



Informatica®
10.4.0

Data Quality Integration for PowerCenter **ガイド**

Informatica Data Quality Integration for PowerCenter ガイド

10.4.0

2019 年 12 月

© 著作権 Informatica LLC 2009, 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、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	4
Informatica のリソース.....	4
Informatica Network.....	4
Informatica ナレッジベース.....	4
Informatica マニュアル.....	5
Informatica 製品可用性マトリックス.....	5
Informatica Velocity.....	5
Informatica Marketplace.....	5
Informatica グローバルカスタマサポート.....	5
 第 1 章 : Informatica Data Quality Integration	6
Data Quality Integration の概要.....	6
Data Quality の特徴および機能.....	7
Integration インストーラ.....	7
PowerCenter へのオブジェクトのエクスポート.....	8
参照データのインストール.....	8
製品統合のルールとガイドライン.....	8
PowerCenter でのデータ品質トランスフォーメーションの動作.....	9
 第 2 章 : インストール	11
インストールの概要.....	11
システム要件.....	11
インストールの前に.....	12
サーバープラグインのインストール.....	12
Windows へのインストール.....	12
UNIX へのインストール.....	13
クライアントプラグインのインストール.....	13
PowerCenter へのプラグインの登録.....	14
 第 3 章 : 参照データの要件	15
参照データの概要.....	15
参照テーブルデータの要件.....	16
アドレス参照データの要件.....	16
ID ポピュレーションデータの要件.....	18
確率的なモデルおよび分類子モデルの要件.....	19
PowerCenter での確率的なモデルおよび分類子モデルファイルの場所.....	19
PowerCenter 統合サービスのメモリ割り当て.....	20
 索引	21

序文

*Informatica Data Quality 統合 (PowerCenter(R)用) ユーザーガイド*を読むと、PowerCenter でデータ品質オブジェクトを設定および使用する方法に関するガイドラインを確認できます。参照データオブジェクトおよびファイルとともにデータ品質トランスフォーメーションを含むマッピングを Informatica Data Quality 環境から PowerCenter 環境にエクスポートできます。このマニュアルでは、データ品質オブジェクトに適用される前提条件を説明し、互換性のある製品リリースに統合コンポーネントをインストールする手順について説明します。

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

Informatica Data Quality Integration

この章では、以下の項目について説明します。

- [Data Quality Integration の概要, 6 ページ](#)
- [Data Quality の特徴および機能, 7 ページ](#)
- [Integration インストーラ, 7 ページ](#)
- [PowerCenter へのオブジェクトのエクスポート, 8 ページ](#)
- [参照データのインストール, 8 ページ](#)
- [製品統合のルールとガイドライン, 8 ページ](#)
- [PowerCenter でのデータ品質トランスフォーメーションの動作, 9 ページ](#)

Data Quality Integration の概要

Informatica のユーザーは、Informatica Data Quality でマッピングを作成し、実行できます。マッピングは、モデルリポジトリに保存します。ユーザーは、PowerCenter で使用するために、マッピングおよび関連する参照データオブジェクトまたはファイルをエクスポートできます。

PowerCenter ユーザーは、PowerCenter リポジトリにマッピングをインポートし、PowerCenter セッションでマッピングを実行することができます。マッピングを使用し、データのコンテンツおよび構造を分析し、強化します。

Data Quality 10.4 で作成したマッピングは、次のバージョンの PowerCenter で使用できます。

- PowerCenter 10.4.0。
- 最新の HotFix がインストールされた PowerCenter 10.2。
- 最新の HotFix がインストールされた PowerCenter 10.1.1。
- PowerCenter 10.1。

最新バージョンの Data Quality および PowerCenter を使用すると、追加のインストールをすることなく、モデルリポジトリに基づくオブジェクトをインポートできます。旧バージョンの PowerCenter を使用している場合は、最新バージョンの Informatica Data Quality 統合プラグインをインストールして登録します。

Data Quality の特徴および機能

Informatica Data Quality では、Informatica アプリケーションを使用してオブジェクトを作成し、モデルリポジトリにオブジェクトを保存し、マッピングを実行します。

最新の製品バージョンを使用していない場合は、PowerCenter 用の Informatica Data Quality プラグインをインストールする必要がある場合があります。

Informatica Data Quality を使用して、次の目標を達成するマッピングを設計し、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。

Integration インストラ

異なるバージョンの Data Quality および PowerCenter を使用してオブジェクトを作成して格納するには、PowerCenter 用の Informatica Data Quality プラグインをインストールします。Informatica からプラグインインストラをダウンロードします。

使用する PowerCenter のバージョン用のプラグインバージョンをインストールします。プラグインには、クライアントインストラとサーバーインストラが含まれています。

マプレットおよびマッピングをリポジトリに表示してセッションに追加するには、PowerCenter クライアントマシンにクライアントインストーラを実行します。ワークフローでセッションを実行する場合は、PowerCenter Integration Service マシンでサーバーインストーラを実行します。

Integration Service マシンにプラグインをインストールした後、そのプラグインを PowerCenter リポジトリに登録します。

PowerCenter へのオブジェクトのエクスポート

Informatica モデルリポジトリからマプレットやマッピングなどのオブジェクトをエクスポートするには、Developer tool を使用します。モデルリポジトリから PowerCenter リポジトリへのオブジェクトのエクスポートや、モデルリポジトリからファイルシステムへのオブジェクトのエクスポートが可能です。XML オブジェクトを PowerCenter リポジトリにインポートするには、PowerCenter Repository Manager を使用します。

- モデルリポジトリからのオブジェクトのエクスポートについては、『*Informatica Developer マッピングガイド*』および『*Informatica Developer Tool ガイド*』を参照してください。
- PowerCenter リポジトリへのオブジェクトのインポートについては、『*PowerCenter リポジトリガイド*』を参照してください。

参照データのインストール

マプレットまたはマッピングが参照データを読み取る場合は、参照データファイルを PowerCenter Integration Service マシンにコピーまたはインストールします。

Developer ツールを使用すると、マプレットまたはマッピングをエクスポートするときに、モデルリポジトリから参照データをエクスポートできます。Developer は、圧縮ファイルに参照データをエクスポートします。ファイルを抽出して、Integration Service マシンに参照データをコピーします。

エクスポートプロセスには、アドレス参照データファイルおよび ID ポピュレーションデータファイルは含まれていません。エクスポートするマプレットまたはマッピングでアドレス参照データファイルまたは ID ポピュレーションデータファイルを読み取る場合は、PowerCenter Integration Service マシンにファイルをインストールします。

製品統合のルールとガイドライン

異なるバージョンの Informatica Data Quality から異なるバージョンの PowerCenter にデータ品質オブジェクトをエクスポートできます。オブジェクトをエクスポートする、Informatica Data Quality のバージョン用の統合プラグインをインストールします。

Informatica Data Quality と PowerCenter を統合する場合、次のルールとガイドラインを考慮します。

- PowerCenter リポジトリの製品バージョンがモデルリポジトリの製品バージョンと一致する場合は、Integration プラグインをインストールしないでください。
- 単一の PowerCenter マシン上で同時に 2 つの Data Quality Integration プラグインを同時に実行できません。PowerCenter は最も最近インストールされたプラグインを使用します。例えば、Data Quality 10.2 Integration プラグインを PowerCenter 10.1.1 マシンにインストールする場合、バージョン 10.1.1 モデル

リポジトリからデータ品質オブジェクトをインポートおよび実行することはできません。10.2 プラグインをインストールするまでは、PowerCenter リポジトリにインポートしたデータ品質オブジェクトを実行し続けることができます。

- PowerCenter リポジトリのデータ品質マプレットやマッピングは編集できません。

PowerCenter でのデータ品質トランスフォーメーションの動作

データ品質マプレットまたはマッピングをモデルリポジトリから PowerCenter リポジトリにインポートすると、データ品質トランスフォーメーションの動作が変化します。

マプレットへの変換

PowerCenter リポジトリへのインポート時、次のトランスフォーメーションはマプレットに変換できます。

- 統合
- 例外
- キージェネレータ
- 一致

これらのトランスフォーメーションを 1 つ以上含むマッピングをインポートすると、PowerCenter によって、各トランスフォーメーションが拡張設定を含んだマプレットに変換されます。

これらのトランスフォーメーションを 1 つ以上含むマプレットをインポートすると、PowerCenter によって、各トランスフォーメーションの設定が複数のトランスフォーメーションへと拡張されます。

PowerCenter では、トランスフォーメーションロジックを完成させるために、データ品質トランスフォーメーションまたはマプレット拡張にトランスフォーメーションが追加されることがあります。例えば、一致トランスフォーメーションは、追加のトランスフォーメーションで PowerCenter によって拡張されます。

ディシジョントランスフォーメーションの考慮事項

ディシジョントランスフォーメーションのトランスフォーメーションスクリプトにシステムパラメータが含まれている場合、そのトランスフォーメーションを含むマッピングまたはマプレットを PowerCenter にエクスポートすることはできません。エクスポート操作では、システムパラメータを PowerCenter が使用できる値に変換することはできません。システムパラメータを使用しているディシジョントランスフォーメーションを含むマッピングまたはマプレットをエクスポートするには、事前にそれらのシステムパラメータを適切な値で置換してください。

一致トランスフォーメーションに関する考慮事項

マッピングでデータベーステーブルに対する ID インデックスデータの読み書きを行う場合は、PowerCenter で ID 照合マッピングを実行できません。PowerCenter で ID 照合分析を実行するには、Developer tool で一時ファイルに対する ID インデックスデータの書き込みを行うように一致トランスフォーメーションを設定します。

複数ストラテジのトランスフォーメーション

Developer tool では、次の各トランスフォーメーション内で複数の Data Transformation ストラテジを定義できます。

- 大文字小文字
- 分類子

- ディシジョン
- キージェネレータ
- ラベラ
- 一致
- マージ
- パーサー
- 標準化

複数ストラテジのトランスフォーメーションを含んだマップレットまたはマッピングをインポートすると、各ストラテジが個別のトランスフォーメーションに変換されます。

第 2 章

インストール

この章では、以下の項目について説明します。

- [インストールの概要, 11 ページ](#)
- [システム要件, 11 ページ](#)
- [インストールの前に, 12 ページ](#)
- [サーバープラグインのインストール, 12 ページ](#)
- [クライアントプラグインのインストール, 13 ページ](#)
- [PowerCenter へのプラグインの登録, 14 ページ](#)

インストールの概要

Data Quality 統合プラグインには、サーバーコンポーネントとクライアントコンポーネントが含まれます。

プラグインのインストールと登録を行うには、以下の手順を実行します。

1. PowerCenter 統合サービスマシン上で Server インストーラを実行します。
2. PowerCenter クライアントマシンで、クライアントインストーラを実行します。
3. IDQZIntegration.xml ファイルを PowerCenter リポジトリに登録します。

システム要件

Informatica Data Quality Integration プラグインのシステム要件は、そのインストール先である PowerCenter アプリケーションおよびサービスと同じです。

PowerCenter の前提条件およびシステム要件については、PowerCenter のインストールマニュアルを参照してください。

インストールの前に

PowerCenter で参照データを使用する場合は、インストーラの次の動作を考慮します。

Integration プラグインのインストーラは、PowerCenter Integration Service マシンにプロパティファイルを書き込みます。プロパティファイルは、データ品質マッピングが読み取ることができる参照データファイルの場所を指定します。

インストーラが書き込むファイルは次のとおりです。

- AD50.cfg。アドレス参照データの設定のプロパティを格納します。
- CLASSIFIER.properties。分類子モデルデータの設定のプロパティを格納します。
- IDQTx.cfg。ID ポピュレーションデータの設定のプロパティを格納します。
- NER.properties。確率的なモデルデータの設定のプロパティを格納します。

インストーラは、次のディレクトリにファイルを追加します。

[PowerCenter_installation]/server/bin

アドレス参照データの前提条件

アドレス参照データをインストールするときは、PowerCenter Integration Service を事前に停止しておきます。データのインストールの終了後、サービスを再起動します。

サーバープラグインのインストール

Informatica Data Quality Integration サーバープラグインインストーラを実行すると、データ品質トランスフォーマーアプリケーションファイルが PowerCenter Integration Service のディレクトリ構造に追加されます。

Windows へのインストール

Windows プラットフォーム上の PowerCenter Integration Service マシンにプラグインを追加するには、Windows サーバーインストーラを実行します。

インストールの前にすべてのアプリケーションを終了します。

1. システムユーザーアカウントを使用してマシンにログインします。
2. Windows 用の Integration サーバーファイルを抽出します。
3. ファイルを抽出したディレクトリに移動し、install.exe を実行します。
4. 言語を選択して、**[OK]** をクリックします。
5. **[ようこそ]** 画面で、**[次へ]** をクリックします。
6. PowerCenter サービスが実行されていないことを確認し、**[OK]** をクリックします。
7. PowerCenter インストールディレクトリのパスを指定します。インストーラから、このディレクトリのデフォルトパスが表示されます。別のパスを選択するには、**[選択]** をクリックします。
8. **[次へ]** をクリックします。
9. インストール前のサマリを確認します。エラーを修正する場合は、**[前へ]** をクリックします。
10. **[インストール]** をクリックします。
11. インストール後のサマリを確認します。

インストールタスクの詳細、およびインストールされるコンポーネントの設定プロパティを確認する場合は、インストールログファイルを参照してください。

12. **【完了】** をクリックします。

UNIX へのインストール

UNIX プラットフォーム上の PowerCenter Integration Service マシンにプラグインを追加するには、UNIX サーバーインストーラを実行します。

1. システムユーザーアカウントを使用してマシンにログインします。
2. 他のすべてのアプリケーションを終了します。
3. UNIX 用の Integration サーバーファイルを抽出します。
4. UNIX シェルを開き、インストールファイルを抽出したディレクトリに移動します。
5. `sh install.bin` と入力します。
6. 英語をインストール言語として使用する場合は、**Enter** を押します。
7. **【ようこそ】** 画面のメッセージを読み、**Enter** を押します。
8. PowerCenter インストールディレクトリの絶対パスを入力するか、**Enter** を押してデフォルトのパスを使用します。
9. インストール前のサマリを確認します。エラーを修正する場合は、`back` と入力して **Enter** を押します。
10. **Enter** を押してインストールを開始します。
11. インストール後のサマリを確認します。

インストールタスクの詳細、およびインストールされるコンポーネントの設定プロパティを確認する場合は、インストールログファイルを参照してください。

12. **Enter** を押してインストールプロセスを完了します。

クライアントプラグインのインストール

Data Quality Integration クライアントインストーラは、PowerCenter クライアントマシンで実行します。

1. 他のすべてのアプリケーションを終了します。
2. Integration クライアントファイルを抽出します。
3. ファイルを抽出したディレクトリを開きます。
4. `install.exe` を実行します。
5. **【ようこそ】** 画面で、**【次へ】** をクリックします。
インストールを続行する前に PowerCenter クライアントを終了するように求めるプロンプトが表示されます。
6. PowerCenter クライアントアプリケーションが実行されていないことを確認し、**【OK】** をクリックします。
7. PowerCenter クライアントディレクトリのパスを指定します。インストーラから、このディレクトリのデフォルトパスが表示されます。別のパスを選択するには、**【選択】** をクリックします。
8. **【次へ】** をクリックします。
9. インストール前のサマリを確認します。エラーを修正する場合は、**【前へ】** をクリックします。
10. **【インストール】** をクリックします。

11. インストール後のサマリを確認します。

インストールタスクの詳細、およびインストールされるコンポーネントの設定プロパティを確認する場合は、インストールログファイルを参照してください。

12. **【完了】** をクリックします。

PowerCenter へのプラグインの登録

データ品質オブジェクトを格納する任意の PowerCenter リポジトリに、プラグインを登録します。

Integration インストーラにより、PowerCenter 統合サービスマシン上の `$INFA_HOME\server\bin\native` ディレクトリに `IDQZIntegration.xml` プラグインファイルが書き込まれます。ファイルにアクセスできることを確認します。必要に応じて、ファイルを PowerCenter 統合サービスマシンにコピーするか、FTP で送信します。

- ▶ プラグインを登録するには、`pmrep RegisterPlugin` コマンドを実行します。コマンドを実行するときに、`-N` オプションを含めます。

`pmrep RegisterPlugin` コマンドの詳細については、『*Informatica コマンドリファレンス*』を参照してください。

注: プラグインを登録するための適切な権限がない場合は、PowerCenter リポジトリサービスの管理者にお問い合わせください。

第 3 章

参照データの要件

この章では、以下の項目について説明します。

- [参照データの概要, 15 ページ](#)
- [参照テーブルデータの要件, 16 ページ](#)
- [アドレス参照データの要件, 16 ページ](#)
- [ID ポピュレーションデータの要件, 18 ページ](#)
- [確率的なモデルおよび分類子モデルの要件, 19 ページ](#)

参照データの概要

データ品質トランスフォーメーションで参照データを読み取れます。PowerCenter リポジトリのトランスフォーメーションで参照データを読み取る場合は、PowerCenter 環境に参照データファイルがあることを確認します。

PowerCenter のデータ品質トランスフォーメーションで使用される可能性がある参照データのタイプは、以下のとおりです。

参照データファイル

標準的なバージョンのビジネス用語および代替バージョンのビジネス用語を含んだテキストファイルです。参照データファイルは編集できます。

アドレス参照データファイル

ある国のすべての住所の完全なデータを含むデータファイルです。アドレス参照データファイルを PowerCenter 統合サービスマシンにインストールします。アドレス参照データファイルは編集できません。

ID ポピュレーションファイル

さまざまな種類の ID を定義するとともに、一致トランスフォーメーションで使用されるアルゴリズムを含んだデータファイルです。ポピュレーションファイルは PowerCenter サービスマシンにインストールします。ポピュレーションファイルは編集できません。

コンテンツセット

1 つ以上のトランスフォーメーションで指定された参照データを含んだデータファイル。コンテンツセットを読み取るマッピングをエクスポートする場合は、エクスポートプロセスがコンテンツセットデータをトランスフォーメーションのメタデータとともにエクスポートします。分類子および確率的なモデルファイルの場合は、このプロセスでファイルデータが Data Quality のディレクトリ構造体からエクスポートされます。

参照テーブルデータの要件

PowerCenter で参照テーブルを読み取るマッピングを実行する前に、そのテーブルのカラムデータを PowerCenter サービスマシンにコピーします。

マッピングをモデルリポジトリからエクスポートする場合、エクスポート操作により、参照テーブルのカラムデータが 1 つ以上のファイルに書き込まれます。エクスポート操作により、当該ファイルは、指定する場所にある圧縮ファイルに追加されます。圧縮ファイルを解凍して、カラムデータファイルを探します。ファイル名は .dic で終わります。このファイルを PowerCenter サービスマシンにコピーします。

PowerCenter ディレクトリ構造体での参照データファイルの場所は、モデルリポジトリでの参照テーブルの場所に対応している必要があります。

次のディレクトリ形式を使用して、参照テーブルデータファイルの正しい場所を判断します。

<PowerCenter_Root_Directory>\services\<Model_Repository_Project_Name>\<Model_Repository_Project_Folder_Name>

参照テーブルデータのエクスポートでのルールとガイドライン

参照テーブルデータを読み取るマッピングをエクスポートする場合、次のルールとガイドラインを考慮します。

- マッピングをモデルリポジトリからエクスポートするには、エクスポートウィザードで **[Informatica]** > **[PowerCenter]** オプションを選択します。**PowerCenter** オプションを選択すると、エクスポート操作により、管理対象の参照テーブルデータと管理対象でない参照テーブルデータの両方がファイルシステムに書き込まれます。
- エクスポート操作により、各参照テーブルの 1 番目のカラムが有効なカラムとして指定されます。
- PowerCenter では、UTF-8 形式の参照データファイルを読み取ります。

環境変数を使用した参照データファイルの場所の定義

Integration Service が参照データファイルを読み取るために使用するパスを設定するために、PowerCenter 統合サービスマシンで *INFA_CONTENT* 環境変数を定義できます。

この環境変数は、PowerCenter Integration Service マシン上の必要な場所にファイルをインストールできない場合に使用します。PowerCenter 統合サービスが読み取れる場所にファイルをインストールし、インストールされたパスが指定された *INFA_CONTENT* 環境変数を作成します。

アドレス参照データの要件

アドレス検証を実行するマッピングでアドレス参照データを読み取ります。アドレス参照データは Informatica で購入し、ダウンロードできます。データファイルを PowerCenter 統合サービスマシンにインストールします。

次の表に、インストールできるアドレス参照データのタイプを示します。

参照データのタイプ	用途
バッチ	全般的なアドレス検証操作に使用します。
CAMEO	コンシューマ人口統計データを住所レコードに追加するために使用します。

参照データのタイプ	用途
認証	<p>住所レコードが、郵便事業者が定義する認証基準を満たすことを確認するために使用します。</p> <p>次の国は認証基準を定義しています。</p> <ul style="list-style-type: none"> - オーストラリア AMAS (Address Matching Approval System/住所照合承認システム) 基準で郵便を認証します。 - カナダ SERP (Software Evaluation And Recognition Program/評価と認識のソフトウェアプログラム) 基準で郵便を認証します。 - フランス SNA (Service National de l'Adresse/国の住所サービス) 基準で郵便を認証します。 - ニュージーランド SendRight 基準で郵便を認証します。 - 米国。CASS (Coding Accuracy Support System/符号化精度支援システム) 基準で郵便を認証します。
提案リスト	<p>不完全な住所レコードについて、有効な代替バージョンを検索するために使用します。提案リストのデータは、単一の住所レコードをリアルタイムで処理するために、アドレス検証マッピングを設定する場合に使用します。</p>
Geocode	<p>緯度と経度の座標を住所レコードに追加するために使用します。</p>
補足	<p>ある住所を含む地理的な領域またはポピュレーションの領域を特定するために使用します。</p> <p>アドレスバリデータトランスフォーメーションでは、以下の国の住所レコードに補足データを追加できます。</p> <ul style="list-style-type: none"> - 日本。アドレスを含む市街地を識別します。 - セルビア。アドレスの通りレベルの郵便番号を識別します。 - 英国。アドレスに配布ポイント ID を追加します。 - 米国。アドレスを含む人口密集地域を識別します。

PowerCenter 統合サービスが、AD50.cfg コンフィギュレーションファイルからアドレス参照データのパスを読み取ります。

PowerCenter インストーラは AD50.cfg ファイルを次のパスに書き込みます。

[PowerCenter_Installation]/server/bin

アドレス参照データをインストールする場合は、AD50.cfg に次のプロパティが設定されていることを確認します。

参照データの場所

ReferenceDataLocation プロパティにアドレス参照データのパスを入力します。

ライセンスキーデータ

LicenseKey プロパティにライセンスデータを入力します。ライセンスが複数ある場合は、各ライセンスをカンマ区切りの文字列で入力してください。

プリロード値

マッピングで読み取る参照データの種類ごとに、プリロード値として、少なくとも 1 つの国の略号を入力します。すべての国にプリロード設定を適用するには、「ALL」と入力します。

Integration Service は、国ごとに異なる方法でデータをロードすることができます。例えば、米国のバッチ/対話データに完全プリロードを指定し、英国のバッチ/対話データに部分プリロードを指定することができます。さらに、Integration Service はデータの種類ごとに異なるプリロード方法を使用することもできます。例えば、米国のバッチ/対話データに完全プリロードを指定し、米国の地理的コーディングデータに部分プリロードを指定することができます。

完全プリロードの設定は、部分プリロードの設定よりも優先され、部分プリロードの設定は、データのプリロードをしない設定よりも優先されます。例えば、データのプリロードをしないに対して「ALL」を入力し、完全プリロードに「USA」を入力した場合、Integration Service によって、すべての米国のデータがメモリにロードされ、他の国のデータはロードされません。プリロード要件がない場合は、使用する予定があるすべての種類の参照データで、データのプリロードをしないに対して「ALL」を入力します。

補足データにはプリロード値を指定しません。

AD50.cfg 内のその他のプロパティも必要に応じて編集できます。

ID ポピュレーションデータの要件

ID 照合分析を実行するマッピングが ID ポピュレーションと呼ばれる参照データファイルを読み取ります。ID ポピュレーションファイルを Informatica から購入してダウンロードします。ポピュレーションファイルを PowerCenter 統合サービスマシンにコピーするか、PowerCenter 統合サービスがアクセスできるマシンの共有ディレクトリにコピーします。

ポピュレーションファイルは、default という名前のディレクトリにコピーまたはインストールする必要があります。

PowerCenter 統合サービスでは、ポピュレーションファイルのディレクトリの場所を、IDQTx.cfg 構成ファイルまたは SSAPR 環境変数から読み取ることができます。PowerCenter 統合サービスは、IDQTx.cfg ファイル内で場所を検索します。IDQTx.cfg ファイルが存在しないか、場所を指定していない場合、PowerCenter 統合サービスは SSAPR 環境変数で場所を検索します。

注: IDQTx.cfg ファイルと SSAPR 環境変数は、default ディレクトリの親ディレクトリのパスを指定します。パスにはディレクトリ名が含まれません。

PowerCenter インストーラは IDQTx.cfg ファイルを次のパスに書き込みます。

```
[PowerCenter_Installation]/server/bin
```

IDQTx.cfg ファイルは、*IdentityReferenceDataLocation* プロパティを使用してポピュレーションファイルのディレクトリの場所を指定します。プロパティを、ディレクトリの現在の場所で更新します。

例えば、以下のパスを *IdentityReferenceDataLocation* プロパティに設定することができます。

```
[PowerCenter_Installation]/services/DQContent/INFA_Content/identity/
```

IdentityReferenceDataLocation プロパティでは、デフォルトディレクトリの絶対パスまたは相対パスを設定できます。ディレクトリへの絶対パスは、PowerCenter 統合サービスマシン、またはサービスがアクセスできるマシンの共有ディレクトリに設定します。ディレクトリへの相対パスは、PowerCenter 統合サービスマシンに設定します。

PowerCenter 統合サービスは以下のディレクトリから相対パスを読み取ります。

```
[PowerCenter_Installation]/server/bin
```

例えば、bin ディレクトリのポピュレーションファイルディレクトリの相対パスを指定するには、以下の相対パスを設定します。

```
./
```

IDQTx.cfg のプロパティ

IdentityReferenceDataLocation プロパティを IDQTx.cfg ファイルに設定するときには、次のプロパティも設定します。

IdentityCacheDir

マッピングの実行時に作成される一時データファイルを格納するディレクトリのパス。このパスは親ディレクトリを識別します。PowerCenter 統合サービスは、一時ファイルを指定した場所の下にあるディレクトリに書き込みます。

プロパティには、以下のデフォルト値があります。

`./identityCache`

IdentityIndexDir

マッピングの実行時に作成される一時インデックスファイルが含まれているディレクトリのパス。ID 照合分析は、照合分析の前にインデックスを使用してレコードをグループにソートします。このパスは親ディレクトリを識別します。PowerCenter 統合サービスは、インデックスファイルを指定した場所の下にあるディレクトリに書き込みます。

プロパティには、以下のデフォルト値があります。

`./identityIndex`

注: Developer ツールのユーザーは、一致トランスフォーメーションにキャッシュディレクトリのパスおよびインデックスディレクトリのパスを設定できます。一致トランスフォーメーションにパスが指定されていない場合、PowerCenter 統合サービスは、IDQTx.cfg ファイルに定義されているパスを使用します。

確率的なモデルおよび分類子モデルの要件

確率的なモデルおよび分類子モデルは、データ文字列内の情報の種類を特定する参照データファイルです。

確率的なモデルまたは分類子モデルのファイルを読み取るマッピングを実行する前に、次のタスクを実行します。

- PowerCenter ディレクトリ構造内でのモデルファイルの場所をコピーする。
- PowerCenter 統合サービスで Java のメモリ割り当てを更新する。

PowerCenter での確率的なモデルおよび分類子モデルファイルの場所

PowerCenter で確率的なモデルおよび分類子モデルを読み取るマッピングを実行する前に、モデルファイルを PowerCenter サービスマシンにコピーします。

マッピングをモデルリポジトリからエクスポートする場合、エクスポート操作により、指定する場所にある圧縮ファイルに各モデルが追加されます。圧縮ファイルを解凍して、マッピングで読み取るモデルファイルを探します。確率的なモデルおよび分類子モデルのデータファイル名は、それぞれ .ner および .classifier で終わります。PowerCenter 統合サービスが読み取れるディレクトリにファイルをコピーします。

デフォルトでは、PowerCenter 統合サービスが以下のディレクトリからモデルファイルを読み取ります。

`[PowerCenter_Installation]/server/bin`

デフォルトの場所にファイルがない場合、PowerCenter 統合サービスはプロパティファイルからファイルの場所を読み取ります。プロパティファイルをモデルファイルへのパスを使用して更新します。

以下の表に、プロパティファイルを示します。

ファイル名	説明
CLASSIFIER.properties	<i>ClassifierFileLocation</i> プロパティが含まれます。このプロパティを更新して、分類子モデルファイルへのパスを特定します。
NER.properties	<i>NERFileLocation</i> プロパティが含まれます。このプロパティを更新して、確率的なモデルファイルへのパスを特定します。

PowerCenter 統合サービスは以下のディレクトリのプロパティファイルを読み取ります。

[PowerCenter_Installation]/server/bin

確率的なモデルおよび分類子モデルのエクスポートでのルールとガイドライン

参照テーブルデータを読み取るマッピングをエクスポートする場合、次のルールとガイドラインを考慮します。

- PowerCenter に使用する確率的なモデルまたは分類子モデルのデータを読み取るマッピングをエクスポートする場合、エクスポートウィザードで **[Informatica] > [オブジェクトメタデータファイルをエクスポート]** オプションを選択して圧縮ファイルを作成します。

エクスポート操作により、マッピングも XML ファイルとしてエクスポートされます。XML ファイルは無視できます。

- トレーニング済みモデルのみをエクスポートするオプションを選択します。確率的なモデルや分類子モデルがトレーニング済みであれば、マッピングで読み取られる論理データが含まれており、モデルのソースデータは含まれません。トレーニング済みモデルファイルを PowerCenter サービスマシンにコピーします。

- 別の操作で、マッピングをモデルリポジトリから PowerCenter リポジトリにエクスポートします。マッピングをモデルリポジトリから PowerCenter リポジトリにエクスポートするには、**[Informatica] > [PowerCenter]** オプションを選択します。

[PowerCenter] オプションを選択した場合、エクスポート操作により、管理対象の参照テーブルからも、管理対象でない参照テーブルからも参照データをコピーできます。

- 絶対パスまたは相対パスで、各プロパティファイル内のモデルファイルの場所を指定できます。相対パスを設定した場合、PowerCenter 統合サービスによって次のディレクトリへのパスが追加されます。

[PowerCenter_Installation]/server/bin

- フォワードスラッシュを使用して、すべてのプラットフォーム上のモデルファイルへのディレクトリパスを定義します。

PowerCenter 統合サービスのメモリ割り当て

PowerCenter 統合サービスでは、分類子モデルファイルや確率的なモデルファイルの読み取りに大量のメモリを使用します。PowerCenter 統合サービスは必要なメモリ量にアクセスできない場合、モデルファイルの読み取りが失敗し、データを処理できません。

PowerCenter 統合サービスのメモリ割り当てを確認します。分類子モデルファイルまたは確率的なモデルファイルを参照するセッションを実行する場合は、最小メモリ割り当てを 512MB に設定します。

メモリの割り当てには、*Java SDK 最大メモリ*というプロセスプロパティを使用できます。プロパティを確認または更新するには、管理者ツールにログインし、セッションタスクを実行する PowerCenter 統合サービスを選択します。*Java SDK 最大メモリ*のプロパティは、**[プロセス]** ビューにあります。

索引

D

Data Quality 統合 [6](#)

I

Informatica Data Quality [6](#)