. スコアの傾向グラフの表示

このタスクでは、State スコアの傾向グラフを表示します。傾向グラフを表示して、スコアを時系列で監視することができます。

1. **【スコアカード】** ワークスペースを開いており、sc_LA_Customer スコアカードが開かれていることを確認します。
2. **【状態】** 行を選択します。
3. **【アクション】** > **【傾向グラフの表示】** をクリックするか、または **【スコアの傾向】** カラムの下矢印をクリックします。
【傾向グラフの詳細】 ダイアログボックスが表示されます。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変わります。
4. グラフ内の任意の円をポイントすると、グラフの下部の **【有効な値】** セクションに有効な値が表示されます。
5. **【閉じる】** をクリックしてスコアカードに戻ります。

スコアカードの作成と実行のまとめ

このレッスンでは、プロファイルの結果からスコアカードを作成できることを学習しました。スコアカードには、プロファイルのカラムが含まれます。スコアカードを実行してカラムのスコアを生成できることを学習し

ました。スコアカードを編集して有効値を設定し、スコアのしきい値を設定しました。スコアの傾向グラフを表示する方法も学習しました。

プロファイル内の CustomerTier カラムと State カラムからスコアカードを作成して、CustomerTier カラムと State カラムのデータ品質を分析しました。スコアカードを実行して、各カラムのスコアを生成しました。スコアカードを編集して、スコアのさまざまな有効値を指定しました。スコアのしきい値を設定し、スコアの傾向グラフを表示しました。

第 8 章

レッスン 7. プロファイルカラムからの参照テーブルの作成

この章では、以下の項目について説明します。

- [プロファイルカラムからの参照テーブルの作成の概要, 41 ページ](#)
- [タスク 1. プロファイルカラムからの参照テーブルの作成, 42 ページ](#)
- [タスク 2. 参照テーブルの編集, 43 ページ](#)
- [プロファイルカラムからの参照テーブルの作成のまとめ, 43 ページ](#)

プロファイルカラムからの参照テーブルの作成の概要

参照テーブルには、ソースデータの標準化に使用できる参照データが含まれます。参照データには、有効な標準値を含めることができます。参照テーブルを作成して、ソースデータの値と有効な標準値の間にリレーションを確立します。

参照テーブルは、プロファイルの結果から作成できます。参照テーブルの作成後、参照テーブルを編集してカラムや行を追加したり、有効な標準値を追加または編集したりすることができます。参照テーブルに加えた変更は監査証跡で表示できます。

内容

HypoStores は、データをプロファイルして異常を明らかにしたり、有効値が含まれるデータを標準化したりする必要があります。あなたは、データ内の有効値を標準化するアナリストです。プロファイルカラムの有効値に基づいて、参照テーブルを作成する必要があります。

目標

このレッスンでは、以下の作業を完了します。

1. カラムの有効値を選択することによって、Profile_LA_Customers_Custom プロファイル内の CustomerTier カラムから参照テーブルを作成します。
2. 参照テーブルを編集して、カラムのさまざまな有効値を設定します。

前提条件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスン 1 から 6 までを完了している。

時間

このレッスンのタスクの所要時間は 15 分です。

タスク 1. プロファイルカラムからの参照テーブルの作成

このタスクでは、参照テーブルを作成し、Profile_LA_Customers プロファイルからその参照テーブルに CustomerTier カラムを追加します。

1. **【ライブラリ】** ワークスペースで、**【アセット】** > **【プロファイル】** をクリックします。
2. **Profile_LA_Customers** プロファイルをクリックして、プロファイル結果をサマリビューで開きます。
3. サマリビューで、参照テーブルに追加する CustomerTier カラムを選択します。右クリックして **【参照テーブルへの追加】** を選択します。
【参照テーブルへの追加】 ダイアログボックスが表示されます。
4. **【参照テーブルの作成】** を選択します。
5. **【次へ】** をクリックします。
6. **【名前】** フィールドに、「**Reftab_CustTier_HypoStores**」と入力します。
7. 説明を入力し、デフォルト値として 0 を設定します。
値が含まれていないテーブルレコードにはこのデフォルト値が使用されます。
8. **【次へ】** をクリックします。
9. **【カラム属性】** セクションで、CustomerTier カラムに次のカラムプロパティを設定します。

プロパティ	説明
名前	CustomerTier
データ型	String
精度	10
説明	参照顧客層の値

10. 必要に応じて、参照テーブルの行の説明カラムを作成します。カラムの名前および精度を入力します。
11. **【プレビュー】** セクションで、CustomerTier カラムの値を確認します。
12. **【次へ】** をクリックします。
Reftab_CustomerTier_HypoStores という参照テーブル名が表示されます。オプションの説明を入力できます。
13. **【保存先】** セクションで、参照テーブルを作成するチュートリアルプロジェクトを選択します。
参照テーブル: パネルに、選択した場所の参照テーブルが一覧表示されます。
14. オプションの監査のメモを入力します。
15. **【完了】** をクリックします。

タスク 2. 参照テーブルの編集

このタスクでは、Reftab_CustomerTier_HypoStores テーブルを編集して顧客層の代替値を追加します。

1. **【ライブラリ】** ワークスペースで、**【アセット】** > **【参照テーブル】** をクリックします。
2. Reftab_CustomerTier_HypoStores 参照テーブルをクリックします。
参照テーブルが **【設計】** ワークスペースに開きます。
3. 行を編集するには、行を選択し、**【アクション】** > **【編集】** をクリックするか、**【編集】** アイコンをクリックします。
【行の編集】 ダイアログボックスが表示されます。オプションとして、複数の行を選択して同じ代替値を各行に追加します。
4. Diamond、Emerald、Gold、Silver、および Bronze 行に、それぞれ代替値 1、2、3、4、5 を入力します。
オプションの監査のメモを入力します。
5. **【適用】** をクリックして、変更を適用します。
6. **【閉じる】** をクリックします。
変更された参照テーブルの値が **【設計】** ワークスペースに表示されます。

プロファイルカラムからの参照テーブルの作成のまとめ

このレッスンでは、プロファイルの結果から参照テーブルを作成してソースデータの有効値を設定する方法を学習しました。

カラムの有効値を選択することによって、プロファイルカラムから参照テーブルを作成しました。参照テーブルを編集して、カラムのさまざまな有効値を設定しました。

第 9 章

レッスン 8. 参照テーブルの作成

この章では、以下の項目について説明します。

- [参照テーブルの作成の概要, 44 ページ](#)
- [タスク 1. 参照テーブルの作成, 45 ページ](#)
- [参照テーブルの作成のまとめ, 45 ページ](#)

参照テーブルの作成の概要

参照テーブルには、ソースデータの標準化に使用できる参照データが含まれます。参照データには、有効な標準値を含めることができます。参照テーブルを作成して、ソースデータの値と有効な標準値の間にリレーションを確立します。

参照テーブルは、参照テーブルエディタを使用して手動で作成できます。参照テーブルを使用して、ソースデータの定義や標準化を行います。参照テーブルを開発者と共有して、Developer ツールの標準化トランスフォーメーションやルックアップトランスフォーメーションで 사용할 ことができます。

内容

HypoStores は、有効な値が含まれるデータを標準化する必要があります。あなたは、データ内の有効値を標準化するアナリストです。参照テーブルを作成して、LA の顧客データを参照する標準の顧客層コードを定義する必要があります。それにより、開発者と参照テーブルを共有できます。

目標

このレッスンでは、以下のタスクを完了します。

- 参照テーブルエディタを使用して参照テーブルを作成し、LA の顧客データを参照する標準の顧客層コードを定義します。

前提条件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスンの 1 から 2 までを完了している。

時間

このレッスンのタスクの完了には、10 分見ておいてください。

タスク 1. 参照テーブルの作成

このタスクでは、Reftab_CustomerTier_Codes 参照テーブルを作成して、顧客層データの有効値を標準化します。

1. **【新規】** > **【参照テーブル】** をクリックします。
新しい参照テーブルウィザードが表示されます。
2. **【参照テーブルエディタを使用します】** を選択します。
3. **【次へ】** をクリックします。
4. 参照テーブルに組み込むカラムごとに、**【新しいカラムの追加】** アイコンをクリックし、各カラムのカラムプロパティを設定します。
カラム名 CustomerID、CustomerTier、Status を追加します。カラムの順序を変更したりカラムを削除したりすることができます。
5. 必要に応じて説明を入力し、デフォルト値を **0** に設定します。
値が含まれていないテーブルレコードにはこのデフォルト値が使用されます。
6. **【次へ】** をクリックします。
7. **【名前】** フィールドに「**Reftab_CustomerTier_Codes**」と入力します。
8. **【フォルダー】** セクションで、チュートリアルプロジェクトの **Customers** フォルダーを選択します。
9. **【完了】** をクリックします。
参照テーブルが **【設計】** ワークスペースに表示されます。
10. **【アクション】** メニューで **【行の追加】** を選択して、各参照テーブルカラムに次の 4 つの値を埋め込みます。
CustomerID = LA1、LA2、LA3、LA4
CustomerTier = 1、2、3、4
Status = Active、Inactive

参照テーブルの作成のまとめ

このレッスンでは、参照テーブルエディタを使用して参照テーブルを作成し、ソースデータで使用する標準の有効値を作成する方法を学習しました。

参照テーブルエディタを使用して参照テーブルを作成し、LA の顧客データの顧客層の値を標準化しました。

パート II: Informatica Developer の基本操作

この部には、以下の章があります。

- [レッスン 1. Informatica Developer の設定, 47 ページ](#)
- [レッスン 2: 物理データオブジェクトのインポート, 51 ページ](#)
- [レッスン 3. データソースでのプロファイルの実行, 59 ページ](#)
- [レッスン 4. データの解析, 63 ページ](#)
- [レッスン 5. データの標準化, 70 ページ](#)
- [レッスン 6. 住所データの検証, 76 ページ](#)

第 10 章

レッスン 1. Informatica Developer の設定

この章では、以下の項目について説明します。

- [Informatica Developer の設定の概要, 47 ページ](#)
- [タスク 1. Informatica Developer の起動, 48 ページ](#)
- [タスク 2. ドメインの追加, 48 ページ](#)
- [タスク 3. モデルリポジトリの追加, 49 ページ](#)
- [タスク 4. プロジェクトの作成, 49 ページ](#)
- [タスク 5. フォルダの作成, 49 ページ](#)
- [タスク 6. デフォルトのデータ統合サービスの選択, 50 ページ](#)
- [Informatica Developer の設定のまとめ, 50 ページ](#)

Informatica Developer の設定の概要

このチュートリアルレッスンを開始する前に、Developer tool を起動して設定する必要があります。Developer tool を設定するには、ドメインを追加します。ドメイン内にあるモデルリポジトリを追加し、作業を格納するプロジェクトとフォルダを作成します。ドメインに複数のサービスが含まれている場合は、デフォルトのデータ統合サービスも選択する必要があります。

Informatica ドメインは、Informatica 環境を定義するノードとサービスの集合です。ドメイン内のサービスには、モデルリポジトリサービスやデータ統合サービスなどがあります。

モデルリポジトリサービスでは、モデルリポジトリが管理されます。モデルリポジトリは、Developer tool で作成するプロジェクトのメタデータを格納するリレーショナルデータベースです。プロジェクトには、Developer tool で作成するオブジェクトが格納されます。プロジェクトには、同じビジネス要件の一部であるオブジェクトなど、関連するオブジェクトを格納するフォルダも含まれます。

データ統合サービスは、Developer tool でデータ統合タスクを実行します。

目標

このレッスンでは、以下の作業を完了します。

- Developer tool を起動し、Developer tool のワークベンチに移動します。
- Developer tool でドメインを追加します。
- プロジェクトを作成できるように、モデルリポジトリを追加します。

- Developer tool で作成するオブジェクトを格納するプロジェクトを作成します。
- プロジェクトに、関連するオブジェクトを格納できるフォルダーを作成します。
- デフォルトのデータ統合サービスを選択して、データ統合タスクを実行します。

前提条件

このレッスンを開始する前に、次の前提条件を確認してください。

- Developer tool をインストールしている。
- ドメインに接続するためのドメイン名、ホスト名、およびポート番号がある。この情報はドメイン管理者から取得することができます。
- ドメイン管理者が Administrator ツールでモデルリポジトリサービスを設定済みである。
- モデルリポジトリサービスにアクセスするためのユーザー名とパスワードがある。この情報はドメイン管理者から取得することができます。
- ドメイン管理者がデータ統合サービスを設定済みである。
- データ統合サービスが実行されている。

時間

このレッスンのタスクの完了には、5～10 分見ておいてください。

タスク 1. Informatica Developer の起動

Developer ツールを起動してチュートリアルを開始します。

1. Developer ツールを起動します。
Developer ツールの【ようこそ】ページが表示されます。
2. 【ワークベンチ】 ボタンをクリックします。
Developer ツールのワークベンチが表示されます。

タスク 2. ドメインの追加

このタスクでは、Developer ツールにドメインを追加して、モデルリポジトリにアクセスします。

1. 【ウィンドウ】 > 【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. 【Informatica】 > 【ドメイン】 を選択します。
3. 【追加】 をクリックします。
【新しいドメイン】 ダイアログボックスが表示されます。
4. ドメイン名、ホスト名、およびポート番号を入力します。
5. 【完了】 をクリックします。
6. 【OK】 をクリックします。

タスク 3. モデルリポジトリの追加

このタスクでは、プロジェクトとフォルダの格納に使用するモデルリポジトリを追加します。

1. **【ファイル】** > **【リポジトリへの接続】** をクリックします。
【リポジトリへの接続】 ダイアログボックスが表示されます。
2. **【参照】** をクリックして、Model Repository Service を選択します。
3. **【OK】** をクリックします。
4. **【次へ】** をクリックします。
5. ユーザ名とパスワードを入力します。
6. 名前空間を選択します。
7. **【完了】** をクリックします。
Object Explorer ビューにモデルリポジトリが表示されます。

タスク 4. プロジェクトの作成

このタスクでは、Developer ツールで作成するオブジェクトを格納するプロジェクトを作成します。このガイドのすべてのチュートリアルで使用するプロジェクトを 1 つ作成します。

1. **【Object Explorer】** ビューで、Model Repository Service を選択します。
2. **【ファイル】** > **【新規】** > **【プロジェクト】** をクリックします。
【新規プロジェクト】 ダイアログボックスが表示されます。
3. 「Tutorial_あなたの名前」をプロジェクト名として入力します。
4. **【完了】** をクリックします。
【Object Explorer】 ビューの **【Model Repository Service】** の下にプロジェクトが表示されます。

タスク 5. フォルダの作成

このタスクでは、関連するオブジェクトを格納するフォルダを作成します。このガイドのすべてのチュートリアルで使用するフォルダを 1 つ作成します。

1. **【Object Explorer】** ビューで、フォルダの追加先のプロジェクトを選択します。
2. **【ファイル】** > **【新規】** > **【フォルダ】** をクリックします。
3. フォルダの名前を入力します。
4. **【完了】** をクリックします。

Developer ツールは、**【Object Explorer】** ビューのプロジェクトの下にフォルダを追加します。プロジェクトを展開して、フォルダを確認してください。

タスク 6. デフォルトのデータ統合サービスの選択

このタスクでは、マッピングを実行したりデータをプレビューしたりできるように、デフォルトの Data Integration Service を選択します。この手順は、ドメイン内に複数のデータ統合サービスがある場合に必要です。ドメインに 1 つのデータ統合サービスが含まれている場合、このサービスはデフォルトとして設定されます。

1. **【ウィンドウ】** > **【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. **【Informatica】** > **【Data Integration Services】** を選択します。
3. ドメインを展開します。
4. Data Integration Service を選択します。
5. **【デフォルトとして設定】** をクリックします。
6. **【OK】** をクリックします。

Informatica Developer の設定のまとめ

このレッスンでは、Informatica ドメインに Model Repository Service と Data Integration Service が含まれることを学習しました。Model Repository Service では、モデルリポジトリが管理されます。モデルリポジトリには、プロジェクトとフォルダが含まれます。Data Integration Service は、データ統合タスクを実行します。

Developer ツールを起動し、Developer ツールを設定しました。Developer ツールにドメインを追加し、モデルリポジトリを追加し、プロジェクトとフォルダを作成しました。デフォルトの Data Integration Service も選択しました。

これで、Developer ツールを使用してこのチュートリアルの中の他のレッスンを完了することができます。

第 11 章

レッスン 2: 物理データオブジェクトのインポート

この章では、以下の項目について説明します。

- [物理データオブジェクトのインポートの概要, 51 ページ](#)
- [タスク 1. Boston_Customers フラットファイルデータオブジェクトのインポート, 52 ページ](#)
- [タスク 2. LA_Customers フラットファイルデータオブジェクトのインポート, 57 ページ](#)
- [タスク 3. All_Customers フラットファイルデータオブジェクトのインポート, 58 ページ](#)
- [物理データオブジェクトのインポートのまとめ, 58 ページ](#)

物理データオブジェクトのインポートの概要

物理データオブジェクトは、フラットファイルまたはリレーショナルデータベーステーブルに基づいたデータの表現です。フラットファイルまたはリレーショナルデータベーステーブルを物理データオブジェクトとしてインポートして、マッピングでソースまたはターゲットとして使用することができます。

内容

HypoStores Corporation では、Los Angeles オフィスと Boston オフィスの顧客データをフラットファイルに格納しています。この顧客データを Developer ツールで使用したいと考えています。そのためには、各フラットファイルを物理データオブジェクトとしてインポートする必要があります。

目標

このレッスンでは、フラットファイルを物理データオブジェクトとしてインポートします。Data Integration Service が正しいディレクトリからソースデータを読み取ることができるようにソースファイルディレクトリを設定することもできます。

前提条件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスン 1 を完了している。

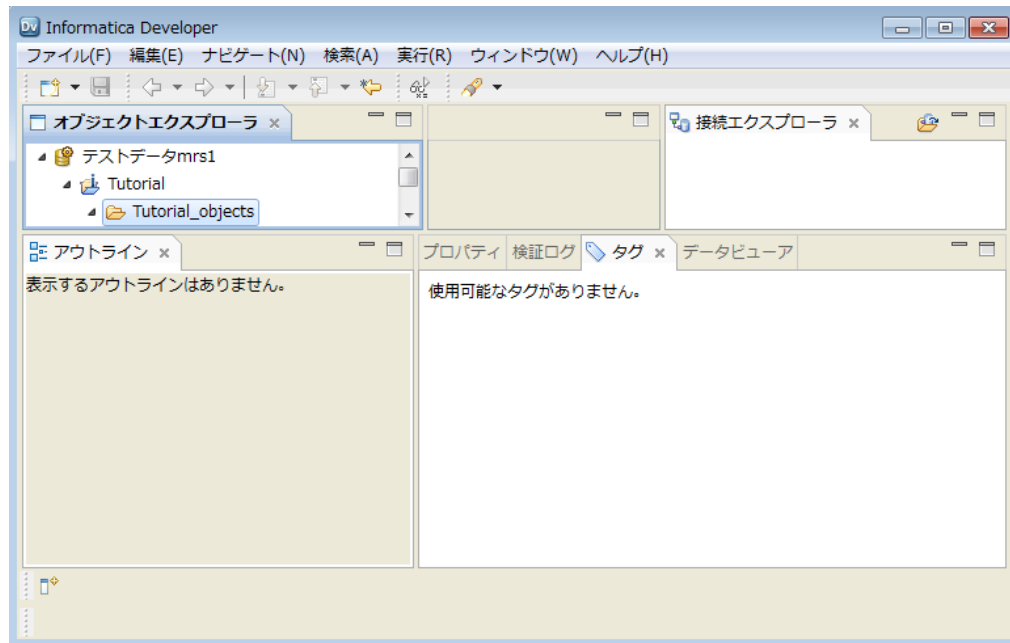
時間

このレッスンのタスクの完了には、10～15 分見ておいてください。

タスク 1. Boston_Customers フラットファイルデー タオブジェクトのインポート

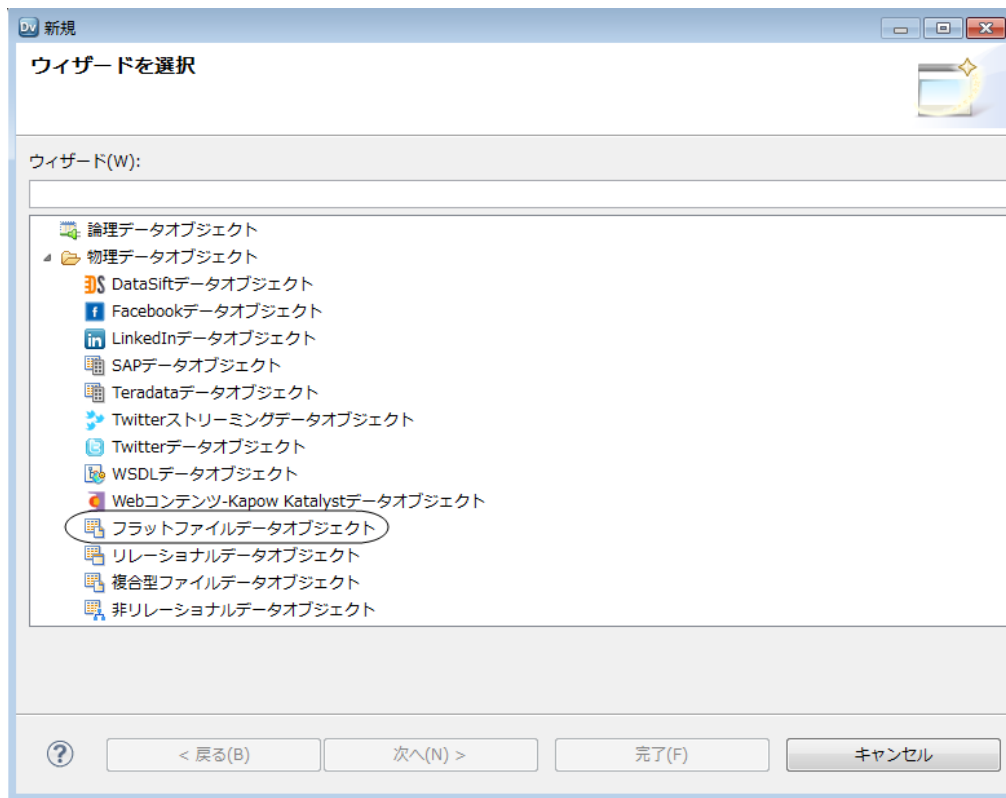
このタスクでは、Boston オフィスの顧客データが含まれるファイルから、物理データオブジェクトをインポートします。

1. **[Object Explorer]** ビューで、Tutorial_Objects フォルダを選択します。



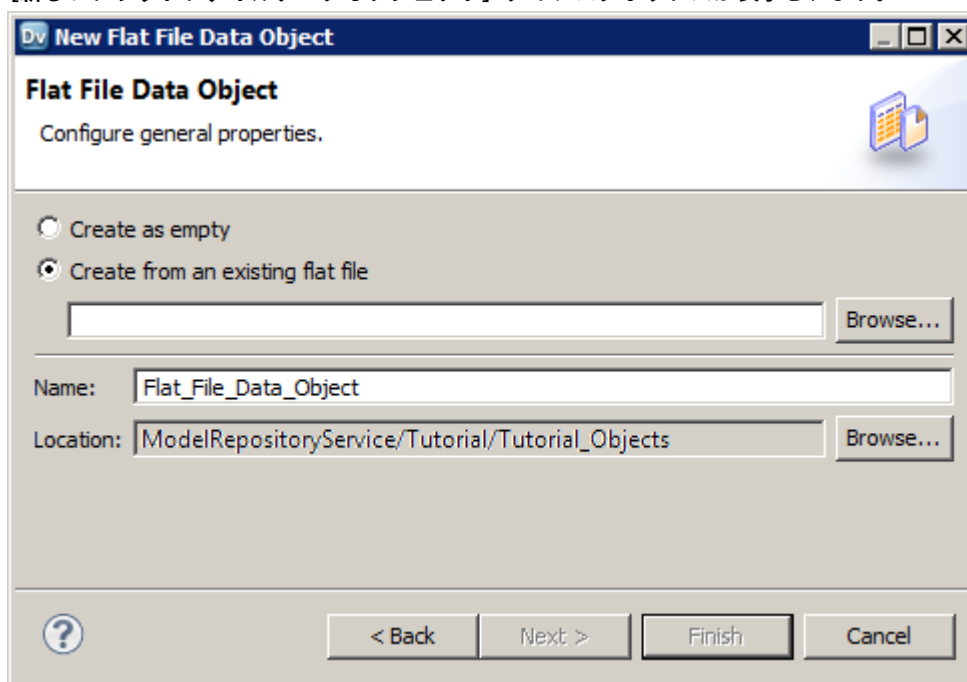
2. Tutorial_Objects フォルダを右クリックし、**[新規]** > **[データオブジェクト]** を選択します。

【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】 > 【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。



4. **【既存のフラットファイルから作成】** を選択します。
5. **【参照】** をクリックし、Developer tool マシンのディレクトリ<Informatica インストールディレクトリ>\clients\DeveloperClient\Tutorials の Boston_Customers.csv に移動します。
6. **【開く】** をクリックします。
ウィザードによって、データオブジェクトに Boston_Customers という名前が付けられます。
7. **【次へ】** をクリックします。
8. コードページが **【MS Windows Latin 1 (ANSI)、Latin 1 のスーパーセット】** に設定され、形式が **【区切り】** に設定されていることを確認します。

【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にはデフォルトコードページ、形式、フラットファイルデータのプレビューが表示されます。

新しいフラットファイル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
コードページと形式を設定します。

コードページ(P): MS Windows Latin 1 (ANSI), superset of Latin1

形式

☒ 区切り(D) (区切り文字で区切られたフィールド)

☐ 固定幅(I) (カラム内でフィールドを整列)

プレビューする最大行数(X): 500

	フィールド1	フィールド2	フィールド3	フィールド4	
1	CustomerID	CustomerTier	Lastname	Firstname	Company
2	10110095	Diamond	Bergeron	Barb	E-AGENCY
3	10110102	Bronze	Brosseau	Derrick	BANK ONE
4	10110105		Anderson	Lauren	KPMG PEAT MARWICK

? < 戻る(B) 次へ(N) > 完了(F) キャンセル

9. **【次へ】** をクリックします。
10. **【1 行目からカラム名をインポート】** を選択します。

【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のフラットファイルデータのプレビューにカラム名が表示されます。

新しいフラットファイル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

区切り文字で区切ったプロパティを設定します。

区切り文字

☐ タブ(T) ☐ セミコロン(M) ☒ カンマ(C)

☐ スペース(P) ☐ その他(H): ...

テキスト修飾子

☒ 引用符なし(O) ☐ 一重引用符(S) ☐ 二重引用符(Q)

プレビューのオプション

☒ 1行目からカラム名をインポート(I) 次の行からインポートを開始(R):

行の区切り文字(L): ¥012 LF (¥n) ☐ 連続した区切り文字を1文字として扱う(U)

エスケープ文字(E): ☐ データ内のエスケープ文字を維持(A)

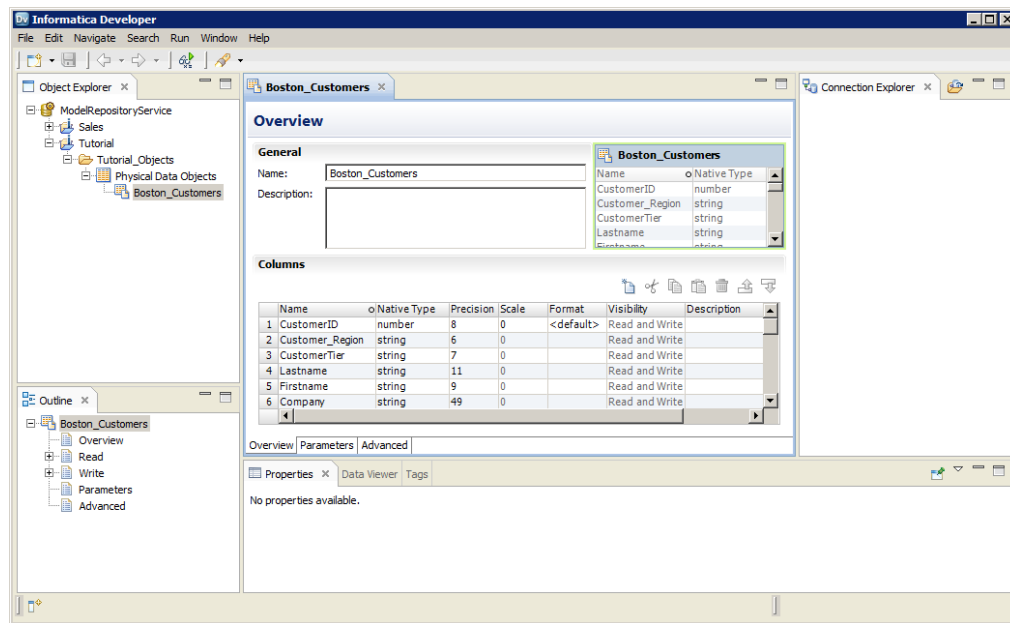
プレビューする最大行数(X):

	CustomerID	CustomerTier	Lastname	Firstname	
1	10110095	Diamond	Bergeron	Barb	E-AGENCY
2	10110102	Bronze	Brosseau	Derrick	BANK ONE
3	10110105		Anderson	Lauren	KPMG PEAT MARWICK

1. [1 行目からカラム名をインポート] オプション
2. カラム名

11. 【完了】をクリックします。

Boston_Customers 物理データオブジェクトが [Tutorial_Objects] フォルダの物理データオブジェクトフォルダの下に表示されます。**【概要】** ビューにファイルコンテンツが表示されます。ファイルコンテンツはエディタで開いています。



12. **【詳細】** ビューをクリックします。

【詳細】 ビューには、物理データオブジェクトのプロパティが表示されます。

13. **【詳細】** ビューで、**【ランタイム: 読み取り】** セクションまでスクロールします。

14. **【ランタイム: 読み取り】** セクションで、**【ソースファイルのディレクトリ】** をデータ統合サービスマシンのディレクトリ<Informatica installation directory>\server\Tutorials に設定します。

データ統合サービスは、データ統合サービスを実行するマシンのサーバーディレクトリでソースファイルを検索します。サーバーのインストールには、チュートリアルファイルのコピーが含まれています。ソースファイルおよびディレクトリに対するアクセス権限を変更しなければ、データ統合サービスはクライアントインストールディレクトリからファイルを読み取ることはできません。

以下の図に、ソースファイルディレクトリの例を示します。

Advanced	
Name	Value
Runtime : Read	
Input type	File
Source type	Direct
Source file name	Boston_Customers.csv
Source file directory	\\MyMachine\Informatica\10.0\server\Tutorials

注: Developer tool マシンは、データ統合サービスを実行するマシンのソースファイルディレクトリにアクセスできる必要があります。Developer tool がソースファイルディレクトリにアクセスできない場合、Developer tool はソースファイルのデータをプレビューできなかつたり、ソースファイルのデータにアクセスするマッピングを実行できなかつたりします。複数のデータ統合サービスを実行する場合、データ統合サービス別に個々のソースファイルディレクトリが存在します。

15. **【データビューア】** ビューをクリックします。

16. **【データビューア】** ビューで、**【実行】** をクリックします。
データ統合サービスが、Boston_Customers ファイルからデータを読み取り、結果を出力ウィンドウに表示します。
17. **【ファイル】** > **【保存】** をクリックし、Boston_Customers 物理データオブジェクトを保存します。

タスク 2. LA_Customers フラットファイルデータオブジェクトのインポート

このタスクでは、Los Angeles オフィスの顧客データが含まれるフラットファイルから、物理データオブジェクトをインポートします。

1. **【Object Explorer】** ビューで、チュートリアル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】** > **【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。
4. **【既存のフラットファイルから作成】** を選択します。
5. **【参照】** をクリックし、次のディレクトリの LA_Customers.csv に移動します。 <Informatica Installation Directory>\clients\DeveloperClient\Tutorials
6. **【開く】** をクリックします。
ウィザードでデータオブジェクトに LA_Customers という名前が付けられます。
7. **【次へ】** をクリックします。
8. コードページが MS Windows Latin 1 (ANSI)、Latin 1 のスーパーセットであることを確認します。
9. 区切り形式であることを確認します。
10. **【次へ】** をクリックします。
11. 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。
12. **【1 行目からカラム名をインポート】** を選択します。
13. **【完了】** をクリックします。
LA_Customers 物理データオブジェクトが、チュートリアルプロジェクトの **【物理データオブジェクト】** の下に表示されます。
14. **【読み取り】** ビューをクリックし、出力トランスフォーメーションを選択します。
15. **【プロパティ】** ビューで **【ランタイム】** タブをクリックします。
16. **【ソースファイルのディレクトリ】** を Data Integration Service マシンの次のディレクトリに設定します。
<Informatica Installation Directory>\server\Tutorials
17. **【ファイル】** > **【保存】** をクリックします。

タスク 3. All_Customers フラットファイルデータオブジェクトのインポート

このタスクでは、Los Angeles オフィスと Boston オフィスの顧客注文データを組み合わせたフラットファイルから、物理データオブジェクトをインポートします。

1. **[Object Explorer]** ビューで、チュートリアル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] > [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ソース] ダイアログボックスが表示されます。
4. **[既存のフラットファイルから作成]** を選択します。
5. **[参照]** をクリックし、次のディレクトリの All_Customers.csv に移動します。 <Informatica Installation Directory>\clients\DeveloperClient\Tutorials
6. **[開く]** をクリックします。
ウィザードでデータオブジェクトに All_Customers という名前が付けられます。
7. **[次へ]** をクリックします。
8. コードページが MS Windows Latin 1 (ANSI)、Latin 1 のスーパーセットであることを確認します。
9. 区切り形式であることを確認します。
10. **[次へ]** をクリックします。
11. 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。
12. **[1 行目からカラム名をインポート]** を選択します。
13. **[完了]** をクリックします。
All_Customers 物理データオブジェクトが、チュートリアルプロジェクトの **[物理データオブジェクト]** の下に表示されます。
14. **[読み取り]** ビューをクリックし、出力トランスフォーメーションを選択します。
15. **[プロパティ]** ビューで **[ランタイム]** タブをクリックします。
16. **[ソースファイルのディレクトリ]** を Data Integration Service マシンの次のディレクトリに設定します。
<Informatica Installation Directory>\server\Tutorials
17. **[ファイル] > [保存]** をクリックします。

物理データオブジェクトのインポートのまとめ

このレッスンでは、物理データオブジェクトがフラットファイルまたはリレーショナルデータベーステーブルに基づいたデータの表現であることを学習しました。

フラットファイルから物理データオブジェクトを作成しました。Data Integration Service が正しいディレクトリからソースデータを読み取ることができるようにソースファイルディレクトリを設定することもできます。

作成したデータオブジェクトは、データ品質のレッスンでマッピングソースとして使用します。

第 12 章

レッスン 3. データソースでのプロファイルの実行

この章では、以下の項目について説明します。

- [データのプロファイリングの概要, 59 ページ](#)
- [タスク 1. 2つのデータソースでの結合分析の実行, 60 ページ](#)
- [タスク 2. 結合分析結果の表示, 61 ページ](#)
- [タスク 3. データソースでのプロファイルの実行, 61 ページ](#)
- [タスク 4. カラムプロファイル結果の表示, 62 ページ](#)
- [データのプロファイリングのまとめ, 62 ページ](#)

データのプロファイリングの概要

プロファイルは、データセットの内容と構造を記述する一連のメタデータです。

プロファイリングとデータ検出は、多くの場合、プロジェクトで最初に行われるステップです。プロファイルを実行してデータの構造を評価し、想定する情報のタイプがデータカラムに入力されていることを確認することができます。プロファイルによってデータの問題が明らかになった場合は、それらの問題を修正する手順をプロジェクト内で定義できます。例えば、想定された長さを超える値がカラムに含まれていることがプロファイルによって明らかになった場合は、問題の値を削除または修正するようにデータ品質プロセスを設計できます。

選択したカラムのデータ品質を分析するプロファイルは、カラムプロファイルと呼ばれます。

注: Developer tool を使用して、プライマリキー、外部キー、および機能依存性リレーションを検出することや、データカラムの結合条件を分析することもできます。

カラムプロファイルは、データに関する次のファクトを提供します。

- カラムごとの重複しない値と NULL 値の数。数および割合として表されます。
- カラムごとのデータのパターン、およびこ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。
- 結合分析プロファイルでの、2つのデータカラム間の重複の度合い。ベン図および割合値として表示されます。結合分析プロファイルを使用して、カラムの結合条件で発生する可能性のある問題を特定します。

カラムプロファイルは、プロジェクトのあらゆる段階で実行して、データ品質を測定したり、データの変更がプロジェクトの目標になっていることを確認したりすることができます。カラムプロファイルをマッピング

のトランスフォーメーションで実行して、そのトランスフォーメーションがデータに与える影響を示すことができます。

内容

HypoStores は、顧客データにエラー、不整合、および重複した情報がないことを確認する必要があります。HypoStores は、データ品質の目標を達成するためのプロセスを設計する前に、ソースデータファイルの品質を測定し、データを処理する準備ができていることを確認する必要があります。

目標

このレッスンでは、以下の作業を完了します。

- Boston_Customers データソースと LA_Customers データソースで、結合分析を実行します。
- 結合分析の結果を表示して、2 つのオフィスのデータを正常にマージできるかどうかを判断します。
- All_Customers データソースで、カラムプロファイルを実行します。
- カラムプロファイリングの結果を表示して、データに含まれる値とパターンを確認します。

前提条件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスンの 1 から 2 までを完了している。

所要時間

- このレッスンの完了には、20 分見ておいてください。

タスク 1.2 つのデータソースでの結合分析の実行

このタスクでは、Boston_Customers および LA_Customers データソースについて結合分析を実行し、結合条件を表示します。

1. チュートリアルフォルダーを選択し、[ファイル] > [新規] > [プロファイル] をクリックします。
2. [エンタープライズ検出プロファイル] を選択します。
3. [次へ] をクリックします。
4. [名前] フィールドに「Tutorial_Profile」と入力します。
5. [完了] をクリックします。
Object Explorer に Tutorial_Profile プロファイルが表示されます。
6. **Boston_Customers** データソースと **LA_Customers** データソースを右側のエディタにドラッグします。
ヒント: 複数のデータオブジェクトを選択するには、Shift キーを押しながら選択します。
7. データオブジェクト名を右クリックし、[結合プロファイル] を選択します。
新しい結合プロファイルウィザードが表示されます。
8. [名前] フィールドに「JoinAnalysis」と入力します。
9. Boston_Customers と LA_Customers がデータオブジェクトとして表示されていることを確認し、[次へ] をクリックします。
10. 両方のデータソースで **CustomerID** カラムが選択されていることを確認します。
ウィザードのペインをスクロールダウンして、両方のデータセットのカラムを表示します。
[次へ] をクリックします。

11. **【追加】** をクリックして、結合条件を追加します。
【結合条件】 ダイアログボックスが表示されます。
12. **【カラム】** セクションで、**【行の追加】** をクリックします。
13. 左側のカラムの最初の行をダブルクリックし、**CustomerID** を選択します。
14. 右側のカラムの最初の行をダブルクリックし、**CustomerID** を選択します。
15. **【OK】** をクリックし、**【完了】** をクリックします。
16. Developer ツールにより、変更内容を保存するように求められたら、**【はい】** をクリックします。
Developer ツールがプロファイルを実行します。

注: プロファイルは閉じないでください。次のタスクでプロファイル結果を表示します。

タスク 2. 結合分析結果の表示

このタスクでは、JoinAnalysis プロファイルの**【結合結果】** ビューに結合分析結果を表示します。

1. エディタの **【JoinAnalysis】** タブをクリックします。
2. **【結合結果】** セクションで、最初の行をクリックします。
【詳細】 セクションに、ベン図と結合分析結果の詳細を表すカラーキーが表示されます。
3. **【行の結合】** カラムで、結合が含まれる行の数としてゼロが表示されていることを確認します。
この値は、CustomerID フィールドに重複がないことを示します。2つのデータソースを正常に結合できます。
4. LA_Customers データオブジェクトの CustomerID 値を表示するには、ベン図内で **LA_Customers** という名前の円をダブルクリックします。

ヒント: ベン図内の円をダブルクリックしてデータ行を表示します。ベン図内で円が交差する場合は、交差をダブルクリックして両方のデータセットに共通のデータ値を表示します。

データビューアに、LA_Customers データオブジェクトに含まれる CustomerID 値が表示されます。

タスク 3. データソースでのプロファイルの実行

このタスクでは、All_Customers データソースでプロファイルを実行して、データの内容と構造を表示します。

1. **【Object Explorer】** ビューで、チュートリアルプロジェクト内のデータオブジェクトを参照します。
2. **All_Customers** データソースを選択します。
3. **【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。
4. **【プロファイル】** を選択します。
5. **【次へ】** をクリックします。
6. **【名前】** フィールドに **All_Customers** と入力します。
7. **【完了】** をクリックします。

エディタに All_Customers プロファイルが開き、プロファイルが実行されます。

タスク 4. カラムプロファイル結果の表示

このタスクでは、All_Customers データオブジェクトのカラムプロファイリング結果を表示し、データに含まれる値とパターンを確認します。

1. **【ウィンドウ】 > 【ビューの表示】 > 【進行状況】** をクリックして、All_Customers プロファイルの進行状況を表示します。
進行状況ビューが開きます。
2. **【進行状況】** ビューで All_Customers プロファイルの実行完了が報告されたら、エディタで **【結果】** ビューをクリックします。
3. **【カラムプロファイリング】** セクションで、**CustomerTier** カラムをクリックします。
【詳細】 セクションに、CustomerTier カラムに含まれているすべての値が表示され、データセットに値が出現する頻度に関する情報が表示されます。
4. **【詳細】** セクションで、**Ruby** をダブルクリックします。
データビューアにより、CustomerTier カラムに値 **Ruby** が含まれるレコードが実行および表示されます。
5. **【カラムプロファイリング】** セクションで、**OrderAmount** カラムをクリックします。
6. **【詳細】** セクションで、**【リストを表示】** をクリックし、**【パターン】** を選択します。
【詳細】 セクションに、OrderAmount カラムで見つかったパターンが表示されます。Pattern カラム内の文字列 **9(5)** は、5 桁の注文金額が含まれるレコードを参照します。文字列 **9(4)** は、4 桁の注文金額が含まれるレコードを参照します。
7. **Pattern** カラムで、文字列 **9(4)** をダブルクリックします。
データビューアにより、OrderAmount カラムに 4 桁の注文金額が含まれるレコードが実行および表示されます。
8. **【詳細】** セクションで、**【リストを表示】** をクリックし、**【統計】** を選択します。
【詳細】 セクションには、平均値、標準偏差、最大および最小長さ、一般的な順に 5 つの値、一般的なでない順に 5 つの値など、OrderAmount カラムの統計が表示されます。

データのプロファイリングのまとめ

このレッスンでは、プロファイルがデータの内容と構造に関する情報を提供することを学習しました。

2 つのデータオブジェクトについて結合分析を実行し、データオブジェクト間の重なり度の合いを表示できることを学習しました。データオブジェクトについてカラムプロファイルを実行し、データオブジェクト内の各カラムに関連する値、パターン、および統計を表示することも学習しました。

Boston_Customers データオブジェクトのデータを LA_Customers データオブジェクトのデータとマージできるかどうかを判断するために、JoinAnalysis プロファイルを作成しました。このプロファイルの結果を表示し、CustomerID カラム内のすべての値が一意であることと、データオブジェクトを正常にマージできることを確認しました。

All_Customers プロファイルを作成し、All_Customers データオブジェクトについてカラムプロファイルを実行しました。All_Customers データオブジェクト内のカラムの値、パターン、および統計を検出するために、このプロファイルの結果を表示しました。最後に、データビューアを実行して選択した値とパターンが含まれる行を表示し、データの品質を確認できるようになりました。

第 13 章

レッスン 4. データの解析

この章では、以下の項目について説明します。

- [データの解析の概要, 63 ページ](#)
- [タスク 1. ターゲットデータオブジェクトの作成, 64 ページ](#)
- [タスク 2. データを解析するためのマッピングの作成, 66 ページ](#)
- [タスク 3. パーサートランスフォーメーションでのプロファイルの実行, 68 ページ](#)
- [タスク 4. マッピングの実行, 68 ページ](#)
- [タスク 5. マッピング出力の表示, 68 ページ](#)
- [データの解析のまとめ, 69 ページ](#)

データの解析の概要

データを解析して、入力フィールドの 1 つまたは複数の要素を識別し、各要素をさまざまな出力フィールドに書き出します。

解析によって、各カラムの情報をより細かく制御できます。例えば、データフィールドに Bob Smith という氏名が含まれているとします。パーサートランスフォーメーションを使用すると、氏名を姓と名に分けて別々のデータカラムに格納することができます。データを新しいカラムに解析したら、カラムごとにカスタムのデータ品質操作を作成できます。

トークンセットを使用してデータカラムを構成文字列に解析するように、パーサートランスフォーメーションを設定することができます。トークンセットは、単語、郵便番号、電話番号、社会保障番号などのデータ要素を識別します。

パーサートランスフォーメーションを使用して、参照テーブルのエントリや入力したカスタムの正規表現に一致するデータを解析することもできます。

内容

HypoStores では、ロサンゼルスオフィスの顧客データファイルの形式をボストンオフィスのデータファイルの形式に合わせたいと考えています。ロサンゼルスオフィスの顧客データでは、FullName カラムに顧客の名前が格納されているのに対し、ボストンオフィスの顧客データでは、FirstName カラムと LastName カラムに分けて顧客の名前が格納されています。ロサンゼルスの FullName カラムのデータを姓と名に解析して、ロサンゼルスのデータの形式をボストンのデータの形式に合わせる必要があります。

目標

このレッスンでは、以下の作業を完了します。

- LA_Customers_tgt というデータオブジェクトを作成し、解析されたデータを格納するように設定する。

- FullName カラムを FirstName カラムと LastName カラムに解析するマッピングを作成する。
- LA_Customers データオブジェクトをマッピングに追加してソースデータに接続する。
- LA_Customers_tgt データオブジェクトをマッピングに追加してターゲットデータオブジェクトを作成する。
- パーサートランスフォーメーションをマッピングに追加し、トークンセットを使用して氏名を姓と名に解析するように設定する。
- ターゲットデータソースを生成する前に、パーサートランスフォーメーションのプロファイルを実行してデータを確認する。
- マッピングを実行して、解析された名前を生成する。
- データビューアを実行してマッピング出力を確認する。

要件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスンの 1 から 2 までを完了している。

時間

このレッスンのタスクの所要時間は 20 分です。

タスク 1. ターゲットデータオブジェクトの作成

このタスクでは、解析した名前を書き込むことができる LA_Customers_tgt データオブジェクトを作成します。

ターゲットデータオブジェクトを作成するには、以下の手順を実行します。

1. LA_Customers.csv ファイルに基づいて、LA_Customers_tgt データオブジェクトを作成します。
2. データオブジェクトの読み取りおよび書き込みのオプションを、ファイルの場所やファイル名も含めて設定します。
3. Firstname カラムおよび Lastname カラムを LA_Customers_tgt データオブジェクトに追加します。

手順 1.LA_Customers_tgt データオブジェクトの作成

このステップでは、LA_Customers.csv ファイルに基づいて LA_Customers_tgt データオブジェクトを作成します。

1. **【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】 ウィンドウが開きます。
2. **【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。
3. **【既存のフラットファイルから作成】** が選択されていることを確認します。
4. **【参照】** をクリックし、次のディレクトリに移動して LA_Customers.csv を選択します。 <Informatica Installation Directory>\clients\DeveloperClient\Tutorials
5. **【開く】** をクリックします。
6. [名前] フィールドに、「LA_Customers_tgt」と入力します。
7. **【次へ】** をクリックします。
8. **【次へ】** をクリックします。
9. **【プレビューのオプション】** セクションで、**【1 行目からカラム名をインポート】** を選択し、**【次へ】** をクリックします。

10. **【完了】** をクリックします。

LA_Customers_tgt データオブジェクトがエディタに表示されます。

手順 2.読み取りオプションおよび書き込みオプションの設定

この手順では、LA_Customers_tgt データオブジェクトの読み取りおよび書き込みのオプションを、ファイルの場所とファイル名を含めて設定します。

1. LA_Customers_tgt データオブジェクトがエディタに表示されていることを確認します。
2. エディタで、**【読み取り】** ビューを選択します。
3. **【ウィンドウ】** > **【ビューの表示】** > **【プロパティ】** をクリックします。
4. **【プロパティ】** ビューで、**【ランタイム】** ビューを選択します。
5. **【値】** カラムで、ソースファイル名をダブルクリックし、「LA_Customers_tgt.csv」と入力します。
6. **【値】** カラムで、ソースファイルディレクトリをダブルクリックして強調表示します。
7. 強調表示された名前を右クリックし、**【コピー】** を選択します。
8. エディタで、**【書き込み】** ビューを選択します。
9. **【プロパティ】** ビューで、**【ランタイム】** ビューを選択します。
10. **【値】** カラムで、**【出力ファイルディレクトリ】** エントリをダブルクリックします。
11. 右クリックして **【貼り付け】** を選択し、**【読み取り】** ビューからコピーしたディレクトリの場所を貼り付けます。
12. **【値】** カラムで、**【ヘッダーのオプション】** エントリをダブルクリックし、**【出力フィールド名】** を選択します。
13. **【値】** カラムで、**【出力ファイル名】** エントリをダブルクリックし、「LA_Customers_tgt.csv」と入力します。
14. **【ファイル】** > **【保存】** をクリックし、データオブジェクトを保存します。

手順 3.データオブジェクトへのカラムの追加

この手順では、Firstname カラムおよび Lastname カラムを LA_Customers_tgt データオブジェクトに追加します。

1. **【Object Explorer】** ビューで、チュートリアルプロジェクト内のデータオブジェクトを参照します。
2. LA_Customers_tgt データオブジェクトをダブルクリックします。
LA_Customers_tgt データオブジェクトがエディタに表示されます。
3. **【概要】** ビューが選択されていることを確認します。
4. **FullName** カラムを選択し、**【新規】** ボタンをクリックしてカラムを追加します。
FullName1 という名前のカラムが表示されます。
5. カラムの名前を Firstname に変更します。 **【精度】** フィールドをクリックし、「30」と入力します。
6. Firstname カラムを選択し、**【新規】** ボタンをクリックしてカラムを追加します。
FirstName1 という名前のカラムが表示されます。
7. カラムの名前を Lastname に変更します。 **【精度】** フィールドをクリックし、「30」と入力します。
8. **【ファイル】** > **【保存】** をクリックし、データオブジェクトを保存します。

タスク 2. データを解析するためのマッピングの作成

このタスクでは、マッピングを作成し、データオブジェクトとパーサトランスフォーメーションを使用するように設定します。

データを解析するマッピングを作成するには、以下の手順を実行します。

1. マッピングを作成します。
2. ソースおよびターゲットのデータオブジェクトをマッピングに追加します。
3. マッピングにパーサトランスフォーメーションを追加します。
4. 顧客の氏名が格納されたソースカラムを解析して姓と名を別々のカラムに格納するようにパーサトランスフォーメーションを設定します。

手順 1. マッピングの作成

この手順では、マッピングを作成して名前を付けます。

1. Object Explorer ビューで、チュートリアルプロジェクトを選択します。
2. **【ファイル】 > 【新規】 > 【マッピング】** をクリックします。
【新しいマッピング】 ウィンドウが開きます。
3. **【名前】** フィールドに、「ParserMapping」と入力します。
4. **【完了】** をクリックします。

マッピングがエディタに表示されます。

手順 2. データオブジェクトのマッピングへの追加

この手順では、LA_Customers データオブジェクトおよび LA_Customers_tgt データオブジェクトをマッピングに追加します。

1. [Object Explorer] ビューで、チュートリアルプロジェクト内のデータオブジェクトを参照します。
2. LA_Customers データオブジェクトを選択し、エディタにドラッグします。
【物理データオブジェクトのマッピングへの追加】 ウィンドウが開きます。
3. **【読み取り】** が選択されていることを確認し、**【OK】** をクリックします。
データオブジェクトがエディタに表示されます。
4. [Object Explorer] ビューで、チュートリアルプロジェクト内のデータオブジェクトを参照します。
5. LA_Customers_tgt データオブジェクトを選択し、エディタにドラッグします。
【物理データオブジェクトのマッピングへの追加】 ウィンドウが開きます。
6. **【書き込み】** を選択し、**【OK】** をクリックします。
データオブジェクトがエディタに表示されます。
7. LA_Customers データオブジェクトの CustomerID、CustomerTier、および FullName ポートを選択します。これらのポートを LA_Customers_tgt データオブジェクトの CustomerID ポートにドラッグします。

ヒント: 複数のポートを選択するには、**Ctrl** キーを押しながら選択します。

LA_Customers データオブジェクトのポートが、LA_Customers_tgt データオブジェクトの対応するポートに接続されます。

手順 3.パーサートランスフォーメーションのマッピングへの追加

この手順では、パーサートランスフォーメーションを ParserMapping マッピングに追加します。

1. ParserMapping マッピングが含まれているエディタを選択します。
2. トランスフォーメーションパレットで、パーサートランスフォーメーションを選択します。
3. エディタをクリックします。
【新しいパーサートランスフォーメーション】ウィンドウが開きます。
4. 【トークンパーサー】が選択されていることを確認し、【完了】をクリックします。
パーサートランスフォーメーションがエディタに表示されます。
5. LA_Customers データオブジェクトの FullName ポートを選択し、そのポートをパーサートランスフォーメーションの Input グループにドラッグします。
FullName ポートがパーサートランスフォーメーションに表示され、データオブジェクトの FullName ポートに接続されます。

手順 4.パーサートランスフォーメーションの設定

この手順では、パーサートランスフォーメーションを設定し、顧客のフルネームが含まれているカラムを解析して、名と姓を別々のカラムに格納します。

1. ParserMapping マッピングが含まれているエディタを選択します。
2. パーサートランスフォーメーションをクリックします。
3. 【ウィンドウ】 > 【ビューの表示】 > 【プロパティ】をクリックします。
4. 【プロパティ】ビューで、【ストラテジ】ビューを選択します。
5. 【新規】をクリックします。新しいストラテジウィザードが表示されます。
6. 入力カラムで選択矢印をクリックし、FullName ポートを選択します。
7. スペース区切り文字 [\s] を選択します。
8. 【次へ】をクリックします。
9. 【トークンセットを使用した解析】操作を選択し、【次へ】をクリックします。
10. 【トークンセット（単一出力のみ）】を選択し、【未定義】トークンセットを選択します。
11. 【出力】フィールドをクリックし、【新規】を選択します。
12. 【操作の出力】ダイアログボックスで、出力名を Undefined_Output に変更します。
13. 【完了】をクリックします。
14. パーサートランスフォーメーションで、Undefined_Output ポートをクリックし、LA_customers_tgt データオブジェクトの FirstName ポートにドラッグします。
ポート間に接続が表示されます。
15. パーサートランスフォーメーションで、OverflowField ポートをクリックし、LA_customers_tgt データオブジェクトの LastName ポートにドラッグします。
ポート間に接続が表示されます。
16. 【ファイル】 > 【保存】をクリックし、マッピングを保存します。

タスク 3. パーサトランスフォーメーションでのプロファイルの実行

このタスクでは、パーサトランスフォーメーションでプロファイルを実行し、フルネームが正しく解析されるようにパーサトランスフォーメーションが設定されていることを確認します。

1. ParserMapping マッピングが含まれているエディタを選択します。
2. パーサ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。
プロファイルが実行され、エディタに表示されます。
3. エディタで、**【結果】** ビューをクリックし、プロファイリング操作の結果を表示します。
4. Undefined_output カラムを選択し、**【詳細】** セクション内のカラムに関する情報を表示します。
Undefined_output カラムに含まれている値が、それぞれの値の頻度と割合の統計とともに、**【詳細】** セクションに表示されます。
5. データを表示し、Undefined_output カラムにファーストネームのみが表示されることを確認します。

タスク 4. マッピングの実行

このタスクでは、マッピングを実行してマッピング出力を作成します。

1. ParserMapping マッピングが含まれているエディタを選択します。
2. **【実行】** > **【マッピングの実行】** をクリックします。
マッピングが実行され、LA_Customers_tgt.csv ファイルに出力が書き込まれます。

タスク 5. マッピング出力の表示

このタスクでは、データビューアを実行してマッピング出力を表示します。

1. Object Explorer ビューで、チュートリアルプロジェクトの LA_Customers_tgt データオブジェクトを特定し、そのデータオブジェクトをダブルクリックします。
データオブジェクトがエディタに表示されます。
2. **【ウィンドウ】** > **【ビューの表示】** > **【データビューア】** をクリックします。
[データビューア] ビューが開きます。
3. [データビューア] ビューで、**【実行】** をクリックします。
データビューアが実行され、データが表示されます。
4. 正しく解析されたデータが FirstName カラムと LastName カラムに表示されていることを確認します。

データの解析のまとめ

このレッスンでは、データの解析によって入力フィールド内のデータ要素を識別し、各要素を新しいカラムに書き出す方法を学習しました。

パーサートランスフォーメーションを使用してデータを解析する方法を学習しました。また、マッピング内の任意のトランスフォーメーションのプロファイルを作成し、そのトランスフォーメーションからの出力を分析することも学習しました。最後に、データビューアを使用してマッピング出力を表示できることを学習しました。

解析した出力を格納するための LA_Customers_tgt データオブジェクトを作成および設定しました。また、データを解析するためのマッピングを作成しました。そのマッピングで、トークンセットを使用してパーサートランスフォーメーションを設定し、ロサンゼルスのお客様ファイルの FullName カラムから姓と名を解析しました。解析したデータを LA_Customers_tgt データオブジェクトの Firstname カラムと Lastname カラムに書き込むようにマッピングを設定しました。また、マッピングを実行する前にプロファイルを実行し、トランスフォーメーションの出力を表示しました。最後に、マッピングを実行し、データビューアを使用して LA_Customers_tgt データオブジェクトの新しいデータカラムを表示しました。

第 14 章

レッスン 5. データの標準化

この章では、以下の項目について説明します。

- [データの標準化の概要, 70 ページ](#)
- [タスク 1. ターゲットデータオブジェクトの作成, 71 ページ](#)
- [タスク 2. データを標準化するためのマッピングの作成, 72 ページ](#)
- [タスク 3. マッピングの実行, 75 ページ](#)
- [タスク 4. マッピング出力の表示, 75 ページ](#)
- [データの標準化のまとめ, 75 ページ](#)

データの標準化の概要

データを標準化すると、データのエラーや不整合が減り、データ品質が向上します。

データ品質を向上させるには、次のような値を含むデータを標準化します。

- 正しくない値
- 情報は正しいものの形式に誤りがある値
- 新しい情報の元になる値

データ内のこのような値を検索するには、標準化トランスフォーメーションを使用します。以下のいずれかの検索操作を選択できます。

- **テキスト:** 入力した任意の文字列を検索します。検索した文字列は削除するか、指定したテキストで置き換えます。
- **参照テーブル:** 選択した参照テーブルに格納されている文字列を検索します。検索した文字列は削除するか、参照テーブルのエントリまたは指定したテキストで置き換えます。

例えば、Street および St. という文字列を含むアドレスデータを、ST. という置換文字を使用して標準化するように標準化トランスフォーメーションを設定することができます。標準化トランスフォーメーションによって、検索語句が ST. という語句に置き換えられ、新しいデータカラムに結果が書き込まれます。

内容

HypoStores の顧客のアドレスデータを、すべてのアドレスで一貫した語句を使用するように標準化する必要があります。All_Customers データオブジェクト内のアドレスデータの形式に一貫性がなく、Street、Boulevard、Avenue、Drive、Park などの一般的な語句がさまざまな形式で入力されています。

目標

このレッスンでは、以下の作業を完了します。

- All_Customers_Stdz_tgt というデータオブジェクトを作成し、標準化されたデータを格納するように設定する。
- アドレスに含まれる Street、Boulevard、Avenue、Drive、および Park という語句を標準化して形式を統一するマッピングを作成する。
- All_Customers データオブジェクトをマッピングに追加してソースデータに接続する。
- All_Customers_Stdz_tgt データオブジェクトをマッピングに追加してターゲットデータオブジェクトを作成する。
- 標準化トランスフォーメーションをマッピングに追加し、アドレスに含まれる語句を標準化するように設定する。
- マッピングを実行して、標準化されたアドレスデータを生成する。
- データビューアを実行してマッピング出力を確認する。

要件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスンの 1 から 2 までを完了している。

時間

このレッスンの所要時間は 15 分です。

タスク 1. ターゲットデータオブジェクトの作成

このタスクでは、標準化したデータを書き込むための All_Customers_Stdz_tgt データオブジェクトを作成します。

ターゲットデータオブジェクトを作成するには、以下の手順を実行します。

1. All_Customers.csv ファイルに基づいて、All_Customers_Stdz_tgt データオブジェクトを作成します。
2. データオブジェクトの読み取りおよび書き込みのオプションを、ファイルの場所やファイル名も含めて設定します。

手順 1.All_Customers_Stdz_tgt データオブジェクトの作成

この手順では、All_Customers.csv ファイルに基づいて、All_Customers_Stdz_tgt データオブジェクトを作成します。

1. **【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】 ウィンドウが開きます。
2. **【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。
3. **【既存のフラットファイルから作成】** が選択されていることを確認します。
4. **【参照】** をクリックし、次のディレクトリに移動して All_Customers.csv を探します。 <Informatica Installation Directory>\clients\DeveloperClient\Tutorials
5. **【開く】** をクリックします。
6. **【名前】** フィールドに、「All_Customers_Stdz_tgt」と入力します。

7. **【次へ】** をクリックします。
8. **【次へ】** をクリックします。
9. **【プレビューのオプション】** セクションで、**【1 行目からカラム名をインポート】** を選択し、**【次へ】** をクリックします。
10. **【完了】** をクリックします。
All_Customers_Stdz_tgt データオブジェクトがエディタに表示されます。

手順 2.読み取りオプションおよび書き込みオプションの設定

この手順では、All_Customers_Stdz_tgt データオブジェクトの読み取りおよび書き込みのオプションを、ファイルの場所やファイル名を含めて設定します。

1. All_Customers_Stdz_tgt データオブジェクトがエディタに表示されていることを確認します。
2. エディタで、**【読み取り】** ビューを選択します。
3. **【ウィンドウ】** > **【ビューの表示】** > **【プロパティ】** をクリックします。
4. **【プロパティ】** ビューで、**【ランタイム】** ビューを選択します。
5. **【値】** カラムで、ソースファイル名をダブルクリックし、「All_Customers_Stdz_tgt.csv」と入力します。
6. **【値】** カラムで、**【ソースファイルのディレクトリ】** エントリをダブルクリックします。
7. 強調表示された名前を右クリックし、**【コピー】** を選択します。
8. エディタで、**【書き込み】** ビューを選択します。
9. **【プロパティ】** ビューで、**【ランタイム】** ビューを選択します。
10. **【値】** カラムで、**【出力ファイルディレクトリ】** エントリをダブルクリックします。
11. 右クリックして **【貼り付け】** を選択し、**【読み取り】** ビューからコピーしたディレクトリの場所を貼り付けます。
12. **【値】** カラムで、**【ヘッダーのオプション】** エントリをダブルクリックし、**【出力フィールド名】** を選択します。
13. **【値】** カラムで、**【出力ファイル名】** エントリをダブルクリックし、「All_Customers_Stdz_tgt.csv」と入力します。
14. **【ファイル】** > **【保存】** をクリックし、データオブジェクトを保存します。

タスク 2. データを標準化するためのマッピングの作成

このタスクでは、マッピングを作成し、データオブジェクトおよび標準化トランスフォーメーションを使用するようにそのマッピングを設定します。

データを標準化するためのマッピングを作成するには、以下の手順を実行します。

1. マッピングを作成します。
2. ソースおよびターゲットのデータオブジェクトをマッピングに追加します。
3. 標準化トランスフォーメーションをマッピングに追加します。
4. 一般的なアドレス用語を一貫性のある形式に標準化するように、標準化トランスフォーメーションを設定します。

手順 1. マッピングの作成

この手順では、マッピングを作成して名前を付けます。

1. **Object Explorer** ビューで、チュートリアルプロジェクトを選択します。
2. **【ファイル】 > 【新規】 > 【マッピング】** をクリックします。
【新しいマッピング】 ウィンドウが開きます。
3. **【名前】** フィールドに、「StandardizerMapping」と入力します。
4. **【完了】** をクリックします。
マッピングがエディタに表示されます。

手順 2. データオブジェクトのマッピングへの追加

このステップでは、All_Customers データオブジェクトと All_Customers_Stdz_tgt データオブジェクトをマッピングに追加します。

1. **【Object Explorer】** ビューで、チュートリアルプロジェクト内のデータオブジェクトを参照します。
2. All_Customers データオブジェクトを選択し、エディタにドラッグします。
【物理データオブジェクトのマッピングへの追加】 ウィンドウが開きます。
3. **【読み取り】** が選択されていることを確認し、**【OK】** をクリックします。
データオブジェクトがエディタに表示されます。
4. **【Object Explorer】** ビューで、チュートリアルプロジェクト内のデータオブジェクトを参照します。
5. All_Customers_Stdz_tgt データオブジェクトを選択し、エディタにドラッグします。
【物理データオブジェクトのマッピングへの追加】 ウィンドウが開きます。
6. **【書き込み】** を選択し、**【OK】** をクリックします。
データオブジェクトがエディタに表示されます。
7. All_Customers データオブジェクトのすべてのポートを選択します。選択したポートを、All_Customers_Stdz_tgt データオブジェクトの CustomerID ポートにドラッグします。
ヒント: 複数のポートを選択するには、Shift キーを押しながら選択します。すべてのポートを選択するには、ポートのリストをスクロールしなければならない場合もあります。
All_Customers データオブジェクトのポートが、All_Customers_Stdz_tgt データオブジェクトの対応するポートに接続されます。

手順 3. 標準化トランスフォーメーションのマッピングへの追加

この手順では、アドレスデータに含まれる文字列を標準化するために、標準化トランスフォーメーションを追加します。

1. StandardizerMapping マッピングが含まれているエディタを選択します。
2. トランスフォーメーションパレットで、標準化トランスフォーメーションを選択します。
3. エディタをクリックします。
NewStandardizer という名前の標準化トランスフォーメーションがマッピング内に表示されます。
4. 標準化トランスフォーメーションの名前を変更するには、トランスフォーメーションのタイトルバーをダブルクリックし、「AddressStandardizer」と入力します。

5. All_Customers データオブジェクトの Address1 ポートを選択し、標準化トランスフォーメーションの【入力】グループにドラッグします。
Address1 という名前のポートが入力グループ内に表示されます。このポートは、All_Customers データオブジェクトの Address1 ポートに接続されます。

注: 出力ポートは、標準化ストラテジを設定する際にトランスフォーメーションに追加します。

手順 4.標準化トランスフォーメーションの設定

このステップでは、ソースデータ内のアドレスに含まれる語句を標準化するように標準化トランスフォーメーションを設定します。

注: ここでは、5 つの標準化操作を定義します。それぞれの操作で、入力カラムの文字列を新しい文字列に置き換えます。

1. StandardizerMapping マッピングが含まれているエディタを選択します。
2. 標準化トランスフォーメーションをクリックします。
3. 【ウィンドウ】 > 【ビューの表示】 > 【プロパティ】 をクリックします。
4. 【プロパティ】 ビューで、【ストラテジ】 を選択します。
5. 【新規】 をクリックします。新しいストラテジウィザードが表示されます。
6. 入力カラムの選択矢印をクリックし、Address1 入力ポートを選択します。
出力フィールドに、出力ポートとして Address1 が表示されます。
7. 空白文字およびカンマ区切り文字 ([\s]と[,]) を選択します。必要に応じて、末尾のスペースを削除するオプションを選択します。
8. 【次へ】 をクリックします。
9. 【カスタム文字列の置換】 操作を選択し、【次へ】 をクリックします。
10. 【プロパティ】 の下にある【新規】 をクリックします。
11. 【カスタム文字列】 フィールドと【置換後の文字列】 フィールドを編集して、以下の表の最初の文字列の組み合わせを入力します。

カスタム文字列	置換後の文字列
STREET	ST.
BOULEVARD	BLVD.
AVENUE	AVE.
DRIVE	DR.
PARK	PK.

12. 手順 9～12 を繰り返して、この表のすべての文字列について標準化操作を定義します。
13. Address1 出力ポートを、All_Customers_Stdz_tgt データオブジェクトの Address1 ポートにドラッグします。
14. 【ファイル】 > 【保存】 をクリックし、マッピングを保存します。

タスク 3. マッピングの実行

このタスクでは、マッピングを実行し、標準化されたアドレスを出力データオブジェクトに書き込みます。

1. StandardizerMapping マッピングが含まれているエディタを選択します。
2. **【実行】** > **【マッピングの実行】** をクリックします。
マッピングが実行され、All_Customers_Stdz_tgt.csv ファイルに出力が書き込まれます。

タスク 4. マッピング出力の表示

このタスクでは、データビューアを実行してマッピング出力を表示し、アドレスデータが正しく標準化されていることを確認します。

1. **Object Explorer** ビューで、チュートリアルプロジェクトの All_Customers_Stdz_tgt データオブジェクトを検索し、そのデータオブジェクトをダブルクリックします。
データオブジェクトがエディタに表示されます。
2. **【ウィンドウ】** > **【ビューの表示】** > **【データビューア】** をクリックします。
[データビューア] ビューが開きます。
3. [データビューア] ビューで、**【実行】** をクリックします。
データビューアにマッピング出力が表示されます。
4. Address1 カラムに正しく標準化されたデータが表示されていることを確認します。例えば、文字列 STREET のすべてのインスタンスが、文字列 ST. で置換されている必要があります。

データの標準化のまとめ

このレッスンでは、データを標準化することにより、そのデータ内のエラーや不整合を削除できることを学習しました。

標準化トランスフォーメーションを使用して、入力カラム内の文字列を標準化できることを学習しました。また、データビューアを使用してマッピング出力を表示することも学習しました。

All_Customers_Stdz_tgt データオブジェクトを作成および設定し、標準化された出力を格納しました。データを標準化するためのマッピングを作成しました。そのマッピングで、標準化トランスフォーメーションを設定し、All_Customers データオブジェクトの Address1 カラムを標準化しました。マッピングを設定し、標準化した出力を All_Customers_Stdz_tgt データオブジェクトに書き込みました。最後に、マッピングを実行し、データビューアを使用して、All_Customers_Stdz_tgt データオブジェクト内の標準化されたデータを表示しました。

第 15 章

レッスン 6. 住所データの検証

この章では、以下の項目について説明します。

- [アドレスデータの検証の概要, 76 ページ](#)
- [タスク 1. ターゲットデータオブジェクトの作成, 77 ページ](#)
- [タスク 2. アドレスを検証するためのマッピングの作成, 79 ページ](#)
- [タスク 3. アドレスバリ Data Transformation の設定, 80 ページ](#)
- [タスク 4. マッピングの実行, 83 ページ](#)
- [タスク 5. マッピング出力の表示, 84 ページ](#)
- [アドレスデータの検証のまとめ, 86 ページ](#)

アドレスデータの検証の概要

アドレスの検証は、郵便アドレスの品質を評価および改善するプロセスです。このプロセスでは、入力アドレスを有効なアドレスの参照データセットと比較することによってアドレスの品質を評価します。また、不正なアドレス値を特定し、参照データセットを使用して正しい値を含むフィールドを作成することにより、アドレスの品質を向上させます。

アドレスは、配達可能な場合に有効とみなされます。アドレスが適切な形式で、実際の番地、市区町村、および郵便番号の情報が含まれていたとしても、そのデータが配達可能なアドレスを構成しなければそのアドレスは無効です。Developer ツールは、アドレス参照データセットを使用して入力アドレスの配達可能性をチェックします。アドレス参照データセットは Informatica から提供されています。

アドレス参照データセットには、国内の配達可能アドレスがすべて記述されたデータが含まれています。アドレス検証プロセスでは、入力アドレスデータに最もよく似たアドレスを参照データセットから探します。よく似たアドレスが参照データセット内で見つかり、無効または不完全なデータ値に対して新しい値が書き込まれます。このプロセスでは、入力アドレスと参照アドレスの間に見つかった一致のタイプを説明する一連の英数字のコードが作成されます。また、アドレスを再構築したり、入力アドレスにない情報（米国のアドレスの 4 桁の郵便番号サフィックスなど）を追加したりすることもできます。

Developer ツールでアドレス検証プロセスを構築するには、アドレスバリデータトランスフォーメーションを使用します。このマルチグループトランスフォーメーションには、入力アドレスで使用するすべてのフィールドに対応する、一連の定義済みの入力ポートおよび出力ポートが含まれています。アドレスバリデータトランスフォーメーションを選択する場合は、デフォルトの参照データセットを設定し、トランスフォーメーションポートを使用して入出力アドレス構造体を作成します。このレッスンでは、トランスフォーメーションを設定し、米国のアドレスデータを検証します。

内容

HypoStores では、自社のダイレクトメールキャンペーンやその他の郵便物が顧客に確実に届くようにするために、正確で完全なアドレスデータを必要としています。アドレスデータが正確で完全であれば、組織の郵送業務のコストも削減されます。さらに、HypoStores では、さまざまな長さのアドレスを格納するのに十分な柔軟性を備えた、印刷可能な形式のアドレスを顧客データに含めることを必要としています。

これらのビジネス要件を満たすために、HypoStores ICC チームは、Developer ツールでアドレス検証マッピングを作成します。

目標

このレッスンでは、以下の作業を完了します。

- 検証済みのアドレスフィールドおよび一致コードを格納するターゲットデータオブジェクトを作成する。
- ソースデータオブジェクト、ターゲットデータオブジェクト、およびアドレスバリデータトランスフォーメーションを使用してマッピングを作成する。
- アドレスバリデータトランスフォーメーションを設定し、顧客のアドレスデータを検証する。
- マッピングを実行してアドレスデータを検証し、一致コード出力を確認してアドレスデータの有効性を検証する。

要件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスン 1 から 2 までを完了している。
- 米国のアドレス参照データがドメイン内にインストールされ、Administrator ツールに登録されていること。米国のアドレスデータがシステムにインストールされていることを確認するには、Informatica 管理者に問い合わせてください。参照データは、Data Quality Content Installer によってインストールされます。

時間

このレッスンには 25 分かかります。

タスク 1. ターゲットデータオブジェクトの作成

このタスクでは、ターゲットデータオブジェクトを作成し、書き込みオプションを設定して、ポートを追加します。

ターゲットデータオブジェクトを作成して設定するには、以下の手順を実行します。

1. All_Customers.csv ファイルに基づいて All_Customers_av_tgt データオブジェクトを作成します。
2. データオブジェクトの読み取りと書き込みのオプションを、ファイルの場所やファイル名を含めて設定します。
3. アドレスバリデータトランスフォーメーションによって生成される一致コードの値を受け取るポートをデータオブジェクトに追加します。

手順 1.All_Customers_av_tgt データオブジェクトの作成

この手順では、All_Customers.csv ファイルに基づいて、All_Customers_av_tgt データオブジェクトを作成します。

1. **【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】 ウィンドウが開きます。
2. **【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。
3. **【既存のフラットファイルから作成】** が選択されていることを確認します。この選択項目の横にある **【参照】** をクリックし、All_Customers.csv ファイルを探して、**【開く】** をクリックします。
4. **【名前】** フィールドに、「All_Customers_av_tgt」と入力します。
5. **【次へ】** をクリックします。
6. **【次へ】** をクリックします。
7. **【プレビューのオプション】** セクションで、**【1 行目からカラム名をインポート】** を選択し、**【次へ】** をクリックします。
8. **【完了】** をクリックします。

All_Customers_av_tgt データオブジェクトがエディタに表示されます。

手順 2.読み取りオプションおよび書き込みオプションの設定

このステップでは、All_Customers_av_tgt データオブジェクトの読み取りと書き込みのオプションを、ターゲットファイルの名前や場所を含めて設定します。

1. All_Customers_av_tgt データオブジェクトがエディタで開いていることを確認します。
2. エディタで、**【読み取り】** ビューを選択します。
3. **【ウィンドウ】 > 【ビューの表示】 > 【プロパティ】** を選択します。
4. **【プロパティ】** ビューで、**【ランタイム】** ビューを選択します。
5. **【値】** カラムで、ソースファイルの名前をダブルクリックし、「All_Customers_av_tgt.csv」と入力します。
6. **【値】** カラムで、ソースファイルのディレクトリパスをダブルクリックして強調表示します。
7. 強調表示したパスと名前を右クリックし、**【コピー】** を選択します。
8. エディタで、**【書き込み】** ビューを選択します。
9. **【プロパティ】** ビューで、**【ランタイム】** ビューを選択します。
10. **【値】** カラムで、**【出力ファイルディレクトリ】** エントリをダブルクリックします。
11. このエントリを右クリックして **【貼り付け】** を選択し、**【読み取り】** ビューからコピーしたパスを追加します。
12. **【値】** カラムで、**【ヘッダーのオプション】** エントリをダブルクリックし、**【出力フィールド名】** を選択します。
13. **【値】** カラムで、**【出力ファイル名】** エントリをダブルクリックし、「All_Customers_av_tgt.csv」と入力します。
14. **【ファイル】 > 【保存】** を選択して、データオブジェクトを保存します。

手順 3.ポートのデータオブジェクトへの追加

このステップでは、アドレスバリデータトランスフォーメーションで一致コードの値をターゲットファイルに書き込めるように、All_Customers_av_tgt データオブジェクトに 2 つのポートを追加します。ポートにはそれぞれ、MailabilityScore と MatchCode という名前を付けます。

MailabilityScore の値は、入力アドレスの配達可能性を示します。MatchCode の値は、トランスフォーメーションで入力アドレスと参照データのアドレスを照合する際の照合のタイプを示します。

1. **[Object Explorer]** ビューで、チュートリアルプロジェクト内のデータオブジェクトを参照します。
2. All_Customers_av_tgt データオブジェクトをダブルクリックします。
All_Customers_av_tgt データオブジェクトがエディタで開きます。
3. **[概要]** が選択されていることを確認します。
4. ポートのリストで一番下にあるポートを選択します。このポートの名前は MiscDate です。
5. **[新規]** をクリックします。
MiscDate1 という名前のポートが表示されます。
6. MiscDate1 ポートの名前を MailabilityScore に変更します。
7. MailabilityScore ポートを選択します。
8. **[新規]** をクリックします。
MailabilityScore1 という名前のポートが表示されます。
9. MailabilityScore1 ポートの名前を MatchCode に変更します。
10. **[ファイル]** > **[保存]** をクリックし、データオブジェクトを保存します。

タスク 2. アドレスを検証するためのマッピングの作成

このタスクでは、マッピングを作成し、データオブジェクトおよびアドレスバリデータトランスフォーメーションを追加します。

マッピングを作成し、必要なオブジェクトを追加するには、以下の手順を実行します。

1. マッピングオブジェクトを作成します。
2. ソースおよびターゲットのデータオブジェクトをマッピングに追加します。
3. アドレスバリデータトランスフォーメーションをマッピングに追加します。

手順 1.マッピングの作成

この手順では、マッピングを作成して名前を付けます。

1. Object Explorer ビューで、チュートリアルプロジェクトを選択します。
2. **[ファイル]** > **[新規]** > **[マッピング]** を選択します。
[新しいマッピング] ウィンドウが開きます。
3. **[名前]** フィールドに、「ValidateAddresses」と入力します。
4. **[完了]** をクリックします。
マッピングがエディタに表示されます。

手順 2. データオブジェクトのマッピングへの追加

この手順では、ソースおよびターゲットのデータオブジェクトをマッピングに追加します。

All_Customers は、マッピングのソースデータオブジェクトです。アドレスバリデータトランスフォーメーションは、このオブジェクトからデータを読み取ります。All_Customers_av_tgt は、マッピングのターゲットデータオブジェクトです。このオブジェクトは、アドレスバリデータトランスフォーメーションからデータを読み取ります。

1. **[Object Explorer]** ビューで、チュートリアルプロジェクト内のデータオブジェクトを参照します。
2. All_Customers データオブジェクトを選択し、エディタにドラッグします。
[物理データオブジェクトのマッピングへの追加] ウィンドウが開きます。
3. **[読み取り]** が選択されていることを確認し、**[OK]** をクリックします。
データオブジェクトがエディタに表示されます。
4. **[Object Explorer]** ビューで、チュートリアルプロジェクト内のデータオブジェクトを参照します。
5. All_Customers_av_tgt データオブジェクトを選択し、エディタにドラッグします。
[物理データオブジェクトのマッピングへの追加] ウィンドウが開きます。
6. **[書き込み]** を選択し、**[OK]** をクリックします。
データオブジェクトがエディタに表示されます。
7. **[保存]** をクリックします。

手順 3. アドレスバリデータトランスフォーメーションのマッピングへの追加

このステップでは、ソースとデータオブジェクトを含むマッピングにアドレスバリデータトランスフォーメーションを追加します。

このステップが完了すると、トランスフォーメーションを設定して、そのポートをデータオブジェクトに接続できるようになります。

1. ValidateAddresses マッピングを含むエディタを選択します。
2. **[トランスフォーメーション]** パレットで、アドレスバリデータトランスフォーメーションを選択します。
3. エディタをクリックします。
アドレスバリデータトランスフォーメーションがエディタに表示されます。

タスク 3. アドレスバリ Data Transformation の設定

このタスクでは、アドレスバリデータトランスフォーメーションを設定し、All_Customers データソースからアドレスを読み取って検証します。

注: アドレスバリデータトランスフォーメーションには、一連の定義済みの入力ポートおよび出力ポートが含まれています。必要なポートを選択し、それらをマッピング内のオブジェクトに接続します。

トランスフォーメーションを設定するには、以下の手順を実行します。

1. アドレス検証のデフォルトの国を選択します。
2. トランスフォーメーションの入力ポートを設定します。
3. トランスフォーメーションの出力ポートを設定します。

4. 未使用のソースポートをデータターゲットに接続します。

手順 1. アドレス検証のデフォルトの国の設定

この手順では、アドレス検証のデフォルトの国を選択します。アドレスバリデータトランスフォーメーションが使用するアドレス参照データファイルは、国別に編成されています。デフォルトの国を選択すると、国の情報が含まれていない入力アドレスにトランスフォーメーションが適用する国データセットが指定されます。

1. エディタでアドレスバリデータトランスフォーメーションを選択します。
2. **【プロパティ】** の下にある **【全般設定】** をクリックします。
3. **【デフォルトの国】** メニューで、**【米国】** を選択します。

手順 2. アドレスバリデータトランスフォーメーションの入力ポートの設定

このステップでは、トランスフォーメーションの入力ポートを選択し、それらのポートを All_Customers_av データオブジェクトに接続します。

アドレスバリデータトランスフォーメーションには、定義済みの入力ポートのグループがいくつか用意されています。入力アドレスのフィールドに対応する入力ポートを選択し、それらのポートをトランスフォーメーションに追加します。

以下の手順で 1 回の操作で複数のポートを選択するときは、Ctrl キーを押しながら選択します。

1. エディタでアドレスバリデータトランスフォーメーションを選択します。
2. **【プロパティ】** で、**【テンプレート】** をクリックします。
3. **【基本モデル】** ポートグループを展開します。
4. **【混合】** 入力ポートグループを展開し、以下のポートを選択します。

ポート名	説明
送付先住所 1	番地、ビル番号などの所在地住所データ。
市区町村の正式表記 1	市区町村の名前。
郵便番号 1	郵便番号。
都道府県 1	都道府県または州の名前。
国名	国の名前または略号。

注: 1 回の操作で複数のポートを選択するには、Ctrl キーを押しながら選択します。

5. ポート名のリストの上にあるツールバーで、**【トランスフォーメーションにポートを追加】** をクリックします。

このツールバーは、**【テンプレート】** を選択すると表示されます。

選択したポートが、マッピングエディタのトランスフォーメーションに表示されます。

6. ソースポートを以下のようにアドレスバリデータトランスフォーメーションのポートに接続します。

ソースポート	アドレスバリデータトランスフォーメーションのポート
Address1	送付先住所 1
City	市区町村の正式表記 1
ZIP	郵便番号 1
状態	都道府県 1
Country	国名

手順 3. アドレスバリデータトランスフォーメーションの出力ポートの設定

この手順では、トランスフォーメーションの出力ポートを選択し、それらのポートを All_Customers_av_tgt データオブジェクトに接続します。

アドレスバリデータトランスフォーメーションには、定義済みの出力ポートのグループが複数含まれています。必要なアドレス構造を定義するポートを選択し、それらのポートをトランスフォーメーションに追加します。

また、各アドレスに対して実行された検証のタイプに関する情報が含まれたポートを選択できます。

- マッピングエディタでアドレスバリデータトランスフォーメーションを選択します。
- 【プロパティ】で、【テンプレート】をクリックします。
- 【基本モデル】ポートグループを展開します。
- 【アドレス要素】出力ポートグループを展開し、以下のポートを選択します。

ポート名	説明
町名の正式表記 1	番地、ビル番号などの所在地住所データ。

5. 【Last Line 要素】出力ポートグループを展開し、以下のポートを選択します。

ポート名	説明
市区町村の正式表記 1	市区町村の名前。
郵便番号 1	郵便番号。
都道府県の略式表記 1	都道府県または州の識別子。

注: 1 回の操作で複数のポートを選択するには、Ctrl キーを押しながら選択します。

6. 【国】出力ポートグループを展開し、以下のポートを選択します。

ポート名	説明
国名 1	国名。

7. **【ステータス情報】** 出力ポートグループを展開し、以下のポートを選択します。

ポート名	説明
郵送可能スコア	郵便配達が正常に行われる確率を表すスコア。
照合コード	入力アドレスと参照データの間の類似度を表すコード。

8. ポート名のリストの上にあるツールバーで、**【トランスフォーメーションにポートを追加】** をクリックします。
このツールバーは、**【テンプレート】** を選択すると表示されます。
9. 以下のように、アドレスバリデータトランスフォーメーションのポートを All_Customers_av_tgt ポートに接続します。

アドレスバリデータトランスフォーメーションのポート	ターゲットポート
町名の正式表記 1	Address1
市区町村の正式表記 1	市区町村
郵便番号 1	ZIP
都道府県の略式表記 1	状態
国名 1	Country
郵送可能スコア	MailabilityScore
照合コード	MatchCode

手順 4.未使用のデータソースポートのデータターゲットへの接続

このステップでは、All_Customers データソースの未使用のポートをデータターゲットに接続します。

- ▶ データソースの未使用のポートをデータターゲットの同じ名前のポートに接続します。

タスク 4. マッピングの実行

このタスクでは、マッピングを実行してマッピング出力を作成します。

1. ValidateAddresses マッピングを含むエディタを選択します。
2. **【実行】** > **【マッピングの実行】** を選択します。

マッピングが実行され、All_Customers_av_tgt.csv ファイルに出力が書き込まれます。

タスク 5。マッピング出力の表示

このタスクでは、データビューアを実行してマッピング出力を表示します。ターゲットデータオブジェクトの郵送可能スコアカラムおよび照合コードカラムに書き込まれた値を調査することにより、検証済みのアドレスの品質を確認します。

照合コードの値は、マッピングでアドレスに対して実行した検証のタイプを表す英数字のコードです。

郵送可能スコアの値は、アドレスの配達可能性を表す 1 桁の値です。

1. Object Explorer ビューで、チュートリアルプロジェクトの All_Customers_av_tgt データオブジェクトを特定し、そのデータオブジェクトをダブルクリックします。

データオブジェクトがエディタに表示されます。

2. **[ウィンドウ]** > **[ビューの表示]** > **[データビューア]** を選択します。

データビューアが開きます。

3. データビューアで、**[実行]** をクリックします。

データビューアにマッピング出力が表示されます。

4. 郵送可能スコアカラムおよび照合コードカラムが表示されるように、マッピング結果全体をスクロールします。

5. 郵送可能スコアカラムの値を確認します。

スコアの範囲は 0 から 5 です。スコアが高いアドレスほど、支障なく配達される可能性が高くなります。

6. 照合コードカラムの値を確認します。

照合コードは英数字のコードです。アルファベットの文字はトランスフォーメーションが実行した検証のタイプを示し、数字は最終的なアドレスの品質を示します。

次の表で、[照合コード] の値について説明します。

コード	説明
A1	アドレスコードルックアップにより、入力コード内に部分的なアドレスまたは完全なアドレスが見つかりました。
A0	アドレスコードにより、入力コード内にアドレスが見つかりませんでした。
C4	修正済み。郵送に関連する要素すべてが確認されています。
C3	修正済み。いくつかの要素が確認できません。
C2	修正済みですが、参照データが存在しないため、配達ステータスが不明です。
C1	修正済みですが、ユーザーの標準化によって誤りが生じたため配達ステータスは不明です。
I4	データを完全に修正できませんが、参照データに 1 つの一致があります。
I3	データを完全に修正できませんが、参照データに複数の一致があります。
I2	データを修正できません。バッチモードで部分的に提案されたアドレスが返されます。

コード	説明
I1	データを修正できません。バッチモードでアドレスを提案できません。
N7	検証エラーです。単一行の検証がロック解除されていないため、検証されませんでした。
N6	検証エラーです。宛先の国で単一行の検証がサポートされていないため、検証されませんでした。
N5	検証エラーです。参照データベースが古すぎるため、検証を行えませんでした。
N4	検証エラーです。参照データが破損しているか不正なフォーマットのため、検証を行えませんでした。
N3	検証エラーです。国データのロック解除ができないため、検証を行えませんでした。
N2	検証エラーです。要求された参照データベースが利用できないため、検証を行えませんでした。
N1	検証エラーです。国が認識できなかったかサポートされていないため、検証を行えませんでした。
Q3	提案リストモード。アドレス検証では、入力アドレスに対応するアドレス参照データから完全なアドレスを 1 つ以上取得できます。
Q2	提案リストモード。アドレス検証では、入力アドレスの要素とアドレス参照データ内の要素を組み合わせ、完全なアドレスを作成できます。
Q1	提案リストモード。アドレスの検証は、完全なアドレスを提案できません。完全なアドレス提案を生成するには、入力アドレスにデータを追加します。
Q0	提案リストモード。提案を生成するための十分な入力データがありません。
RB	略称によって国が認識されました。ISO の 2 文字と ISO の 3 文字の国コードを認識します。ドイツを表す「GER」のような一般的な略称も認識できます。
RA	トランスフォーメーション内の国を強制的に適用設定によって国が認識されました。
R9	トランスフォーメーション内のデフォルトの国設定によって国が認識されました。
R8	国名によって国が認識されました。
R7	国名によって国が認識されましたが、トランスフォーメーションによって国データ内にエラーが識別されました。
R6	地域データによって国が認識されました。
R5	都道府県のデータによって国が認識されました。

コード	説明
R4	主要な都市データによって国が認識されました。
R3	住所フォーマットによって国が認識されました。
R2	スクリプトから国が認識されました。
R1	使用できる一致が複数あるため、国が認識されませんでした。
R0	国が認識されません。
S4	解析モード。住所は完全に解析されました。
S3	解析モード。住所は解析され、複数の結果が返されました。
S1	解析モード。入力フォーマットの不一致のため解析エラーが発生しました。
V4	検証済み。入力されたデータは正確です。アドレス検証が郵送に関連する要素すべてを確認した結果、入力は完全に一致しました。
V3	検証済み。入力データは正しいですが、一部あるいは全部の要素が標準化されていたか、入力データに古い名前あるいは外名が含まれています。
V2	検証済み。入力データは正確ですが、参照データが完全ではないため、一部の要素を検証できません。
V1	検証済み。入力されたデータは正確ですが、ユーザーによる標準化が配達可能性を低下させています。たとえば、ポストコードの長さが短すぎます。

アドレスデータの検証のまとめ

このレッスンでは、アドレスの検証では、入力アドレスデータが参照データと比較され、最も正確と考えられるアドレスが返されることを学習しました。

アドレス検証プロセスでは、各アドレスの品質に関するステータス情報も返されることを学習しました。

Administrator ツールのユーザーが Data Quality Content Installer を実行してアドレス参照データをインストールすることを学習しました。

また、アドレスバリデータトランスフォーメーションはマルチグループトランスフォーメーションであること、ポートグループからのトランスフォーメーションのために入力および出力ポートを選択することも学習しました。選択した入力ポートによって、検証されるアドレスの内容が決まります。出力ポートによって、最終的なアドレスレコードの内容が決まります。

付録 A

FAQ（よくある質問）

この付録では、以下の項目について説明します。

- [Informatica Analyst の FAQ（よくある質問）, 87 ページ](#)
- [Informatica Developer の FAQ（よくある質問）, 87 ページ](#)

Informatica Analyst の FAQ（よくある質問）

ここでは、Informatica Analyst についてのよくある質問とその答えを紹介します。

1つのユーザーアカウントで、Administrator ツール、Developer tool、および Analyst ツールにアクセスできますか。

はい、できます。3つのツールすべてにアクセスするためのユーザー権限を付与できます。クライアントアプリケーションごとに別々のユーザーアカウントを作成する必要はありません。

個人の参照データはどこに格納されるのですか。

参照データオブジェクトの作成と共有には、Developer tool と Analyst ツールを使用できます。モデルリポジトリには、参照データオブジェクトのメタデータが保存されます。参照データデータベースには、参照テーブルデータ値が格納されます。コンテンツ管理サービスで参照データデータベースを設定します。

Informatica Developer の FAQ（よくある質問）

ここでは、Informatica Developer についてのよくある質問とその答えを紹介します。

マップレットとルールの違いは何ですか。

マップレットをルールとして検証できます。ルールは、プロファイルの実行時など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。マップレットが次の要件を満たすときは、マップレットをルールとして検証できます。

- 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が含まれている。
- アクティブなトランスフォーメーションが含まれていない。
- 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。

私には Data Engineering 製品ライセンスがあります。Developer tool を使用して PowerCenter へオブジェクトをエクスポートできますか。

いいえ。Engineering 製品は PowerCenter と統合しません。

PowerCenter のソースおよびターゲットと Developer tool の物理データオブジェクトの違いは何ですか。

PowerCenter では、マッピングソースとして含めるソース定義を作成します。マッピングターゲットとして含めるターゲット定義を作成します。Developer tool では、マッピングソースまたはターゲットとして使用できる物理データオブジェクトを作成します。

Developer tool でのマッピングと PowerCenter でのマッピングの違いは何ですか。

PowerCenter でのマッピングは、ソースとターゲットの間でデータを移動する方法を指定します。Developer tool でのマッピングは、マッピング入力とマッピング出力の間でデータを移動する方法を指定します。

PowerCenter でのマッピングには、1 つまたは複数のソース定義、ソース修飾子、およびターゲット定義を含める必要があります。PowerCenter でのマッピングには、ショートカット、トランスフォーメーション、およびマップレットも含めることができます。

Developer tool でのマッピングには、マッピング入力とマッピング出力を含める必要があります。Developer tool でのマッピングには、トランスフォーメーションとマップレットも含めることができます。

Developer tool には次のタイプのマッピングがあります。

- ソースとターゲットの間でデータを移動するマッピング。このタイプのマッピングは、ショートカットを使用できず、ソース修飾子が使用されないという点においてのみ、PowerCenter でのマッピングと異なります。
- 論理データオブジェクトマッピング。論理データオブジェクトモデル内のマッピング。論理データオブジェクトマッピングには、マッピング入力として論理データオブジェクトを、マッピング出力としてデータオブジェクトを含めることができます。または、マッピング入力として1つまたは複数の物理データオブジェクトを、マッピング出力として論理データオブジェクトを含めることができます。
- 仮想テーブルマッピング。SQL データサービス内のマッピング。マッピング入力としてデータオブジェクトが、マッピング出力として仮想テーブルが含まれます。
- 仮想ストアードプロシージャマッピング。SQL データサービス内で一連のビジネスロジックを定義します。マッピング入力として入力パラメータトランスフォーメーションまたは物理データオブジェクトが、マッピング出力として出力パラメータトランスフォーメーションまたは物理データオブジェクトが含まれます。

PowerCenter でのマップレットと Developer tool でのマップレットの違いは何ですか。

PowerCenter でのマップレットと Developer tool でのマップレットは、一連のトランスフォーメーションが含まれる再利用可能なオブジェクトです。複数のマッピングでトランスフォーメーションロジックを再利用できます。

PowerCenter マップレットには、マップレット入力としてソース定義または入力トランスフォーメーションを含めることができます。マップレット出力として出力トランスフォーメーションを含める必要があります。

Developer tool マップレットには、マップレット入力としてデータオブジェクトまたは入力トランスフォーメーションを含めることができます。マップレット出力としてデータオブジェクトまたは出力トランスフォーメーションを含めることができます。Developer tool でのマッピングには、次の機能もあります。

- マップレットをルールとして検証できます。プロファイルのルールを使用します。
- マップレットには、他のマップレットを含めることができます。

索引

A

Analyst ツールの設定
概要 [19](#)

D

Developer tool の設定
概要 [47](#)

か

カスタムプロファイルの作成
概要 [28](#)
カラムからの参照テーブルの作成
概要 [41](#)

さ

参照テーブル
概要 [44](#)

し

式ルールの作成
概要 [32](#)

す

スコアカードの作成
概要 [35](#)

て

データオブジェクトの作成
概要 [22](#)
データのプロファイリング
概要 [59](#)
デフォルトプロファイルの作成
概要 [25](#)

ふ

物理データオブジェクトのインポート
概要 [51](#)