



Informatica® Data Quality
10.1.1 HotFix 2

Guia de Introdução

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, o logotipo Informatica, PowerCenter e PowerExchange 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.

Consulte as patentes em <https://www.informatica.com/legal/patents.html>.

ISENÇÃO DE RESPONSABILIDADE: a Informatica LLC fornece esta documentação no estado em que se encontra, sem garantia de qualquer tipo, expressa ou implícita, incluindo, mas não limitando-se, as garantias implícitas de não infração, comercialização ou uso para um determinado propósito. A Informatica LLC não garante que este software ou documentação não contenha erros. As informações fornecidas neste software ou documentação podem incluir imprecisões técnicas ou erros tipográficos. As informações deste software e documentação estão sujeitas a alterações a qualquer momento sem aviso prévio.

AVISOS

Este produto da Informatica (o "Software") traz determinados drivers (os "drivers da DataDirect") da DataDirect Technologies, uma empresa em funcionamento da Progress Software Corporation ("DataDirect"), que estão sujeitos aos seguintes termos e condições:

1. OS DRIVERS DA DATADIRECT SÃO FORNECIDOS NO ESTADO EM QUE SE ENCONTRAM, SEM GARANTIA DE QUALQUER TIPO, EXPRESSA OU IMPLÍCITA, INCLUINDO, MAS NÃO LIMITANDO-SE, AS GARANTIAS IMPLÍCITAS DE COMERCIALIZAÇÃO, ADEQUAÇÃO A UMA FINALIDADE ESPECÍFICA E NÃO INFRAÇÃO.
2. EM NENHUM CASO, A DATADIRECT OU SEUS FORNECEDORES TERCEIRIZADOS SERÃO RESPONSÁVEIS, EM RELAÇÃO AO CLIENTE FINAL, POR QUAISQUER DANOS DIRETOS, INDIRETOS, INCIDENTAIS, ESPECIAIS, CONSEQUENCIAIS OU DE MAIS QUE POSSAM ADVIR DO USO DE DRIVERS ODBC, SENDO OU NÃO ANTERIORMENTE INFORMADOS DAS POSSIBILIDADES DE TAIS DANOS. ESTAS LIMITAÇÕES SE APLICAM A TODAS AS CAUSAS DE AÇÃO, INCLUINDO, SEM LIMITAÇÕES, QUEBRA DE CONTRATO, QUEBRA DE GARANTIA, NEGLIGÊNCIA, RESPONSABILIDADE RIGOROSA, DETURPAÇÃO E OUTROS ATOS ILÍCITOS.

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: 2018-06-09

Conteúdo

Prefácio.....	7
Recursos da Informatica.	7
Rede da Informatica.	7
Base de Dados de Conhecimento da Informatica.	7
Documentação da Informatica.	7
Matrizes de Disponibilidade de Produto Informatica.	8
Informatica Velocity.	8
Informatica Marketplace.	8
Suporte global a clientes Informatica.	8
 Capítulo 1: Visão Geral da Introdução.....	 9
Visão Geral do Domínio Informatica.	9
Disponibilidade de Recursos.	11
Introdução ao Informatica Analyst.	11
Visão Geral do Informatica Developer.	12
Página de Boas-vindas do Informatica Developer.	13
Folhas de Cola.	13
Qualidade de Dados e Criação de Perfil.	14
A História do Tutorial.	14
A Estrutura do Tutorial.	15
Tutorial do Informatica Analyst.	15
Informatica Developer tool.	16
Pré-requisitos do Tutorial.	17
 Parte I: Introdução ao Informatica Analyst.....	 18
 Capítulo 2: Lição 1. Configurar o Informatica Analyst.....	 19
Visão Geral da Configuração do Informatica Analyst.	19
Tarefa 1. Fazer login no Informatica Analyst.	20
Tarefa 2. Criar um Projeto.	20
Tarefa 3. Criar uma Pasta.	21
Resumo da Configuração do Informatica Analyst.	21
 Capítulo 3: Lição 2. Criar Objetos de Dados.....	 22
Visão Geral da Criação de Objetos de Dados.	22
Tarefa 1. Criar os Objetos de Dados de Arquivo Simples.	23
Tarefa 2. Exibir as Propriedades de Objeto de Dados.	23
Resumo da Criação de Objetos de Dados.	24

Capítulo 4: Lição 3. Criando Perfis Padrão.	25
Criando Perfis Padrão - Visão Geral.	25
Tarefa 1. Criar e Executar um Perfil Padrão.	26
Tarefa 2. Visualizar os resultados do perfil na exibição resumida.	26
Criando Perfis Padrão - Resumo.	27
Capítulo 5: Lição 4. Criar Perfis Personalizados.	28
Criar Visão Geral de Perfis Personalizados.	28
Tarefa 1. Criar um Perfil Personalizado.	29
Tarefa 2. Executar o Perfil.	30
Tarefa 3. Fazer uma Busca Detalhada nos Resultados do Perfil.	30
Criando Perfis Personalizados - Resumo.	31
Capítulo 6: Lição 5. Criar Regras de Expressão.	32
Criando Regras de Expressão - Visão Geral.	32
Tarefa 1. Criar Regras de Expressão e Executar o Perfil.	33
Tarefa 2. Exibir a Saída de Regra de Expressão.	33
Tarefa 3. Editar as regras de expressão.	34
Criar Resumo de Regras de Expressão.	34
Capítulo 7: Lição 6. Criar e Executar Scorecards.	35
Visão Geral da Criação e Execução de Scorecards.	35
Tarefa 1. Criar um Scorecard dos Resultados de Perfil.	36
Tarefa 2. Executar o Scorecard.	37
Tarefa 3. Exibir o Scorecard.	37
Tarefa 4. Editar o Scorecard.	38
Tarefa 5. Configurar Limites.	39
Tarefa 6. Exibir Gráficos de Tendência de Pontuação.	39
Resumo da Criação e Execução de Scorecards.	39
Capítulo 8: Lição 7. Criar Tabelas de Referência de Colunas de Perfil.	40
Visão Geral da Criação de Tabelas de Referência de Colunas de Perfil.	40
Tarefa 1. Criar uma Tabela de Referência das Colunas de Perfil.	41
Tarefa 2. Editar a Tabela de Referência.	42
Resumo da Criação de Tabelas de Referência de Colunas de Perfil.	42
Capítulo 9: Lição 8. Criar Tabelas de Referência.	43
Visão Geral da Criação de Tabelas de Referência.	43
Tarefa 1. Criar uma Tabela de Referência.	44
Resumo da Criação de Tabelas de Referência.	44

Parte II: Introdução ao Informatica Developer.....	45
Capítulo 10: Lição 1. Configurando o Informatica Developer.....	46
Configurando o Informatica Developer - Visão Geral.	46
Tarefa 1. Iniciar o Informatica Developer.	47
Tarefa 2. Adicionar um Domínio.	47
Tarefa 3. Adicionar um Repositório do Modelo.	48
Tarefa 4. Criar um Projeto.	48
Tarefa 5. Criar uma Pasta.	48
Tarefa 6. Selecionar um Data Integration Service padrão.	49
Configurando o Informatica Developer - Resumo.	49
Capítulo 11: Lição 2: Importando Objetos de Dados Físicos.....	50
Importando Objetos de Dados Físicos - Visão Geral.	50
Tarefa 1. Importar o Objeto de Dados de Arquivo Simples Boston_Customers.	51
Tarefa 2. Importar o Objeto de Dados de Arquivo Simples LA_Customers.	57
Tarefa 3. Importar o Objeto de Dados de Arquivo Simples All_Customers.	58
Importando Objetos de Dados Físicos - Resumo.	58
Capítulo 12: Lição 3. Executar um Perfil em Dados de Origem.....	59
Visão Geral da Criação de Perfil de Dados.	59
Tarefa 1. Executar uma Análise de Associação em Duas Fontes de Dados.	60
Tarefa 2. Exibir Resultados de Análise de Associação.	61
Tarefa 3. Executar um Perfil em uma Fonte de Dados.	61
Tarefa 4. Exibir Resultados de Criação de Perfil de Coluna.	62
Resumo da Criação de Perfil.	63
Capítulo 13: Lição 4. Analisando dados.....	64
Visão Geral da Análise de Dados.	64
Tarefa 1. Criar um Objeto de Dados de Destino.	65
Etapa 1. Criar um Objeto de Dados LA_Customers_tgt.	65
Etapa 2. Configurar Opções de Leitura e Gravação.	66
Etapa 3. Adicionar Colunas ao Objeto de Dados.	66
Tarefa 2. Criar um Mapeamento para Analisar Dados.	67
Etapa 1. Criar um Mapeamento.	67
Etapa 2. Adicionar Objetos de Dados ao Mapeamento.	67
Etapa 3. Adicionar uma Transformação do Analisador ao Mapeamento.	68
Etapa 4. Configurar a Transformação do Analisador.	68
Tarefa 3. Executar um Perfil na Transformação do Analisador.	69
Tarefa 4. Executar o Mapeamento.	69
Tarefa 5. Exibir a Saída de Mapeamento.	69
Analisando dados de Resumo.	70

Capítulo 14: Lição 5. Padronizar Dados 71

Visão Geral da Padronização de Dados.	71
Tarefa 1. Criar um Objeto de Dados de Destino.	72
Etapa 1. Criar um objeto de dados All_Customers_Stdz_tgt.	72
Etapa 2. Configurar Opções de Leitura e Gravação.	73
Tarefa 2. Criar um Mapeamento para Padronizar Dados.	73
Etapa 1. Criar um Mapeamento.	74
Etapa 2. Adicionar Objetos de Dados ao Mapeamento.	74
Etapa 3. Adicionar uma Transformação do Padronizador ao Mapeamento.	74
Etapa 4. Configurar a Transformação do Padronizador.	75
Tarefa 3. Executar o Mapeamento.	76
Tarefa 4. Exibir a Saída de Mapeamento.	76
Resumo de Padronização de Dados.	76

Capítulo 15: Lição 6. Validar Dados de Endereço. 77

Visão Geral da Validação de Dados de Endereço.	77
Tarefa 1. Criar um Objeto de Dados de Destino.	78
Etapa 1. Criar o Objeto de Dados All_Customers_av_tgt.	79
Etapa 2. Configurar Opções de Leitura e Gravação.	79
Etapa 3. Adicionar Portas ao Objeto de Dados.	80
Tarefa 2. Criar um Mapeamento para Validar Endereços.	80
Etapa 1. Criar um Mapeamento.	80
Etapa 2. Adicionar Objetos de Dados ao Mapeamento.	81
Etapa 3. Adicionar uma Transformação do Validador de Endereço ao Mapeamento.	81
Tarefa 3. Configurar a Nova Transformação do Validador de Endereço.	81
Etapa 1. Definir o País Padrão para Validação de Endereço.	82
Etapa 2. Configurar as Portas de Entrada de Transformação do Validador de Endereço.	82
Etapa 3. Configurar as Portas de Saída de Transformação do Validador de Endereço.	83
Etapa 4. Conectar Portas de Fonte de Dados Não Utilizadas ao Destino de Dados.	84
Tarefa 4. Executar o Mapeamento.	84
Tarefa 5. Exibir a Saída de Mapeamento.	85
Resumo da Validação dos Dados de Endereço.	87

Apêndice A: Perguntas Frequentes..... 89

Perguntas Frequentes do Informatica Analyst.	89
Perguntas Frequentes sobre o Informatica Developer.	89

Índice..... 91

Prefácio

O *Guia de Introdução* do Data Quality foi desenvolvido para analistas e desenvolvedores de qualidade de dados. Ele fornece tutoriais para ajudar usuários iniciantes a usar o Informatica Developer e o Informatica Analyst. Este guia presume que você compreende os conceitos de qualidade de dados, os conceitos de arquivo simples e banco de dados relacional e os mecanismos de banco de dados em seu ambiente.

Recursos da Informatica

Rede da Informatica

A Rede da Informatica hospeda o Suporte Global a Clientes da Informatica, a Base de Dados de Conhecimento da Informatica e outros recursos de produtos. Para acessar a Rede da Informatica, visite <https://network.informatica.com>.

Como membro, você pode:

- Acessar todos os seus recursos Informatica em um só lugar.
- Pesquisar a Base de Dados de Conhecimento em busca de recursos de produtos, incluindo documentações, perguntas frequentes e práticas recomendadas.
- Visualizar informações sobre disponibilidade de produtos.
- 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 pesquisar a Rede da Informatica em busca de recursos de produtos, como documentações, artigos de instruções, práticas recomendadas e PAMs.

Para acessar a Base de Dados de Conhecimento, visite <https://kb.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

Para obter a documentação mais recente do seu produto, navegue pela Base de Dados de Conhecimento da Informatica em https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx

Em caso de dúvidas, comentários ou ideias sobre esta documentação, entre em contato com a equipe de Documentação da Informatica pelo e-mail infa_documentation@informatica.com.

Matrizes de Disponibilidade de Produto Informatica

As Matrizes de Disponibilidade de Produto (PAMs) indicam as versões dos sistemas operacionais, os bancos de dados e outros tipos de fontes e destinos de dados com os quais uma versão de produto é compatível. Se você for membro da Rede da Informatica, poderá acessar PAMs 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. Desenvolvido com base na experiência no mundo real de centenas de projetos de gerenciamento de dados, o Informatica Velocity representa o conhecimento coletivo de nossos consultores, que trabalharam com organizações de todo o mundo para planejar, desenvolver, implantar e manter soluções de gerenciamento de dados bem-sucedidas.

Se você for membro da Rede da Informatica, poderá acessar 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 aumentam, ampliam ou aprimoram suas implementações da Informatica. Aproveitando qualquer uma das centenas de soluções fornecidas por desenvolvedores e parceiros da Informatica, você pode melhorar sua produtividade e agilizar o tempo de implementação nos seus projetos. Você pode acessar o Informatica Marketplace através do link <https://marketplace.informatica.com>.

Suporte global a clientes Informatica

Você pode entrar em contato com um Centro de Suporte Global por telefone ou via Suporte Online na 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:

<http://www.informatica.com/us/services-and-training/support-services/global-support-centers>.

Se você for membro da Rede da Informatica, poderá usar o Suporte Online em <http://network.informatica.com>.

CAPÍTULO 1

Visão Geral da Introdução

Este capítulo inclui os seguintes tópicos:

- [Visão Geral do Domínio Informatica, 9](#)
- [Introdução ao Informatica Analyst, 11](#)
- [Visão Geral do Informatica Developer, 12](#)
- [A História do Tutorial, 14](#)
- [A Estrutura do Tutorial, 15](#)

Visão Geral do Domínio Informatica

O Informatica possui uma arquitetura orientada a serviços que oferece a capacidade de escalar serviços e compartilhar recursos entre várias máquinas. O domínio Informatica é a unidade principal para gerenciamento e administração de serviços.

Você pode fazer login no Informatica Administrator depois que você instalar o Informatica. Use a ferramenta Administrator para gerenciar o domínio e configurar os serviços de aplicativo antes de você poder acessar o restante dos clientes de aplicativo.

O domínio Informatica contém os seguintes componentes:

- Clientes de Aplicativo Um grupo de clientes que você usa para acessar a funcionalidade Informatica subjacente. Clientes de aplicativo fazem solicitações para o Gerenciador de Serviços ou serviços de aplicativo.
- Serviços de aplicativo. Grupo de serviços que representam a funcionalidade baseada em servidor do representante. Um domínio Informatica pode conter um subconjunto de serviços de aplicativo. Você cria e configura os serviços de aplicativo que os clientes de aplicativo exigem.

Os serviços de aplicativo incluem serviços do sistema que podem ter uma única instância no domínio. Quando você cria o domínio, os serviços de sistema são criados para você. Você pode configurar e ativar um serviço do sistema para usar a funcionalidade que o serviço oferece.
- Depósito de perfil. Um banco de dados relacional que o Serviço de Integração de Dados usa para armazenar os resultados do perfil.
- Data warehouse de referência. Um banco de dados relacional que armazena valores de dados de referência para os objetos da tabela de referência no repositório do modelo.
- Repositórios. Um grupo de bancos de dados relacionais que armazena metadados sobre objetos e processos necessários para gerenciar as solicitações do usuário de clientes de aplicativo.

- Gerenciador de Serviços. Um serviço que é criado no domínio para gerenciar todas as operações de domínio. O Gerenciador de Serviços executa os serviços de aplicativo e executa funções de domínio, incluindo autenticação, autorização e log.
- Banco de dados do fluxo de trabalho. Um banco de dados relacional que armazena metadados em tempo de execução para fluxos de trabalho.

A tabela a seguir lista os clientes de aplicativo, incluindo a ferramenta Administrator, e os serviços de aplicativo e os repositórios que o cliente exige:

Cliente de Aplicativo	Serviços de Aplicativo	Repositórios
Informatica Analyst	<ul style="list-style-type: none"> - Serviço Analyst - Serviço do Gerenciamento de Conteúdo - Serviço de Integração de Dados - Serviço de Repositório do Modelo - Serviço de Pesquisa 	Repositório do modelo
Informatica Developer	<ul style="list-style-type: none"> - Serviço Analyst - Serviço do Gerenciamento de Conteúdo - Serviço de Integração de Dados - Serviço de Repositório do Modelo 	Repositório do modelo
Metadata Manager	<ul style="list-style-type: none"> - Serviço do Metadata Manager - Serviço de Integração do PowerCenter - Serviço do Repositório do PowerCenter 	<ul style="list-style-type: none"> - Repositório do Metadata Manager - Repositório do PowerCenter
Cliente do PowerCenter®	<ul style="list-style-type: none"> - Serviço de Integração do PowerCenter - Serviço do Repositório do PowerCenter 	Repositório do PowerCenter
Console do Hub de Serviços da Web	<ul style="list-style-type: none"> - Serviço de Integração do PowerCenter - Serviço do Repositório do PowerCenter - Hub de Serviços da Web 	Repositório do PowerCenter

Os seguintes serviços de aplicativo não são acessados por um cliente do aplicativo Informatica:

- Serviço do Ouvinte do PowerExchange®. Gerencia o Ouvinte do PowerExchange para a movimentação de dados em massa e a captura de dados de alterações. O Serviço de Integração do PowerCenter se conecta ao Ouvinte do PowerExchange por meio do Serviço do Ouvinte.
- Serviço do Agente de Log do PowerExchange. Gerencia o Agente de Log do PowerExchange para Linux, UNIX e Windows, para capturar dados de alteração e gravá-los nos arquivos de Log do Agente de Log do PowerExchange. Os dados de alteração podem se originar de logs de recuperação DB2, Oracle logs de redo, um banco de dados de distribuição do Microsoft SQL Server ou fontes de dados em um sistema i5/OS ou z/OS.
- Serviço SAP BW. Escuta solicitações RFC do SAP BI e solicita que o Serviço de Integração do PowerCenter execute fluxos de trabalho para extrair ou carregar no SAP BI.

Disponibilidade de Recursos

Os produtos Informatica usam um conjunto comum de aplicativos. Os recursos do produto que você pode usar dependem da sua licença de produto.

A tabela a seguir descreve as opções de licenciamento e os recursos de aplicativo disponíveis com cada opção:

Opção de Licenciamento	Recursos do Informatica Developer	Recursos do Informatica Analyst
Data Quality	<ul style="list-style-type: none">- Criar e executar mapeamentos- Criação e execução de regras.- Criação de perfil que inclui o uso do perfil de descoberta empresarial e a descoberta de chave primária, de chave externa e de dependência funcional- Administrar resultados de perfil inferidos- Execução de Scorecard- Exportação de objetos para o PowerCenter.	<ul style="list-style-type: none">- Criação de perfil, incluindo descoberta empresarial- Usar a pesquisa de descoberta para descobrir onde existem dados e metadados nos repositórios de criação de perfil- Execução de Scorecard- Administrar resultados de perfil inferidos- Criar e executar regras de criação de perfil- Gerenciamento de tabela de referência- Criar especificações de regra e compilar especificações de regra em maplets- Gerenciamento de registro duplicado ou inválido
Data Services	<ul style="list-style-type: none">- Criação de modelos de objeto de dados lógicos.- Criação e execução de mapeamentos com transformações de Data Services.- Criação de serviços de dados SQL.- Criação de serviços da Web- Exportação de objetos para o PowerCenter.	<ul style="list-style-type: none">- Gerenciamento de tabela de referência
Opção de Data Services e de Criação de Perfil	<ul style="list-style-type: none">- Criação de modelos de objeto de dados lógicos.- Criação e execução de mapeamentos com transformações de Data Services.- Criação de serviços de dados SQL.- Criação de serviços da Web- Exportação de objetos para o PowerCenter.- Criação e execução de regras com transformações de Serviço de Dados.- Criação de Perfil	<ul style="list-style-type: none">- Gerenciamento de tabela de referência

Introdução ao Informatica Analyst

O Informatica Analyst é um novo aplicativo baseado na Web que os analistas podem usar para analisar, limpar, padronizar, criar perfil e pontuar dados em uma empresa.

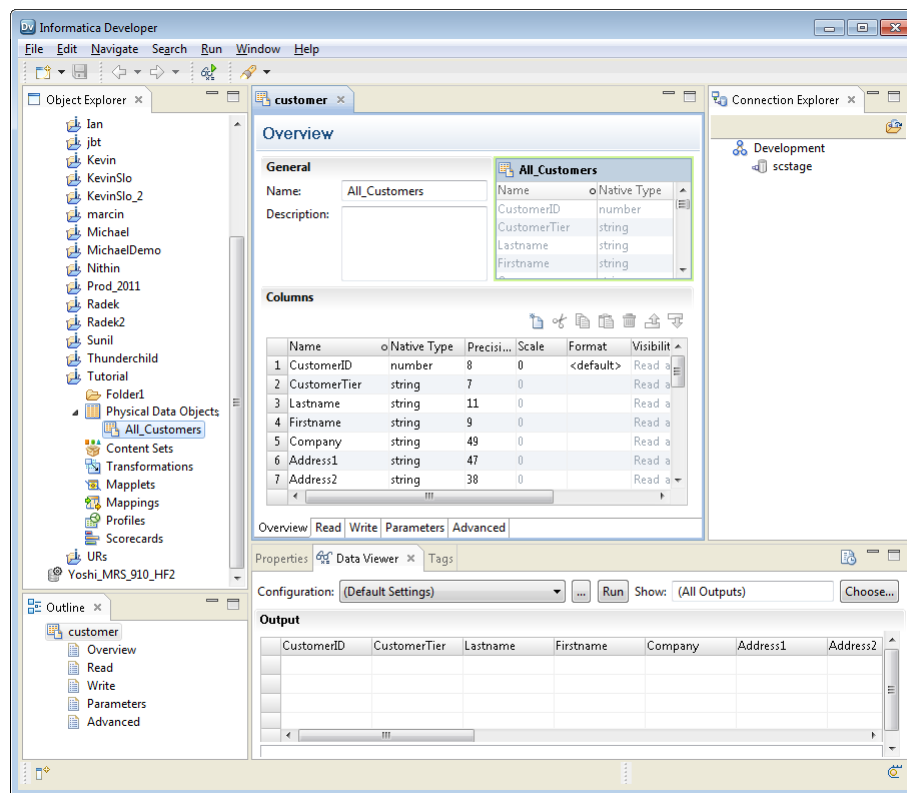
Dependendo da licença, analistas de negócios e desenvolvedores usam a ferramenta Analyst para colaboração orientada por dados. Você pode executar a criação de perfil de regra e dados de coluna, a

execução de scorecard e o gerenciamento de registros duplicados e inválidos. Também é possível gerenciar dados de referência e fornecer os dados aos desenvolvedores em uma solução de qualidade de dados.

Visão Geral do Informatica Developer

O Informatica Developer é um aplicativo cliente que os desenvolvedores usam para criar e implementar soluções de dados.

A figura a seguir mostra a ferramenta Developer:



A ferramenta Developer pode mostrar as seguintes exibições:

Exibição Object Explorer

Exibe projetos e pastas, e os objetos dentro desses projetos e pastas.

Exibição Explorador de Conexões

Mostra conexões com bancos de dados relacionais.

Exibição Estrutura de Tópicos

Mostra objetos que dependem de um objeto selecionado na exibição **Object Explorer**.

Exibição Folhas de Cola

Mostra a folha de cola que você abrir. Para abrir uma folha de cola, clique em **Ajuda > Folhas de Cola** e selecione uma folha de cola.

Exibição Visualizador de Dados

Exibe os dados de origem, os resultados de perfil e uma visualização do resultado de uma transformação.

Exibição Dependências de Objetos

Exibe dependências de objetos quando você exibe, modifica ou exclui um objeto.

Exibição Alertas

Exibe alertas de status de conexão.

Exibição Log de Validação

Exibe erros de validação de objetos.

A ferramenta Desenvolvedor também pode exibir outras exibições. Você pode ocultar exibições e movê-las para outra localização no workbench da ferramenta Developer. Clique em **Janela > Mostrar Exibição** para selecionar as exibições que você deseja exibir.

Página de Boas-vindas do Informatica Developer

Quando você abre a ferramenta Developer pela primeira vez, a página de Boas-vindas é exibida. Use a página de Boas-vindas para saber mais sobre a ferramenta Developer, configurá-la e começar a trabalhar com ela.

A página de Boas-vindas exibe as seguintes opções:

- Visão geral. Clique no botão Visão Geral para obter uma visão geral das soluções de qualidade de dados e serviços de dados.
- Primeiras Etapas. Clique no botão Primeiras Etapas para saber mais sobre como configurar a ferramenta Developer e acessar as lições sobre o Informatica Data Quality e o Informatica Data Services.
- Tutoriais. Clique no botão Tutoriais para ver lições de tutorial para soluções de qualidade de dados e serviços de dados.
- Recursos da Web. Clique no botão Recursos da Web para acessar um link até mysupport.informatica.com, onde você pode acessar a Biblioteca de Recursos da Informatica. A Biblioteca de Recursos da Informatica contém artigos sobre o Informatica Data Quality, o Informatica Data Services e outros produtos da Informatica.
- Novidades. Clique no botão Novidades para exibir os recursos mais recentes na Developer tool.

Folhas de Cola

A ferramenta Developer inclui folhas de cola como parte da ajuda online. Uma folha de cola é um guia passo a passo que ajuda você a concluir uma ou mais tarefas na ferramenta Developer.

Depois de preencher uma folha de cola, você concluirá as tarefas e verá os resultados. Por exemplo, depois de concluir uma folha de cola para importar e visualizar um objeto de dados relacionais, você importou uma tabela de banco de dados relacional e visualizou os dados na ferramenta Developer.

Para acessar folhas de cola, clique em **Ajuda > Folhas de Cola**.

Qualidade de Dados e Criação de Perfil

Use os recursos de qualidade de dados na ferramenta Developer para analisar o conteúdo e a estrutura dos seus dados. Você pode aprimorar os dados de forma a atender às suas necessidades de negócios.

Use a ferramenta Developer para projetar e executar processos que concretizem 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 ferramenta Desenvolvedor.
- Colabore com os usuários da Informatica. As tabelas de dados de referência e regras que você adiciona ao repositório do Modelo estão disponíveis para usuários na ferramenta Desenvolvedor 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.

A História do Tutorial

A HypoStores Corporation é uma organização de varejo nacional com sede em Boston e lojas em vários estados. Ela integra regularmente os dados operacionais das lojas de todo o país com o repositório de dados na sede. Recentemente, a empresa abriu uma loja em Los Angeles.

A sede inclui uma equipe ICC central de administradores, desenvolvedores e arquitetos responsáveis por fornecer uma camada comum de serviços de dados para todos os aplicativos de BI e compostos. Os

aplicativos de BI incluem um sistema de CRM que contém os arquivos mestres de dados de clientes usados para cobrança e marketing.

A HypoStores Corporation precisa realizar as seguintes tarefas para integrar dados da operação de Los Angeles com os dados na sede em Boston:

- Examinar os dados de Boston e Los Angeles em relação a problemas de qualidade de dados.
- Analisar as informações dos dados de Los Angeles.
- Padronizar as informações de endereço entre os dados de Boston e Los Angeles.
- Validar a precisão das informações de endereço postal nos dados para fins de CRM.

A Estrutura do Tutorial

O *Guia de Introdução* contém tutoriais que incluem lições e tarefas.

Lições

Cada lição apresenta conceitos que o ajudam a compreender as tarefas a serem feitas na lição. A lição fornece requisitos comerciais do histórico geral. Os objetivos da lição descrevem as tarefas que você executará para atender aos requisitos comerciais. Cada lição fornece um tempo estimado de conclusão. Quando concluir as tarefas na lição, você poderá rever o resumo da lição.

Se o ambiente na ferramenta não estiver configurado, a primeira lição em cada tutorial ajuda você a fazê-lo.

Tarefas

As tarefas fornecem instruções detalhadas. Execute as tarefas na ordem listada para concluir a lição.

Tutorial do Informatica Analyst

Nesse tutorial, um analista se conecta na ferramenta Analyst, cria projetos e pastas, cria perfis e regras, pontua dados e cria tabelas de referência.

As lições que você pode realizar dependem de você ter os produtos Informatica Data Quality ou Informatica Data Services.

A tabela a seguir descreve as lições que você pode executar, dependendo do seu produto:

Lição	Descrição	Produto
Lição 1. Configurando o Informatica Analyst	Fazer login na ferramenta Analyst e criar um projeto e uma pasta para as lições do tutorial.	Data Quality Data Services
Lição 2. Criando Objetos de Dados	Importar um arquivo simples como um objeto de dados e visualizar dados.	Data Quality
Lição 3. Criando Perfis Rápidos	Criar um perfil rápido para obter rapidamente um ideia da qualidade de dados.	Data Quality
Lição 4. Criando Perfis Personalizados	Criar um perfil personalizado para configurar as colunas e as opções de busca detalhada e amostragem.	Data Quality
Lição 5. Criando Regras de Expressão	Criar regras de expressão para modificar e criar perfil dos valores de coluna.	Data Quality

Lição	Descrição	Produto
Lição 6. Criando e Executando Scorecards	Criar e executar um scorecard para medir o progresso da qualidade de dados ao longo do tempo.	Data Quality
Lição 7. Criar Tabelas de Referência de Resultados de Perfil	Criar uma tabela de referência que você possa usar para padronizar dados de origem.	Data Quality Data Services
Lição 8. Criando Tabelas de Referência	Criar uma tabela de referência para estabelecer relações entre dados de origem e valores válidos e padrão.	Data Quality Data Services

Informatica Developer tool

Neste tutorial, use a ferramenta Developer para realizar várias operações de qualidade de dados.

Os usuários do Informatica Data Quality usam a Developer tool para criar e executar processos que aprimoram a qualidade de dados. Os usuários do Informatica Data Quality também usam a Developer tool para criar e executar perfis que analisam o conteúdo e a estrutura dos dados.

Conclua as seguintes lições no tutorial de qualidade de dados:

Lição 1. Configurando o Informatica Developer

Crie uma conexão a um repositório do Modelo que é gerenciado por um Serviço de Repositório do Modelo em um domínio. Crie um projeto e pasta para armazenar o trabalho para as lições no tutorial. Selecionar um Serviço de Integração de Dados padrão.

Lição 2. Importando Objetos de Dados Físicos

Você não definir processos de qualidade de dados para os arquivos de dados do cliente associados a esses objetos.

Lição 3. Dados de criação de perfil

A criação de perfil de dados revela o conteúdo e a estrutura de seus dados.

A criação de perfil inclui análise de associação, um formulário de análise que determina se uma associação válida é possível entre duas colunas de dados.

Lição 4. Analisando dados

A análise aprimora os registros de dados e melhora a estrutura de registro. Também pode encontrar informações úteis em seus dados e gerar novas informações de dados atuais.

Lição 5. Padronizar Dados

A padronização remove erros de dados e inconsistências encontrados durante a criação de perfil.

Lição 6. Validar Dados de Endereço

A validação de endereço avalia a precisão e a capacidade de entrega de endereços postais e corrige erros e omissões de endereços.

Pré-requisitos do Tutorial

Antes de começar as lições do tutorial, o domínio Informatica deve ser executado com pelo menos um nó configurado.

O instalador incluirá arquivos do tutorial que você usará para concluir as lições. Todos os arquivos estão localizados em ambas as instalações do cliente e do servidor:

- Os arquivos do tutorial estão disponíveis na seguinte localização do caminho de instalação da ferramenta Developer:

`<Diretório de Instalação da Informatica>\clients\DeveloperClient\Tutorials`

- Os arquivos do tutorial estão disponíveis na seguinte localização do caminho de instalação de serviços:

`<Diretório de Instalação da Informatica>\server\Tutorials`

Você precisa dos seguintes arquivos para as lições do tutorial:

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

Parte I: Introdução ao Informatica Analyst

Esta parte contém os seguintes capítulos:

- [Lição 1. Configurar o Informatica Analyst, 19](#)
- [Lição 2. Criar Objetos de Dados, 22](#)
- [Lição 3. Criando Perfis Padrão, 25](#)
- [Lição 4. Criar Perfis Personalizados, 28](#)
- [Lição 5. Criar Regras de Expressão, 32](#)
- [Lição 6. Criar e Executar Scorecards, 35](#)
- [Lição 7. Criar Tabelas de Referência de Colunas de Perfil, 40](#)
- [Lição 8. Criar Tabelas de Referência, 43](#)

CAPÍTULO 2

Lição 1. Configurar o Informatica Analyst

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Configuração do Informatica Analyst, 19](#)
- [Tarefa 1. Fazer login no Informatica Analyst, 20](#)
- [Tarefa 2. Criar um Projeto, 20](#)
- [Tarefa 3. Criar uma Pasta, 21](#)
- [Resumo da Configuração do Informatica Analyst, 21](#)

Visão Geral da Configuração do Informatica Analyst

Antes de iniciar as lições deste tutorial, configure a ferramenta Analyst. Para configurar a ferramenta Analyst, faça login na ferramenta Analyst e crie um projeto e uma pasta para armazenar seu trabalho.

O domínio Informatica é um conjunto de nós e serviços que definem o ambiente Informatica. Os serviços no domínio incluem o Serviço Analyst e o Serviço de Repositório do Modelo. O Serviço Analyst executa a ferramenta Analyst, e o Serviço de Repositório do Modelo gerencia o repositório do Modelo. Quando você trabalha na ferramenta Analyst, esta armazena os ativos criados no repositório do Modelo.

Para poder criar ativos na ferramenta Analyst, você precisa criar um projeto. Um projeto contém ativos na ferramenta Analyst. Um projeto também pode conter pastas que armazenam ativos relacionados, como objetos de dados que fazem parte do mesmo requisito de negócios.

Objetivos

Nesta lição, você concluirá as seguintes tarefas:

- Faça login na ferramenta Analyst.
- Crie um projeto para armazenar os objetos criados na ferramenta Analyst.
- Crie uma pasta no projeto que possa armazenar ativos relacionados.

Pré-requisitos

Antes de iniciar esta lição, verifique os seguintes pré-requisitos:

- Um administrador foi configurado no Serviço de Repositório do Modelo e um Serviço Analyst na ferramenta Administrator.

- Você tem o nome de host e número de porta para a ferramenta Analyst.
- Também tem um nome de usuário e senha para acessar o Serviço Analyst. Obtenha essas informações de um administrador.

Tempo

Defina 5 a 10 minutos para concluir nesta lição.

Tarefa 1. Fazer login no Informatica Analyst

Faça login na ferramenta Analyst para iniciar o tutorial.

1. Inicie o navegador Microsoft Internet Explorer ou Google Chrome.
2. No campo Endereço, insira a URL do Informatica Analyst:
`http[s]://<fully qualified host name>:<port number>/analyst`
3. Se o domínio usa a autenticação LDAP ou nativo, insira seu nome de usuário e senha na página de login.
4. Selecione **Nativo** ou o nome de um domínio de segurança específico.
 O campo Domínio de Segurança é exibido quando o domínio Informatica usa autenticação LDAP ou Kerberos. Se você não souber a que domínio de segurança sua conta de usuário pertence, entre em contato com o administrador do domínio Informatica.
5. Clique em **Entrar**.
 A ferramenta Analyst é exibida no espaço de trabalho **Iniciar**.

Tarefa 2. Criar um Projeto

Nesta tarefa, você criará um projeto para conter os ativos criados na ferramenta Analyst. Crie um projeto de tutorial para conter a pasta do projeto.

1. No cabeçalho **Gerenciar**, clique em **Projetos**.
 O espaço de trabalho **Projetos** é exibido.
2. No menu **Ações**, clique em **Novo > Projeto**.
 A janela **Novo Projeto** é exibida.
3. Insira seu nome usando "Tutorial_" como prefixo para o nome do projeto.
4. Clique em **OK**.

Tarefa 3. Criar uma Pasta

Nesta tarefa, você criará uma pasta para armazenar ativos relacionados. Você pode criar uma pasta em um projeto ou outra pasta. Crie uma pasta denominada Customers para armazenar os ativos relacionados ao projeto de qualidade de dados.

1. No painel **Projetos**, selecione o projeto de tutorial.
2. No menu **Ações**, clique em **Nova > Pasta**.
A janela **Nova Pasta** é exibida.
3. Insira Customers como o nome da pasta.
4. Clique em **OK**.
A pasta do tutorial é exibida no projeto de tutorial.

Resumo da Configuração do Informatica Analyst

Nesta lição, você aprendeu que a ferramenta Analyst armazena ativos em projetos e pastas. Um repositório do Modelo contém os projetos e pastas. O Serviço Analyst gerencia a ferramenta Analyst. O Serviço de Repositório do Modelo gerencia o repositório do Modelo. O Serviço Analyst e o Serviço de Repositório do Modelo são serviços de aplicativo no domínio Informatica.

Você se conectou à ferramenta Analyst e criou um projeto e uma pasta.

Agora, você poderá usar a ferramenta Analyst para concluir outras lições deste tutorial.

CAPÍTULO 3

Lição 2. Criar Objetos de Dados

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Criação de Objetos de Dados, 22](#)
- [Tarefa 1. Criar os Objetos de Dados de Arquivo Simples, 23](#)
- [Tarefa 2. Exibir as Propriedades de Objeto de Dados, 23](#)
- [Resumo da Criação de Objetos de Dados, 24](#)

Visão Geral da Criação de Objetos de Dados

Na ferramenta Analyst, um objeto de dados é uma representação de dados com base em um arquivo simples ou tabela de banco de dados relacional. Você cria um arquivo simples ou objeto de tabela e executa um perfil em relação aos dados no arquivo simples ou tabela de banco de dados relacional. Quando criar um objeto de dados de arquivo simples na ferramenta Analyst, você pode fazer upload do arquivo para o cache do arquivo simples no computador que executa a ferramenta Analyst ou pode especificar a localização da rede onde o arquivo simples está armazenado.

História

A HypoStores mantém os dados de clientes de Los Angeles em arquivos simples. A HypoStores precisa criar um perfil, analisar os dados e executar tarefas de qualidade de dados.

Objetivos

Nesta lição, você concluirá as seguintes tarefas:

1. Fazer o upload do arquivo simples para a localização de cache do arquivo simples e criar um objeto de dados.
2. Visualizar os dados para o objeto de dados de arquivo simples.

Pré-requisitos

Antes de iniciar esta lição, verifique os seguintes pré-requisitos:

- Você concluiu a lição 1 deste tutorial.
- Você tem o arquivo simples LA_Customers.csv. Localize esse arquivo na pasta <Diretório Raiz de Instalação>\<Versão de Lançamento>\clients\DeveloperClient\Tutorials.

Tempo

Reserve de 5 a 10 minutos para concluir esta tarefa.

Tarefa 1. Criar os Objetos de Dados de Arquivo Simples

Nessa tarefa, um objeto de dados de arquivo simples do arquivo LA_Customers será criado.

1. Na ferramenta Analyst, clique em **Novo > Objeto de Dados de Arquivo Simples**.
O assistente para **Adicionar Arquivo Simples** é exibido.
2. Selecione **Procurar e Fazer Upload** e clique em **Procurar**.
3. Vá para a localização de LA_Customers.csv e clique em **Abrir**.
4. Clique em **Avançar**.
O painel **Escolher tipo de importação** exibe as opções **Delimitado** e **Delimitado**. A opção padrão é **Delimitado**.
5. Clique em **Avançar**.
6. Em **Especifique os delimitadores e qualificadores de texto usados nos seus dados**, selecione **Aspas duplas** como qualificador de texto.
7. Em **Especifique as linhas para importação**, selecione **Importe a partir da primeira linha** para importar nomes de coluna da primeira linha que não estiver em branco.
O painel **Visualizar** é atualizado para mostrar os títulos de colunas da primeira linha.
8. Clique em **Avançar**.
O painel **Atributo de Coluna** mostra o tipo de dados, precisão, escala e formato para cada coluna.
9. Clique em **Avançar**.
O campo **Nome** exibe LA_Customers.
10. Selecione o projeto **Tutorial_** e a pasta **Customers**.
11. Clique em **Concluir**.
O objeto de dados é exibido no conteúdo da pasta Clientes.

Tarefa 2. Exibir as Propriedades de Objeto de Dados

Nesta tarefa, você pode exibir as propriedades do objeto de dados LA_Customers.

1. Clique em **Abrir** para abrir o espaço de trabalho da **Biblioteca**.
2. No espaço de trabalho **Biblioteca**, clique em **Objetos de Dados** no painel **Ativos**.
Uma lista de objetos de dados aparece no painel **Objetos de Dados**.
3. Clique no arquivo simples **LA_Customers**.
O painel **Visualização de Dados** é exibido com os dados recuperados do objeto de dados LA_Customers.
4. No painel **Visualização de Dados**, consulte a estrutura e o conteúdo do objeto de dados LA_Customers.
A ferramenta Analyst exibe as primeiras 100 linhas do objeto de dados de arquivo simples.
5. Clique em **Propriedades**.
O painel **Propriedades** mostra o nome, a descrição e a localização do objeto de dados. Você também pode ver os nomes de coluna e as propriedades de coluna do objeto de dados.

Resumo da Criação de Objetos de Dados

Nessa lição, você aprendeu que os objetos de dados são representações de dados com base em um arquivo simples ou uma origem de banco de dados relacional. Também aprendeu que pode criar um objeto de dados de arquivo simples e visualizar os dados.

Ainda carregou um arquivo simples e criou um objeto de dados de arquivo simples, visualizou os dados do objeto de dados e exibiu as propriedades do objeto de dados.

Depois de criar um objeto de dados, crie um perfil padrão para o objeto de dados na Lição 3 e crie um perfil personalizado para o objeto de dados na Lição 4.

CAPÍTULO 4

Lição 3. Criando Perfis Padrão

Este capítulo inclui os seguintes tópicos:

- [Criando Perfis Padrão - Visão Geral, 25](#)
- [Tarefa 1. Criar e Executar um Perfil Padrão, 26](#)
- [Tarefa 2. Visualizar os resultados do perfil na exibição resumida, 26](#)
- [Criando Perfis Padrão - Resumo, 27](#)

Criando Perfis Padrão - Visão Geral

Um perfil é a análise de qualidade de dados com base no conteúdo e na estrutura de dados. Um perfil padrão é um perfil que você cria com opções padrão. Use um perfil padrão para obter resultados de perfil sem configurar todas as colunas e opções para um perfil.

Crie e execute um perfil padrão para analisar a qualidade de dados quando você inicia um projeto de qualidade de dados. Ao criar um objeto de perfil padrão, selecione o objeto de dados e a coluna de objeto de dados que você deseja analisar. Um perfil padrão ignora a coluna de perfil e a configuração de opção. A ferramenta Analyst realiza a criação de perfil no arquivo simples dinâmico para o objeto de dados de arquivo simples.

História

A HypoStores deseja incorporar os dados do recém adquirido escritório de Los Angeles ao data warehouse. Antes que os dados possam ser incorporados ao data warehouse, eles precisam ser limpos. Você é o analista responsável pela avaliação da qualidade de dados e por passar as informações para o desenvolvedor responsável pela limpeza de dados. Você deseja exibir os resultados do perfil de maneira rápida e obter uma ideia básica da qualidade de dados.

Objetivos

Nesta lição, você concluirá as seguintes tarefas:

1. Criar e executar um perfil padrão para o objeto de dados de arquivo simples LA_Customers.
2. Exibir os resultados do perfil.

Pré-requisitos

Antes de iniciar esta lição, verifique o seguinte pré-requisito:

- Você concluiu as lições 1 e 2 neste tutorial.

Tempo

Defina 5 a 10 minutos para concluir nesta lição.

Tarefa 1. Criar e Executar um Perfil Padrão

Nessa tarefa, você criará um perfil rápido para todas as colunas no objeto de dados e usará opções padrão de amostragem e busca detalhada.

1. No espaço de trabalho **Biblioteca**, selecione um objeto de dados no painel **Ativos**.
2. Clique com o botão direito do mouse no objeto de dados e selecione **Criar Perfil**.
O assistente de **Novo Perfil** é exibido.
3. A opção **Origem única** é selecionada por padrão. Clique em **Avançar**.
4. Na tela **Especificar Propriedades Gerais**, o nome, a descrição e a localização são preenchidos por padrão.
5. Clique em **Avançar**.
6. Na tela **Selecionar Origem**, o objeto de dados é exibido por padrão. Você pode exibir as colunas no painel **Escolher Colunas**.
7. Clique em **Avançar**.
8. Na tela **Especificar Configurações**, as seguintes opções estão selecionadas por padrão:
 - **Executar perfil de coluna**
 - **Todas as linhas** no painel **Executar perfil em**
 - **Dinâmico** no painel **Busca Detalhada**
 - **Excluir tipos de dados e domínios de dados aprovados da inferência de tipo de dados e de domínio de dados nas execuções de perfil subsequentes.**
 - Conexão **Nativa** no painel **Ambiente de tempo de execução**.
9. Clique em **Avançar**.
10. Na tela **Especificar Regras e Filtros**, clique em **Salvar e Executar** para criar e executar o perfil.
A ferramenta Analyst cria e executa o perfil. Os resultados do perfil aparecem na exibição resumida.

Tarefa 2. Visualizar os resultados do perfil na exibição resumida

Nessa tarefa, a exibição resumida é usada para o perfil LA_Customers para obter uma rápida visão geral dos resultados do perfil.

1. No painel **Biblioteca > Ativos > Perfis**, clique no perfil **LA_Customers**.
Os resultados do perfil aparecem na exibição resumida.
2. Na exibição resumida, clique em **Colunas** no painel **Filtrar por** para visualizar os resultados do perfil para colunas.
3. Mova o ponteiro sobre os gráficos de barras horizontais para ver os valores em porcentagens.
4. Na seção **% De Nulos, Distintos e Não Distintos**, você pode exibir os valores nulos, os valores distintos e os valores não distintos em porcentagens para uma coluna.
5. Na seção **Padrão**, é possível exibir vários padrões na coluna como gráficos de barras horizontais. Você pode exibir os caracteres padrão e o número de padrões semelhantes em uma coluna como porcentagem ao mover o ponteiro sobre o gráfico de barras.

6. Na seção **Tamanho**, você pode exibir os comprimentos mínimo e máximo dos valores na coluna.
7. Na seção **Valor**, você pode exibir os valores mínimo e máximo em uma coluna.
8. Na seção **Tipo de Dados**, você pode exibir todos os tipos de dados inferidos e documentados para uma coluna ao mover o ponteiro sobre os valores.
9. Na seção **Domínio de Dados**, você pode exibir todos os domínios de dados inferidos para uma coluna ao mover o ponteiro sobre os valores.
10. Para exibir os dados de exceção, clique em **Ações > Detectar Exceção** para detectar exceções nos resultados do perfil.
11. Clique nos filtros **Atípico padrão** ou **Atípico de frequência de valor** para visualizar os valores atípicos nos resultados do perfil.
12. Clique em um nome de coluna para ver os resultados de perfil dessa coluna na exibição detalhada.

Criando Perfis Padrão - Resumo

Nessa lição, você aprendeu que um perfil padrão mostra os resultados de perfil sem configurar todas as opções de amostragem de colunas e linha para um perfil. Também aprendeu que você pode criar e executar um perfil padrão para analisar a qualidade de dados quando inicia um projeto de qualidade de dados. Aprendeu, ainda, que a ferramenta Analyst realiza a criação de perfil no arquivo simples dinâmico para o objeto de dados de arquivo simples.

Você criou um perfil padrão e analisou os resultados do perfil. Também obteve mais informações sobre as colunas no perfil, incluindo valores nulos e tipos de dados. Você também usou os padrões e valores de coluna para identificar problemas de qualidade de dados.

Depois de analisar os resultados de um perfil rápido, você pode concluir as seguintes tarefas:

- Criar um perfil personalizado para excluir colunas do perfil e incluir somente as colunas de seu interesse.
- Criar uma regra de expressão para criar colunas virtuais e fazer um perfil delas.
- Criar uma tabela de referência para incluir os valores válidos de uma coluna.

CAPÍTULO 5

Lição 4. Criar Perfis Personalizados

Este capítulo inclui os seguintes tópicos:

- [Criar Visão Geral de Perfis Personalizados, 28](#)
- [Tarefa 1. Criar um Perfil Personalizado, 29](#)
- [Tarefa 2. Executar o Perfil, 30](#)
- [Tarefa 3. Fazer uma Busca Detalhada nos Resultados do Perfil, 30](#)
- [Criando Perfis Personalizados - Resumo, 31](#)

Criar Visão Geral de Perfis Personalizados

Um perfil é a análise de qualidade de dados com base no conteúdo e na estrutura de dados. Um perfil personalizado é um perfil que você cria quando deseja configurar as colunas, opções de amostragem e opções de busca detalhada para obter uma criação de perfil mais rápida. Configure as opções de amostragem para selecionar as linhas de amostra na origem. Configure as opções de busca detalhada para fazer uma busca detalhada nas linhas de dados na fonte de dados ou nos dados de fase. Você pode escolher executar o perfil em um ambiente nativo ou Hive.

Crie e execute um perfil para analisar a qualidade de dados quando você iniciar um projeto de qualidade de dados. Quando você cria um objeto de perfil, comece selecionando o objeto de dados e as colunas de objeto de dados nas quais deseja executar um perfil.

História

A HypoStores precisa incorporar os dados do recém adquirido escritório de Los Angeles ao data warehouse. A HypoStores deseja acessar a qualidade dos dados de camada do cliente no arquivo de dados de cliente de LA. Você é o analista responsável pela avaliação da qualidade de dados e por transmitir as informações para o desenvolvedor responsável pela limpeza de dados.

Objetivos

Nesta lição, você concluirá as seguintes tarefas:

1. Criar um perfil personalizado para o objeto de dados de arquivo simples e excluir as colunas com valores nulos.
2. Executar o perfil para analisar o conteúdo e a estrutura da coluna CustomerTier.
3. Fazer uma busca detalhada nas linhas dos resultados do perfil.

Pré-requisitos

Antes de iniciar esta lição, verifique o seguinte pré-requisito:

- Você concluiu as lições 1, 2, 3 neste tutorial.

Tempo

Defina 5 a 10 minutos para concluir nesta lição.

Tarefa 1. Criar um Perfil Personalizado

Nessa tarefa, você cria um perfil personalizado. Ao criar um perfil personalizado, você seleciona o objeto de dados e as colunas nas quais deseja executar um perfil. Você também pode configurar a amostragem e as opções de busca detalhada.

1. Clique em **Novo > Perfil**.
O assistente de **Novo Perfil** é exibido.
2. A opção **Origem única** é selecionada por padrão. Clique em **Avançar**.
3. Na tela **Especificar Propriedades Gerais**, defina as seguintes opções:
 - No campo Nome, insira **Profile_LA_Customers**.
 - No campo Localização, selecione a pasta **Clientes**.
4. Clique em **Avançar**.
5. Na tela **Selecionar Origem**, clique em **Escolher**.
A caixa de diálogo **Escolher Objeto de Dados** é exibida.
6. Na caixa de diálogo **Escolher Objeto de Dados**, selecione **LA_Customers**. Clique em **OK**.
7. Na tela **Selecionar Origem**, desmarque as colunas Address2, Address3 e City2.
8. Clique em **Avançar**.
9. Na tela **Especificar Configurações**, defina as seguintes opções:
 - **Executar perfil de coluna**.
 - Selecione a opção **Amostra aleatória** no painel **Executar perfil em**.
 - Selecione a opção **Excluir tipos de dados e domínios de dados aprovados da inferência de tipo de dados e de domínio de dados nas execuções de perfil subsequentes**.
 - Selecione a opção **Preparado** no painel **Busca Detalhada**.
10. Clique em **Avançar**.
11. Na tela **Especificar Regras e Filtros**, clique em **Salvar e Concluir** para criar e executar o perfil.
A ferramenta Analyst cria o perfil e o exibe no espaço de trabalho **Descoberta**. Você precisa executar o perfil para exibir os resultados.

Tarefa 2. Executar o Perfil

Nessa tarefa, você executará um perfil para executar a criação de perfil no objeto de dados e exibir os resultados do perfil. A ferramenta Analyst executa a criação de perfil no arquivo simples preparado para o objeto de dados de arquivo simples.

1. Verifique se você está no espaço de trabalho **Descoberta**.
Você pode ver o perfil **Profile_LA_Customers_Custom** no espaço de trabalho.
2. Clique em **Profile_LA_Customers_Custom** no espaço de trabalho.
3. É exibida a tela de perfil, na qual você pode optar por editar ou executar o perfil. Clique em **Executar**.
4. Os resultados do perfil aparecem na exibição resumida.

Tarefa 3. Fazer uma Busca Detalhada nos Resultados do Perfil

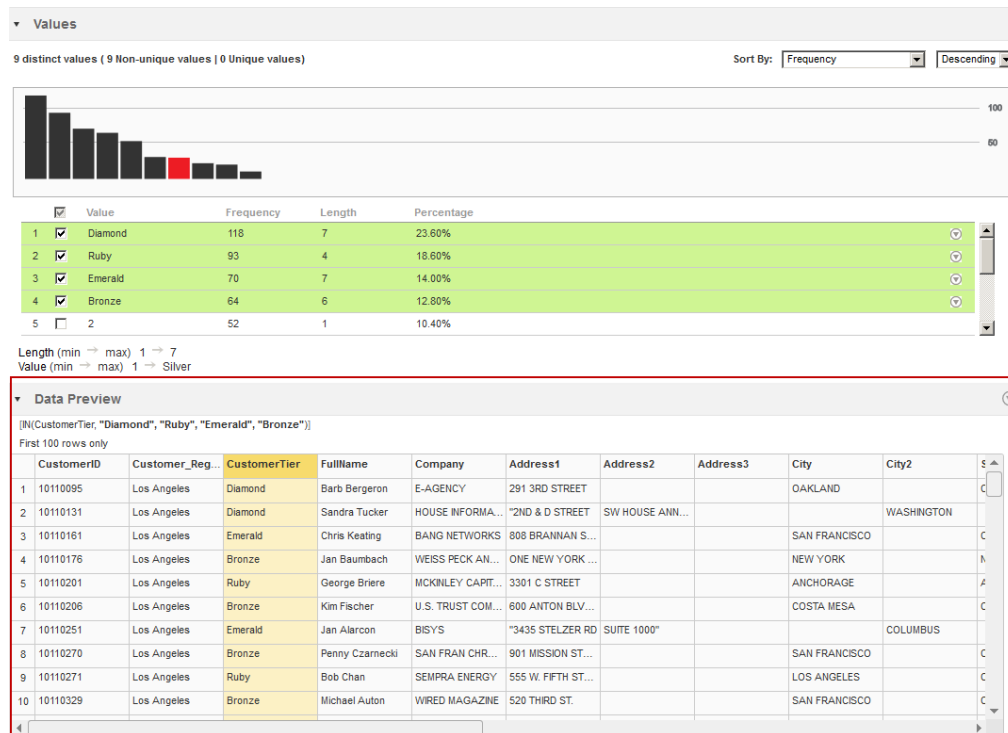
Nessa tarefa, faça uma busca detalhada nos valores de coluna CustomerTier para ver as linhas subjacentes no objeto de dados do perfil.

1. Verifique se você está na exibição resumida dos resultados do perfil **Profile_LA_Customers**.
2. Clique na coluna **CustomerTier**.
Os resultados do perfil para a coluna aparecem na exibição detalhada.

3. Na exibição detalhada, selecione os valores Diamante, Rubi, Esmeralda e Bronze. Clique com o botão direito do mouse nos valores do painel **Valores** e selecione **Busca Detalhada**.

As linhas das colunas com um valor de Diamante, Rubi, Esmeralda e Bronze são exibidas no painel **Visualização de Dados**.

A seguinte imagem mostra os resultados da busca detalhada no painel **Visualização de Dados** quando você faz uma busca detalhada nos valores Diamante, Rubi, Esmeralda ou Bronze:



O painel **Visualização de Dados** mostra as 100 primeiras linhas da coluna selecionada. O título do painel **Visualização de Dados** mostra a lógica usada para a coluna de origem.

Criando Perfis Personalizados - Resumo

Nessa lição, você aprendeu que pode configurar as colunas cujos perfis foram criados e que pode configurar as opções de busca detalhada e amostragem. Também aprendeu que pode fazer uma busca detalhada para ver as linhas subjacentes dos valores de coluna e que pode configurar as colunas que estão incluídas quando os valores de coluna são exibidos.

Você criou um perfil personalizado que incluía a coluna CustomerTier, executou o perfil e fez uma busca detalhada nas linhas subjacentes da coluna CustomerTier nos resultados.

Use o objeto de perfil personalizado para criar uma regra de expressão na lição 5.

CAPÍTULO 6

Lição 5. Criar Regras de Expressão

Este capítulo inclui os seguintes tópicos:

- [Criando Regras de Expressão - Visão Geral, 32](#)
- [Tarefa 1. Criar Regras de Expressão e Executar o Perfil, 33](#)
- [Tarefa 2. Exibir a Saída de Regra de Expressão, 33](#)
- [Tarefa 3. Editar as regras de expressão, 34](#)
- [Criar Resumo de Regras de Expressão, 34](#)

Criando Regras de Expressão - Visão Geral

As regras de expressão usam funções de expressão e colunas de origem para definir a lógica da regra. Você pode criar regras de expressão e adicioná-las a um perfil na ferramenta Analyst. Uma regra de expressão pode ser associada a um ou mais perfis.

A saída de uma regra de expressão é uma coluna virtual no perfil. A ferramenta Analyst faz o perfil da coluna virtual quando você executa o perfil.

Você pode usar regras de expressão para validar as colunas de origem ou criar colunas de origem adicionais com base no valor das colunas de origem.

História

A HypoStores deseja incorporar os dados do recém adquirido escritório de Los Angeles ao data warehouse. A HypoStores deseja analisar os nomes de clientes e separá-los em nome e sobrenome. A HypoStores deseja usar regras de expressão para analisar uma coluna que contém o nome e o sobrenome em colunas virtuais separadas e, em seguida, fazer perfil das colunas. A HypoStores também quer criar regras disponíveis para outros analistas que precisem analisar a saída das regras.

Objetivos

Nesta lição, você concluirá as seguintes tarefas:

1. Criar regras de expressão para separar a coluna FullName em colunas de nome e de sobrenome. Você cria uma regra que separa o primeiro nome do nome completo. E outra regra que separa o sobrenome do primeiro nome. Essas regras são criadas para o perfil Profile_LA_Customers.
2. Executar o perfil e exibir a saída das regras no perfil.

3. Editar as regras para torná-las úteis para outros usuários da ferramenta Analyst.

Pré-requisitos

Antes de iniciar esta lição, verifique o seguinte pré-requisito:

- Você concluiu as Lições de 1 a 4.

Tempo

Reserve de 10 a 15 minutos para concluir esta lição.

Tarefa 1. Criar Regras de Expressão e Executar o Perfil

Nessa tarefa, duas regras de expressão para analisar a coluna FullName em duas colunas virtuais chamadas FirstName e LastName serão criadas. Os nomes de regra são FirstName e LastName.

1. No espaço de trabalho **Biblioteca**, clique no perfil **Profile_LA_Customers**.
Os resultados do perfil aparecem na exibição resumida.
2. Clique em **Editar**.
O assistente de **Perfil** é exibido.
3. Clique na tela **Especificar Regras e Filtros**.
4. No painel **Regras**, clique em **Ações > Criar Regra**.
5. No campo **Nome**, insira **FirstName**.
6. Na seção **Expressão**, insira a seguinte expressão para separar o primeiro nome da coluna **Nome**:

```
SUBSTR(FullName,1, INSTR(FullName, ' ', -1, 1) - 1)
```
7. Clique em **Validar**.
8. Clique em **OK**.
9. Repita as etapas de 4 a 8 para criar uma regra denominada **LastName**. Insira a seguinte expressão para separar o sobrenome da coluna **Nome**:

```
SUBSTR(FullName, INSTR(FullName, ' ', -1, 1), LENGTH(FullName))
```
10. Clique em **Salvar e Executar** para salvar e executar o perfil.

Tarefa 2. Exibir a Saída de Regra de Expressão

Nessa tarefa, você vê a saída de regras de expressão que separa o nome e o sobrenome depois de executar um perfil.

1. Na exibição resumida, clique em **Editar**.
O assistente de perfil é exibido
2. No assistente de perfil, clique em **Selecionar Origem**.
3. Na tela **Selecionar Origem**, marque a caixa de seleção ao lado do **Nome** na barra de ferramentas para limpar todas as colunas.

Uma das colunas é selecionada por padrão, pois você precisa selecionar pelo menos uma coluna na seção **Colunas**.

4. Selecione a coluna **FullName** e as regras **FirstName** e **LastName**.
5. Desmarque qualquer outra coluna que esteja selecionada.
6. Clique em **Salvar e Executar**.
Os resultados do perfil aparecem na exibição resumida.
7. Clique na regra **FirstName**. Os resultados do perfil para a regra aparecem na exibição detalhada.
8. Selecione um valor no painel **Valores**. Clique com o botão direito do mouse no valor e clique em **Busca Detalhada**.

Os valores para a coluna **FullName** e as regras **FirstName** e **LastName** são exibidos no painel **Visualização de Dados** junto com outros valores de colunas. Observe que a ferramenta Analyst separa a coluna **FullName** em nome e sobrenome.

Tarefa 3. Editar as regras de expressão

Nesta tarefa, as regras de expressão passarão a ser reutilizáveis e ficarão disponíveis para todos os usuários da ferramenta Analyst.

1. Na exibição resumida do perfil **Profile_LA_Customers**, clique em **Editar**.
O assistente de perfil é exibido.
2. Clique na tela **Especificar Regras e Filtros**.
3. Na tela **Especificar Regras e Filtros**, selecione a regra **FirstName** e clique em **Ações > Editar Regra**.
A caixa de diálogo **Editar Regra** é exibida.
4. Selecione a opção **Deseja salvar essa regra como uma regra reutilizável?** e escolha uma localização para salvar a regra.
5. Clique em **OK**.
6. Selecione a regra **LastName** e repita as etapas de 3 a 5.
7. Clique em **Salvar e Concluir** para salvar o perfil.

Qualquer usuário da ferramenta Analyst pode usar as regras **FirstName** e **LastName** para dividir uma coluna com nome e sobrenome em colunas separadas.

Criar Resumo de Regras de Expressão

Nessa lição, você aprendeu que as regras de expressão usam funções de expressão e colunas de origem para definir a lógica da regra. Também aprendeu que a saída de uma regra de expressão é uma coluna virtual no perfil. A ferramenta Analyst inclui a coluna virtual quando você executa o perfil.

Você criou duas regras de expressão, adicionou-as a um perfil e executou o perfil. Também exibiu a saída das regras e as colocou à disposição para todos os usuários da ferramenta Analyst.

CAPÍTULO 7

Lição 6. Criar e Executar Scorecards

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Criação e Execução de Scorecards, 35](#)
- [Tarefa 1. Criar um Scorecard dos Resultados de Perfil, 36](#)
- [Tarefa 2. Executar o Scorecard, 37](#)
- [Tarefa 3. Exibir o Scorecard, 37](#)
- [Tarefa 4. Editar o Scorecard, 38](#)
- [Tarefa 5. Configurar Limites, 39](#)
- [Tarefa 6. Exibir Gráficos de Tendência de Pontuação, 39](#)
- [Resumo da Criação e Execução de Scorecards, 39](#)

Visão Geral da Criação e Execução de Scorecards

Um scorecard é a representação gráfica de valores válidos para uma coluna ou saída de uma regra nos resultados do perfil. Use scorecards para medir e monitorar o progresso da qualidade de dados ao longo do tempo.

Para criar um scorecard, adicione colunas a partir do perfil a um scorecard como métricas, atribua pesos para as métricas e configure os limites de pontuação. Você pode adicionar filtros a scorecards com base nos dados de origem. Para executar um scorecard, selecione os valores válidos para a métrica e execute o scorecard para ver as pontuações das métricas.

Os scorecards exibem a frequência de valor para colunas em um perfil como pontuações. As pontuações refletem a porcentagem de valores válidos para uma métrica.

História

A HypoStores deseja incorporar os dados do recém adquirido escritório de Los Angeles ao data warehouse. Antes que a organização mescle os dados, eles desejam verificar se os dados em diferentes camadas e estados de cliente sofreram uma análise de qualidade de dados. Você é o analista responsável pelo monitoramento do andamento da execução da análise de qualidade de dados e deseja criar um scorecard das colunas de perfil de estado e camada de cliente, configurar os limites de qualidade de dados e exibir os gráficos de tendência de pontuação para determinar como as pontuações melhoram ao longo do tempo.

Objetivos

Nessa lição, você executará as seguintes tarefas:

1. Criar um scorecard dos resultados do perfil `Profile_LA_Customers_Custom` para exibir as pontuações das colunas `Estado` e `CustomerTier`
2. Executar o scorecard para gerar as pontuações para as colunas `CustomerTier` e `Estado`.
3. Exibir o scorecard para ver as pontuações de cada coluna.
4. Editar o scorecard para especificar valores válidos diferentes para as pontuações.
5. Configure os limites de pontuação e execute o scorecard.
6. Exibir gráficos de tendência de pontuação para determinar como as pontuações melhoram ao longo do tempo.

Pré-requisitos

Antes de iniciar esta lição, verifique o seguinte pré-requisito:

- Você concluiu as lições 1 a 5 neste tutorial.

Tempo

Reserve 15 minutos para concluir as tarefas nesta lição.

Tarefa 1. Criar um Scorecard dos Resultados de Perfil

Nessa tarefa, um scorecard do perfil `Profile_LA_Customers_Custom` para pontuar os valores de coluna `Estado` e `CustomerTier` será criado.

1. No espaço de trabalho **Biblioteca**, clique no perfil **Profile_LA_Customers**.
A exibição resumida dos resultados do perfil é aberta.
2. Na exibição resumida, selecione a coluna **CustomerTier**, clique nela com o botão direito do mouse e selecione **Adicionar ao > Scorecard**.
O assistente **Adicionar ao Scorecard** é exibido.
3. No assistente **Adicionar ao Scorecard**, a opção **Novo Scorecard** é selecionada por padrão. Clique em **Avançar**.
4. Na tela **Etapa 2 de 8**, insira **sc_LA_Customer** para o nome do scorecard e navegue até a pasta **Clientes** para a localização do scorecard.
5. Clique em **Avançar**.
6. Na tela **Etapa 3 de 8**, selecione as colunas **CustomerTier** e **State** para adicioná-las ao scorecard.
7. Clique em **Avançar**.
8. Na tela **Etapa 4 de 8**, você pode criar, editar ou excluir filtros para as métricas. Neste tutorial, não criaremos um filtro de scorecard. Clique em **Avançar**.
9. Na tela **Etapa 5 de 8**, selecione a métrica **CustomerTier** no painel **Métricas**.
10. No painel **Pontuar usando: Valores**, selecione todos os valores e clique no botão **Adicionar Tudo** para movê-los até seção **Valores Válidos**.
Use a tecla **Shift** para selecionar vários valores.

11. No painel **Métricas**, selecione a métrica **Estado** e escolha valores com códigos de estado de duas letras na seção **Pontuação usando: Valores**.
12. Clique no botão **Adicionar** para mover os valores para a seção **Valores Válidos**.
Você pode ver o número total de valores válidos e a porcentagem de valores válidos na parte superior da seção.
13. Para cada métrica na seção **Métricas**, aceite as configurações padrão para os limites de pontuação na seção **Limites de Métrica**.
14. Clique em **Avançar**.
15. Na tela **Etapa 6 de 8**, você tem a opção de selecionar um grupo de métricas para adicionar as métricas. Por padrão, a ferramenta Analyst adiciona as métricas para o grupo de métricas **Padrão**.
16. Clique em **Avançar**.
17. Na tela **Etapa 7 de 8**, clique duas vezes na coluna **Peso** da métrica **CustomerTier** no painel **Padrão - Métricas**.
Quando você executa um scorecard, a ferramenta Analyst calcula a média ponderada para cada grupo de métrica com base na pontuação e peso da métrica que você atribui a cada métrica.
18. Insira um peso para as métricas **CustomerTier** e **Estado**.
19. Clique em **Avançar**.
20. Na tela **Etapa 8 de 8**, a opção **Nativo** é selecionada por padrão. Clique em **Salvar** para criar o scorecard. O scorecard é exibido no espaço de trabalho **Scorecards**.

Tarefa 2. Executar o Scorecard

Nessa tarefa, o scorecard sc_LA_Customer será executado para gerar as pontuações para as colunas CustomerTier e Estado.

1. No espaço de trabalho **Biblioteca**, clique em **Ativos > Scorecards**.
2. Selecione um scorecard no painel **Scorecards**.
3. Clique em **Ações > Abrir**.
O scorecard é exibido.
4. Clique em **Ações > Executar Scorecard**.
A caixa de diálogo **Executar Scorecard** é exibida.
5. Verifique as configurações na caixa de diálogo e clique em **Executar**.
O espaço de trabalho **Scorecard** mostra as pontuações das colunas CustomerTier e Estado.

Tarefa 3. Exibir o Scorecard

Nessa tarefa, o scorecard sc_LA_Customer será exibido para que você veja as pontuações para as colunas CustomerTier e Estado.

1. Selecione a linha **Estado** que contém a pontuação de Estado que você deseja exibir.

Na seção **sc_LA_Customer - métricas**, é possível visualizar as seguintes propriedades do scorecard:

- Nome do scorecard.
 - Número total de linhas no scorecard.
 - Número de linhas que não são válidas.
 - Pontuação junto com um gráfico de barras horizontais.
 - Tendência de pontuação. Você pode clicar na tendência de pontuação para visualizar uma representação gráfica na tela **Detalhe do Gráfico de Tendência**.
 - Peso da métrica.
 - Custo de dados inválidos.
 - Tendência de custo.
 - Objeto de dados. Clique no objeto de dados para exibir a visualização de dados desse objeto no espaço de trabalho **Descoberta**.
 - Nome da coluna ou regra.
 - Tipo de origem.
 - Ícone de busca detalhada.
2. Clique no ícone de busca detalhada na linha Estado.
As pontuações que não são válidas para a coluna Estado aparecem na seção **Linhas Inválidas** do painel **Busca Detalhada**.
 3. Selecione **Linhas Válidas** para exibir as pontuações que são válidas para a coluna Estado.
 4. Clique no ícone de busca detalhada na linha **CustomerTier**.
Todas as pontuações para a coluna **CustomerTier** são válidas.

Tarefa 4. Editar o Scorecard

Nesta tarefa, edite o scorecard sc_LA_Customer para especificar o valor Rubi como inválido para a pontuação de CustomerTier.

1. Verifique se você está no espaço de trabalho **Scorecard** e se o scorecard sc_LA_Customer está aberto.
2. Selecione **Ações > Editar > Métricas**.
A caixa de diálogo **Editar Scorecard** é exibida.
3. Na seção **Métricas**, selecione **CustomerTier**.
4. Na seção **Pontuar usando: Valores**, mova **Rubi** da seção **Valores Válidos** para a seção **Valores Disponíveis**.
Aceite as configurações padrão na seção **Limites de Métrica**.
5. Clique em **Salvar e Executar** para salvar as alterações no scorecard e executá-lo.
6. Exiba a pontuação de CustomerTier novamente.
A pontuação de CustomerTier é alterada para a porcentagem 81,4.

Tarefa 5. Configurar Limites

Nessa tarefa, os limites para a pontuação Estado no scorecard sc_LA_Customer são configurados para determinar as faixas aceitáveis dos dados na coluna Estado. Valores com código de duas letras, como CA, são aceitáveis, e códigos com mais do que duas letras, como Califa, não são aceitáveis.

1. Verifique se você está no espaço de trabalho **Scorecard** e se o scorecard sc_LA_Customer está aberto.
2. Selecione **Ações > Editar > Métricas**.
A caixa de diálogo **Editar Scorecard** é exibida.
3. Na seção **Métricas**, selecione **Estado**.
4. Na seção **Limites de Métrica**, insira as seguintes faixas para as pontuações Boa ou Inaceitável: 90 a 100% Boa; 0 a 50% Inaceitável; 51% a 89% Aceitável.
Os limites representam o menor limite das faixas aceitável e boa.
5. Clique em **Salvar e Executar** para salvar as alterações no scorecard e executá-lo.
No painel **Scorecard**, exiba as alterações para a porcentagem de pontuação e a pontuação exibida como uma barra para a pontuação Estado.

Tarefa 6. Exibir Gráficos de Tendência de Pontuação

Nessa tarefa, o gráfico de tendência para a pontuação Estado será exibido. Você pode exibir os gráficos de tendência para monitorar as pontuações ao longo do tempo.

1. Verifique se você está no espaço de trabalho **Scorecard** e se o scorecard sc_LA_Customer está aberto.
2. Selecione a linha **Estado**.
3. Clique em **Ações > Mostrar Gráfico de Tendência** ou clique na seta sob a coluna **Tendência de Pontuação**.
A caixa de diálogo **Detalhe do Gráfico de Tendência** é exibida. Você pode exibir os limites **Bom**, **Aceitável** e **Inaceitável** para a pontuação. Os limites mudam a cada vez que você executar o scorecard depois de editar os valores das pontuações no scorecard.
4. Aponte para qualquer círculo no gráfico para exibir os valores válidos na seção **Valores Válidos** na parte inferior do gráfico.
5. Clique em **Fechar** para retornar ao scorecard.

Resumo da Criação e Execução de Scorecards

Nessa lição, você aprendeu que pode criar um scorecard dos resultados de um perfil. Um scorecard contém as colunas de um perfil. Também aprendeu que pode executar um scorecard para gerar pontuações para colunas. Você editou um scorecard para configurar os valores válidos e definir os limites de pontuações. Ainda aprendeu como para exibir o gráfico de tendência de pontuação.

Você criou um scorecard das colunas CustomerTier e Estado em um perfil para analisar a qualidade de dados das colunas de camada de cliente e estado. Também executou o scorecard para gerar pontuações para cada coluna. E ainda editou o scorecard para especificar valores válidos diferentes para as pontuações. Você configurou os limites para uma pontuação e exibiu o gráfico de tendência de pontuação.

CAPÍTULO 8

Lição 7. Criar Tabelas de Referência de Colunas de Perfil

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Criação de Tabelas de Referência de Colunas de Perfil, 40](#)
- [Tarefa 1. Criar uma Tabela de Referência das Colunas de Perfil, 41](#)
- [Tarefa 2. Editar a Tabela de Referência, 42](#)
- [Resumo da Criação de Tabelas de Referência de Colunas de Perfil, 42](#)

Visão Geral da Criação de Tabelas de Referência de Colunas de Perfil

Uma tabela de referência contém dados de referência que você pode usar para padronizar dados de origem. Os dados de referência podem incluir valores válidos e padrão. Crie tabelas de referência para estabelecer relações entre os valores dos dados de origem e os valores válidos e padrão.

Você pode criar uma tabela de referência dos resultados de perfil. Depois de criar uma tabela de referência, edite-a para adicionar colunas ou linhas e adicionar ou editar valores válidos e padrão. Você pode exibir as alterações feitas em uma tabela de referência em uma trilha de auditoria.

História

A HypoStores deseja criar um perfil de dados para descobrir anomalias e padronizar os dados com valores válidos. Você é o analista responsável pela padronização dos valores válidos nos dados. Você deseja criar uma tabela de referência com base em valores válidos de colunas de perfil.

Objetivos

Nesta lição, você concluirá as seguintes tarefas:

1. Criar uma tabela de referência da coluna CustomerTier no perfil Profile_LA_Customers_Custom selecionando valores válidos para as colunas.
2. Editar a tabela de referência para configurar valores válidos diferentes para as colunas.

Pré-requisitos

Antes de iniciar esta lição, verifique o seguinte pré-requisito:

- Você concluiu as lições 1 a 6 neste tutorial.

Tempo

Reserve 15 minutos para concluir as tarefas nesta lição.

Tarefa 1. Criar uma Tabela de Referência das Colunas de Perfil

Nessa tarefa, uma tabela de referência é criada e adicionada à coluna CustomerTier no perfil Profile_LA_Customers da tabela de referência.

1. No espaço de trabalho **Biblioteca**, clique em **Ativos > Perfis**.
2. Clique no perfil **Profile_LA_Customers** para abrir os resultados de perfil na exibição resumida.
3. Na exibição resumida, selecione a coluna CustomerTier que você deseja adicionar à tabela de referência. Clique com o botão direito do mouse e selecione **Adicionar à Tabela de Referência**.
A caixa de diálogo **Adicionar à Tabela de Referência** é exibida.
4. Selecione **Criar uma tabela de referência**.
5. Clique em **Avançar**.
6. No campo **Nome**, insira **Reftab_CustTier_HypoStores**.
7. Insira uma descrição e defina 0 como o valor padrão.
A ferramenta Analyst usa o valor padrão para qualquer registro de tabela que não contém um valor.
8. Clique em **Avançar**.
9. Na seção **Atributos de Coluna**, configure as seguintes propriedades de coluna para a coluna CustomerTier:

Propriedade	Descrição
Nome	CustomerTier
Tipo de dados	String
Precisão	10
Descrição	Referência de valores de camada de cliente

10. Como opção, escolha a criação de uma coluna de descrição para as linhas da tabela de referências. Digite o nome e a precisão da coluna.
11. Verifique os valores de coluna CustomerTier na seção **Visualização**.
12. Clique em **Avançar**.
O nome de tabela de referência Reftab_CustomerTier_HypoStores será exibido. Você pode inserir uma descrição opcional.
13. No painel **Salvar em**, selecione o projeto de tutorial onde você deseja criar a tabela de referência.
O painel **Tabelas de Referência** lista as tabelas de referência na localização selecionada.
14. Insira uma nota de auditoria opcional.
15. Clique em **Concluir**.

Tarefa 2. Editar a Tabela de Referência

Nessa tarefa, a tabela Reftab_CustomerTier_HypoStores será editada para adicionar valores alternativos às camadas de cliente.

1. No espaço de trabalho **Biblioteca**, clique em **Ativos > Tabelas de Referência**.
2. Clique na tabela de referência Reftab_CustomerTier_HypoStores.
A tabela de referência é aberta no espaço de trabalho **Design**.
3. Para editar uma linha, selecione-a e clique em **Ações > Editar** ou clique no ícone **Editar**.
A caixa de diálogo **Editar Linha** é exibida. Opcionalmente, selecione várias linhas para adicionar o mesmo valor alternativo para cada linha.
4. Insira os seguintes valores alternativos para as linhas Diamante, Esmeralda, Ouro, Prata e Bronze: 1, 2, 3, 4, 5.
Insira uma nota de auditoria opcional.
5. Clique em **Aplicar** para aplicar as alterações.
6. Clique em **Fechar**.
Os valores alterados da tabela de referência aparecem no espaço de trabalho **Design**.

Resumo da Criação de Tabelas de Referência de Colunas de Perfil

Nessa lição, você aprendeu como criar tabelas de referência dos resultados de um perfil para configurar os valores válidos para os dados de origem.

Você criou uma tabela de referência de uma coluna de perfil selecionando valores válidos para as colunas. Também editou a tabela de referência para configurar valores válidos diferentes para as colunas.

CAPÍTULO 9

Lição 8. Criar Tabelas de Referência

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Criação de Tabelas de Referência, 43](#)
- [Tarefa 1. Criar uma Tabela de Referência, 44](#)
- [Resumo da Criação de Tabelas de Referência, 44](#)

Visão Geral da Criação de Tabelas de Referência

Uma tabela de referência contém dados de referência que você pode usar para padronizar dados de origem. Os dados de referência podem incluir valores válidos e padrão. Crie tabelas de referência para estabelecer relações entre os valores dos dados de origem e os valores válidos e padrão.

Você pode criar manualmente uma tabela de referência usando o editor de tabela de referência. Use a tabela de referência para definir e padronizar os dados de origem. A tabela de referência pode ser compartilhada com um desenvolvedor para uso nas transformações de Pesquisa e Padronizador na ferramenta Developer.

História

A HypoStores deseja padronizar dados com os valores válidos. Você é o analista responsável pela padronização dos valores válidos nos dados. Você deseja criar uma tabela de referência para definir os códigos de camada de cliente padrão que fazem referência aos dados de cliente de LA. A tabela de referência pode ser compartilhada com um desenvolvedor.

Objetivos

Nessa lição, você concluirá a seguinte tarefa:

- Criar uma tabela de referência usando o editor de tabela de referência para definir os códigos de camada de cliente padrão que fazem referência aos dados de cliente de LA.

Pré-requisitos

Antes de iniciar esta lição, verifique o seguinte pré-requisito:

- Você concluiu as lições 1 e 2 neste tutorial.

Tempo

Defina 10 minutos para concluir a tarefa nesta lição.

Tarefa 1. Criar uma Tabela de Referência

Nessa tarefa, a tabela de referência Reftab_CustomerTier_Codes será criada para padronizar os valores válidos dos dados de camada do cliente.

1. Clique em **Novo > Tabela de Referência**.
O assistente de **Nova Tabela de Referência** é exibido.
2. Selecione **Usar o editor da tabela de referência**.
3. Clique em **Avançar**.
4. Para cada coluna que você deseja incluir na tabela de referência, clique no ícone **Adicionar Nova Coluna** e configure as propriedades de coluna para cada coluna.
Adicione os seguintes nomes de coluna: CustomerID, CustomerTier e Status. As colunas podem ser reorganizadas ou excluídas.
5. Insira uma descrição opcional e defina o valor padrão como **0**.
A ferramenta Analyst usa o valor padrão para qualquer registro de tabela que não contenha um valor.
6. Clique em **Avançar**.
7. No campo **Nome**, digite **Reftab_CustomerTier_Codes**.
8. Na seção **Pastas**, selecione a pasta **Clientes** no projeto de tutorial.
9. Clique em **Concluir**.
A tabela de referência é exibida no espaço de trabalho **Design**.
10. No menu **Ações**, selecione **Adicionar Linha** para preencher cada coluna da tabela de referência com os seguintes quatro valores:
CustomerID = LA1, LA2, LA3, LA4
CustomerTier = 1, 2, 3, 4.
Status= Ativo, Inativo

Resumo da Criação de Tabelas de Referência

Nessa lição, você aprendeu como criar tabelas de referência usando o editor de tabela de referência para criar valores válidos padrão para usar com dados de origem.

Você criou uma tabela de referência usando o editor de tabela de referência para padronizar os valores de camada de cliente nos dados de clientes de LA.

Parte II: Introdução ao Informatica Developer

Esta parte contém os seguintes capítulos:

- [Lição 1. Configurando o Informatica Developer, 46](#)
- [Lição 2: Importando Objetos de Dados Físicos, 50](#)
- [Lição 3. Executar um Perfil em Dados de Origem, 59](#)
- [Lição 4. Analisando dados, 64](#)
- [Lição 5. Padronizar Dados , 71](#)
- [Lição 6. Validar Dados de Endereço, 77](#)

CAPÍTULO 10

Lição 1. Configurando o Informatica Developer

Este capítulo inclui os seguintes tópicos:

- [Configurando o Informatica Developer - Visão Geral, 46](#)
- [Tarefa 1. Iniciar o Informatica Developer, 47](#)
- [Tarefa 2. Adicionar um Domínio, 47](#)
- [Tarefa 3. Adicionar um Repositório do Modelo, 48](#)
- [Tarefa 4. Criar um Projeto, 48](#)
- [Tarefa 5. Criar uma Pasta, 48](#)
- [Tarefa 6. Selecionar um Data Integration Service padrão, 49](#)
- [Configurando o Informatica Developer - Resumo, 49](#)

Configurando o Informatica Developer - Visão Geral

Antes de iniciar as lições deste tutorial, inicie e configure a ferramenta Developer. Para configurar a ferramenta Developer, adicione um domínio. Adicione um repositório do Modelo que está no domínio e crie um projeto e uma pasta para armazenar seu trabalho. Você também seleciona um Data Integration Service padrão.

O domínio Informatica é um conjunto de nós e serviços que definem o ambiente Informatica. Os serviços no domínio incluem o Model Repository Service e o Data Integration Service.

O Model Repository Service gerencia o repositório do Modelo. O repositório do Modelo é um banco de dados relacional que armazena os metadados para projetos criados na ferramenta Developer. Um projeto armazena os objetos que você cria na ferramenta Developer. Um projeto também pode conter pastas que armazenam objetos relacionados, como objetos que fazem parte do mesmo requisito de negócios.

O Data Integration Service realiza tarefas de integração de dados na ferramenta Developer.

Objetivos

Nesta lição, você concluirá as seguintes tarefas:

- Iniciar a ferramenta Developer e acessar o workbench da ferramenta Developer.
- Adicionar um domínio na ferramenta Developer.
- Adicionar um repositório do Modelo para poder criar um projeto.

- Criar um projeto para armazenar os objetos criados na ferramenta Developer.
- Criar uma pasta no projeto que pode armazenar objetos relacionados.
- Selecionar um Data Integration Service padrão para realizar tarefas de integração de dados.

Pré-requisitos

Antes de iniciar esta lição, verifique os seguintes pré-requisitos:

- Você instalou a ferramenta Developer.
- Você possui um nome de domínio, um nome de host e um número de porta para se conectar a um domínio. Você pode obter essas informações de um administrador de domínio.
- Um administrador de domínio foi configurado para um Model Repository Service na ferramenta Administrator.
- Você tem um nome de usuário e uma senha para acessar o Model Repository Service. Você pode obter essas informações de um administrador de domínio.
- Um administrador de domínio configurou um Data Integration Service.
- O Data Integration Service está em execução.

Tempo

Reserve de 5 a 10 minutos para concluir as tarefas nesta lição.

Tarefa 1. Iniciar o Informatica Developer

Inicie a ferramenta Developer para começar o tutorial.

1. Inicie a ferramenta Developer.
A página de **Boas-vindas** da ferramenta Developer é exibida.
2. Clique no botão **Workbench**.
O workbench da ferramenta Developer é exibido.

Tarefa 2. Adicionar um Domínio

Nesta tarefa, você adicionará um domínio na ferramenta Developer para acessar um repositório do Modelo.

1. Clique em **Janela > Preferências**.
A caixa de diálogo **Preferências** é exibida.
2. Selecione **Informatica > Domínios**.
3. Clique em **Adicionar**.
A caixa de diálogo **Novo Domínio** é exibida.
4. Digite o nome do domínio, o nome do host e o número da porta.
5. Clique em **Concluir**.
6. Clique em **OK**.

Tarefa 3. Adicionar um Repositório do Modelo

Nesta tarefa, você adicionará o repositório do Modelo que deseja usar para armazenar projetos e pastas.

1. Clique em **Arquivo > Conectar com o Repositório**.
A caixa de diálogo **Conectar com o Repositório** é exibida.
2. Clique em **Procurar** para selecionar um Model Repository Service.
3. Clique em **OK**.
4. Clique em **Avançar**.
5. Insira seu nome de usuário e senha.
6. Selecione um espaço de nome.
7. Clique em **Concluir**.
O repositório do Modelo aparece na exibição **Object Explorer**.

Tarefa 4. Criar um Projeto

Nesta tarefa, você criará um projeto para armazenar os objetos criados na ferramenta Developer. Você pode criar um projeto para todos os tutoriais deste guia.

1. Na exibição **Object Explorer**, selecione um Model Repository Service.
2. Clique em **Arquivo > Novo > Projeto**.
A caixa de diálogo **Novo Projeto** é exibida.
3. Insira seu nome usando "Tutorial_" como prefixo para o nome do projeto.
4. Clique em **Concluir**.
O projeto aparece abaixo de Model Repository Service na exibição **Object Explorer**.

Tarefa 5. Criar uma Pasta

Nesta tarefa, você criará uma pasta para armazenar objetos relacionados. Você pode criar uma pasta para todos os tutoriais deste guia.

1. Na exibição **Object Explorer**, selecione o projeto ao qual você deseja adicionar a pasta.
2. Clique em **Arquivo > Nova > Pasta**.
3. Insira um nome para a pasta.
4. Clique em **Concluir**.
A ferramenta Developer adiciona a pasta ao projeto na exibição **Object Explorer**. Expanda o projeto para ver a pasta.

Tarefa 6. Selecionar um Data Integration Service padrão

Nessa tarefa, um Data Integration Service padrão será selecionado para que você possa executar mapeamentos e visualizar dados.

1. Clique em **Janela > Preferências**.
A caixa de diálogo **Preferências** é exibida.
2. Selecione **Informatica > Data Integration Services**.
3. Expanda o domínio.
4. Selecione um Data Integration Service.
5. Clique em **Definir como Padrão**.
6. Clique em **OK**.

Configurando o Informatica Developer - Resumo

Nessa lição, você aprendeu que o domínio Informatica inclui o Model Repository Service e o Data Integration Service. O Model Repository Service gerencia o repositório do Modelo. Um repositório do Modelo contém projetos e pastas. O Data Integration Service realiza tarefas de integração de dados.

Você iniciou e configurou a ferramenta Developer. Você adicionou um domínio à ferramenta Developer, adicionou um repositório do Modelo e criou um projeto e uma pasta. Você também selecionou um Data Integration Service padrão.

Agora, pode usar a ferramenta Developer para concluir outras lições deste tutorial.

CAPÍTULO 11

Lição 2: Importando Objetos de Dados Físicos

Este capítulo inclui os seguintes tópicos:

- [Importando Objetos de Dados Físicos - Visão Geral, 50](#)
- [Tarefa 1. Importar o Objeto de Dados de Arquivo Simples Boston_Customers, 51](#)
- [Tarefa 2. Importar o Objeto de Dados de Arquivo Simples LA_Customers, 57](#)
- [Tarefa 3. Importar o Objeto de Dados de Arquivo Simples All_Customers, 58](#)
- [Importando Objetos de Dados Físicos - Resumo, 58](#)

Importando Objetos de Dados Físicos - Visão Geral

Um objeto de dados físicos é uma representação de dados com base em um arquivo simples ou em uma tabela de banco de dados relacional. Você pode importar um arquivo simples ou uma tabela de banco de dados relacional como um objeto de dados físicos a ser usado como origem ou destino em um mapeamento.

História

A HypoStores Corporation armazena dados de clientes dos escritórios de Los Angeles e de Boston em arquivos simples. Você deseja trabalhar com esses dados do cliente na ferramenta Developer. Para fazer isso, importe cada arquivo simples como um objeto de dados físicos.

Objetivos

Nesta lição, você importará arquivos simples como objetos de dados físicos. Também configurou o diretório de arquivos de origem para que o Data Integration Service possa ler os dados de origem a partir do diretório correto.

Pré-requisitos

Antes de iniciar esta lição, verifique o seguinte pré-requisito:

- Você concluiu a lição 1 deste tutorial.

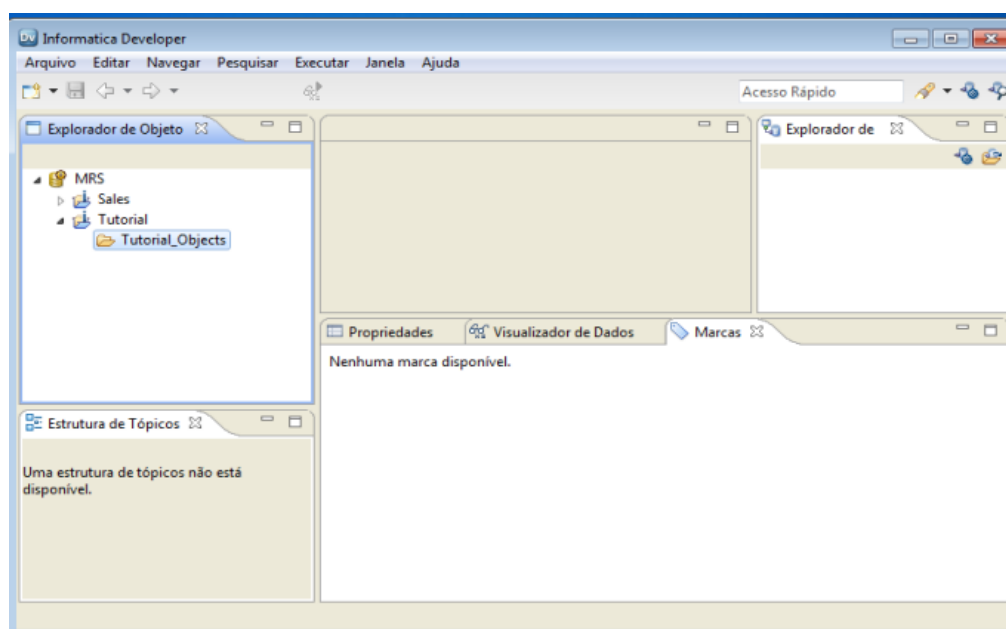
Tempo

Reserve de 10 a 15 minutos para concluir as tarefas nesta lição.

Tarefa 1. Importar o Objeto de Dados de Arquivo Simples Boston_Customers

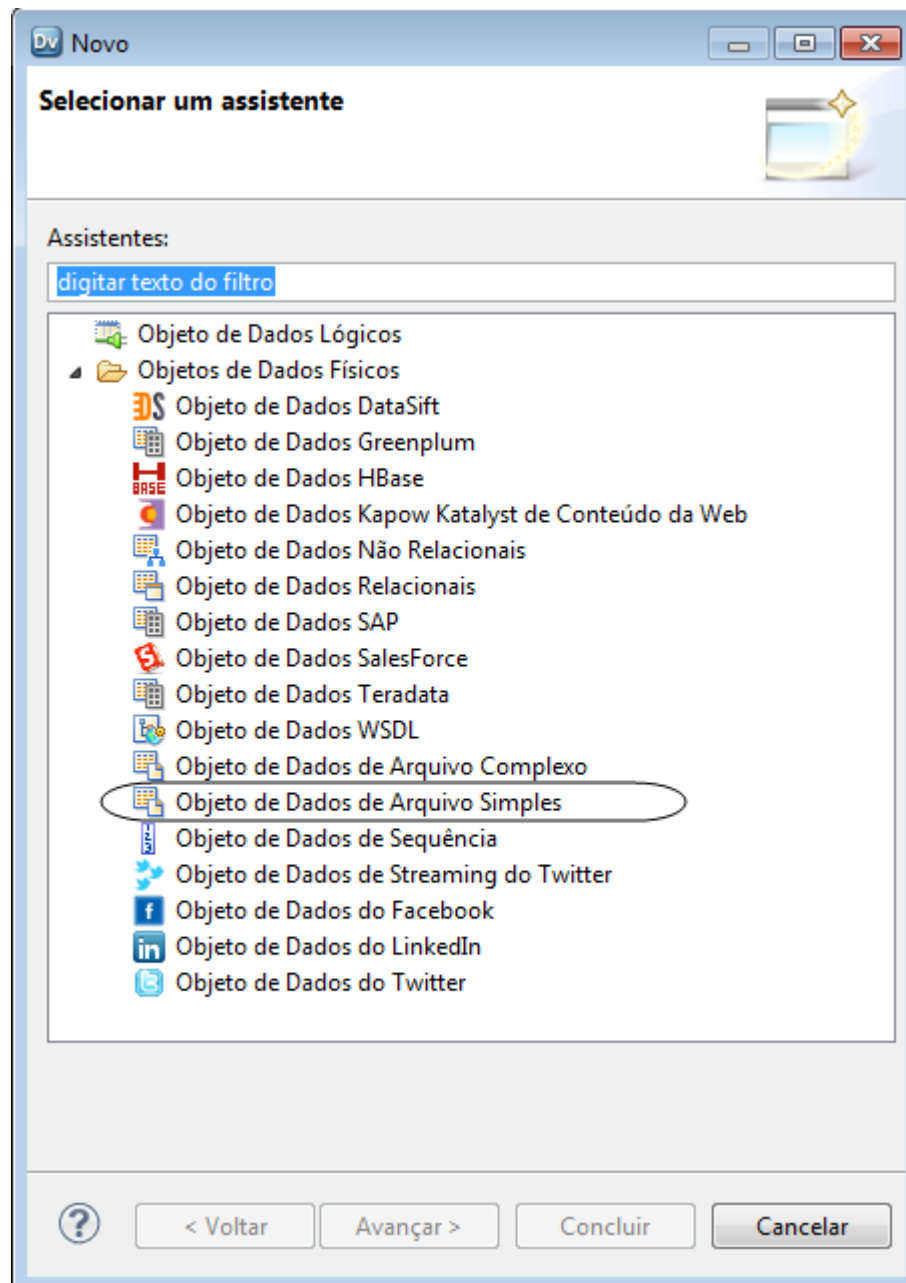
Nesta tarefa, você importará um objeto de dados físicos de um arquivo que contém dados de clientes do escritório de Boston.

1. Na exibição **Object Explorer**, selecione a pasta de Tutorial_Objects.



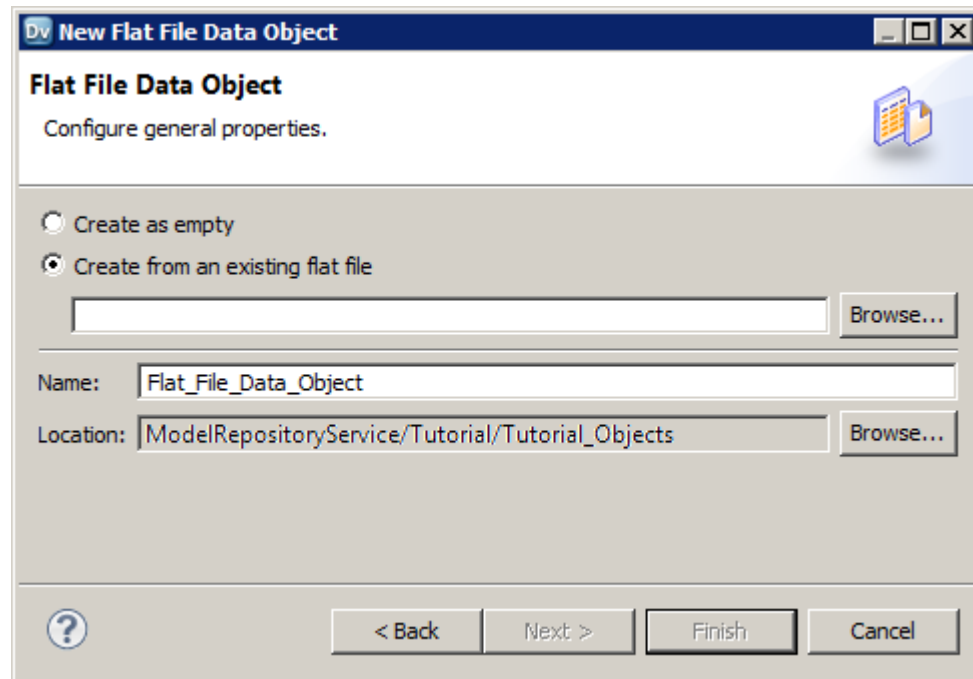
2. Clique com o botão direito do mouse na pasta Tutorial_Objects e selecione **Novo > Objeto de Dados**.

A caixa de diálogo **Novo** é exibida.



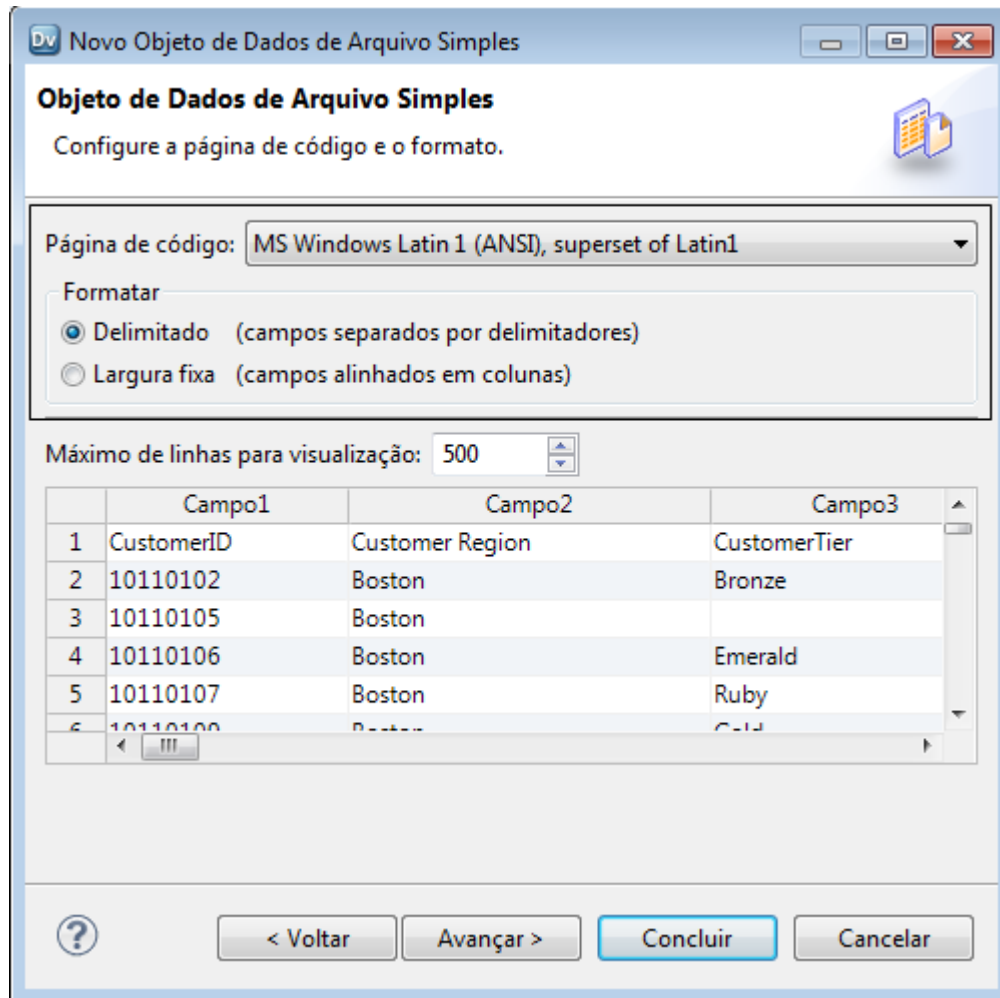
3. Selecione **Objetos de Dados Físicos > Objeto de Dados de Arquivo Simples** e clique em **Avançar**.

A caixa de diálogo **Novo Objeto de Dados de Arquivo Simples** é exibida.



4. Selecione **Criar a partir de um arquivo simples existente**.
5. Clique em **Procurar** e navegue até Boston_Customers.csv no seguinte diretório da máquina da ferramenta Developer: <Diretório de instalação da Informatica >\clients\DeveloperClient\Tutorials
6. Clique em **Abrir**.
O assistente nomeia o objeto de dados como "Boston_Customers".
7. Clique em **Avançar**.
8. Verifique se a página de código está definida como **MS Windows Latino 1 (ANSI)**, **superconjunto de Latino 1** e se o formato está definido como **Delimitado**.

A caixa de diálogo **Novo Objeto de Dados de Arquivo Simples** mostra a página de código padrão, o formato e uma visualização dos dados de arquivo simples.



Novo Objeto de Dados de Arquivo Simples

Objeto de Dados de Arquivo Simples

Configure a página de código e o formato.

Página de código: MS Windows Latin 1 (ANSI), superset of Latin1

Formatar

☒ Delimitado (campos separados por delimitadores)

☐ Largura fixa (campos alinhados em colunas)

Máximo de linhas para visualização: 500

	Campo1	Campo2	Campo3
1	CustomerID	Customer Region	CustomerTier
2	10110102	Boston	Bronze
3	10110105	Boston	
4	10110106	Boston	Emerald
5	10110107	Boston	Ruby
6	10110108	Boston	Gold

< Voltar Avançar > Concluir Cancelar

9. Clique em **Avançar**.
10. Selecione **Importar nomes de coluna da primeira linha**.

A caixa de diálogo **Novo Objeto de Dados de Arquivo Simples** mostra os nomes de coluna na visualização dos dados de arquivo simples.

Novo Objeto de Dados de Arquivo Simples

Objeto de Dados de Arquivo Simples

Configurar propriedades de formato delimitado.

Delimitadores

☐ Tabulação ☐ Ponto-e-vírgula ☒ Virgula

☐ Espaço ☐ Outros:

Qualificador de texto

☒ Sem aspas ☐ Aspas simples ☐ Aspas duplas

Opções de visualização

☒ Importar nomes de coluna da primeira linha Iniciar importação na linha: 1

Delimitador de linha: \012 LF (\n) ☐ Tratar delimitadores consecutivos como um

Caractere de escape: ☐ Manter caractere de escape nos dados

Máximo de linhas para visualização: 500

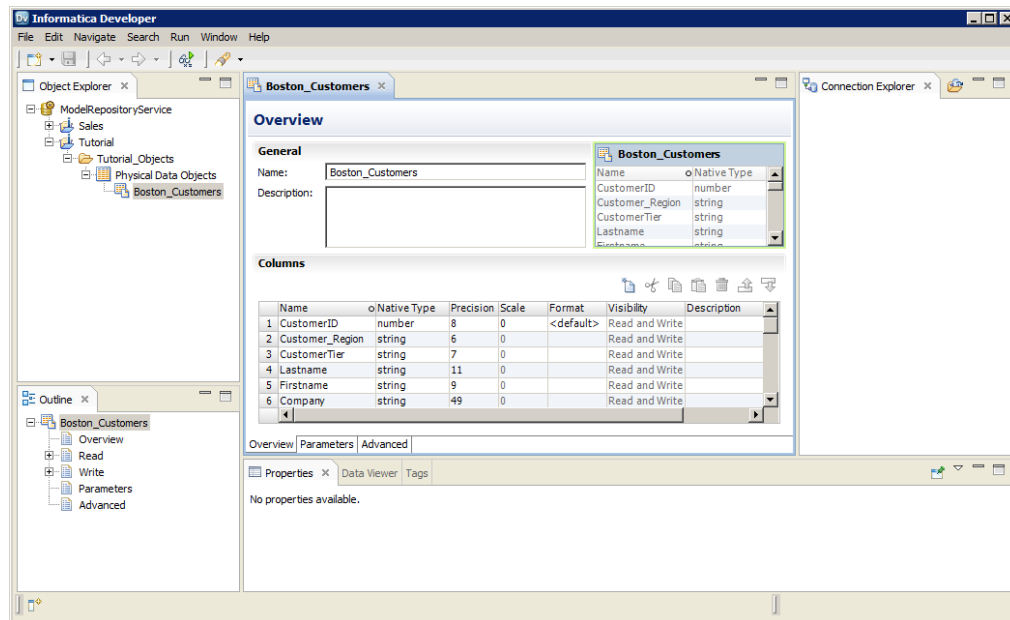
	CustomerID	Customer_Region	CustomerTier	Last Name
1	10110102	Boston	Bronze	Brosseau
2	10110105	Boston		Anderson
3	10110106	Boston	Emerald	Boonstra
4	10110107	Boston	Ruby	Chan
5	10110109	Boston	Gold	Acheson

< Voltar Avançar > Concluir Cancelar

1. A opção **Importar nomes de coluna na primeira linha**
2. Nomes de coluna

11. Clique em **Concluir**.

O objeto de dados físicos Boston_Customers é exibido na pasta Objetos de Dados Físicos sob a pasta Tutorial_Objects. A exibição **Visão Geral** mostra o conteúdo do arquivo e é aberta no editor.



12. Clique na exibição **Avançado**.

A exibição **Avançado** mostra propriedades para o objeto de dados físicos.

13. Na exibição **Avançado**, role até a seção **Tempo de execução: Leitura**.

14. Na seção **Tempo de execução: Leitura**, defina **Diretório do arquivo de origem** no seguinte diretório da máquina do Serviço de Integração de Dados: <diretório de instalação do Informatica>\server \Tutorials

O Serviço de Integração de Dados procura o arquivo de origem no diretório do servidor na máquina que executa o Serviço de Integração de Dados. A instalação do servidor contém uma cópia dos arquivos do tutorial. O Serviço de Integração de Dados não pode ler arquivos do diretório de instalação de cliente, a não ser que você altere as permissões de acesso no arquivo de origem e no diretório.

A seguinte figura mostra um exemplo do diretório de arquivo de origem:

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

Nota: A máquina da ferramenta Developer deve ter acesso ao diretório de arquivo de origem na máquina que executa o Serviço de Integração de Dados. Se a ferramenta Developer não puder acessar o diretório do arquivo de origem, a ferramenta Developer não poderá visualizar os dados no arquivo de origem ou executar mapeamentos que acessem os dados no arquivo de origem. Se você executar vários Serviços de Integração de Dados, haverá um diretório de arquivo de origem separado para cada Serviço de Integração de Dados.

15. Clique na exibição **Visualizador de Dados**.

16. Na exibição **Visualizador de Dados**, clique em **Executar**.
O Serviço de Integração de Dados lê os dados do arquivo Boston_Customers e mostra os resultados na janela Saída.
17. Clique em **Arquivo > Salvar** para salvar o objeto de dados físicos Boston_Customers.

Tarefa 2. Importar o Objeto de Dados de Arquivo Simples LA_Customers

Nesta tarefa, você importará um objeto de dados físicos de um arquivo simples que contém dados de clientes do escritório de Los Angeles.

1. Na exibição **Object Explorer**, selecione o projeto do tutorial.
2. Clique em **Arquivo > Novo > Objeto de Dados**.
A caixa de diálogo **Novo** é exibida.
3. Selecione **Objetos de Dados Físicos > Objeto de Dados de Arquivo Simples** e clique em **Avançar**.
A caixa de diálogo **Novo Objeto de Dados de Arquivo Simples** é exibida.
4. Selecione **Criar de um Arquivo Simples Existente**.
5. Clique em **Procurar** e navegue até LA_Customers.csv no seguinte diretório: <Diretório de Instalação da Informatica>\clients\DeveloperClient\Tutorials
6. Clique em **Abrir**.
O assistente nomeia o objeto de dados LA_Customers.
7. Clique em **Avançar**.
8. Verifique se a página de código é MS Windows Latin 1 (ANSI), superconjunto de Latin 1.
9. Verifique se o formato é delimitado.
10. Clique em **Avançar**.
11. Verifique se o delimitador está definido como vírgula.
12. Selecione **Importar nomes de coluna da primeira linha**.
13. Clique em **Concluir**.
O objeto de dados físicos LA_Customers será exibido em **Objetos de Dados Físicos** no projeto de tutorial.
14. Clique na exibição **Leitura** e selecione a transformação de Saída.
15. Clique na guia **Tempo de Execução** da exibição **Propriedades**.
16. Defina o Diretório de Arquivo de Origem para o seguinte diretório no computador Data Integration Service: <Diretório de Instalação da Informatica>\server\Tutorials
17. Clique em **Arquivo > Salvar**.

Tarefa 3. Importar o Objeto de Dados de Arquivo Simples All_Customers

Nessa tarefa, um objeto de dados físicos de um arquivo simples que combina os dados de pedidos de clientes dos escritórios de Los Angeles e Boston será importado.

1. Na exibição **Object Explorer**, selecione o projeto do tutorial.
2. Clique em **Arquivo > Novo > Objeto de Dados**.
A caixa de diálogo **Novo** é exibida.
3. Selecione **Objetos de Dados Físicos > Objeto de Dados de Arquivo Simples** e clique em **Avançar**.
A caixa de diálogo **Nova Fonte de Dados de Arquivo Simples** é exibida.
4. Selecione **Criar de um Arquivo Simples Existente**.
5. Clique em **Procurar** e navegue para All_Customers.csv no seguinte diretório: <Diretório de Instalação da Informatica>\clients\DeveloperClient\Tutorials.
6. Clique em **Abrir**.
O assistente nomeia o objeto de dados All_Customers.
7. Clique em **Avançar**.
8. Verifique se a página de código é MS Windows Latin 1 (ANSI), superconjunto de Latin 1.
9. Verifique se o formato é delimitado.
10. Clique em **Avançar**.
11. Verifique se o delimitador está definido como vírgula.
12. Selecione **Importar nomes de coluna da primeira linha**.
13. Clique em **Concluir**.
O objeto de dados físicos All_Customers é exibido em **Objetos de Dados Físicos** no projeto de tutorial.
14. Clique na exibição **Leitura** e selecione a transformação de Saída.
15. Clique na guia **Tempo de Execução** da exibição **Propriedades**.
16. Defina o Diretório de Arquivo de Origem para o seguinte diretório no computador Data Integration Service: <Diretório de Instalação da Informatica>\server\Tutorials
17. Clique em **Arquivo > Salvar**.

Importando Objetos de Dados Físicos - Resumo

Nesta lição, você aprendeu que objetos de dados físicos são representações de dados com base em um arquivo simples ou em uma tabela de banco de dados relacional.

Também criou objetos de dados físicos de arquivos simples. Também configurou o diretório de arquivos de origem para que o Data Integration Service possa ler os dados de origem a partir do diretório correto.

Use os objetos de dados como origens de mapeamento nas lições de qualidade de dados.

CAPÍTULO 12

Lição 3. Executar um Perfil em Dados de Origem

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Criação de Perfil de Dados, 59](#)
- [Tarefa 1. Executar uma Análise de Associação em Duas Fontes de Dados, 60](#)
- [Tarefa 2. Exibir Resultados de Análise de Associação, 61](#)
- [Tarefa 3. Executar um Perfil em uma Fonte de Dados, 61](#)
- [Tarefa 4. Exibir Resultados de Criação de Perfil de Coluna, 62](#)
- [Resumo da Criação de Perfil, 63](#)

Visão Geral da Criação de Perfil de Dados

Um perfil é um conjunto de metadados que descreve o conteúdo e a estrutura de um conjunto de dados.

A criação de perfil e a descoberta de dados geralmente são a primeira etapa de um projeto. Você pode executar um perfil para avaliar a estrutura de dados e verificar se as colunas de dados estão preenchidas com os tipos de informações que você espera. Se um perfil revela problemas nos dados, é possível definir as etapas no seu projeto para corrigir esses problemas. Por exemplo, se um perfil revela que uma coluna contém valores de maior tamanho que o esperado, você pode criar processos de qualidade de dados para remover ou corrigir os valores problemáticos.

Um perfil que analisa a qualidade de dados de colunas selecionadas é chamado de um perfil da coluna.

Nota: Você também pode usar a Developer tool para descobrir a chave primária, a chave externa e os relacionamentos de dependência funcional, e para analisar as condições de associação em colunas de dados.

Um perfil da coluna oferece os seguintes fatos sobre os dados:

- O número de valores distintos e nulos em cada coluna, expresso como um número e uma porcentagem.
- Os padrões de dados em cada coluna e as frequências com que esses valores ocorrem.
- As estatísticas sobre os valores de coluna, como o número máximo e mínimo de tamanhos de valores e o primeiro e o último valores em cada coluna.
- Para perfis de análise de associação, o grau de sobreposição entre duas colunas de dados, exibido como um diagrama de Venn e como um valor de porcentagem. Use perfis de análise de associação para identificar possíveis problemas com a coluna condições de associação.

Você pode executar um perfil da coluna em qualquer fase do projeto para medir a qualidade de dados e para verificar se as alterações nos dados atendem aos objetivos do seu projeto. Também pode executar um perfil da coluna na transformação em um mapeamento para indicar o efeito que a transformação terá nos dados.

História

A HypoStores deseja verificar se os dados de clientes não contêm erros, inconsistências e informações duplicadas. Antes da HypoStores desenvolver os processos para apresentar os objetivos de qualidade de dados, ela precisa mensurar a qualidade dos seus arquivos de dados de origem e confirmar se os dados estão prontos para processamento.

Objetivos

Nesta lição, você concluirá as seguintes tarefas:

- Executar uma análise de associação na fonte de dados `Boston_Customers` e na fonte de dados `LA_Customers`.
- Exibir os resultados de análise de associação para determinar se você pode ou não ter êxito ao mesclar dados de dois escritórios.
- Executar um perfil da coluna na fonte de dados `All_Customers`.
- Exibir os resultados da criação de perfil de coluna para observar os valores e padrões contidos nos dados.

Pré-requisitos

Antes de iniciar esta lição, verifique o seguinte pré-requisito:

- Você concluiu as lições 1 e 2 neste tutorial.

Tempo Necessário

- Defina 20 minutos para concluir esta lição.

Tarefa 1. Executar uma Análise de Associação em Duas Fontes de Dados

Nesta tarefa, você realizará uma análise de associação nas fontes de dados `Boston_Customers` e `LA_Customers` para exibir as condições de associação.

1. Selecione a pasta do tutorial e clique em **Arquivo > Novo > Perfil**.
2. Selecione **Perfil de Descoberta Empresarial**.
3. Clique em **Avançar**.
4. No campo **Nome**, insira **Tutorial_Profile**.
5. Clique em **Concluir**.
O perfil `Tutorial_Profile` é exibido no Object Explorer.
6. Arraste as fontes de dados **`Boston_Customers`** e **`LA_Customers`** até o editor à direita.
Sugestão: Mantenha pressionada a tecla Shift para selecionar vários objetos de dados.
7. Clique com o botão direito do mouse em um nome de objeto de dados e selecione **Perfil de Associação**.
O assistente de **Novo Perfil de Junção** é exibido.
8. No campo **Nome**, insira **JoinAnalysis**.

9. Verifique se Boston_Customers e LA_Customers são exibidos como objetos de dados e clique em **Avançar**.
10. Verifique se a coluna **CustomerID** está selecionada em ambas as fontes de dados. Role o painel de assistente para exibir as colunas em ambos os conjuntos de dados. Clique em **Avançar**.
11. Clique em **Adicionar** para adicionar condições de associação. A caixa de diálogo **Condição de Junção** é exibida.
12. Na seção **Colunas**, clique em **Adicionar linha**.
13. Clique duas vezes na primeira linha na coluna esquerda e selecione **CustomerID**.
14. Clique duas vezes na primeira linha na coluna direita e selecione **CustomerID**.
15. Clique em **OK** e em **Concluir**.
16. Se a ferramenta Developer perguntar se você deseja salvar as alterações, clique em **Sim**. A ferramenta Developer executa o perfil.

Nota: Não feche o perfil. Exiba os resultados do perfil na próxima tarefa.

Tarefa 2. Exibir Resultados de Análise de Associação

Nesta tarefa, você exibirá os resultados da análise de associação na exibição Resultado da Associação do perfil JoinAnalysis.

1. Clique na guia **JoinAnalysis** do editor.
2. Na seção **Resultado da Associação**, clique na primeira linha. A seção **Detalhes** exibe um diagrama de Venn e a chave de cor que detalha os resultados da análise de associação.
3. Verifique se a coluna **Linhas de Associação** mostra zero como o número de linhas que contêm uma associação. Esse valor indica que os campos CustomerID não têm duplicatas. Você pode mesclar com êxito as duas fontes de dados.
4. Para exibir os valores de CustomerID do objeto de dados LA_Customers, clique duas vezes no círculo **LA_Customers** do diagrama de Venn.
Sugestão: Clique duas vezes nos círculos do diagrama de Venn para exibir as linhas de dados. Se os círculos fizerem interseção no diagrama de Venn, clique duas vezes na interseção para exibir valores de dados comuns a ambos os conjuntos de dados.
O Visualizador de Dados exibe os valores de CustomerID do objeto de dados LA_Customers.

Tarefa 3. Executar um Perfil em uma Fonte de Dados

Nesta tarefa, você executará um perfil na fonte de dados All_Customers para exibir o conteúdo e a estrutura dos dados.

1. Na exibição **Object Explorer**, navegue até os objetos de dados no seu projeto de tutorial.

2. Selecione a fonte de dados **All_Customers**.
3. Clique em **Arquivo > Novo > Perfil**.
A caixa de diálogo **Novo** é exibida.
4. Selecione **Perfil**.
5. Clique em **Avançar**.
6. No campo **Nome**, insira **All_Customers**.
7. Clique em **Concluir**.
O perfil All_Customers é aberto no editor e executado.

Tarefa 4. Exibir Resultados de Criação de Perfil de Coluna

Nesta tarefa, você exibirá os resultados da criação de perfil de coluna para o objeto de dados All_Customers e examinará os valores e padrões contidos nos dados.

1. Clique em **Janela > Mostrar Exibição > Andamento** para exibir o andamento do perfil All_Customers.
A exibição Andamento se abre.
2. Quando a exibição Andamento informar que a execução do perfil All_Customers foi concluída, clique na exibição **Resultados** no editor.
3. Na seção **Criação de Perfil de Coluna**, clique na coluna **CustomerTier**.
A seção **Detalhes** mostra todos os valores contidos na coluna CustomerTier, bem como informações sobre com que frequência eles ocorrem no conjunto de dados.
4. Na seção **Detalhes**, clique duas vezes em **Rubi**.
O Visualizador de Dados é executado e exibe os registros em que a coluna CustomerTier contém o valor **Rubi**.
5. Na seção **Criação de Perfil de Coluna**, clique na coluna **OrderAmount**.
6. Na seção **Detalhes**, clique na lista **Mostrar** e selecione **Padrões**.
A seção **Detalhes** mostra os padrões encontrados na coluna OrderAmount. A string **9 (5)** na coluna Padrão se refere aos registros que contêm quantidades de pedidos com cinco dígitos. A string **9 (4)** se refere a registros que contêm quantidades de pedidos com quatro dígitos.
7. Na coluna **Padrão**, clique duas vezes na string **9 (4)**.
O Visualizador de Dados é executado e exibe os registros em que a coluna OrderAmount contém uma quantidade de pedido com quatro dígitos.
8. Na seção **Detalhes**, clique na lista **Mostrar** e selecione **Estatísticas**.
A seção **Detalhes** mostra as estatísticas da coluna OrderAmount, incluindo o valor médio, o desvio padrão, os comprimentos máximo e mínimo, os cinco valores mais comuns e os cinco valores menos comuns.

Resumo da Criação de Perfil

Nessa lição, você aprendeu que um perfil fornece informações sobre o conteúdo e a estrutura dos dados.

Também aprendeu que é possível executar uma análise de associação nos dois objetos de dados e exibir o grau de sobreposição entre os objetos de dados. Você aprendeu que pode executar um perfil da coluna em um objeto de dados e exibir os valores, os padrões e as estatísticas que se relacionam a cada coluna no objeto de dados.

Você criou o perfil `JoinAnalysis` para determinar se os dados do objeto de dados `Boston_Customers` podem se mesclar aos dados no objeto de dados `LA_Customers`. Também exibiu os resultados do perfil e determinou que todos os valores na coluna `CustomerID` são exclusivos e que é possível mesclar os objetos de dados com êxito.

Você criou o perfil `All_Customers` e executou um perfil da coluna no objeto de dados `All_Customers`. Ainda exibiu os resultados do perfil para descobrir os valores, os padrões e as estatísticas para colunas no objeto de dados `All_Customers`. Por fim, executou o Visualizador de Dados para exibir linhas com os valores e padrões que você selecionou, permitindo que a qualidade de dados fosse verificada.

CAPÍTULO 13

Lição 4. Analisando dados

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Análise de Dados, 64](#)
- [Tarefa 1. Criar um Objeto de Dados de Destino, 65](#)
- [Tarefa 2. Criar um Mapeamento para Analisar Dados, 67](#)
- [Tarefa 3. Executar um Perfil na Transformação do Analisador, 69](#)
- [Tarefa 4. Executar o Mapeamento, 69](#)
- [Tarefa 5. Exibir a Saída de Mapeamento, 69](#)
- [Analisando dados de Resumo, 70](#)

Visão Geral da Análise de Dados

Analise dados para identificar um ou mais elementos de dados em um campo de entrada e para gravar cada elemento em um campo de saída diferente.

A análise permite que você tenha maior controle sobre as informações em cada coluna. Por exemplo, considere um campo de dados que contém um nome completo, `Bob Smith`. Você pode usar a transformação do Analisador para dividir o nome completo em colunas de dados separadas para o nome e o sobrenome. Depois de você analisar os dados em novas colunas, poderá criar operações de qualidade de dados personalizados para cada coluna.

Você pode configurar a transformação do Analisador para usar *Conjuntos de Tokens* para analisar colunas de dados em strings do componente. Um conjunto de tokens identifica elementos de dados, como palavras, CEPs, números de telefone e CPFs.

Também pode usar a transformação do Analisador para analisar dados que correspondam às entradas da tabela de referência ou expressões regulares personalizadas que você inserir.

História

A HypoStores deseja formatar os arquivos de dados de clientes do escritório de Los Angeles para que correspondam ao formato de arquivos de dados do escritório de Boston. Os dados do cliente do escritório de Los Angeles armazenam o nome do cliente em uma coluna `FullName`, enquanto os dados do cliente do escritório de Boston armazenam o nome do cliente em colunas separadas `FirstName` e `LastName`. A HypoStores precisa analisar os dados de coluna `FullName` de Los Angeles em nomes e sobrenomes para que o formato dos dados de Los Angeles correspondam ao formato dos dados de Boston.

Objetivos

Nesta lição, você concluirá as seguintes tarefas:

- Criar e configurar um objeto de dados `LA_Customers_tgt` para conter os dados analisados.
- Criar um mapeamento para analisar a coluna `FullName` em colunas `FirstName` e `LastName` separadas.
- Adicionar o objeto de dados `LA_Customers` ao mapeamento para estabelecer conexão com os dados de origem.
- Adicionar o objeto de dados `LA_Customers_tgt` para o mapeamento para criar um objeto de dados de destino.
- Adicionar uma transformação do Analisador para o mapeamento e configurá-lo para usar um conjunto de tokens para analisar nomes completos em nomes e sobrenomes.
- Executar um perfil na transformação do Analisador para analisar os dados antes de gerar a fonte de dados de destino.
- Executar o mapeamento para gerar os nomes analisados.
- Executar o Visualizador de Dados para exibir a saída de mapeamento.

Pré-requisitos

Antes de iniciar esta lição, verifique o seguinte pré-requisito:

- Você concluiu as lições 1 e 2 neste tutorial.

Tempo

Reserve 20 minutos para concluir as tarefas nesta lição.

Tarefa 1. Criar um Objeto de Dados de Destino

Nessa tarefa, um objeto de dados `LA_Customers_tgt`, em que os nomes analisados podem ser gravados, será criado.

Para criar um objeto de dados de destino, execute as etapas a seguir:

1. Crie um objeto de dados `LA_Customers_tgt` com base no arquivo `LA_Customers.csv`.
2. Configure as opções de leitura e gravação para o objeto de dados, incluindo as localizações e nomes de arquivo.
3. Adicione as colunas `Firstname` e `Lastname` ao objeto de dados `LA_Customers_tgt`.

Etapas 1. Criar um Objeto de Dados `LA_Customers_tgt`

Nesta etapa, um objeto de dados `LA_Customers_tgt` com base no arquivo `LA_Customers.csv` será criado.

1. Clique em **Arquivo > Novo > Objeto de Dados**.
A janela **Novo** é exibida.
2. Selecione **Objeto de Dados de Arquivo Simples** e clique em **Avançar**.
3. Verifique se a opção **Criar a partir de um arquivo simples existente** está selecionada.
4. Clique em **Procurar** e navegue até `LA_Customers.csv` no seguinte diretório: `<Diretório de Instalação da Informatica>\clients\DeveloperClient\Tutorials`

5. Clique em **Abrir**.
6. No campo **Nome**, insira `LA_Customers_tgt`.
7. Clique em **Avançar**.
8. Clique em **Avançar**.
9. Na seção **Visualizar Opções**, selecione **Importar nomes de coluna da primeira linha** e clique em **Avançar**.
10. Clique em **Concluir**.
O objeto de dados `LA_Customers_tgt` é exibido no Editor.

Etapa 2. Configurar Opções de Leitura e Gravação

Nesta etapa, as opções de leitura e gravação do objeto de dados `LA_Customers_tgt`, incluindo as localizações e nomes de arquivo, são configuradas.

1. Verifique se o objeto de dados `LA_Customers_tgt` está aberto no editor.
2. No editor, selecione a exibição **Leitura**.
3. Clique em **> Mostrar Exibição > Propriedades**.
4. Na exibição **Propriedades**, selecione a exibição **Tempo de Execução**.
5. Na coluna **Valor**, clique duas vezes no nome do arquivo de origem e insira `LA_Customers_tgt.csv`.
6. Na coluna **Valor**, clique duas vezes para destacar o diretório do arquivo de origem.
7. Clique com o botão direito no nome destacado e selecione **Copiar**.
8. No editor, selecione a exibição **Gravar**.
9. Na exibição **Propriedades**, selecione a exibição **Tempo de Execução**.
10. Na coluna **Valor**, clique duas vezes na entrada **Diretório do Arquivo de Saída**.
11. Clique duas vezes e selecione **Colar** para colar a localização do diretório que você copiou da exibição **Leitura**.
12. Na coluna **Valor**, clique duas vezes na entrada **Opções de Cabeçalho** e selecione **Nomes de Campo de Saída**.
13. Na coluna **Valor**, clique duas vezes na entrada **Nome de Arquivo de Saída** e insira `LA_Customers_tgt.cs`.
14. Clique em **Arquivo > Salvar** para salvar o objeto de dados.

Etapa 3. Adicionar Colunas ao Objeto de Dados

Nesta etapa, você adicionará as colunas `Firstname` e `Lastname` ao objeto de dados `LA_Customers_tgt`.

1. Na exibição do Object Explorer, navegue para os objetos de dados em seu projeto de tutorial.
2. Clique duas vezes no objeto de dados `LA_Customers_tgt`.
O objeto de dados `LA_Customers_tgt` será aberto no editor.
3. Verifique se a exibição **Visão Geral** está selecionada.
4. Selecione a coluna **FullName** e clique no botão **Nova** para adicionar uma coluna.
Uma coluna chamada `FullName1` será exibida.
5. Renomeie a coluna para `Firstname`. Clique no campo **Precisão** e digite "30."
6. Selecione a coluna `Firstname` e clique no botão **Nova** para adicionar uma coluna.

Uma coluna chamada `FirstName1` será exibida.

7. Renomeie a coluna para `LastName`. Clique no campo **Precisão** e digite "30."
8. Clique em **Arquivo > Salvar** para salvar o objeto de dados.

Tarefa 2. Criar um Mapeamento para Analisar Dados

Nessa tarefa, um mapeamento é criado e configurado para usar objetos de dados e uma transformação do Analisador.

Para criar um mapeamento para analisar dados, conclua as etapas a seguir:

1. Crie um mapeamento.
2. Adicione objetos de dados de origem e destino ao mapeamento.
3. Adicionar uma transformação do Analisador ao mapeamento.
4. Configurar a transformação do Analisador para analisar a coluna de origem que contém o nome completo do cliente em colunas de destino separadas contendo o nome e sobrenome.

Etapa 1. Criar um Mapeamento

Nesta etapa, o nome do mapeamento será criado.

1. Na exibição do Object Explorer, selecione o projeto do tutorial.
2. Clique em **Arquivo > Novo > Mapeamento**.
A janela **Novo Mapeamento** é exibida.
3. No campo **Nome**, digite `ParserMapping`.
4. Clique em **Concluir**.
O mapeamento é aberto no editor.

Etapa 2. Adicionar Objetos de Dados ao Mapeamento

Nesta etapa, o objeto de dados `LA_Customers` e o objeto de dados `LA_Customers_tgt` serão adicionados ao mapeamento.

1. Na exibição do Object Explorer, navegue para os objetos de dados em seu projeto de tutorial.
2. Selecione o objeto de dados `LA_Customers` e arraste-o para o editor.
A janela **Adicionar Objeto de Dados Físicos ao Mapeamento** será exibida.
3. Verifique se a opção **Leitura** está selecionada e clique em **OK**.
O objeto de dados será exibido no editor.
4. Na exibição do Object Explorer, navegue para os objetos de dados em seu projeto de tutorial.
5. Selecione o objeto de dados `LA_Customers_tgt` e arraste-o para o editor.
A janela **Adicionar Objeto de Dados Físicos ao Mapeamento** será exibida.
6. Selecione **Gravar** e clique em **OK**.
O objeto de dados será exibido no editor.

7. Selecione as portas `CustomerID`, `CustomerTier` e `FullName` no objeto de dados `LA_Customers`. Arraste as portas para a porta `CustomerID` no objeto de dados `LA_Customers_tgt`.

Sugestão: Mantenha pressionada a tecla **CTRL** para selecionar várias portas.

As portas do objeto de dados `LA_Customers` se conectam às portas correspondentes no objeto de dados `LA_Customers_tgt`.

Etapa 3. Adicionar uma Transformação do Analisador ao Mapeamento

Nesta etapa, uma transformação do Analisador ao mapeamento `ParserMapping` será adicionada.

1. Selecione o editor que contém o mapeamento `ParserMapping`.
2. Na paleta de Transformação, selecione a transformação do Analisador.
3. Clique no editor.
A janela **Nova Transformação do Analisador** é exibida.
4. Verifique se a opção **Analisador de Token** está selecionada e clique em **Concluir**.
A transformação do Analisador é exibida no editor.
5. Selecione a porta `FullName` no objeto de dados `LA_Customers` e arraste-a para o grupo **Entrada** da transformação do Analisador.
A porta `FullName` é exibida na transformação do Analisador e está conectada à porta `FullName` no objeto de dados.

Etapa 4. Configurar a Transformação do Analisador

Nesta etapa, a transformação do Analisador será configurada para analisar a coluna que contém o nome completo do cliente em colunas separadas que contenham o nome e o sobrenome.

1. Selecione o editor que contém o mapeamento `ParserMapping`.
2. Clique na transformação do Analisador.
3. Clique em **> Mostrar Exibição > Propriedades**.
4. Na exibição **Propriedades**, selecione a exibição **Estratégias**.
5. Clique em **Novo**. O assistente **Nova Estratégia** será exibido.
6. Clique na seta de seleção na coluna **Entradas** e selecione a porta `FullName`.
7. Selecione o delimitador de espaço de caractere `[\s]`.
8. Clique em **Avançar**.
9. Selecione a operação **Analisar Usando o Conjunto de Token** e clique em **Avançar**.
10. Selecione **Conjuntos de Tokens Fixos (Somente Saída Única)** e escolha o conjunto de token **Indefinido**.
11. Clique no campo **Saídas** e selecione **Novo**.
12. Na caixa de diálogo **Saídas de Operação**, altere o nome de saída para `Undefined_Output`.
13. Clique em **Concluir**.
14. Na transformação do Analisador, clique na porta `Undefined_Output` e arraste-a para a porta `FirstName` no objeto de dados `LA_customers_tgt`.
Uma conexão é exibida entre as portas.

15. Na transformação do Analisador, clique na porta `OverflowField` e arraste-a para a porta `LastName` no objeto de dados `LA_customers_tgt`.
Uma conexão é exibida entre as portas.
16. Clique em **Arquivo > Salvar** para salvar o mapeamento.

Tarefa 3. Executar um Perfil na Transformação do Analisador

Nessa tarefa, um perfil na transformação do Analisador será executado para verificar se você configurou a transformação do Analisador para analisar o nome completo corretamente.

1. Selecione o editor que contém o mapeamento `ParserMapping`.
2. Clique com o botão direito na transformação do Analisador e selecione **Criar Perfil Agora**.
O perfil será executado e aberto no editor.
3. No editor, clique na exibição **Resultados** para exibir o resultado da operação de criação de perfil.
4. Selecione a coluna `Undefined_output` para exibir informações sobre a coluna na seção **Detalhes**.
Os valores contidos na coluna `Undefined_output` são exibidos na seção **Detalhes**, juntamente com as estatísticas de frequência e porcentagem para cada valor.
5. Exiba os dados e verifique se somente os nomes são exibidos na coluna `Undefined_output`.

Tarefa 4. Executar o Mapeamento

Nessa tarefa, o mapeamento será executado para criar a saída de mapeamento.

1. Selecione o editor que contém o mapeamento `ParserMapping`.
2. Clique em **Executar > Executar Mapeamento**.
O mapeamento é executado e grava a saída no arquivo `LA_Customers_tgt.csv`.

Tarefa 5. Exibir a Saída de Mapeamento

Nessa tarefa, o Visualizador de Dados será executado para exibir a saída de mapeamento.

1. Na exibição do Object Explorer, localize o objeto de dados `LA_Customers_tgt` em seu projeto de tutorial e clique duas vezes no objeto de dados.
O objeto de dados será aberto no editor.
2. Clique em **> Mostrar Exibição > Visualizador de Dados**.
A exibição do Visualizador de Dados será aberta.
3. Na exibição do Visualizador de Dados, clique em **Executar**.
O Visualizador de Dados é executado e exibe os dados.

4. Verifique se as colunas `FirstName` e `LastName` exibem corretamente os dados analisados.

Analizando dados de Resumo

Nessa lição, você aprendeu que a análise de dados identifica os elementos de dados em um campo de entrada e grava cada elemento em uma nova coluna.

Também aprendeu que você usa a transformação do Analisador para analisar dados. Aprendeu, ainda, que você pode criar um perfil para uma transformação em um mapeamento para analisar a saída da transformação. Por fim, aprendeu que é possível exibir a saída de mapeamento usando o Visualizador de Dados.

Você criou e configurou o objeto de dados `LA_Customers_tgt` para conter a saída analisada. Também criou um mapeamento para analisar os dados. Nesse mapeamento, você configurou uma transformação do Analisador com um conjunto de tokens para analisar nomes e sobrenomes da coluna `FullName` no arquivo de cliente de Los Angeles. Configurou o mapeamento para gravar os dados analisados nas colunas `Firstname` e `Lastname` no objeto de dados `LA_Customers_tgt`. Você também executou um perfil para exibir a saída da transformação antes de executar o mapeamento. Por fim, executou o mapeamento e usou o Visualizador de Dados para exibir as novas colunas de dados no objeto de dados `LA_Customers_tgt`.

CAPÍTULO 14

Lição 5. Padronizar Dados

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Padronização de Dados, 71](#)
- [Tarefa 1. Criar um Objeto de Dados de Destino, 72](#)
- [Tarefa 2. Criar um Mapeamento para Padronizar Dados, 73](#)
- [Tarefa 3. Executar o Mapeamento, 76](#)
- [Tarefa 4. Exibir a Saída de Mapeamento, 76](#)
- [Resumo de Padronização de Dados, 76](#)

Visão Geral da Padronização de Dados

A padronização de dados melhora a qualidade de dados por remover erros e inconsistências nos dados.

Para melhorar a qualidade de dados, padronize os dados que contenham os seguintes tipos de valores:

- Valores incorretos
- Valores com as informações corretas no formato incorreto
- Valores dos quais você deseja gerar novas informações

Use a transformação do Padronizador para procurar esses valores nos dados. Você pode escolher um dos seguintes tipos de operação de pesquisa:

- **Texto.** Pesquise strings personalizadas que você inserir. Remova essas strings ou as substitua por texto personalizado.
- **Tabela de referência.** Pesquise strings contidas em uma tabela de referência que você selecionar. Remova essas strings ou as substitua por entradas de tabela de referência ou texto personalizado.

Por exemplo, você pode configurar a transformação do Padronizador para padronizar dados de endereço que contenham as strings personalizadas `Rua` e `R.` usando a string de substituição `R.` A transformação do Padronizador substitui os termos de pesquisa com o termo `R.` e grava o resultado em uma nova coluna de dados.

História

A HypoStores precisa padronizar os dados de endereço do cliente de forma que todos os endereços usem termos consistentemente. Os dados de endereço no objeto de dados `All_Customers` contêm entradas formatadas de modo inconsistente para termos comuns como `Rua`, `Bulevar`, `Avenida`, `Estrada` e `Parque`.

Objetivos

Nesta lição, você concluirá as seguintes tarefas:

- Criar e configurar um objeto de dados `All_Customers_Stdz_tgt` para conter dados padronizados.
- Criar um mapeamento para padronizar os termos de endereço `Rua, Bulevar, Avenida, Estrada e Parque` para um formato consistente.
- Adicionar o objeto de dados `All_Customers` ao mapeamento para estabelecer conexão com os dados de origem.
- Adicionar o objeto de dados `All_Customers_Stdz_tgt` ao mapeamento para criar um objeto de dados de destino.
- Adicionar uma transformação do Padronizador ao mapeamento e configurá-la para padronizar os termos de endereço.
- Executar o mapeamento para gerar dados de endereço padronizados.
- Executar o Visualizador de Dados para exibir a saída de mapeamento.

Pré-requisitos

Antes de iniciar esta lição, verifique o seguinte pré-requisito:

- Você concluiu as lições 1 e 2 neste tutorial.

Tempo

Defina 15 minutos para concluir esta lição.

Tarefa 1. Criar um Objeto de Dados de Destino

Nessa tarefa, um objeto de dados `All_Customers_Stdz_tgt` em que você pode gravar dados padronizados será criado.

Para criar um objeto de dados de destino, execute as etapas a seguir:

1. Crie um objeto de dados `All_Customers_Stdz_tgt` com base no arquivo `All_Customers.csv`.
2. Configure as opções de leitura e gravação para o objeto de dados, incluindo as localizações e nomes de arquivo.

Etapa 1. Criar um objeto de dados `All_Customers_Stdz_tgt`

Nesta etapa, um objeto de dados `All_Customers_Stdz_tgt` com base no arquivo `All_Customers.csv` será criado.

1. Clique em **Arquivo > Novo > Objeto de Dados**.
A janela **Novo** é exibida.
2. Selecione **Objeto de Dados de Arquivo Simples** e clique em **Avançar**.
3. Verifique se a opção **Criar a partir de um arquivo simples existente** está selecionada.
4. Clique em **Procurar** e navegue até `All_Customers.csv` no seguinte diretório: `<Diretório de Instalação da Informatica>\clients\DeveloperClient\Tutorials`
5. Clique em **Abrir**.
6. No campo **Nome**, insira `All_Customers_Stdz_tgt`.

7. Clique em **Avançar**.
8. Clique em **Avançar**.
9. Na seção **Visualizar Opções**, selecione **Importar nomes de coluna da primeira linha** e clique em **Avançar**.
10. Clique em **Concluir**.
O objeto de dados `All_Customers_Stdz_tgt` é exibido no Editor.

Etapa 2. Configurar Opções de Leitura e Gravação

Nesta etapa, as opções de leitura e gravação para o objeto de dados `All_Customers_Stdz_tgt` será configurada, incluindo as localizações e nomes de arquivo.

1. Verifique se o objeto de dados `All_Customers_Stdz_tgt` está aberto no editor.
2. No editor, selecione a exibição **Leitura**.
3. Clique em **> Mostrar Exibição > Propriedades**.
4. Na exibição **Propriedades**, selecione a exibição **Tempo de Execução**.
5. Na coluna **Valor**, clique duas vezes no nome do arquivo de origem e insira `All_Customers_Stdz_tgt.csv`.
6. Na coluna **Valor**, clique duas vezes na entrada **Diretório do Arquivo de Origem**.
7. Clique com o botão direito no nome destacado e selecione **Copiar**.
8. No editor, selecione a exibição **Gravar**.
9. Na exibição **Propriedades**, selecione a exibição **Tempo de Execução**.
10. Na coluna **Valor**, clique duas vezes na entrada **Diretório do Arquivo de Saída**.
11. Clique duas vezes e selecione **Colar** para colar a localização do diretório que você copiou da exibição **Leitura**.
12. Na coluna **Valor**, clique duas vezes na entrada **Opções de Cabeçalho** e selecione **Nomes de Campo de Saída**.
13. Na coluna **Valor**, clique duas vezes na entrada **Nome de Arquivo de Saída** e insira `All_Customers_Stdz_tgt.csv`.
14. Clique em **Arquivo > Salvar** para salvar o objeto de dados.

Tarefa 2. Criar um Mapeamento para Padronizar Dados

Nessa tarefa, um mapeamento é criado e configurado para usar objetos de dados e uma transformação do Padronizador.

Para criar um mapeamento para padronizar dados, conclua as etapas a seguir:

1. Crie um mapeamento.
2. Adicione objetos de dados de origem e destino ao mapeamento.
3. Adicione uma transformação do Padronizador ao mapeamento.
4. Configure a transformação do Padronizador para padronizar termos de endereço comuns aos formatos consistentes.

Etapa 1. Criar um Mapeamento

Nesta etapa, o nome do mapeamento será criado.

1. Na exibição do **Object Explorer**, selecione o seu projeto de tutorial.
2. Clique em **Arquivo > Novo > Mapeamento**.
A janela **Novo Mapeamento** é exibida.
3. No campo **Nome**, insira `StandardizerMapping`.
4. Clique em **Concluir**.
O mapeamento é aberto no editor.

Etapa 2. Adicionar Objetos de Dados ao Mapeamento

Nesta etapa, você adicionará o objeto de dados `All_Customers` e o objeto de dados `All_Customers_Stdz_tgt` ao mapeamento.

1. Na exibição do **Object Explorer**, navegue para os objetos de dados em seu projeto de tutorial.
2. Selecione o objeto de dados `All_Customers` e arraste-o para o editor.
A janela **Adicionar Objeto de Dados Físicos ao Mapeamento** será exibida.
3. Verifique se a opção **Leitura** está selecionada e clique em **OK**.
O objeto de dados será exibido no editor.
4. Na exibição do **Object Explorer**, navegue para os objetos de dados em seu projeto de tutorial.
5. Selecione o objeto de dados `All_Customers_Stdz_tgt` e arraste-o para o editor.
A janela **Adicionar Objeto de Dados Físicos ao Mapeamento** será exibida.
6. Selecione **Gravar** e clique em **OK**.
O objeto de dados será exibido no editor.
7. Selecione todas as portas no objeto de dados `All_Customers`. Arraste as portas para a porta `CustomerID` no objeto de dados `All_Customers_Stdz_tgt`.
Sugestão: Mantenha pressionada a tecla Shift para selecionar várias portas. Talvez seja necessário rolar a lista de portas para selecionar todas elas.
As portas do objeto de dados `All_Customers` se conectam com as portas correspondentes no objeto de dados `All_Customers_Stdz_tgt`.

Etapa 3. Adicionar uma Transformação do Padronizador ao Mapeamento

Nesta etapa, uma transformação do Padronizador será adicionada para padronizar strings nos dados de endereço.

1. Selecione o editor que contém o mapeamento `StandardizerMapping`.
2. Na paleta de Transformação, selecione a transformação do Padronizador.
3. Clique no editor.
Uma transformação do Padronizador chamada de `NewStandardizer` será exibida no mapeamento.
4. Para renomear a transformação do Padronizador, clique duas vezes na barra de título da transformação e insira `AddressStandardizer`.

5. Selecione a porta `Address1` no objeto de dados `All_Customers` e a arraste para o grupo **Entrada** da transformação do Padronizador.

Uma porta denominada `Address1` será exibida no grupo de entrada. A porta se conecta com a porta `Address1` no objeto de dados `All_Customers`.

Nota: Você adicionará uma porta de saída para a transformação quando configurar uma estratégia de padronização.

Etapa 4. Configurar a Transformação do Padronizador

Nesta etapa, a transformação do Padronizador para padronizar os termos de endereço nos dados de origem será configurada.

Nota: Você definirá cinco operações de padronização nesta tarefa. Cada operação substitui uma string na coluna de entrada com uma nova string.

1. Selecione o editor que contém o mapeamento `StandardizerMapping`.
2. Clique em transformação do Padronizador.
3. Clique em **> Mostrar Exibição > Propriedades**.
4. Na exibição Propriedades, selecione **Estratégias**.
5. Clique em **Novo**. O assistente Nova Estratégia será exibido.
6. Clique na seta na coluna Entradas e selecione a porta de entrada `Address1`.
O campo Saída mostra `Address1` como a porta de saída.
7. Selecione o espaço de caractere e os delimitadores de vírgula `[\s]` e `[,]`. Opcionalmente, selecione as opções para remover espaços em branco.
8. Clique em **Avançar**.
9. Selecione a operação **Substituir strings personalizadas** e clique em **Avançar**.
10. Em **Propriedades**, clique em **Novo**.
11. Edite os campos **Strings Personalizadas** e **Substituir por** de forma que eles contenham o primeiro par de strings da tabela a seguir:

Strings Personalizadas	Substituir por
STREET	ST.
BOULEVARD	BLVD.
AVENUE	AVE.
DRIVE	DR.
PARK	PK.

12. Repita as etapas de 9 até 12 para definir as operações de padronização para todas as strings na tabela.
13. Arraste a porta de saída `Address1` para a porta `Address1` no objeto de dados `All_Customers_Stdz_tgt`.
14. Clique em **Arquivo > Salvar** para salvar o mapeamento.

Tarefa 3. Executar o Mapeamento

Nessa tarefa, o mapeamento para gravar endereços padronizados no objeto de dados de saída será executado.

1. Selecione o editor que contém o mapeamento `StandardizerMapping`.
2. Clique em **Executar** > **Executar Mapeamento**.
O mapeamento é executado e grava a saída no arquivo `All_Customers_Stdz_tgt.csv`.

Tarefa 4. Exibir a Saída de Mapeamento

Nessa tarefa, o visualizador de dados para exibir a saída de mapeamento e verificar se os dados de endereço estão corretamente padronizados é executado.

1. Na exibição do **Object Explorer**, localize o objeto de dados `All_Customers_Stdz_tgt` em seu projeto de tutorial e clique duas vezes no objeto de dados.
O objeto de dados será aberto no editor.
2. Clique em > **Mostrar Exibição** > **Visualizador de Dados**.
A exibição do Visualizador de Dados será aberta.
3. Na exibição do Visualizador de Dados, clique em **Executar**.
O Visualizador de Dados exibe a saída de mapeamento.
4. Verifique se a coluna `Address1` exibe corretamente os dados padronizados. Por exemplo, todas as instâncias da string `STREET` devem ser substituídas com a string `ST`.

Resumo de Padronização de Dados

Nessa lição, você aprendeu que pode padronizar dados para remover erros e inconsistências.

Também aprendeu que pode usar uma transformação do Padronizador para padronizar strings em uma coluna de entrada. Você também aprendeu que pode exibir a saída de mapeamento usando o Visualizador de Dados.

Criou e configurou o objeto de dados `All_Customers_Stdz_tgt` para conter a saída padronizada. Você criou um mapeamento para padronizar os dados. Nesse mapeamento, você configurou uma transformação do Padronizador para padronizar a coluna `Address1` no objeto de dados `All_Customers`. Você configurou o mapeamento para gravar a saída padronizada no objeto de dados `All_Customers_Stdz_tgt`. Por fim, você executou o mapeamento e usou o Visualizador de Dados para exibir os dados padronizados no objeto de dados `All_Customers_Stdz_tgt`.

CAPÍTULO 15

Lição 6. Validar Dados de Endereço

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Validação de Dados de Endereço, 77](#)
- [Tarefa 1. Criar um Objeto de Dados de Destino, 78](#)
- [Tarefa 2. Criar um Mapeamento para Validar Endereços, 80](#)
- [Tarefa 3. Configurar a Nova Transformação do Validador de Endereço, 81](#)
- [Tarefa 4. Executar o Mapeamento, 84](#)
- [Tarefa 5. Exibir a Saída de Mapeamento, 85](#)
- [Resumo da Validação dos Dados de Endereço, 87](#)

Visão Geral da Validação de Dados de Endereço

A validação de endereço é o processo de avaliação e aprimoramento de qualidade de endereços postais. Ela avalia a qualidade do endereço comparando endereços de entrada com um conjunto de dados de referência de endereços válidos. Também melhora a qualidade do endereço identificando valores de endereço incorretos e usando o conjunto de dados de referência para criar campos que contêm os valores corretos.

Um endereço é válido quando ele tem capacidade de recebimento de entrega. Um endereço pode ser bem formatado e conter informações reais de rua, cidade e CEP, mas se os dados não resultam em um endereço com capacidade de recebimento de entrega, então não é válido. A ferramenta Developer usa conjuntos de dados de referência de endereço para verificar a capacidade de recebimento de entrega de endereços de entrada. A Informatica fornece conjuntos de dados de referência de endereço.

Um conjunto de dados de referência de endereço contém dados que descrevem todos os endereços com capacidade de recebimento de entrega em um país. O processo de validação de endereço pesquisa o conjunto de dados de referência para o endereço que mais se assemelha aos dados de endereço de entrada. Se o processo localiza uma correspondência próxima no conjunto de dados de referência, ele grava novos valores nos valores de dados incorretos ou incompletos. O processo cria um conjunto de códigos alfanuméricos que descrevem o tipo de correspondência encontrada entre o endereço de entrada e os endereços de referência. Ele também pode reestruturar o endereço e adicionar informações que estão ausentes do endereço de entrada, como um CEP de quatro dígitos para um endereço nos Estados Unidos.

Use a transformação do Validador de Endereço para criar os processos de validação de endereço na ferramenta Developer. Essa transformação de vários grupos contém um conjunto de portas de entrada e saída predefinidas que correspondem a todos os campos possíveis em um endereço de entrada. Quando

you configure a transformation of the Address Validator, select the set of reference data and you can create an input and output address structure using the transformation ports. In this lesson, you will configure the transformation for address data from the United States.

History

HypoStores needs to correct and fill in the address data to ensure that its direct mail and other correspondence items reach the customers. Correcting and completing address data also reduces the cost of mailing operations for the organization. In addition, HypoStores needs the customer data to include the addresses in a printable format that is flexible enough to include addresses of different sizes.

To meet these commercial requirements, the HypoStores ICC team will create a validation mapping of address data in the Developer tool.

Objectives

In this lesson, you will complete the following tasks:

- Create a target data object that contains the validated address fields and correspondence codes.
- Create a mapping with an origin data object, a target data object and an Address Validator transformation.
- Configure the Address Validator transformation to validate the customer address data.
- Execute the mapping to validate the address data and analyze the output code of correspondence to verify the validity of the address data.

Prerequisites

Before starting this lesson, verify the following prerequisites:

- You completed lessons 1 and 2 in this tutorial.
- The United States reference address data is installed in the domain and registered with the Administrator tool. Contact the Informatica administrator to verify if the United States address data is installed in the system. The reference data is installed through the Data Quality Content Installer.

Time

Allocate 25 minutes to complete this lesson.

Task 1. Create a Target Data Object

In this task, you will create a target data object, configure the save options and add ports.

To create and configure the target data object, complete the following steps:

1. Create a target data object `All_Customers_av_tgt` with base on the `All_Customers.csv` file.
2. Configure the read and save options for the data object, including the file locations and file names.
3. Add ports to the data object to receive the correspondence code values generated by the Address Validator transformation.

Etapa 1. Criar o Objeto de Dados All_Customers_av_tgt

Nesta etapa, o objeto de dados All_Customers_av_tgt com base no arquivo All_Customers.csv será criado.

1. Clique em **Arquivo > Novo > Objeto de Dados**.
A janela **Novo** é exibida.
2. Selecione **Objeto de Dados de Arquivo Simples** e clique em **Avançar**.
3. Verifique se a opção **Criar a partir de um arquivo simples existente** está selecionada. Clique em **Procurar** ao lado da seleção, localize o arquivo All_Customers.csv e clique em **Abrir**.
4. No campo **Nome**, insira All_Customers_av_tgt.
5. Clique em **Avançar**.
6. Clique em **Avançar**.
7. Na seção **Visualizar Opções**, selecione **Importar nomes de coluna da primeira linha** e clique em **Avançar**.
8. Clique em **Concluir**.
O objeto de dados All_Customers_av_tgt é exibido no editor.

Etapa 2. Configurar Opções de Leitura e Gravação

Nesta etapa, as opções para o objeto de dados All_Customers_av_tgt, incluindo o nome e a localização do arquivo de destino, são configuradas para leitura e gravação.

1. Verifique se o objeto de dados All_Customers_av_tgt está aberto no editor.
2. No editor, selecione a exibição **Leitura**.
3. Selecione **> Mostrar Exibição > Propriedades**.
4. Na exibição **Propriedades**, selecione a exibição **Tempo de Execução**.
5. Na coluna **Valor**, clique duas vezes no nome do arquivo de origem e tipo All_Customers_av_tgt.csv.
6. Na coluna **Valor**, clique duas vezes para destacar o caminho do diretório do arquivo de origem.
7. Clique com o botão direito no caminho e nome destacados e selecione **Copiar**.
8. No editor, selecione a exibição **Gravar**.
9. Na exibição **Propriedades**, selecione a exibição **Tempo de Execução**.
10. Na coluna **Valor**, clique duas vezes na entrada **Diretório do Arquivo de Saída**.
11. Clique com o botão direito nessa entrada e selecione **Colar** para adicionar o caminho copiado da exibição **Ler**.
12. Na coluna **Valor**, clique duas vezes na entrada **Opções de Cabeçalho** e selecione **Nomes de Campo de Saída**.
13. Na coluna **Valor**, clique duas vezes na entrada **Nome de arquivo de saída** e digite All_Customers_av_tgt.csv.
14. Selecione **Arquivo > Salvar** para salvar o objeto de dados.

Etapa 3. Adicionar Portas ao Objeto de Dados

Nesta etapa, duas portas serão adicionadas ao objeto de dados `All_Customers_av_tgt` de forma que a transformação do Validador de Endereço possa gravar os valores do código de correspondência no arquivo de destino. Nomeie as portas `MailabilityScore` e `MatchCode`.

O valor `MailabilityScore` descreve a capacidade de entrega do endereço de entrada. O valor `MatchCode` descreve o tipo de correspondência que a transformação faz entre o endereço de entrada e os endereços de dados de referência.

1. Na exibição do **Object Explorer**, navegue para os objetos de dados em seu projeto de tutorial.
2. Clique duas vezes no objeto de dados `All_Customers_av_tgt`.
O objeto de dados `All_Customers_av_tgt` será aberto no editor.
3. Verifique se a opção **Visão Geral** está selecionada.
4. Selecione a porta final na lista de porta. Esta porta é denominada `MiscDate1`.
5. Clique em **Novo**.
Uma porta denominada `MiscDate1` é exibida.
6. Renomeie a porta `MiscDate1` para `MailabilityScore`.
7. Selecione a porta `MailabilityScore`.
8. Clique em **Novo**.
Uma porta denominada `MailabilityScore1` é exibida.
9. Renomeie a porta `MailabilityScore1` para `MatchCode`.
10. Clique em **Arquivo > Salvar** para salvar o objeto de dados.

Tarefa 2. Criar um Mapeamento para Validar Endereços

Nessa tarefa, você criará um mapeamento e adicionará objetos de dados e uma transformação do Validador de Endereço.

Para criar o mapeamento e adicionar os objetos de que você precisa, conclua as etapas a seguir:

1. Crie um objeto de mapeamento.
2. Adicione objetos de dados de origem e destino ao mapeamento.
3. Adicione uma transformação do Validador de Endereço ao mapeamento.

Etapa 1. Criar um Mapeamento

Nesta etapa, o nome do mapeamento será criado.

1. Na exibição do **Object Explorer**, selecione o projeto do tutorial.
2. Selecione **Arquivo > Novo > Mapeamento**.
A janela **Novo Mapeamento** é exibida.
3. No campo **Nome**, insira `ValidateAddresses`.

4. Clique em **Concluir**.
O mapeamento é aberto no editor.

Etapa 2. Adicionar Objetos de Dados ao Mapeamento

Nesta etapa, os objetos de dados de origem e destino são adicionados ao mapeamento.

`All_Customers` é o objeto de dados de origem do mapeamento. A transformação do Validador de Endereço lê os dados deste objeto. `All_Customers_av_tgt` é o objeto de destino de dados do mapeamento. Este objeto lê dados da transformação do Validador de Endereço.

1. Na exibição do **Object Explorer**, navegue para os objetos de dados em seu projeto de tutorial.
2. Selecione o objeto de dados `All_Customers` e arraste-o para o editor.
A janela **Adicionar Objeto de Dados Físicos ao Mapeamento** será exibida.
3. Verifique se a opção **Leitura** está selecionada e clique em **OK**.
O objeto de dados será exibido no editor.
4. Na exibição do **Object Explorer**, navegue para os objetos de dados em seu projeto de tutorial.
5. Selecione o objeto de dados `All_Customers_av_tgt` arraste-o para o editor.
A janela **Adicionar Objeto de Dados Físicos ao Mapeamento** será exibida.
6. Selecione **Gravar** e clique em **OK**.
O objeto de dados será exibido no editor.
7. Clique em **Salvar**.

Etapa 3. Adicionar uma Transformação do Validador de Endereço ao Mapeamento

Nesta etapa, uma transformação do Validador de Endereço é adicionada ao mapeamento que contém a origem e os objetos de dados.

Quando esta etapa for concluída, você poderá configurar a transformação e conectar as portas aos objetos de dados.

1. Selecione o editor que contém o mapeamento `ValidateAddresses`.
2. Na paleta de Transformação, selecione a transformação do Validador de Endereço.
3. Clique no editor.
A transformação do Validador de Endereço é exibida no editor.

Tarefa 3. Configurar a Nova Transformação do Validador de Endereço

Nessa tarefa, a transformação do Validador de Endereço para leitura e validação dos endereços da fonte de dados `All_Customers` será configurada.

Nota: A transformação do Validador de Endereço contém uma série de portas de entrada e saída predefinidas. Selecione as portas que você precisa e vincule-as aos objetos no mapeamento.

Para configurar a transformação, execute estas etapas:

1. Selecione o país padrão para validação de endereço.
2. Configure as portas de entrada da transformação.
3. Configure as portas de saída da transformação.
4. Conecte as portas de origem não utilizadas ao destino de dados.

Etapa 1. Definir o País Padrão para Validação de Endereço

Nesta etapa, você seleciona o país padrão para a validação de endereço. Os arquivos de dados de referência de endereço utilizados pela transformação de Validador de Endereço são organizados por país. Ao selecionar o país padrão, você identifica o conjunto de dados do país que a transformação aplica a qualquer endereço de entrada que não contenha informações de país.

1. Selecione a transformação do Validador de Endereço no editor.
2. Em **Propriedades**, clique em **Configurações Gerais**.
3. No menu **País Padrão**, selecione `Estados Unidos`.

Etapa 2. Configurar as Portas de Entrada de Transformação do Validador de Endereço

Nesta etapa, você selecionará as portas de entrada da transformação e conectará essas portas ao objeto de dados `All_Customers_av`.

A transformação do Validador de Endereço contém vários grupos de portas de entrada predefinidos. Selecione as portas de entrada que correspondem aos campos em seu endereço de entrada e adicione-as à transformação.

Mantenha pressionada a tecla Ctrl ao selecionar portas nas etapas a seguir para selecionar várias portas em uma única operação.

1. Selecione a transformação do Validador de Endereço no editor.
2. Em **Propriedades**, clique em **Modelos**.
3. Expanda o grupo de portas **Modelo Básico**.
4. Expanda o grupo de portas de entrada **Híbrido** e selecione as portas a seguir:

Nome da Porta	Descrição
Linha de Endereço de Entrega 1	Dados de endereço de rua, como o nome da rua e o número do edifício.
Localidade Completa 1	Nome da cidade.
CEP 1	CEP.
Província 1	Nome da província ou estado.
Nome do País	Nome ou abreviação do país.

Nota: Mantenha pressionada a tecla Ctrl para selecionar várias portas em uma única operação.

5. Na barra de ferramentas acima da lista de nomes de portas, clique em **Adicionar porta para a transformação**.

A barra de ferramentas fica visível quando você seleciona **Modelos**.

As portas selecionadas são exibidas na transformação no Editor de Mapeamento.

6. Conecte as portas de origem às portas de transformação do Validador de Endereço como a seguir:

Porta de Origem	Porta de Transformação do Validador de Endereço
Address1	Linha de Endereço de Entrega 1
City	Localidade Completa 1
CEP	CEP 1
Estado	Província 1
País	Nome do País

Etapa 3. Configurar as Portas de Saída de Transformação do Validador de Endereço

Nesta etapa, as portas de saída de transformação são selecionadas e conectadas ao objeto de dados `All_Customers_av_tgt`.

A transformação do Validador de Endereço contém vários grupos de portas de saída predefinidos. Selecione as portas que definem a estrutura de endereço que você precisa e as adicione à transformação.

Você também pode selecionar as portas que contêm informações sobre o tipo de validação obtido para cada endereço.

1. Selecione a transformação do Validador de Endereço no editor de mapeamento.
2. Em **Propriedades**, clique em **Modelos**.
3. Expanda o grupo de portas **Modelo Básico**.
4. Expanda o grupo de portas de saída **Elementos de Endereço** e selecione a seguinte porta:

Nome da Porta	Descrição
Rua Completa 1	Dados de endereço de rua, como o nome da rua e o número do edifício.

5. Expanda o grupo de portas de saída **Elementos da Última Linha** e selecione as seguintes portas:

Nome da Porta	Descrição
Localidade Completa 1	Nome da cidade.
CEP 1	CEP.
Abreviação da Província 1	Identificador de províncias ou estado.

Nota: Mantenha pressionada a tecla Ctrl para selecionar várias portas em uma única operação.

6. Expanda o grupo de portas de saída **País** e selecione a seguinte porta:

Nome da Porta	Descrição
Nome do País 1	Nome do país.

7. Expanda o grupo de portas de saída **Informações de Status** e selecione as seguintes portas:

Nome da Porta	Descrição
Pontuação de Capacidade de Envio por Correio	Pontuação que representa a possibilidade bem-sucedida de entrega de correspondência.
Código de Correspondência	Código que representa o grau de similaridade entre o endereço de entrada e os dados de referência.

8. Na barra de ferramentas acima da lista de nomes de portas, clique em **Adicionar porta para a transformação**.

A barra de ferramentas fica visível quando você seleciona **Modelos**.

9. Conecte as portas da transformação do Validador de Endereço com as portas `All_Customers_av_tgt` conforme abaixo:

Porta de Transformação do Validador de Endereço	Porta de Destino
Rua Completa 1	Address1
Localidade Completa 1	City
CEP 1	CEP
Abreviação da Província 1	Estado
Nome do País 1	País
Pontuação de Capacidade de Envio por Correio	MailabilityScore
Código de Correspondência	MatchCode

Etapa 4. Conectar Portas de Fonte de Dados Não Utilizadas ao Destino de Dados

Nesta etapa, as portas não usadas na fonte de dados `All_Customers` serão conectadas ao destino de dados.

- Conecte as portas não usadas na fonte de dados às portas com os mesmos nomes no destino de dados.

Tarefa 4. Executar o Mapeamento

Nessa tarefa, o mapeamento será executado para criar a saída de mapeamento.

1. Selecione o editor que contém o mapeamento `ValidateAddresses`.
2. Selecione **Executar > Executar Mapeamento**.

O mapeamento é executado e grava a saída no arquivo `All_Customers_av_tgt.csv`.

Tarefa 5. Exibir a Saída de Mapeamento

Nessa tarefa, o Visualizador de Dados será executado para exibir a saída de mapeamento. Consulte a qualidade dos seus endereços validados examinando os valores gravados nas colunas Pontuação de Capacidade de Envio por Correio e Código de Correspondência no objeto de dados de destino.

O valor do Código de Correspondência é um código alfanumérico que representa o tipo de validação que o mapeamento executou no endereço.

O valor de Pontuação de Capacidade de Envio por Correio é um valor de um dígito que resume a capacidade de entrega do endereço.

1. Na exibição do Object Explorer, localize o objeto de dados `All_Customers_av_tgt` no seu projeto de tutorial e clique duas vezes nesse objeto.

O objeto de dados será aberto no editor.

2. Selecione **Janela > Mostrar Exibição > Visualizador de Dados**.

O Visualizador de Dados será aberto.

3. No Visualizador de Dados, clique em **Executar**.

O Visualizador de Dados exibe a saída de mapeamento.

4. Role pelos resultados do mapeamento para que as colunas Pontuação de Capacidade de Envio por Correio e Código de Correspondência fiquem visíveis.

5. Analise os valores na coluna Pontuação de Capacidade de Envio por Correio.

As pontuações podem variar de 0 a 5. Endereços com pontuações maiores são mais prováveis de receberem uma entrega com êxito.

6. Analise os valores na coluna de Código de Correspondência.

O Código de Correspondência é um código alfanumérico. O caractere alfabético indica o tipo de validação que a transformação executou e o dígito indica a qualidade do endereço final.

A seguinte tabela descreve os valores de Código de Correspondência:

Código	Descrição
A1	A pesquisa de código de endereço encontrou um endereço parcial ou um endereço completo para o código de entrada.
A0	A pesquisa de código de endereço não encontrou endereços para o código de entrada.
C4	Corrigido. Todos os elementos postais relevantes são verificados.
C3	Corrigido. Alguns elementos não podem ser verificados.
C2	Corrigido, mas o status de entrega não está claro devido à ausência de dados de referência.
C1	Corrigido, mas o status de entrega não está claro porque a padronização do usuário introduziu erros.
I4	Os dados não puderam ser corrigidos completamente, mas há uma única correspondência com um endereço nos dados de referência.

Código	Descrição
I3	Os dados não puderam ser corrigidos completamente, e há várias correspondências com endereços nos dados de referência.
I2	Os dados não podem ser corrigidos. O modo em lote retorna endereços sugeridos parcialmente.
I1	Os dados não podem ser corrigidos. O modo em lote não pode sugerir um endereço.
N7	Erro de validação. A validação não ocorreu porque a validação de uma única linha não está desbloqueada.
N6	Erro de validação. A validação não ocorreu porque a validação de uma única linha não é compatível com o país de destino.
N5	Erro de validação. A validação não ocorreu porque o banco de dados de referência está desatualizado.
N4	Erro de validação. A validação não ocorreu porque o banco de dados de referência está corrompido ou mal formatado.
N3	Erro de validação. A validação não ocorreu porque não foi possível desbloquear os dados do país.
N2	Erro de validação. A validação não ocorreu porque o banco de dados de referência exigido não está disponível.
N1	Erro de validação. A validação não ocorreu porque o país não foi reconhecido ou não é compatível.
Q3	Modo de Lista de Sugestões. A validação de endereço pode recuperar um ou mais endereços completos dos dados de referência de endereço que correspondem ao endereço de entrada.
Q2	Modo de Lista de Sugestões. A validação de endereço pode combinar os elementos do endereço de entrada e os elementos dos dados de referência de endereço para criar um endereço completo.
Q1	Modo de Lista de Sugestões. A validação de endereço não pode sugerir um endereço completo. Para gerar uma sugestão de endereço completo, adicione os dados ao endereço de entrada.
Q0	Modo de Lista de Sugestões. Não há dados de entrada suficientes para gerar uma sugestão.
RB	País reconhecido com base na abreviação. Reconhece os códigos de país ISO de dois e três caracteres. Também pode reconhecer abreviações comuns, como "GER" para Alemanha.
RA	País reconhecido com base na configuração Forçar País da transformação.
R9	País reconhecido com base na configuração País Padrão da transformação.

Código	Descrição
R8	País reconhecido pelo nome de país.
R7	País reconhecido pelo nome de país, mas a transformação identificou erros nos dados do país.
R6	País reconhecido pelos dados de território.
R5	País reconhecido pelos dados de província.
R4	País reconhecido pelos dados da cidade principal.
R3	País reconhecido pelo formato de endereço.
R2	País reconhecido por um script.
R1	País não reconhecido porque várias correspondências estão disponíveis.
R0	País não reconhecido.
S4	Modo de análise. O endereço foi analisado perfeitamente.
S3	Modo de análise. O endereço foi analisado com vários resultados.
S1	Modo de análise. Ocorreu um erro de análise devido a um formato de entrada incompatível.
V4	Verificado. Os dados de entrada estão corretos. A validação de endereço verificou todos os elementos postais relevantes. As entradas são perfeitamente correspondentes.
V3	Verificado. Os dados de entrada estão corretos, mas alguns ou todos os elementos foram padronizados, ou a entrada contém nomes ou exônimos desatualizados.
V2	Verificado. Os dados de entrada estão corretos, mas alguns elementos não podem ser verificados devido a dados de referência incompletos.
V1	Verificado. Os dados de entrada estão corretos, mas a padronização do usuário afetou negativamente a capacidade de entrega. Por exemplo, o tamanho do código postal é muito pequeno.

Resumo da Validação dos Dados de Endereço

Nessa lição, você aprendeu que a validação de endereço compara os dados do endereço de entrada com dados de referência e retorna a versão mais precisa possível do endereço.

Aprendeu, ainda, que o processo de validação de endereço também retorna informações de status sobre a qualidade de cada endereço.

Você aprendeu que os usuários da ferramenta Administrator executam o Instalador de Conteúdo de Data Quality para instalar dados de referência de endereço.

Também aprendeu que a transformação do Validador de Endereço é uma transformação de vários grupos, e que você seleciona as portas de entrada e saída para a transformação dos grupos de porta. As portas de entrada que você seleciona determinam o conteúdo de endereço que é validado. As portas de saída determinam o conteúdo de um registro de endereço.

APÊNDICE A

Perguntas Frequentes

Este apêndice inclui os seguintes tópicos:

- [Perguntas Frequentes do Informatica Analyst, 89](#)
- [Perguntas Frequentes sobre o Informatica Developer, 89](#)

Perguntas Frequentes do Informatica Analyst

Consulte as Perguntas Frequentes para responder às perguntas que você possa ter sobre o Informatica Analyst.

Posso usar uma única conta de usuário para acessar as ferramentas Administrator, Developer e Analyst?

Sim. Você pode fornecer uma permissão de usuário para acessar as três ferramentas. Não é necessário criar contas de usuário separadas para cada aplicativo cliente.

Onde meus dados de referência estão armazenados?

Você pode usar as ferramentas Developer e Analyst para criar e compartilhar objetos de dados de referência. O repositório do Modelo armazena os metadados de objeto de dados de referência. O banco de dados de dados de referência armazena valores de dados da tabela de referência. Configure o banco de dados de referência no Serviço do Gerenciamento de Conteúdo.

Perguntas Frequentes sobre o Informatica Developer

Consulte as perguntas frequentes para responder às perguntas que você possa ter sobre o Informatica Developer.

Qual é a diferença entre uma origem e um destino no PowerCenter e um objeto de dados físicos na ferramenta Developer?

No PowerCenter, você cria uma definição de origem a ser incluída como uma origem de mapeamento. Você cria uma definição de destino a ser incluída como um destino de mapeamento. Na ferramenta Developer, você cria um objeto de dados físicos que pode ser usado como origem ou destino de mapeamento.

Qual é a diferença entre um mapeamento na ferramenta Developer e um mapeamento no PowerCenter?

Um mapeamento do PowerCenter especifica como mover dados entre origens e destinos. Um mapeamento da ferramenta Developer especifica como mover dados entre a entrada e a saída de mapeamento.

Um mapeamento do PowerCenter deve incluir uma ou mais definições de origem, qualificadores de origem e definições de destino. Um mapeamento do PowerCenter também pode incluir atalhos, transformações e mapplets.

Um mapeamento da ferramenta Developer deve incluir a entrada e a saída de mapeamento. Um mapeamento da ferramenta Developer também pode incluir transformações e mapplets.

A ferramenta Developer tem os seguintes tipos de mapeamentos:

- Mapeamento que move dados entre origens e destinos. Esse tipo de mapeamento é diferente de um mapeamento do PowerCenter somente por não poder utilizar atalhos e por não utilizar um qualificador de origem.
- Mapeamento de objeto de dados lógicos. Um mapeamento em um modelo de objeto de dados lógicos. Um mapeamento de objeto de dados lógicos pode conter um objeto de dados lógicos como a entrada de mapeamento e um objeto de dados como a saída de mapeamento. Ou, pode conter um ou mais objetos de dados físicos como a entrada de mapeamento e o objeto de dados lógicos como a saída de mapeamento.
- Mapeamento de tabela virtual. Um mapeamento em um serviço de dados SQL. Ele contém um objeto de dados como a entrada de mapeamento e uma tabela virtual como a saída de mapeamento.
- Mapeamento de procedimento armazenado virtual. Define um conjunto de lógica de negócios em um serviço de dados SQL. Ele contém um objeto de dados físicos ou uma transformação de Parâmetro de Entrada como entrada de mapeamento e um objeto de dados físicos ou uma transformação de Parâmetro de Saída como saída de mapeamento.

Qual é a diferença entre um mapplet no PowerCenter e um mapplet na ferramenta Developer?

Um mapplet no PowerCenter e na ferramenta Developer é um objeto reutilizável que contém um conjunto de transformações. Você pode reutilizar a lógica de transformação em vários mapeamentos.

Um mapplet do PowerCenter pode conter definições de origem ou transformações de Entrada como entrada de mapplet. Ele deve conter transformações de Saída como saída de mapplet.

Um mapplet da ferramenta Developer pode conter objetos de dados ou transformações de Entrada como entrada de mapplet. Ele pode conter objetos de dados ou transformações de Saída como saída de mapplet. Um mapeamento na ferramenta Developer também inclui os seguintes recursos:

- É possível validar um mapplet como uma regra. Você usa uma regra em um perfil.
- Um mapplet pode conter outros mapplets.

Qual é a diferença entre um mapplet e uma regra?

É possível validar um mapplet como uma regra. Uma regra é uma lógica de negócios que define condições aplicadas a dados de origem quando você executa um perfil. É possível validar um mapplet como uma regra quando esse mapplet atende aos requisitos a seguir:

- Ele contém uma transformação de Entrada e de Saída.
- O mapplet não contém transformações ativas.
- Ele não especifica cardinalidade entre grupos de entrada.

ÍNDICE

C

- configurando a ferramenta Analyst
 - visão geral [19](#)
- configurando a Developer tool
 - visão geral [46](#)
- criando objetos de dados
 - visão geral [22](#)
- criando perfis padrão
 - visão geral [25](#)
- criando perfis personalizados
 - visão geral [28](#)
- criando regras de expressão
 - visão geral [32](#)
- criando scorecards
 - visão geral [35](#)
- criando tabelas de referência a partir de colunas
 - visão geral [40](#)

D

- dados de criação de perfil
 - visão geral [59](#)

I

- importando um objeto de dados físicos
 - visão geral [50](#)

T

- tabelas de referência
 - visão geral [43](#)