



Informatica® Data Quality Integration for
PowerCenter

10.1.1 HotFix 2

ユーザーガイド

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、Informatica ロゴ、および PowerCenter 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。

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

免責: 本文書は、一切の保証を伴わない「現状渡し」で提供されるものとし、Informatica LLC は他社の権利の非侵害、市場性および特定の目的への適合性の黙示の保証などを含めて、一切の明示的および黙示的保証の責任を負いません。Informatica LLC では、本ソフトウェアまたはドキュメントに誤りのないことを保証していません。本ソフトウェアまたはドキュメントに記載されている情報には、技術的に不正確な記述や誤植が含まれる場合があります。本ソフトウェアまたはドキュメントの情報は、予告なしに変更されることがあります。

NOTICES

この Informatica 製品（以下「ソフトウェア」）には、Progress Software Corporation（以下「DataDirect」）の事業子会社である DataDirect Technologies からの特定のドライバ（以下「DataDirect ドライバ」）が含まれています。DataDirect ドライバには、次の用語および条件が適用されます。

1. DataDirect ドライバは、特定物として現存するままの状態を提供され、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任を負わないものとします。国または地域によっては、法律の強行規定により、保証責任の制限が禁じられる場合、強行規定の制限を受けるものとします。
2. DataDirect または第三者は、予見の有無を問わず発生した ODBC ドライバの使用に関するいかなる直接的、間接的、偶発的、特別、あるいは結果的損害に対しても責任を負わないものとします。本制限事項は、すべての訴訟原因に適用されます。訴訟原因には、契約違反、保証違反、過失、厳格責任、詐称、その他の不法行為を含みますが、これらに限るものではありません。

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: 2018-07-03

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	4
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
データ品質統合の概要	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
統合マトリックス	8
製品統合のルールとガイドライン	10
PowerCenter でのデータ品質トランスフォーメーションの動作	10
 第 2 章 : インストール	 14
インストールの概要	14
システム要件	14
インストールの前に	15
サーバープラグインのインストール	15
Windows へのインストール	15
UNIX へのインストール	16
クライアントプラグインのインストール	16
PowerCenter へのプラグインの登録	17
 第 3 章 : 参照データの要件	 18
参照データの概要	18
参照データファイルの要件	19
アドレス参照データの要件	20
ID ポピュレーションデータの要件	21
確率的なモデルおよび分類子モデルの要件	22
確率的なモデルおよび分類子モデルファイルの場所	23
PowerCenter 統合サービスのメモリ割り当て	23
 索引	 24

序文

PowerCenter(R)ユーザーは Data Quality 環境からインポートしたマッピングおよび参照データオブジェクトを使用して作業できます。『Informatica Data Quality 統合 (PowerCenter 用) ユーザーガイド』はデータ品質マッピングの実行および関連する参照データオブジェクトの管理を行う PowerCenter ユーザー向けです。

Informatica のリソース

Informatica Network

Informatica Network は、Informatica グローバルカスタマサポート、Informatica ナレッジベースなどの製品リソースをホストします。Informatica Network には、<https://network.informatica.com> からアクセスしてください。

メンバーは以下の操作を行うことができます。

- 1 つの場所からすべての Informatica のリソースにアクセスできます。
- ドキュメント、FAQ、ベストプラクティスなどの製品リソースをナレッジベースで検索できます。
- 製品の提供情報を表示できます。
- 自分のサポート事例を確認できます。
- 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

ドキュメント、ハウツー記事、ベストプラクティス、PAM などの製品リソースを Informatica Network で検索するには、Informatica ナレッジベースを使用します。

ナレッジベースには、<https://kb.informatica.com> から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

使用している製品の最新のドキュメントを取得するには、https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx にある Informatica ナレッジベースを参照してください。

このマニュアルに関する質問、コメント、ご意見の電子メールの送付先は、Informatica マニュアルチーム (infa_documentation@informatica.com) です。

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 Network メンバである場合は、PAM (<https://network.informatica.com/community/informatica-network/product-availability-matrices>) にアクセスできます。

Informatica Velocity

Informatica Velocity は、Informatica プロフェッショナルサービスによって開発されたヒントおよびベストプラクティスのコレクションです。数多くのデータ管理プロジェクトの経験から開発された Informatica Velocity には、世界中の組織と協力して優れたデータ管理ソリューションの計画、開発、展開、および維持を行ってきた弊社コンサルタントの知識が集約されています。

Informatica Network メンバである場合は、Informatica Velocity リソース (<http://velocity.informatica.com>) にアクセスできます。

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 製品を強化したり拡張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。Informatica の開発者およびパートナーの何百ものソリューションを利用して、プロジェクトで実装にかかる時間を短縮したり、生産性を向上させたりできます。Informatica Marketplace には、<https://marketplace.informatica.com> からアクセスできます。

Informatica グローバルカスタマサポート

Informatica Network の電話またはオンラインサポートからグローバルカスタマサポートに連絡できます。

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

Informatica Network メンバである場合は、オンラインサポート (<http://network.informatica.com>) を使用できます。

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。

- [データ品質統合の概要, 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](#)
- [統合マトリクス, 8 ページ](#)
- [製品統合のルールとガイドライン, 10 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, 10 ページ](#)

データ品質統合の概要

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。ユーザーはマッピングとその他のオブジェクトをモデルリポジトリに保存します。ユーザーは、PowerCenter で使用するオブジェクトをエクスポートできます。PowerCenter のユーザーは、オブジェクトを PowerCenter リポジトリにインポートし、そのオブジェクトをセッションに追加できます。

オブジェクトを使用し、データのコンテンツおよび構造を分析し、強化します。

ユーザーが Data Quality 10.1.1 HotFix 2 で作成したオブジェクトは、次のバージョンの PowerCenter で使用できます。

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

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、モデルリポジトリ保存されていたプロ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 Integration 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 ツール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ファイルを読み取る場合は、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、圧縮ファイルとして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。

また、Data Quality コンテンツインストーラを使用して、参照データを Integration Service マシンに追加することもできます。コンテンツインストーラを実行してアドレス参照データおよび ID ポピュレーションデータファイルをインストールします。

統合マトリクス

Data Quality および PowerCenter 間の統合レベルは、対応製品のバージョンによって異なります。複数のバージョンの Data Quality から複数のバージョンの PowerCenter にオブジェクトおよび参照データをエクスポート

ートできます。オブジェクトも参照データも PowerCenter から Data Quality にエクスポートすることはできません。

統合可能な製品バージョンを次の表に示します。

製品名	PowerCenter 10.1.1 HotFix 2	PowerCenter 10.1.1 HotFix 1	PowerCenter 10.1.1	PowerCenter 10.1	PowerCenter 10.0	PowerCenter 9.6.1
Data Quality 10.1.1 HotFix 2	○	×	×	プラグインを インストール	×	プラグインを インストール
Data Quality 10.1.1 HotFix 1	×	○	×	プラグインを インストール	×	プラグインを インストール
Data Quality 10.1.1	×	×	○	プラグインを インストール	×	プラグインを インストール
Data Quality 10.1	×	×	×	○	×	プラグインを インストール
Data Quality 10.0	×	×	×	×	○	プラグインを インストール
Data Quality 9.6.1	×	×	×	×	×	○

表の値は、次の種類の統合を表します。

○

Data Quality 環境から PowerCenter 環境へオブジェクトおよびファイルをエクスポートできます。プラグインのインストールまたは登録は不要です。指定バージョンへアップグレードした PowerCenter リポジトリのデータ品質オブジェクトを実行できます。

プラグインをインストール

Data Quality 環境から PowerCenter 環境へオブジェクトおよびファイルをエクスポートできます。使用している PowerCenter のバージョンに対応するクライアントおよびサーバープラグインをインストールして、登録します。指定バージョンへアップグレードした PowerCenter リポジトリのデータ品質オブジェクトを実行できます。

×

PowerCenter 環境へエクスポートする Data Quality 環境で定義したオブジェクトやファイルは使用できません。

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 9.6.1 Integration プラグインを PowerCenter 9.5.1 マシンにインストールする場合、バージョン 9.5.1 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。9.6.1 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

データ品質マッピング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。

統合トランスフォーメーションに関する考慮事項

Informatica Data Quality は、バージョン 9.5.1 統合トランスフォーメーションのデフォルトのストラテジを更新しました。バージョン 9.5.1 以降では、統合トランスフォーメーションは最大行 ID を持つレコードを存続レコードとして選択します。バージョン 9.5 以前では、統合トランスフォーメーションは存続レコードの識別のため指定した列の最も頻出する空白以外の値を使用します。

バージョン 9.5.1 以降への統合トランスフォーメーションを含む PowerCenter リポジトリをアップグレードする場合、アップグレードでは最も頻出する空白以外の値を存続設定として保持します。9.5.1 モデルリポジトリまたはそれ以降のリポジトリからのデフォルト設定を使用する統合トランスフォーメーションを含むマッピングをインポートする場合、トランスフォーメーションで最大行 ID を使用し存続を識別します。

マッピングへの変換

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ッピングに変換できます。

- 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ッピングに変換されます。

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ッ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ップレットを 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ッ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。

例外トランスフォーメーションに関する考慮事項

PowerCenter では、バージョン 9.5 より古い PowerCenter リポジトリから発生した例外トランスフォーメーションは使用できません。PowerCenter リポジトリをバージョン 9.5 以降にアップグレードし、リポジトリに例外トランスフォーメーションのあるマッピングが含まれている場合は、そのマッピングを再度エクスポートします。必要に応じて、Developer tool でマッピングを再度作成します。

例外トランスフォーメーションを含むマッピングを実行すると、Analyst ツールで不良レコードテーブルまたは重複レコード出力テーブルを更新できます。Analyst ツールで不良レコードテーブルを更新する予定がある場合は、マッピングを実行するセッションで高精度を有効にしておく必要があります。

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。

パーティションの設定

グリッドでデータ品質マッピングが含まれているセッションタスクを実行する前に、グリッドノードが作成された PowerCenter のバージョンを識別します。PowerCenter 9.5.1 HotFix 2 以前のバージョンの PowerCenter でノードが作成された場合は、マッピングに失敗することがあります。

グリッド上のすべてのノードでデータ品質マッピングを実行できることを確認するには、次のいずれかのタスクを実行します。

- タスクをグリッド上のノードに分散させる PowerCenter 統合サービスの詳細プロパティを確認します。タスクを分散させるときにリソース要件を無視するよう各 PowerCenter 統合サービスを設定します。詳細プロパティはデフォルトで、PowerCenter 統合サービスがリソース要件を無視するように指定します。
- PowerCenter 9.5.1 HotFix 2 以前のバージョンの PowerCenter で作成したノードを再作成します。ノードを再作成した場合、PowerCenter は現在のノード構成ファイルをそのノード適用します。

グリッド上でセッションタスクを実行すると、Data Transformation マネージャ（DTM）プロセスはセッションスレッドをグリッドの各ノードに分散させます。トランスフォーメーションの中には、グリッドを越えてパーティションできないものがあります。グリッドを越えてトランスフォーメーションをパーティション化できない場合、DTM は、トランスフォーメーションスレッド用の単一パーティショングループを作成して、単一ノードでスレッドを実行します。

次の表に、データ品質トランスフォーメーションのパーティションの設定を示します。

トランスフォーメーション	パーティションの設定
アドレス検証	グリッドをまたがる
関連付け	パーティション化不可
AV レポート	ローカルでパーティション可能
大文字小文字変換プログラム	グリッドをまたがる
分類子	ローカルでパーティション可能
クラスタ	パーティション化不可
比較	グリッドをまたがる
統合	パーティション化不可
ディンジョン	グリッドをまたがる
例外	パーティション化不可
実行ポイント	グリッドをまたがる
キージェネレータ	パーティション化不可
キーストア	パーティション化不可
ラベラ	グリッドをまたがる
マージ	グリッドをまたがる
ペアジェネレータ	パーティション化不可
パーサー（確率的解析）	ローカルでパーティション可能
パーサー	グリッドをまたがる

トランスフォーメーション	パーティションの設定
トークン標準化	グリッドをまたがる
ウェイトベースアナライザ	グリッドをまたがる

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, 14 ページ](#)
- [システム要件, 14 ページ](#)
- [インストールの前に, 15 ページ](#)
- [サーバープラグインのインストール, 15 ページ](#)
- [クライアントプラグインのインストール, 16 ページ](#)
- [PowerCenter へのプラグインの登録, 17 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

インストーラは、bin ディレクトリでプロパティファイルを検出すると、そのファイルのバックアップコピーを作成してから、新しいファイルをインストールします。インストール後に現在の参照データ設定を保持するには、古いファイルおよび新しいファイルのコンテンツをマージします。

バージョン 9.5.1 以降の PowerCenter をインストールする場合は、サーバーインストールプロセスがデフォルトでプロパティファイルを bin ディレクトリに追加します。

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ション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, 18 ページ](#)
- [参照データファイルの要件, 19 ページ](#)
- [アドレス参照データの要件, 20 ページ](#)
- [ID ポピュレーションデータの要件, 21 ページ](#)
- [確率的なモデルおよび分類子モデルの要件, 22 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。

参照データファイルの要件

データ品質マッピングは、モデルリポジトリに定義された参照データオブジェクトを読み取れます。

Informatica Data Quality のユーザーが、参照データオブジェクトを読み取るマッピングをエクスポートすると、エクスポートプロセスが参照データを 1 つ以上の参照データファイルとしてファイルシステムに書き込みます。PowerCenter のマッピングを使用する場合は、Integration Service が読み取れる場所にファイルがインストールされることを確認する必要があります。

参照データファイルの場所はマッピングのエクスポート時に指定できます。ファイルを PowerCenter ディレクトリ構造体にエクスポートするか、ファイルシステムにファイルをエクスポートしてから、それらのファイルを PowerCenter 統合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、モデルリポジトリ構造体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。ディレクトリ構造は、次の形式に従っている必要があります。

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

注: PowerCenter では、UTF-8 形式の参照データファイルを読み取ります。Developer ツールでは、参照データが UTF-8 形式で PowerCenter にエクスポートされます。

Data Quality Content Installer からの参照データのインストール

Data Quality Content Installer ファイルのセットには、PowerCenter ディレクトリ構造体書き込む参照データファイルが含まれています。

コンテンツインストーラを実行するときに、PowerCenter インストールディレクトリを特定します。デフォルトでは、Content Installer が次のパスをルートに追加します。

\services\DQContent\INFA_Content\

モデルリポジトリ構造体に一致するようにパスを編集するか、Content Installer の実行後、参照データファイルを必須パスにコピーします。

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 統合サービスマシン上の必要な場所に参照データ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コピーします。

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	一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。
認証	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。 次の国は認証基準を定義しています。 <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	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。
ジオコード	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。
補足	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。 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。 <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。

また、Data Quality Content Installer を使用して、アドレス参照データをインストールすることもできます。PowerCenter 統合サービスマシンで Content Installer を実行し、ファイルをインストールします。

Data Quality コンテンツインストーラにより、PowerCenter ディレクトリ構造に /av/ ディレクトリが作成され、そのディレクトリにアドレス参照データが書き込まれます。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。

注: Informatica Data Quality 9.x では、AD50.cfg ファイルが読み取れません。Informatica Data Quality 9.x ユーザーは、Administrator ツールを通じて参照データの場所、ライセンスキーの情報、およびプリロード値を入力します。

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。

また、Data Quality Content Installer を使用してポピュレーションファイルをインストールすることもできます。PowerCenter 統合サービスマシンで Content Installer を実行し、ファイルをインストールします。

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ファイルを読み取るマッピングを実行する前に、それらのファイルを読み取るために PowerCenter を設定します。

以下の設定タスクを実行します。

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

確率的なモデルおよび分類子モデルファイルの場所

確率的なモデルまたは分類子モデルデータを読み取るマッピングをエクスポートするときには、エクスポートされたモデルデータファイルのディレクトリの場所を指定します。PowerCenter ディレクトリ構造体にファイルをエクスポートするか、ファイルシステムにファイルをエクスポートしてから、PowerCenter サービスマシンにコピーします。

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

絶対パスまたは相対パスを設定すると、各プロパティファイル内のモデルファイルの場所を特定できます。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

注: 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

I

Informatica Data Quality [6](#)

て

データ品質の統合 [6](#)