



Informatica®

10.4.0

Guia de Integração do Data Quality para o PowerCenter

Este software e a documentação são fornecidos somente sob um contrato de licença separado, contendo restrições sobre uso e divulgação. Não está permitida de forma alguma a reprodução ou a transmissão de qualquer parte deste documento (seja por meio eletrônico, fotocópia, gravação ou quaisquer outros meios) sem o consentimento prévio da Informatica LLC.

DIREITOS DO GOVERNO DOS ESTADOS UNIDOS Programas, softwares, bancos de dados, bem como a documentação e os dados técnicos relacionados, distribuídos a clientes do Governo dos EUA são "softwares de computador comerciais" ou "dados técnicos comerciais", de acordo com o Regulamento de Aquisição Federal aplicável e os regulamentos suplementares específicos da agência. Como tal, a utilização, duplicação, divulgação, modificação e adaptação estão sujeitas às restrições e aos termos de licença estabelecidos no contrato governamental aplicável e, na medida do que for aplicável pelos termos do contrato governamental, aos direitos adicionais estabelecidos no FAR 52.227-19, Licença de Software de Computador Comercial.

Informatica, PowerCenter e o logotipo Informatica são marcas comerciais ou marcas registradas da Informatica LLC nos Estados Unidos e em muitas jurisdições por todo o mundo. Uma lista atual das marcas comerciais da Informatica está disponível na Internet em <https://www.informatica.com/trademarks.html>. Os nomes de outras companhias e produtos podem ser nomes ou marcas comerciais de seus respectivos proprietários.

Partes deste software e/ou documentação estão sujeitas a copyright detido por terceiros. Os avisos de terceiros necessários são incluídos no produto.

As informações contidas neste documento estão sujeitas a alteração sem aviso prévio. Se você encontrar quaisquer problemas nesta documentação, informe-os em infa_documentation@informatica.com.

Os produtos Informatica apresentam garantias segundo os termos e condições dos acordos em que são fornecidos. A INFORMATICA FORNECE AS INFORMAÇÕES NESTE DOCUMENTO "COMO ESTÃO" SEM GARANTIA DE QUALQUER TIPO, EXPRESSA OU IMPLÍCITA, INCLUINDO, SEM QUAISQUER GARANTIAS DE COMERCIALIZAÇÃO, ADEQUAÇÃO A UM DETERMINADO FIM E QUALQUER GARANTIA OU CONDIÇÃO DE NÃO-VIOLAÇÃO.

Data da Publicação: 2020-02-07

Conteúdo

Prefácio.....	4
Recursos da Informatica.	4
Rede da Informatica.	4
Base de Dados de Conhecimento da Informatica.	4
Documentação da Informatica.	5
Matrizes de Disponibilidade de Produto da Informatica.	5
Informatica Velocity.	5
Informatica Marketplace.	5
Suporte Global a Clientes da Informatica.	5
 Capítulo 1: Integração do Informatica Data Quality.....	6
Visão Geral de Integração da Qualidade de Dados.	6
Recursos e Funcionalidade da Qualidade de Dados.	7
Instaladores de Integração.	8
Exportação de Objetos para o PowerCenter.	8
Instalação de Dados de Referência.	8
Regras e Diretrizes para a Integração de Produtos.	9
Comportamento das Transformações de Qualidade de Dados no PowerCenter.	9
 Capítulo 2: Instalação	11
Visão Geral da Instalação.	11
Requisitos de Sistema.	11
Antes da Instalação.	12
Instalando o Plug-in de Servidor.	12
Instalando no Windows.	12
Instalando no UNIX.	13
Instalando o Plug-in Cliente.	13
Registrando o Plug-In no PowerCenter.	14
 Capítulo 3: Requisitos de Dados de Referência.....	15
Visão Geral dos Dados de Referência.	15
Requisitos de Dados da Tabela de Referência.	16
Requisitos de Dados de Referência de Endereço.	17
Requisitos de Dados de Preenchimento de Identidade.	18
Requisitos do Modelo de Probabilidade e do Modelo de Classificador.	19
Localização de Arquivos do Modelo Probabilístico e do Modelo de Classificador no PowerCenter.	20
Alocação de Memória do Serviço de Integração do PowerCenter.	21
 Índice.....	22

Prefácio

Leia o *Guia do Usuário de Integração de Informatica Data Quality para PowerCenter®* para encontrar diretrizes sobre como configurar e usar objetos de qualidade de dados com o PowerCenter. Você pode exportar mapeamentos com transformações de qualidade de dados junto com objetos e arquivos de dados de referência de um ambiente do Informatica Data Quality para um ambiente do PowerCenter. O guia descreve os pré-requisitos aplicáveis aos objetos de qualidade de dados e descreve as etapas para instalar o componente de integração em versões compatíveis do produto.

Recursos da Informatica

A Informatica oferece uma variedade de recursos de produtos através da Rede da Informatica e outros portais on-line. Use os recursos para obter o máximo de seus produtos e soluções da Informatica e para aprender com outros usuários da Informatica e especialistas no assunto.

Rede da Informatica

A Rede da Informatica é a porta de entrada para muitos recursos, incluindo a Base de Dados de Conhecimento da Informatica e o Suporte Global a Clientes da Informatica. Para acessar a Rede da Informatica, visite <https://network.informatica.com>.

Como membro da Rede da Informatica, você tem as seguintes opções:

- Pesquisar por recursos do produto na Base de Dados de Conhecimento.
- Visualizar informações sobre disponibilidade de produtos.
- Criar e revisar seus casos de suporte.
- Encontrar a sua Rede de Grupo de Usuários da Informatica local e colaborar com seus colegas.

Base de Dados de Conhecimento da Informatica

Use a Base de Dados de Conhecimento da Informatica para encontrar recursos de produtos, como artigos de instruções, práticas recomendadas, tutoriais em vídeo e respostas a perguntas frequentes.

Para pesquisar na Base de Dados de Conhecimento, visite <https://search.informatica.com>. Em caso de dúvidas, comentários ou ideias sobre a Base de Dados de Conhecimento, entre em contato com a equipe da Base de Dados de Conhecimento da Informatica em KB_Feedback@informatica.com.

Documentação da Informatica

Use o Portal de Documentação da Informatica para explorar uma extensa biblioteca de documentação para versões de produtos atuais e recentes. Para explorar o Portal de Documentação, visite <https://docs.informatica.com>.

Em caso de dúvidas, comentários ou ideias sobre a documentação do produto, entre em contato com a equipe da Documentação da Informatica em infa_documentation@informatica.com.

Matrizes de Disponibilidade de Produto da Informatica

As Matrizes de Disponibilidade de Produto (PAMs) indicam as versões dos sistemas operacionais, os bancos de dados e tipos de fontes e destinos de dados com os quais uma versão de produto é compatível. Veja as PAMs da Informatica em <https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Velocity

O Informatica Velocity é uma coleção de dicas e práticas recomendadas desenvolvidas pelos Serviços Profissionais da Informatica e baseada em experiências reais de centenas de projetos de gerenciamento de dados. O Informatica Velocity representa o conhecimento coletivo dos consultores da Informatica que trabalham com organizações em todo o mundo para planejar, desenvolver, implantar e manter soluções de gerenciamento de dados bem-sucedidas.

Encontre os recursos do Informatica Velocity em <http://velocity.informatica.com>. Se você tiver dúvidas, comentários ou ideias sobre o Informatica Velocity, entre em contato com os Serviços Profissionais da Informatica em ips@informatica.com.

Informatica Marketplace

O Informatica Marketplace é um fórum onde você pode encontrar soluções que ampliam e aprimoram suas implementações da Informatica. Aproveite as centenas de soluções dos desenvolvedores e parceiros da Informatica no Marketplace para melhorar sua produtividade e agilizar o tempo de implementação em seus projetos. Encontre o Informatica Marketplace em <https://marketplace.informatica.com>.

Suporte Global a Clientes da Informatica

Você pode entrar em contato com um Centro de Suporte Global por telefone ou por meio da Rede da Informatica.

Para descobrir o número de telefone local do Suporte Global a Clientes da Informatica, visite o site da Informatica no seguinte link: <https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>.

Para encontrar recursos de suporte on-line na Rede da Informatica, visite <https://network.informatica.com> e selecione a opção eSupport.

CAPÍTULO 1

Integração do Informatica Data Quality

Este capítulo inclui os seguintes tópicos:

- [Visão Geral de Integração da Qualidade de Dados, 6](#)
- [Recursos e Funcionalidade da Qualidade de Dados, 7](#)
- [Instaladores de Integração, 8](#)
- [Exportação de Objetos para o PowerCenter, 8](#)
- [Instalação de Dados de Referência, 8](#)
- [Regras e Diretrizes para a Integração de Produtos, 9](#)
- [Comportamento das Transformações de Qualidade de Dados no PowerCenter, 9](#)

Visão Geral de Integração da Qualidade de Dados

Os usuários do Informatica podem criar e executar mapeamentos no Informatica Data Quality. Os usuários salvam os mapeamentos no repositório do Modelo. Os usuários podem exportar os mapeamentos e qualquer objeto de dados de referência ou arquivo relacionado para uso no PowerCenter.

Como um usuário do PowerCenter, você pode importar os mapeamentos para o repositório do PowerCenter e executá-los nas sessões do PowerCenter. Use os mapeamentos para analisar e aprimorar o conteúdo e a estrutura dos dados.

Você pode usar os mapeamentos que os usuários criam no Data Quality 10.4 com as seguintes versões do PowerCenter:

- PowerCenter 10.4.0.
- PowerCenter 10.2 com o hotfix mais recente instalado.
- PowerCenter 10.1.1 com o hotfix mais recente instalado.
- PowerCenter 10.1.

Se você usar a versão atual do Data Quality e do PowerCenter, poderá importar objetos que têm origem no repositório do Modelo sem nenhuma instalações adicionais. Se você usar uma versão anterior do PowerCenter, instale e registre a versão atual do plug-in de Integração de Informatica Data Quality.

Recursos e Funcionalidade da Qualidade de Dados

O Informatica Data Quality usa aplicativos Informatica para criar objetos, salvar objetos no repositório do Modelo e executar mapeamentos.

Talvez seja necessário instalar o Plug-in do Informatica Data Quality no PowerCenter se você não usou as versões de produto atuais.

Use o Informatica Data Quality para projetar e executar mapeamentos que obtêm os seguintes objetivos:

- Crie dados de perfil de dados. A criação de perfil de dados revela o conteúdo e a estrutura de seus dados. A criação de perfil de dados é uma etapa essencial em qualquer projeto de dados, pois é capaz de identificar pontos fortes e fracos nos dados e ajudar você a definir seu plano de projeto.
- Crie scorecards para analisar a qualidade dos dados. Um scorecard é uma representação gráfica das medidas de qualidade em um perfil.
- Padronize os valores de dados. Padronize os dados para remover erros e inconsistências encontrados na execução de um perfil. Você pode padronizar variações em pontuação, formatação e ortografia. Por exemplo, você pode garantir que os valores de cidade, estado e CEP sejam consistentes.
- Analise os registros. Analise registros de dados para aprimorar a estrutura de registro e gerar mais informações a partir de seus dados. É possível dividir um único campo de dados de formulário livre em campos que contêm tipos de informações diferentes. Você também pode adicionar informações aos seus registros. Por exemplo, você pode sinalizar registros do cliente como clientes pessoais ou comerciais.
- Valide endereços postais. A validação de endereço avalia e aprimora a precisão e a capacidade de entrega dos seus dados de endereço postal. A validação de endereço corrige erros em endereços e completa endereços parciais comparando registros de endereço com dados de referência de transportadoras postais nacionais. A validação de endereço pode ainda adicionar informações postais que aceleram a entrega de correspondências e reduz os custos de postagem.
- Localize registros duplicados. A análise de registros duplicados compara um conjunto de registros com outro para encontrar valores semelhantes ou correspondentes nas colunas de dados selecionadas. Você define o nível de similaridade que indica uma boa correspondência entre os valores do campo. Além disso, você também pode definir o peso relativo fixo para cada coluna em cálculos de correspondência. Por exemplo, você pode priorizar informações de sobrenome em relação às de nome.
- Crie e execute regras de qualidade de dados. A Informatica fornece regras pré-criadas que você pode executar ou editar para atender os objetivos do seu projeto. Você pode criar regras na Developer tool.
- Colaborar com usuários da Informatica. As tabelas de dados de referência e regras que você adiciona ao repositório do Modelo estão disponíveis para usuários na Developer tool e na ferramenta Analyst. Os usuários podem colaborar em projetos, e usuários diferentes podem assumir a propriedade de objetos em estágios diferentes de um projeto.
- Exporte mapeamentos para o PowerCenter. Você pode exportar mapeamentos para o PowerCenter reutilizar os metadados para integração física de dados ou para criar serviços Web.

Instaladores de Integração

Se você usar versões diferentes da Qualidade de Dados e do PowerCenter para criar e armazenar objetos, instale o Plug-in da Qualidade de Dados da Informatica para o PowerCenter. Baixe os instaladores de plug-in da Informatica.

Instale a versão de plug-in para a versão do PowerCenter que você deseja usar. O plug-in tem instaladores de Cliente e Servidor.

Execute o instalador do Cliente na máquina cliente do PowerCenter para exibir mapplets e mapeamentos no repositório e adicioná-los às sessões. Execute o instalador do Servidor na máquina com o Serviço de Integração do PowerCenter para executar as sessões em fluxos de trabalho.

Após instalar o plug-in na máquina do Serviço de Integração, registre o plug-in com o repositório do PowerCenter.

Exportação de Objetos para o PowerCenter

Use a ferramenta Desenvolvedor para exportar objetos, como mapplets e mapeamentos, do repositório do Modelo do Informatica. É possível exportar objetos do repositório do Modelo para o repositório do PowerCenter, bem como exportar objetos do repositório do Modelo para o sistema de arquivos. Use o PowerCenter Repository Manager para importar objetos XML para o repositório do PowerCenter.

- Para obter informações sobre como exportar objetos do repositório do Modelo, consulte o *Guia de Mapeamento do Informatica Developer* e o *Guia da Informatica Developer Tool*.
- Para obter informações sobre como importar objetos para o repositório do PowerCenter, consulte o *Guia do Repositório do PowerCenter*.

Instalação de Dados de Referência

Se um mapplet ou mapeamento ler dados de referência, copie ou instale os arquivos de dados de referência na máquina do Serviço de Integração do PowerCenter.

Você pode usar a ferramenta Desenvolvedor para exportar os dados de referência do repositório do Modelo quando exporta um mapplet ou mapeamento. O Desenvolvedor exporta os dados de referência em um arquivo compactado. Você extrai o arquivo e copia os dados de referência para a máquina do Serviço de Integração.

O processo de exportação não inclui arquivos de dados de referência de endereço nem arquivos de preenchimento de identidade. Se o mapplet ou mapeamento exportado ler arquivos de dados de referência de endereço ou arquivos de dados de preenchimento de identidade, instale os arquivos na máquina do Serviço de Integração do PowerCenter.

Regras e Diretrizes para a Integração de Produtos

Você pode exportar objetos de qualidade de dados de diferentes versões do Informatica Data Quality para diferentes versões do PowerCenter. Instale o plug-in de Integração na versão do Informatica Data Quality da qual você exporta os objetos.

Considere as seguintes regras e diretrizes ao integrar o Informatica Data Quality e o PowerCenter:

- Não instale o plug-in de Integração se a versão de produto do repositório do PowerCenter corresponder à versão de produto do repositório do Modelo.
- Não é possível executar dois plug-ins de Integração do Data Quality simultaneamente em uma única máquina do PowerCenter. O PowerCenter usa o plug-in que você instalou mais recentemente. Por exemplo, se você instalar o plug-in de Integração de Data Quality 10.2 em uma máquina do PowerCenter 10.1.1, não poderá mais importar e executar objetos de qualidade de dados de um repositório do Modelo da versão 10.1.1. Você pode continuar a executar qualquer objeto de qualidade de dados importado para o repositório do PowerCenter antes da instalação do plug-in 10.2.
- Não é possível editar mapplets ou mapeamentos de qualidade de dados em um repositório do PowerCenter.

Comportamento das Transformações de Qualidade de Dados no PowerCenter

Durante a importação de um mapeamento ou mapplet de qualidade de dados do repositório do Modelo para o repositório do PowerCenter, o comportamento das transformações de qualidade de dados pode mudar.

Conversão em Mapplets

As seguintes transformações podem ser convertidas em mapplets na importação para o repositório do PowerCenter:

- Consolidação
- Exceção
- Gerador de Chave
- Correspondência

Durante a importação de um mapeamento que contenha uma ou mais das transformações, o PowerCenter converte cada uma delas em um mapplet que contém a configuração expandida.

Quando você importa um mapplet que contém uma ou mais das transformações, o PowerCenter expande a configuração de cada transformação em várias transformações.

O PowerCenter pode adicionar transformações à transformação de qualidade de dados ou à expansão de mapplet para concluir a lógica de transformação. Por exemplo, o PowerCenter expande a transformação Correspondência com outras transformações.

Considerações da transformação de Decisão

Você não poderá exportar um mapeamento ou mapplet que contém uma transformação de Decisão para o PowerCenter se o script da transformação incluir um parâmetro do sistema. A operação de exportação não pode converter o parâmetro do sistema em um valor que pode ser usado pelo PowerCenter. Antes de exportar um mapeamento ou mapplet com uma transformação de Decisão que utiliza um parâmetro do sistema, substitua o parâmetro por um valor apropriado.

Considerações sobre a Transformação de Correspondência

Você não poderá executar um mapeamento de correspondência de identidade no PowerCenter se esse mapeamento fizer a leitura ou a gravação de dados de índice de identidade em tabelas de banco de dados. Para executar a análise de correspondência de identidade no PowerCenter, configure a transformação de Correspondência na Developer tool de forma a gravar dados de índice de identidade em arquivos temporários.

Transformações de Várias Estratégias

Você pode definir várias estratégias de transformação de dados em cada uma das transformações a seguir na Developer tool:

- Caso
- Classificador
- Decisão
- Gerador de Chave
- Rotulador
- Correspondência
- Mesclar
- Analisador
- Padronizador

Durante a importação de um mapplet ou de um mapeamento que contenha uma transformação de várias instâncias, cada estratégia é convertida em uma transformação separada.

CAPÍTULO 2

Instalação

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Instalação, 11](#)
- [Requisitos de Sistema, 11](#)
- [Antes da Instalação, 12](#)
- [Instalando o Plug-in de Servidor, 12](#)
- [Instalando o Plug-in Cliente, 13](#)
- [Registrando o Plug-In no PowerCenter, 14](#)

Visão Geral da Instalação

O plug-in Data Quality Integration tem um componente de servidor e um componente cliente.

Execute as seguintes etapas para instalar e registrar o plug-in:

1. Execute o instalador do Servidor na máquina com o Serviço de Integração do PowerCenter.
2. Execute o instalador do Cliente em uma máquina do cliente do PowerCenter.
3. Registre o arquivo `IDQZIntegration.xml` com o repositório do PowerCenter.

Requisitos de Sistema

O plug-in de Integração do Informatica Data Quality tem os mesmos requisitos de sistema dos aplicativos e serviços do PowerCenter nos quais é instalado.

Para obter informações sobre os pré-requisitos do PowerCenter e os requisitos de sistema, consulte a documentação de instalação do PowerCenter.

Antes da Instalação

Considere o comportamento do instalador a seguir se você usar dados de referência no PowerCenter.

O instalador do plug-in de Integração grava arquivos de propriedades na máquina do Serviço de Integração do PowerCenter. Os arquivos de propriedades especificam as localizações dos arquivos de dados de referência que os mapeamentos de qualidade de dados podem ler.

O instalador grava os seguintes arquivos:

- AD50.cfg. Armazena as propriedades de configuração para dados de referência de endereço.
- CLASSIFIER.properties. Armazena as propriedades de configuração para dados do modelo de classificador.
- IDQTx.cfg. Armazena as propriedades de configuração para dados de preenchimento de identidade.
- NER.properties. Armazena as propriedades de configuração para dados do modelo de probabilidade.

O instalador adiciona os arquivos ao seguinte diretório:

```
[PowerCenter_installation]/server/bin
```

Pré-requisito de Dados de Referência de Endereço

Antes de instalar dados de referência de endereço, interrompa o Serviço de Integração do PowerCenter. Reinicie o serviço depois de instalar os dados.

Instalando o Plug-in de Servidor

O instalador do plug-in de Servidor de Integração do Informatica Data Quality adiciona arquivos de aplicativo de transformação de qualidade de dados à estrutura de diretório do Serviço de Integração do PowerCenter.

Instalando no Windows

Execute o instalador do Windows Server para adicionar o plug-in a uma máquina com o Serviço de Integração do PowerCenter em uma plataforma Windows.

Feche todos os aplicativos antes de instalar.

1. Faça logon na máquina com uma conta de sistema do usuário.
2. Extraia o arquivo do Servidor de Integração para Windows.
3. Vá para o diretório em que você extraiu os arquivos e execute `install.exe`.
4. Selecione um idioma e clique em **OK**.
5. Na tela **Bem-vindo**, clique em **Avançar**.
6. Verifique se os serviços do PowerCenter não estão em execução e clique em **OK**.
7. Especifique o caminho para o diretório do PowerCenter. O instalador fornece um caminho padrão para esse diretório. Para selecionar outro caminho, clique em **Fechar**.
8. Clique em **Avançar**.
9. Examine o resumo de pré-instalação. Clique em **Anterior** se desejar corrigir algum erro.
10. Clique em **Instalar**.
11. Examine o resumo de pós-instalação.

Para obter mais informações sobre as tarefas de instalação e para exibir as propriedades de configuração dos componentes instalados, consulte os arquivos de log de instalação.

12. Clique em **Concluído**.

Instalando no UNIX

Execute o instalador do UNIX Server para adicionar o plug-in a uma máquina com o Serviço de Integração do PowerCenter em uma plataforma UNIX.

1. Faça logon na máquina com uma conta de sistema do usuário.
2. Feche todos os outros aplicativos.
3. Extraia o arquivo do Integration Server for UNIX.
4. Abra um shell UNIX e navegue para o diretório em que você extraiu os arquivos de instalação.
5. Digite `sh install.bin`.
6. Para aceitar o inglês como o idioma de instalação, pressione **Enter**.
7. Leia o texto de Boas-vindas e pressione **Enter**.
8. Digite um caminho absoluto para o diretório de instalação do PowerCenter ou pressione **Enter** para aceitar o caminho padrão.
9. Examine o resumo de pré-instalação. Para corrigir um erro, digite `back` e pressione **Enter**.
10. Pressione **Enter** para iniciar a instalação.
11. Examine o resumo de pós-instalação.

Para obter mais informações sobre as tarefas de instalação e para exibir as propriedades de configuração dos componentes instalados, consulte os arquivos de log de instalação.

12. Pressione **Enter** para concluir o processo de instalação.

Instalando o Plug-in Cliente

Execute o instalador do Data Quality Integration Client nas máquinas Cliente do PowerCenter.

1. Feche todos os outros aplicativos.
2. Extraia o arquivo de Cliente de Integração.
3. Abra o diretório no qual você extraiu os arquivos.
4. Execute `install.exe`.
5. Na tela **Bem-vindo**, clique em **Avançar**.
O instalador exibe um prompt para fechar os aplicativos cliente do PowerCenter antes de continuar a instalação.
6. Verifique se não há nenhum aplicativo cliente do PowerCenter em execução e clique em **OK**.
7. Especifique o caminho para o diretório do Cliente do PowerCenter. O instalador fornece um caminho padrão para esse diretório. Para selecionar outro caminho, clique em **Fechar**.
8. Clique em **Avançar**.
9. Examine o resumo de pré-instalação. Clique em **Anterior** se desejar corrigir algum erro.
10. Clique em **Instalar**.

11. Examine o resumo de pós-instalação.

Para obter mais informações sobre as tarefas de instalação e para exibir as propriedades de configuração dos componentes instalados, consulte os arquivos de log de instalação.

12. Clique em **Concluído**.

Registrando o Plug-In no PowerCenter

Registre o plug-in em qualquer repositório do PowerCenter que deverá conter objetos de qualidade de dados.

O instalador de Integração grava o arquivo de plug-in `IDQZIntegration.xml` no diretório `$INFA_HOME\server\bin\native` na máquina do Serviço de Integração do PowerCenter. Verifique se você consegue acessar esse arquivo. Se necessário, copie ou envie esse arquivo por FTP para a máquina do Serviço de Integração do PowerCenter.

- Para registrar o plug-in, execute o comando `pmrep RegisterPlugin`. Inclua a opção `-N` quando executar o comando.

Para obter mais informações sobre o comando `pmrep RegisterPlugin`, consulte a *Referência de Comandos do Informatica*.

Nota: Se você não tiver os privilégios corretos para registrar o plug-in, contate o administrador do Serviço de Repositório do PowerCenter.

CAPÍTULO 3

Requisitos de Dados de Referência

Este capítulo inclui os seguintes tópicos:

- [Visão Geral dos Dados de Referência, 15](#)
- [Requisitos de Dados da Tabela de Referência, 16](#)
- [Requisitos de Dados de Referência de Endereço, 17](#)
- [Requisitos de Dados de Preenchimento de Identidade, 18](#)
- [Requisitos do Modelo de Probabilidade e do Modelo de Classificador, 19](#)

Visão Geral dos Dados de Referência

As transformações de qualidade de dados pode ler dados de referência. Se qualquer transformação no repositório do PowerCenter ler dados de referência, verifique se os arquivos de dados de referência estão presentes no ambiente do PowerCenter.

As transformações de qualidade de dados no PowerCenter podem usar os seguintes tipos de dados de referência:

Arquivos de dados de referência

Os arquivos de texto que contêm as versões padrão de termos comerciais e versões alternativas dos termos. É possível editar os arquivos de dados de referência.

Arquivos de dados de referência de endereço

Arquivos de dados que contêm dados completos para todos os endereços postais em um país. Instale os arquivos de dados de referência de endereço na máquina do Serviço de Integração do PowerCenter. Não é possível editar arquivos de dados de referência de endereço.

Arquivos de preenchimento de identidade

Arquivos de dados que definem tipos diferentes de identidade e contêm algoritmos usados pela transformação de Correspondência. Instale os arquivos de preenchimento na máquina de Serviços do PowerCenter. Não é possível editar arquivos de preenchimento.

Conjuntos de conteúdo

Os arquivos de dados que contêm dados de referência especificados por uma ou mais transformações. Quando você exporta um mapeamento que lê conjuntos de conteúdo, o processo de exportação exporta os dados do conjunto de conteúdo com os metadados de transformação. No caso dos arquivos do

modelo de classificador e de probabilidade, o processo exporta os dados da estrutura de diretório da Qualidade de Dados.

Requisitos de Dados da Tabela de Referência

Antes de executar um mapeamento que lê tabelas de referência no PowerCenter, copie os dados da coluna das tabelas para a máquina de serviços do PowerCenter.

Ao exportar o mapeamento do repositório do Modelo, a operação de exportação grava os dados da coluna da tabela de referência em um ou mais arquivos. A operação de exportação adiciona os arquivos a um arquivo compactado em uma localização que você especifica. Extraia o arquivo compactado e encontre os arquivos de dados da coluna. Os nomes dos arquivos terminam em `.dic`. Copie os arquivos para a máquina de serviços do PowerCenter.

As localizações do arquivo de dados de referência na estrutura de diretório do PowerCenter devem corresponder às localizações da tabela de referência no repositório do Modelo.

Use o seguinte formato de diretório para determinar as localizações corretas para os arquivos de dados da tabela de referência:

```
<Diretório_Raiz_do_PowerCenter>\services\<Nome _do_Projeto_do_Repositório_do_Modelo>\<Nome_da_Pasta_do_Projeto_do_Repositório_do_Modelo>
```

Regras e Diretrizes para Exportação de Dados da Tabela de Referência

Considere as seguintes regras e diretrizes ao exportar um mapeamento que lê dados de tabelas de referência:

- Para exportar o mapeamento do repositório do Modelo, escolha a opção **Informatica > PowerCenter** no assistente de exportação. Quando você escolhe a opção **PowerCenter**, a operação de exportação grava os dados da tabela de referência gerenciada e os dados da tabela de referência não gerenciada no sistema de arquivos.
- A operação de exportação especifica a primeira coluna em cada tabela de referência como a coluna válida.
- O PowerCenter lê arquivos de dados de referência no formato UTF-8.

Usando uma Variável de Ambiente para Definir a Localização do Arquivo de Dados de Referência

Você pode definir uma variável de ambiente `INFA_CONTENT` na máquina do Serviço de Integração do PowerCenter para definir o caminho que o Serviço de Integração usa para ler arquivos de dados de referência.

Use essa variável de ambiente se você não puder instalar os arquivos na localização exigida na máquina do Serviço de Integração do PowerCenter. Instale os arquivos em um local que o Serviço de Integração do PowerCenter possa ler e crie a variável de ambiente `INFA_CONTENT` com o caminho instalado.

Requisitos de Dados de Referência de Endereço

Um mapeamento que executa a validação de endereços lê dados de referência de endereço. Você adquiriu e fez download de dados de referência de endereço da Informatica. Instale os arquivos de dados para a máquina do Serviço de Integração do PowerCenter.

A tabela a seguir descreve os tipos de dados de referência de endereço que você pode instalar:

Tipo de dados de referência	Quando usar
Lote	Use para operações gerais de validação de endereço.
CAMEO	Use para adicionar dados demográficos de consumidores aos registros de endereço.
Certificado	Use para verificar se os registros de endereço atendem aos padrões de certificação definidos por um serviço de correio. Os seguintes países definem padrões de certificação: <ul style="list-style-type: none">- Austrália. Certifica a correspondência de acordo com o padrão AMAS (Sistema de Aprovação de Correspondência de Endereço).- Canadá. Certifica a correspondência de acordo com o padrão SERP (Programa de Análise e Reconhecimento de Software) padrão.- França. Certifica a correspondência de acordo com o padrão SNA (Serviço Nacional de Gerenciamento de Endereços).- Nova Zelândia. Certifica a correspondência de acordo com o padrão SendRight.- Estados Unidos. Certifica a correspondência de acordo com o padrão CASS (Sistema de Suporte para Precisão de Codificação).
Lista de sugestões	Use para encontrar versões válidas de um registro de endereço incompleto. Use dados da lista de sugestões ao configurar um mapeamento de validação de endereço para processar registros de endereço simples em tempo real.
Codificação Geográfica	Use para adicionar coordenadas de latitude e de longitude aos registros de endereço.
Complementares	Use para identificar a área geográfica ou a área populacional que contém um endereço. A transformação do Validador de Endereço pode adicionar dados suplementares a registros de endereço dos seguintes países: <ul style="list-style-type: none">- Japão. Identifica o distrito urbano que contém o endereço.- Sérvia. Identifica o código postal no nível da rua para o endereço.- Reino Unido. Adiciona um identificador do ponto de entrega ao endereço.- Estados Unidos. Identifica o centro populacional que contém o endereço.

O Serviço de Integração do PowerCenter lê o caminho para os dados de referência de endereço no arquivo de configuração AD50.cfg.

O instalador do PowerCenter grava o arquivo AD50.cfg no seguinte caminho:

```
[Instalação_do_PowerCenter]/server/bin
```

Quando você instala dados de referência de endereço, verifique se as seguintes propriedades estão definidas no AD50.cfg:

Localização dos Dados de Referência

Digite o caminho para os dados de referência de endereço na propriedade `ReferenceDataLocation`.

Chave de licença de dados

Insira a licença dados na propriedade `LicenseKey`. Se você tiver mais de uma licença, digite cada licença em uma cadeia de caracteres separados por vírgula.

Valores de pré-carregamento

Insira pelo menos uma abreviação de país como valor de pré-carregamento para cada tipo de dados de referência que um mapeamento lê. Digite ALL para aplicar uma configuração de pré-carregamento para todos os países.

O Serviço de Integração pode usar um método diferente para carregar dados para cada país. Por exemplo, você pode especificar o pré-carregamento total de dados em lotes/interativo dos Estados Unidos e o pré-carregamento parcial de dados em lotes/interativo do Reino Unido. O Serviço de Integração também pode usar um método de pré-carregamento diferente para cada tipo de dados. Por exemplo, você pode especificar o pré-carregamento total de dados em lotes/interativo dos Estados Unidos e o pré-carregamento parcial dos dados de geocodificação dos Estados Unidos.

As configurações de pré-carregamento total substituem as configurações de pré-carregamento parcial, e as configurações de pré-carregamento parcial substituem as configurações que indicam nenhum pré-carregamento de dados. Por exemplo, se você não inserir ALL para nenhum pré-carregamento de dados e inserir USA para o pré-carregamento total, o Serviço de Integração carregará todos os dados dos Estados Unidos na memória e não carregará os dados de nenhum outro país. Se você não tiver um requisito de pré-carregamento, digite ALL para o não pré-carregamento de dados de nenhum tipo de dado de referência que você planeje usar.

Você não especifica um valor de pré-carregamento para dados suplementares.

Se desejar, edite outras propriedades no arquivo `AD50.cfg`.

Requisitos de Dados de Preenchimento de Identidade

Um mapeamento que executa análise de correspondência de identidade lê arquivos de dados de referência denominados preenchimentos de identidade. Adquira e faça download dos arquivos de preenchimento de identidade da Informatica. Copie os arquivos de preenchimento na máquina do Serviço de Integração do PowerCenter ou em um diretório compartilhado em uma máquina que o Serviço de Integração do PowerCenter possa acessar.

Você deve copiar ou instalar os arquivos de preenchimento em um diretório com o nome `padrão`.

O Serviço de Integração do PowerCenter pode ler a localização do diretório do arquivo de preenchimento por meio do arquivo de configuração `IDQTx.cfg` ou da variável de ambiente `SSAPR`. O Serviço de Integração do PowerCenter procura a localização no arquivo `IDQTx.cfg`. Se o arquivo `IDQTx.cfg` não estiver presente ou não especificar uma localização, o Serviço de Integração do PowerCenter procurará a localização na variável de ambiente `SSAPR`.

Nota: O arquivo `IDQTx.cfg` e a variável de ambiente `SSAPR` especificam o caminho para o diretório pai do diretório `padrão`. O caminho não inclui o nome de diretório.

O instalador do PowerCenter grava o arquivo `IDQTx.cfg` no seguinte caminho:

```
[Instalação_do_PowerCenter]/server/bin
```

O arquivo `IDQTx.cfg` usa a propriedade `IdentityReferenceDataLocation` para especificar a localização do diretório de arquivos de preenchimento. Atualize a propriedade com a localização atual do diretório.

Por exemplo, você pode definir o seguinte caminho na propriedade *IdentityReferenceDataLocation*:

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

Você pode definir um caminho absoluto ou relativo para o diretório padrão na propriedade *IdentityReferenceDataLocation*. Defina um caminho absoluto para um diretório na máquina do Serviço de Integração do PowerCenter ou para um diretório compartilhado em uma máquina que o serviço possa acessar. Defina um caminho relativo para um diretório na máquina do Serviço de Integração do PowerCenter.

O Serviço de Integração do PowerCenter lê um caminho relativo no seguinte diretório:

```
[Instalação_do_PowerCenter]/server/bin
```

Por exemplo, para especificar um caminho relativo a um diretório de arquivo de preenchimento no diretório *bin*, defina o seguinte caminho relativo:

```
./
```

Propriedades do IDQTx.cfg

Quando você define a propriedade *IdentityReferenceDataLocation* no arquivo *IDQTx.cfg*, também pode definir as seguintes propriedades:

IdentityCacheDir

Caminho para o diretório que armazena os arquivos de dados temporários criados quando o mapeamento é executado. O caminho identifica um diretório pai. O Serviço de Integração do PowerCenter grava os arquivos temporários nos diretórios abaixo da localização que você especificar.

A propriedade tem o seguinte valor padrão:

```
./identityCache
```

IdentityIndexDir

Caminho para o diretório que contém os arquivos de índice temporários criados quando o mapeamento é executado. A análise de correspondência de identidade usa o índice para classificar os registros em grupos antes da análise de correspondência. O caminho identifica um diretório pai. O Serviço de Integração do PowerCenter grava os arquivos de índice nos diretórios abaixo da localização que você especificar.

A propriedade tem o seguinte valor padrão:

```
./identityIndex
```

Nota: Um usuário da ferramenta Developer pode definir o caminho do diretório de cache e o caminho do diretório de índice na transformação de Correspondência. O Serviço de Integração do PowerCenter usará o caminho no arquivo *IDQTx.cfg* se a transformação de Correspondência não especificar o caminho.

Requisitos do Modelo de Probabilidade e do Modelo de Classificador

Modelos probabilísticos e modelos de classificador são arquivos de dados de referência que identificam os tipos de informações em uma string de dados.

Antes de executar um mapeamento que lê um arquivo de modelo probabilístico ou de modelo de classificador, execute as seguintes tarefas:

- Copie as localizações do arquivo de modelo na estrutura de diretórios do PowerCenter.

- Atualize a alocação de memória do Java no Serviço de Integração do PowerCenter.

Localização de Arquivos do Modelo Probabilístico e do Modelo de Classificador no PowerCenter

Antes de executar um mapeamento que leia um modelo probabilístico ou um modelo de classificador no PowerCenter, copie o arquivo de modelo para a máquina de serviços do PowerCenter.

Quando você exporta o mapeamento de repositório do Modelo, a operação de exportação adiciona cada modelo a um arquivo compactado em uma localização especificada por você. Extraia o arquivo compactado e encontre os arquivos de modelo que o mapeamento lê. Os nomes dos arquivos de dados do modelo probabilístico e do modelo classificador terminam em `.ner` e `.classifier` respectivamente. Copie os arquivos para um diretório que o Serviço de Integração do PowerCenter possa ler.

Por padrão, o Serviço de Integração do PowerCenter lê os arquivos de modelo no seguinte diretório:

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

Se os arquivos não estiverem presentes na localização padrão, o Serviço de Integração do PowerCenter lerá a localização dos arquivos nos arquivos de propriedade. Você deve atualizar os arquivos de propriedade com os caminhos para os arquivos de modelo.

A tabela a seguir descreve os arquivos de propriedade:

Nome do Arquivo	Descrição
CLASSIFIER.properties	Contém a propriedade <i>ClassifierFileLocation</i> . Atualize a propriedade para identificar o caminho para os arquivos de modelo de classificador.
NER.properties	Contém a propriedade <i>NERFileLocation</i> . Atualize a propriedade para identificar o caminho para os arquivos de modelo probabilístico.

O Serviço de Integração do PowerCenter lê os arquivos de propriedades do seguinte diretório:

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

Regras e diretrizes para exportação do modelo probabilístico e do modelo de classificador

Considere as seguintes regras e diretrizes ao exportar um mapeamento que lê dados de tabelas de referência:

- Quando você exporta um mapeamento que lê dados de modelo probabilístico ou de modelo de classificador para uso no PowerCenter, selecione a opção **Informatica > Exportar Arquivo de Metadados do Objeto** no assistente de exportação para criar o arquivo compactado.

A operação de exportação também exporta o mapeamento como um arquivo XML. É possível desconsiderar o arquivo XML.

- Selecione a opção para exportar somente modelos treinados. Um modelo probabilístico ou modelo de classificador treinado contém os dados lógicos que o mapeamento lê e não contém os dados de origem para o modelo. Copie os arquivos de modelo treinado para a máquina de serviços do PowerCenter.
- Exporte o mapeamento do repositório do Modelo para o repositório do PowerCenter em uma operação separada. Para exportar o mapeamento do repositório do Modelo para o repositório do PowerCenter, selecione a opção **Informatica > PowerCenter**.

Ao selecionar a opção **PowerCenter**, a operação de exportação pode copiar dados de referência de tabelas de referência gerenciadas e tabelas de referência não gerenciadas.

- Você pode definir um caminho absoluto ou um caminho relativo para identificar as localizações de arquivos de modelo em cada arquivo de propriedades. Se você definir um caminho relativo, o Serviço de Integração do PowerCenter adicionará o caminho para o seguinte diretório:

`[PowerCenter_Installation]/server/bin`

- Use barras invertidas para definir os caminhos de diretório para os arquivos de modelo em todas as plataformas.

Alocação de Memória do Serviço de Integração do PowerCenter

Um Serviço de Integração do PowerCenter usa grandes quantidades de memória para ler um arquivo de modelo de classificador ou um arquivo de modelo probabilístico. Se o Serviço de Integração do PowerCenter não puder acessar a memória necessária, ele falhará ao ler o arquivo de modelo e não poderá processar os dados.

Consulte a alocação de memória no Serviço de Integração do PowerCenter. Se você executar sessões que fazem referência a um arquivo de modelo de classificador ou a um arquivo de modelo probabilístico, defina a alocação de memória como pelo menos 512 MB.

Você pode usar a propriedade de processo *Memória Máxima para Java SDK* para alocar a memória. Para analisar ou atualizar a propriedade, faça login na ferramenta Administrator e selecione o Serviço de Integração do PowerCenter que executa as tarefas de sessão. Você encontrará a propriedade *Memória Máxima do Java SDK* na exibição **Processos**.

ÍNDICE

I

Informatica Data Quality [6](#)

Integração com o Data Quality [6](#)