



Informatica®

10.4.0

Guia de Introdução à Criação de Perfil

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.

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

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.

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

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

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

Data da Publicação: 2020-02-04

Conteúdo

Prefácio.....	6
Recursos da Informatica.	6
Rede da Informatica.	6
Base de Dados de Conhecimento da Informatica.	6
Documentação da Informatica.	6
Matrizes de Disponibilidade de Produto da Informatica.	7
Informatica Velocity.	7
Informatica Marketplace.	7
Suporte Global a Clientes da Informatica.	7
 Capítulo 1: Visão Geral da Introdução.....	8
Visão Geral da Criação de Perfil.	8
Introdução ao Informatica Analyst.	9
Interface do Usuário do Informatica Analyst.	10
Interface do Usuário do Informatica Developer.	12
Página de Boas-vindas do Informatica Developer.	13
Folhas de Cola.	13
Tarefas do Informatica Developer.	14
História do Tutorial.	14
A Estrutura do Tutorial.	14
Pré-requisitos do tutorial.	15
Tutorial do Informatica Analyst.	15
Tutorial do Informatica Developer.	16
 Parte I: Introdução ao Informatica Analyst.....	17
 Capítulo 2: Lição 1. Configurar o Informatica Analyst.....	18
Visão Geral da Configuração do Informatica Analyst.	18
Tarefa 1. Fazer logon no Informatica Analyst.	19
Tarefa 2. Criar um Projeto.	19
Tarefa 3. Criar uma Pasta.	20
Resumo da Configuração do Informatica Analyst.	20
 Capítulo 3: Lição 2. Criar Objetos de Dados.....	21
Visão Geral da Criação de Objetos de Dados.	21
Tarefa 1. Criar o Objeto de Dados de Arquivo Simples.	22
Tarefa 2. Exibir as Propriedades de Objeto de Dados.	22
Resumo da Criação de Objetos de Dados.	23

Capítulo 4: Lição 3. Criando Perfis Padrão.	24
Criando Perfis Padrão - Visão Geral.	24
Tarefa 1. Criar e Executar um Perfil Padrão.	25
Tarefa 2. Visualizar os resultados do perfil na exibição resumida.	25
Tarefa 3. Visualizar os Resultados do Perfil na Exibição Detalhada.	26
Criando Perfis Padrão - Resumo.	27
 Capítulo 5: Lição 4. Criar Perfis Personalizados.	 28
Visão Geral da Criação 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 Serviço de Integração de Dados 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
Índice.....	64

Prefácio

Siga as instruções no *Guia de Introdução à Criação de Perfil* para aprender como criar e executar perfis na Developer tool e na ferramenta Analyst. O guia fornece um tutorial para ajudar usuários iniciantes a aprender como usar o *Informatica Developer* e o *Informatica Analyst*.

O guia foi escrito para analistas de dados e desenvolvedores.

Recursos da Informatica

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

Rede da Informatica

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

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

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

Base de Dados de Conhecimento da Informatica

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

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

Documentação da Informatica

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

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

Matrizes de Disponibilidade de Produto da Informatica

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

Informatica Velocity

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

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

Informatica Marketplace

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

Suporte Global a Clientes da Informatica

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

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

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

CAPÍTULO 1

Visão Geral da Introdução

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Criação de Perfil, 8](#)
- [Introdução ao Informatica Analyst, 9](#)
- [Interface do Usuário do Informatica Developer, 12](#)
- [História do Tutorial, 14](#)
- [A Estrutura do Tutorial, 14](#)

Visão Geral da Criação de Perfil

Use a criação de perfil para localizar o conteúdo, a qualidade e a estrutura das fontes de dados de um aplicativo, esquema ou empresa. O conteúdo da fonte de dados inclui frequências de valor e tipos de dados. A estrutura de fonte de dados inclui chaves e dependências funcionais.

Como parte do processo de descoberta, você pode criar e executar perfis. Um perfil é um objeto de repositório que localiza e analisa todas as irregularidades nas fontes de dados da empresa e problemas ocultos em dados que colocam os projetos em risco. A execução de um perfil em qualquer fonte de dados da empresa oferece uma boa compreensão das vantagens e desvantagens de seus dados e metadados.

Você pode usar o Informatica Analyst e o Informatica Developer para analisar os dados e metadados de origem. Analistas e desenvolvedores podem usar essas ferramentas para colaborar, identificar problemas de qualidade dos dados e analisar relacionamentos de dados. Com base na função do seu cargo, você pode usar os recursos da ferramenta Analyst ou da Developer tool. O grau de criação de perfil que você pode executar difere com base na ferramenta usada.

Você pode executar as seguintes tarefas tanto na Developer tool quanto na ferramenta Analyst:

- Execute a criação de perfil de coluna. O processo inclui descobrir o número de valores exclusivos, valores nulos e padrões de dados em uma coluna.
- Execute a descoberta do domínio de dados. Você pode descobrir características de dados críticas dentro de uma empresa.
- Faça curadoria dos resultados do perfil incluindo tipos de dados, domínios de dados, chaves primárias e chaves externas.
- Crie scorecards para monitorar a qualidade dos dados.
- Escolha um perfil do sistema operacional para criar e executar perfis de coluna, perfis de descoberta empresarial e scorecards com base nas permissões do usuário do sistema operacional que você define nesse perfil do sistema operacional.

- Use bloqueios de ativos de repositório para impedir que outros usuários sobrescrevam trabalhos.
- Use o sistema de controle de versão para salvar várias versões de um perfil.
- Crie e atribua marcas aos objetos de dados.
- Pesquise o significado do nome de um objeto como um termo comercial no Business Glossary Desktop. Por exemplo, você pode pesquisar o significado de um nome de coluna ou nome de perfil para entender seu requisito de negócios e a implementação atual.

Você pode executar as seguintes tarefas na Developer tool:

- Descobrir o grau de possíveis associações entre duas colunas de dados de uma fonte de dados.
- Determinar a porcentagem de dados de sobreposição em pares de colunas dentro de uma fonte de dados ou várias fontes de dados.
- Comparar os resultados da criação de perfil de coluna.
- Gerar um objeto de mapeamento a partir de um perfil.
- Descobrir chaves primárias em uma fonte de dados.
- Descobrir chaves externas em um conjunto de uma ou mais fontes de dados.
- Descobrir dependência funcional entre colunas de uma fonte de dados.
- Executar tarefas de descoberta de dados em um grande número de fontes de dados em várias conexões. As tarefas de descoberta de dados incluem perfil de coluna, inferência de relacionamentos de chave primária e chave externa, descoberta do domínio de dados e geração de um resumo gráfico consolidado dos relacionamentos de dados.

Você pode executar as seguintes tarefas na ferramenta Analyst:

- Executar a descoberta empresarial em um grande número de fontes de dados em várias conexões. Você pode exibir um resumo consolidado dos resultados da descoberta de metadados da coluna e domínios de dados.
- Executar a pesquisa de descoberta para encontrar onde os dados e metadados existem na empresa. Você pode procurar ativos específicos, como objetos de dados, regras e perfis. A pesquisa de descoberta localiza ativos e identifica relacionamentos com outros ativos nos bancos de dados e esquemas da empresa.
- Exiba os resultados do perfil de uma execução de perfil histórico.
- Compare os resultados de perfil para duas execuções de perfil em um perfil de coluna.
- Exiba a linhagem para cada métrica e grupo de métricas do scorecard.
- Exiba o painel de scorecards.
- Adicione comentários a um perfil ou a colunas em um perfil.
- Atribua marcas a um perfil ou a colunas em um perfil.
- Atribua termos comerciais a colunas em um perfil.

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.

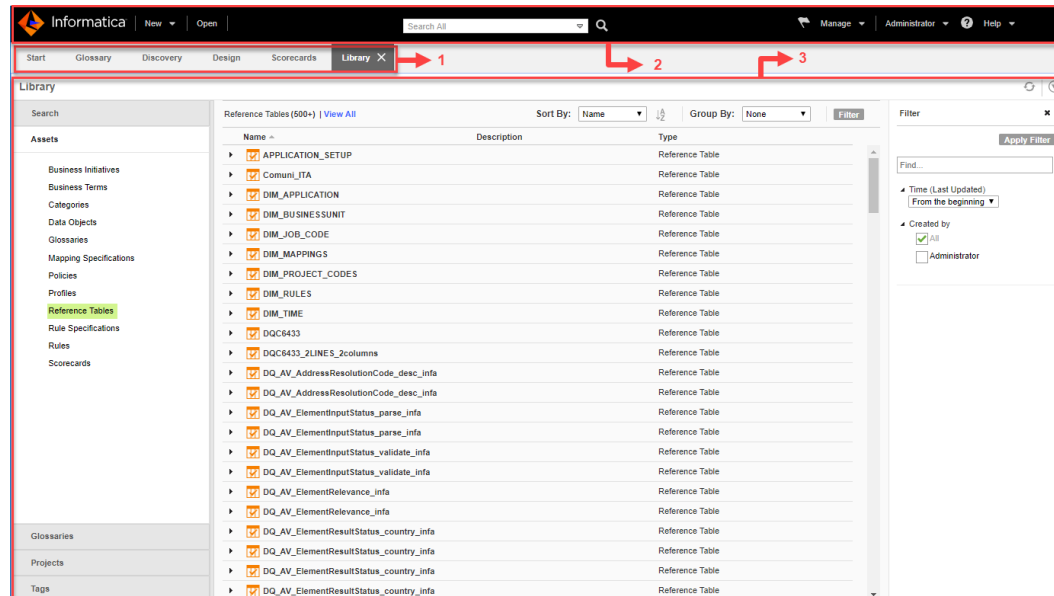
Dependendo da licença, analistas de negócios e desenvolvedores usam a ferramenta Analyst para colaboração orientada por dados. Você pode executar a criação de perfil de regra e dados de coluna, a

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.

Interface do Usuário do Informatica Analyst

A interface de usuário da ferramenta Analyst consiste em vários espaços de trabalho que você usa para criar as soluções de integração de dados, equalidade de dados, federação de dados. Um espaço de trabalho é uma página da Web que você pode acessar com base na licença e no privilégio.

A seguinte imagem mostra a interface do usuário da ferramenta Analyst:



1. Painel de acesso do espaço de trabalho
2. Área de cabeçalho
3. Espaço de trabalho

A ferramenta Analyst abre no espaço de trabalho **Iniciar**. O espaço de trabalho **Iniciar** lista os espaços de trabalho em que você possui a licença e o privilégio para uso por meio dos painéis de acesso de espaço de trabalho.

A ferramenta Analyst exibe os seguintes menus de cabeçalho:

Novo

Crie ativos nos espaços de trabalho **Glossário**, **Descoberta** e **Design**.

Abrir

Abra o espaço de trabalho **Biblioteca**.

Alerta de notificações

Exiba notificações para ativos de Glossário.

Gerenciar

Abra os espaços de trabalho temporários e as **Notificações**. Você pode abrir os espaços de trabalho **Conexões**, **Domínios de Dados**, **Status do Trabalho**, **Projetos** e **Segurança do Business Glossary**.

Nome de usuário

Defina as preferências do usuário para alterar a senha e faça logoff da ferramenta Analyst.

Ajuda

Acesse a ajuda no espaço de trabalho atual.

A ferramenta Analyst contém os seguintes espaços de trabalho permanentes:

Início

Acesse outros espaços de trabalho que você tenha licença para acessar por meio do painel de acesso de espaço de trabalho. Se você tiver a licença para executar gerenciamento de exceções, suas tarefas aparecerão no painel **Minhas Tarefas** do espaço de trabalho.

Glossário

Defina e descreva conceitos comerciais que são importantes para a sua organização. Você pode criar e gerenciar termos comerciais, categorias, glossários e diretivas.

Descoberta

Analise a qualidade de dados e os metadados em sistemas de origem. Você pode criar e gerenciar perfis, objetos de dados de arquivo simples e objetos de dados da tabela.

Design

Crie uma lógica comercial que ajude na colaboração entre analistas e desenvolvedores. Você pode criar e gerenciar as especificações de mapeamento, as tabelas de referência e as definições de regra.

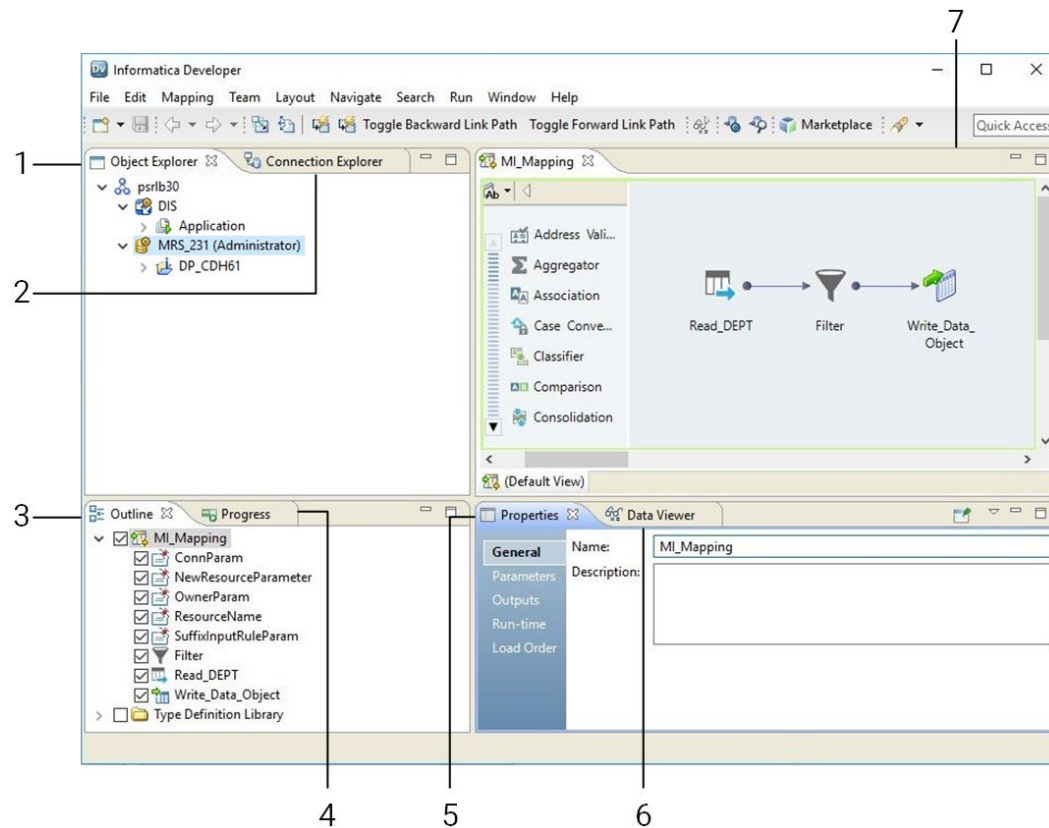
Scorecards

Abra, edite e execute scorecards que você criou com os resultados do perfil. Você pode adicionar métricas, pesquisar colunas detalhadamente, adicionar filtros de scorecard e exibir gráficos de tendências de um scorecard.

Interface do Usuário do Informatica Developer

A interface do usuário da ferramenta Desenvolvedor consiste em um workbench com várias exibições que você usa para criar soluções de integração de dados, qualidade de dados e federação de dados .

A seguinte imagem mostra a interface do usuário da ferramenta Developer:



1. Exibição Object Explorer
2. Exibição Estrutura de Tópicos
3. Exibição Propriedades
4. Exibição Visualizador de Dados
5. Exibição Marcas
6. Exibição Dependências de Objetos
7. Exibição Alertas
8. Exibição Explorador de Conexões
9. Editor

A ferramenta Developer pode mostrar as seguintes exibições:

Exibição Object Explorer

Exibe o domínio e os objetos em tempo de design e tempo de execução no domínio. Objetos em tempo de design são armazenados em projetos e pastas no repositório do modelo. Objetos de tempo de execução são armazenados como parte de um aplicativo em tempo de execução em um Serviço de Integração de Dados.

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. Você também pode visualizar as mensagens de serviços da Web ou executar uma consulta SQL na exibição Visualizador de Dados.

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

A página de boas-vindas aparece da primeira vez em que você abre a Developer tool. Use a página de boas-vindas para aprender a configurar e começar a trabalhar na Developer tool.

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 Developer tool 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 obter um link à Base de Dados de Conhecimento da Informatica. 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.
- Workbench. Clique no botão Workbench para começar a trabalhar na Developer tool.

Clique em **Ajuda > Bem-vindo** para acessar a página de boas-vindas depois de fechá-la.

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**.

Tarefas do Informatica Developer

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 concretizam os seguintes objetivos:

- Execute um perfil em dados de origem. 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.
- 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.
- Colaborar com usuários da Informatica. As regras que você adiciona ao repositório do Modelo estão disponíveis para os usuários nas ferramentas Developer e Analyst. Os usuários podem colaborar em projetos, e usuários diferentes podem assumir a propriedade de objetos em estágios diferentes de um projeto.

História do Tutorial

A Hypostores Corporation é uma organização nacional de varejo com matriz em Boston e lojas em diversos estados. Ela integra dados operacionais de lojas em todo o país com o armazenamento de dados da matriz regularmente. A Hypostores abriu recentemente uma loja em Los Angeles.

A matriz inclui uma equipe de ICC central com administradores, desenvolvedores e arquitetos responsáveis por fornecer uma camada de serviços de dados comuns para todos os aplicativos de composição e de BI. 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 quer criar o perfil dos dados de Boston e Los Angeles antes de integrar os conjuntos de dados. As operações de perfil identificam problemas de qualidade de dados que a Hypostores pode corrigir antes da integração.

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.

Pré-requisitos do tutorial

Antes de começar as lições do tutorial, o domínio Informatica deve estar em execução com pelo menos um nó configurado. Você deve verificar se o Serviço Analyst, o Serviço de Repositório do Modelo e o Serviço de Integração de Dados estão em execução.

O instalador inclui arquivos do tutorial que você usará para concluir as lições. Você pode encontrar todos os arquivos em ambas as instalações do cliente e 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
- Customer_Order.xsd
- LA_customers.csv
- pedidos.csv

Tutorial do Informatica Analyst

Neste tutorial, um analista faz login na ferramenta Analyst, cria projetos e pastas, cria perfis e regras, cria scorecards sobre dados e cria tabelas de referência.

A seguinte tabela descreve as lições que você pode realizar:

Lição	Descrição
Lição 1. Configurando o Informatica Analyst	Fazer login na ferramenta Analyst e criar um projeto e uma pasta para as lições do tutorial.
Lição 2. Criando Objetos de Dados	Importar um arquivo simples como um objeto de dados e visualizar dados.
Lição 3. Criando Perfis Padrão	Criar um perfil padrão para obter rapidamente um ideia da qualidade de dados.
Lição 4. Criando Perfis Personalizados	Criar um perfil personalizado para configurar colunas e opções de busca detalhada e amostragem.
Lição 5. Criando Regras de Expressão	Criar regras de expressão para modificar e criar perfil dos valores de coluna.
Lição 6. Criando e Executando Scorecards	Criar e executar um scorecard para medir o progresso da qualidade de dados ao longo do tempo.

Lição	Descrição
Lição 7. Criando Tabelas de Referência a Partir de Colunas de Perfil	Criar uma tabela de referência que você possa usar para padronizar dados de origem.
Lição 8. Criando Tabelas de Referência	Criar uma tabela de referência para estabelecer relações entre dados de origem e valores válidos e padrão.

Tutorial do Informatica Developer

Neste tutorial, use a Developer tool para realizar várias operações de qualidade de dados. Você cria uma conexão com um repositório do Modelo, cria projetos e pastas, importa um arquivo simples como um objeto de dados físicos, cria um perfil de descoberta corporativa, executa uma análise de associação em duas fontes de dados e visualiza os resultados do perfil.

A seguinte tabela descreve as lições que você pode realizar:

Lição	Descrição
Lição 1. Configurando o Informatica Developer	Adicione um domínio e crie uma conexão a um repositório do Modelo que é gerenciado por um Serviço de Repositório do Modelo no domínio. Crie um projeto e pasta para armazenar o trabalho para as lições no tutorial. Selecione um Serviço de Integração de Dados padrão para executar mapeamentos e visualizar dados.
Lição 2. Importando Objetos de Dados Físicos	Importe um arquivo simples para criar um objeto de dados físicos.
Lição 3. Executar um Perfil em Dados de Origem	Crie um perfil de descoberta corporativa, execute uma análise de associação em duas fontes de dados e visualize os resultados do perfil. Você pode executar a análise de associação em um perfil para determinar se uma associação válida é possível entre duas colunas de dados.

Parte I: Introdução ao Informatica Analyst

Esta parte contém os seguintes capítulos:

- [Lição 1. Configurar o Informatica Analyst, 18](#)
- [Lição 2. Criar Objetos de Dados, 21](#)
- [Lição 3. Criando Perfis Padrão, 24](#)
- [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, 18](#)
- [Tarefa 1. Fazer login no Informatica Analyst, 19](#)
- [Tarefa 2. Criar um Projeto, 19](#)
- [Tarefa 3. Criar uma Pasta, 20](#)
- [Resumo da Configuração do Informatica Analyst, 20](#)

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, 21](#)
- [Tarefa 1. Criar o Objeto de Dados de Arquivo Simples, 22](#)
- [Tarefa 2. Exibir as Propriedades de Objeto de Dados, 22](#)
- [Resumo da Criação de Objetos de Dados, 23](#)

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 o Objeto de Dados de Arquivo Simples

Nesta tarefa, você usa o assistente para **Adicionar Arquivo Simples** para criar objetos de dados de arquivo simples a partir do arquivo de dados LA_Customers.

1. No Navegador de **Projetos**, selecione a pasta Clientes no projeto de tutorial.
Nota: Você deve selecionar o projeto ou pasta onde deseja criar o objeto de dados de arquivo simples antes de criá-lo.
2. Clique em **Novo > Objeto de Dados de Arquivo Simples**.
O assistente **Adicionar Arquivo Simples** é exibido no espaço de trabalho **Descoberta**.
3. Selecione **Procurar e Fazer Upload** e clique em **Procurar**.
4. Vá para a localização de LA_Customers.csv e clique em **Abrir**.
5. Clique em **Avançar**.
A página **Escolher tipo de importação** exibe as opções **Delimitado** e **Largura fixa**. A opção padrão é **Delimitado**.
6. Clique em **Avançar**.
7. Em **Especificar os delimitadores e qualificadores de texto usados nos dados**, selecione **Aspas Duplas** como um qualificador de texto
8. Em **Especifique as linhas para importação**, selecione **Importar da primeira linha** para importar nomes de coluna na primeira linha que não estão vazias.
9. Clique em **Avançar**.
A seção **Atributo de Coluna** mostra o nome, o tipo de dados, a precisão, a escala e o formato de cada coluna.
10. Clique em **Avançar**.
O campo **Nome** exibe **LA_Customers**.
11. Opcionalmente, altere o nome do arquivo e adicione uma descrição.
12. Na seção **Pastas**, selecione a pasta **Clientes**.
13. Clique em **Concluir**.
O objeto de dados LA_Customers é exibido no espaço de trabalho **Descoberta** com a visualização de dados dos dados de origem.

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, 24](#)
- [Tarefa 1. Criar e Executar um Perfil Padrão, 25](#)
- [Tarefa 2. Visualizar os resultados do perfil na exibição resumida, 25](#)
- [Tarefa 3. Visualizar os Resultados do Perfil na Exibição Detalhada, 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. Na tela **Especificar Propriedades Gerais**, o nome, a descrição e a localização são preenchidos por padrão.
4. Clique em **Avançar**.
5. Na tela **Selecionar Origem**, o objeto de dados é exibido por padrão. Você pode exibir as colunas no painel **Escolher Colunas**.
6. Clique em **Avançar**.
7. 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.**
 - Opção **Nativo** no painel **Ambiente de tempo de execução**.
8. Clique em **Avançar**.
9. 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.

Tarefa 3. Visualizar os Resultados do Perfil na Exibição Detalhada

Nessa tarefa, você usa a exibição detalhada do perfil LA_Customers para obter uma rápida visão geral dos resultados desse perfil.

1. Clique em um nome de coluna na exibição resumida para ver os resultados de perfil dessa coluna na exibição detalhada.
Os resultados do perfil para a coluna aparecem na exibição detalhada.
2. No painel **Geral**, você pode visualizar os valores nulos, distintos e não distintos como um gráfico de barras horizontais e em porcentagens.
3. No painel **Valores**, você pode visualizar todos os valores no painel superior. Ao percorrer com o controle deslizante sobre os valores no painel superior, todos os valores dentro desse controle aparecem no painel inferior.
4. No painel **Visualização de Dados**, clique em **Ações > Mostrar Visualização** para exibir os dados de origem.
5. No painel **Padrões**, é possível visualizar os padrões deduzidos na coluna como um gráfico de barras horizontais e em porcentagens.
6. No painel **Tipos de dados**, é possível visualizar os tipos de dados deduzidos na coluna como um gráfico de barras horizontais e em porcentagens.
7. No painel **Domínios de Dados**, você pode visualizar os domínios de dados deduzidos.

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:

- [Visão Geral da Criação 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](#)

Visão Geral da Criação 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 Hadoop.

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

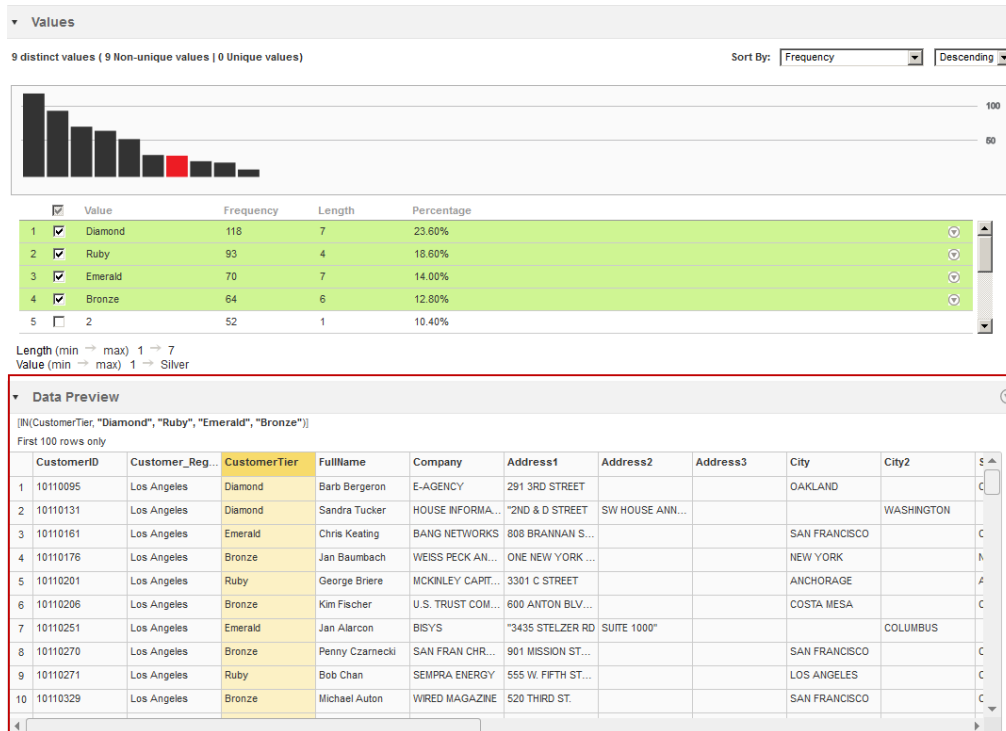
Nessa tarefa, faça uma busca detalhada nos valores de coluna CustomerTier para ver as linhas subjacentes no objeto de dados do perfil.

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](#)

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 Serviço de Integração de Dados 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 Developer tool. Para configurar a Developer tool, 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 deverá selecionar um Serviço de Integração de Dados padrão se o domínio incluir mais de um serviço.

O domínio Informatica é um conjunto de nós e serviços que definem o ambiente Informatica. Os serviços no domínio incluem o Serviço de Repositório do Modelo e o Serviço de Integração de Dados.

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

O Serviço de Integração de Dados realiza tarefas de integração de dados na Developer tool.

Objetivos

Nesta lição, você concluirá as seguintes tarefas:

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

- Criar um projeto para armazenar os objetos criados na Developer tool.
- Criar uma pasta no projeto que pode armazenar objetos relacionados.
- Selecionar um Serviço de Integração de Dados padrão para realizar tarefas de integração de dados.

Pré-requisitos

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

- Você instalou a Developer tool.
- 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 Serviço de Repositório do Modelo na ferramenta Administrator.
- Você tem um nome de usuário e uma senha para acessar o Serviço de Repositório do Modelo. Você pode obter essas informações de um administrador de domínio.
- Um administrador de domínio configurou um Serviço de Integração de Dados.
- O Serviço de Integração de Dados 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 Serviço de Integração de Dados padrão

Nessa tarefa, um Serviço de Integração de Dados padrão será selecionado para que você possa executar mapeamentos e visualizar dados. Essa etapa será necessária se houver mais de um Serviço de Integração de Dados no domínio. Se o domínio contiver um Serviço de Integração de Dados, este serviço será definido como o padrão.

1. Clique em **Janela > Preferências**.
A caixa de diálogo **Preferências** é exibida.
2. Selecione **Informatica > Serviços de Integração de Dados**.
3. Expanda o domínio.
4. Selecione um Serviço de Integração de Dados.
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.

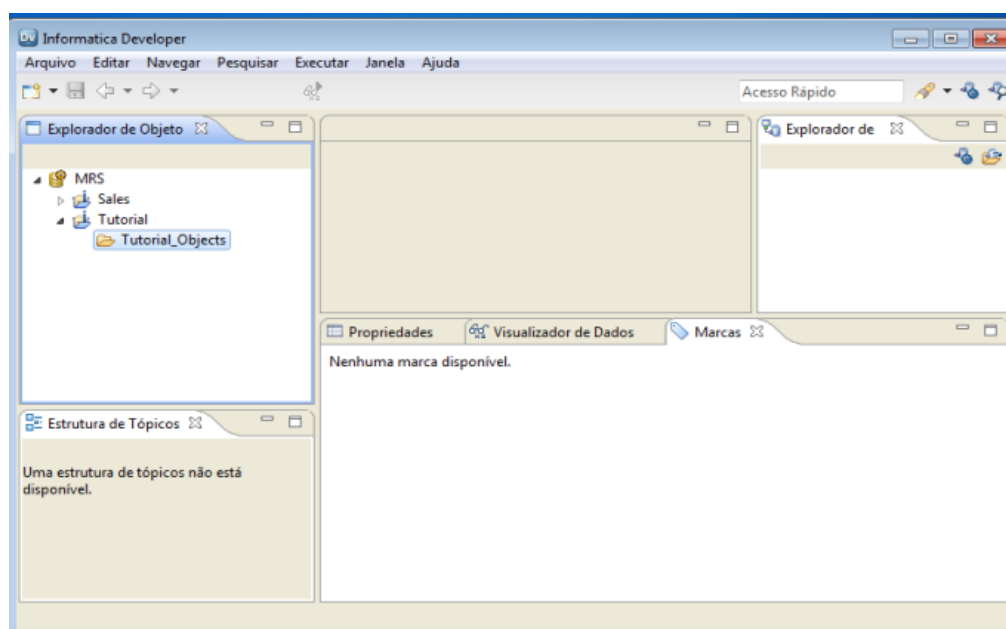
Tempo

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

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

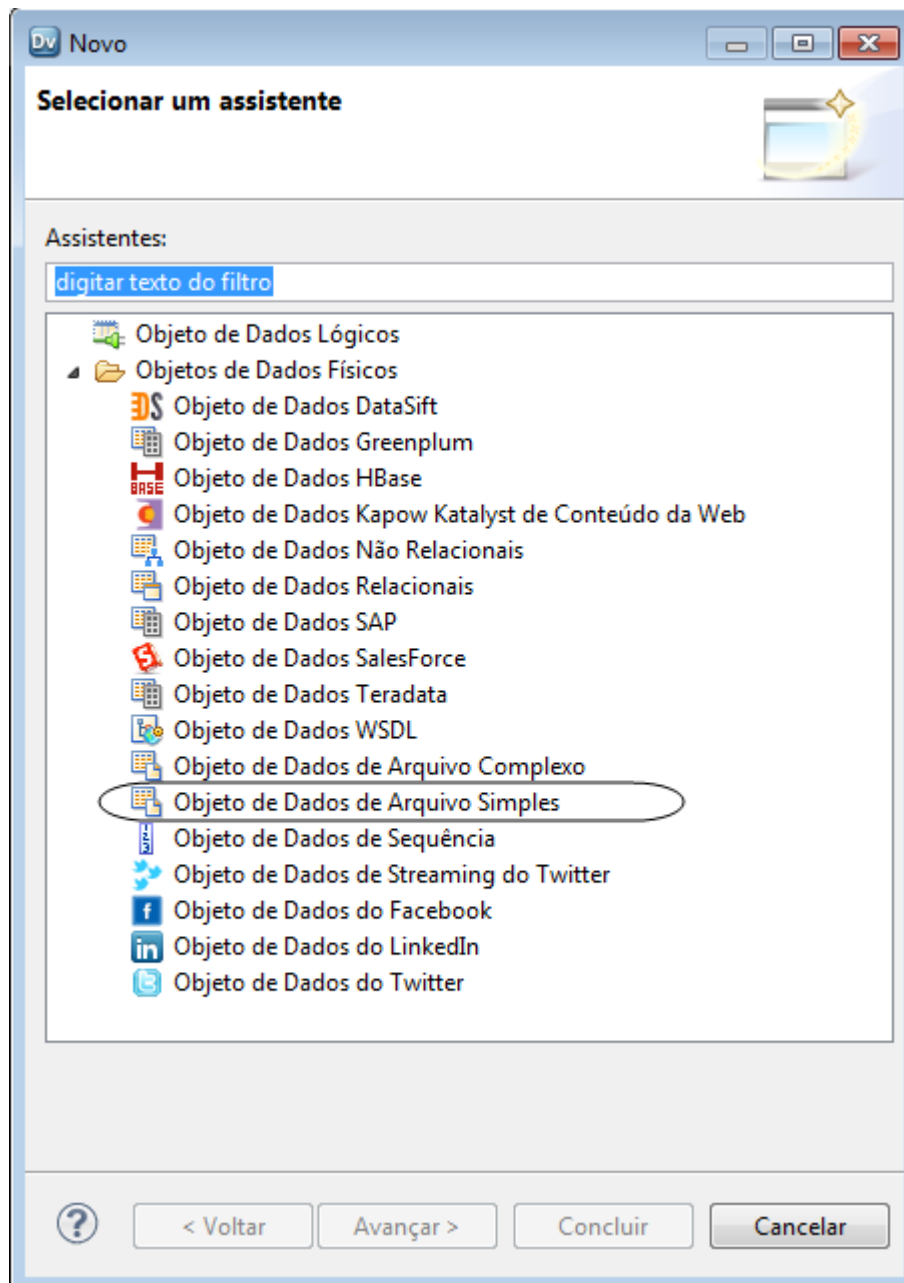
Nesta tarefa, você importará um objeto de dados físicos de um arquivo que contém dados de clientes do escritório de Boston.

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



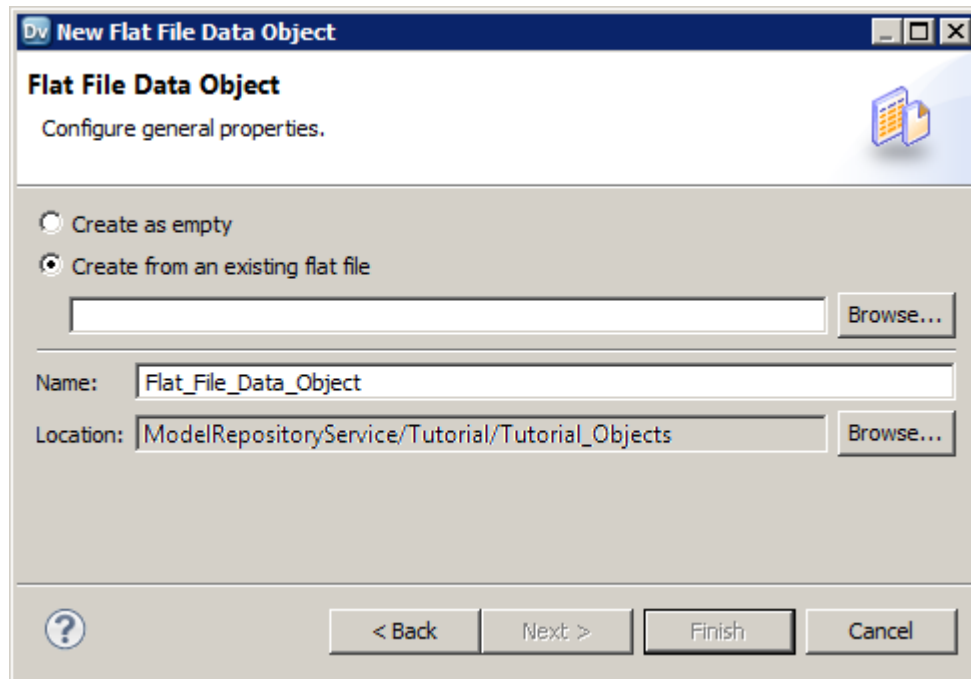
2. Clique com o botão direito do mouse na pasta Tutorial_Objects e selecione **Novo > Objeto de Dados**.

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)

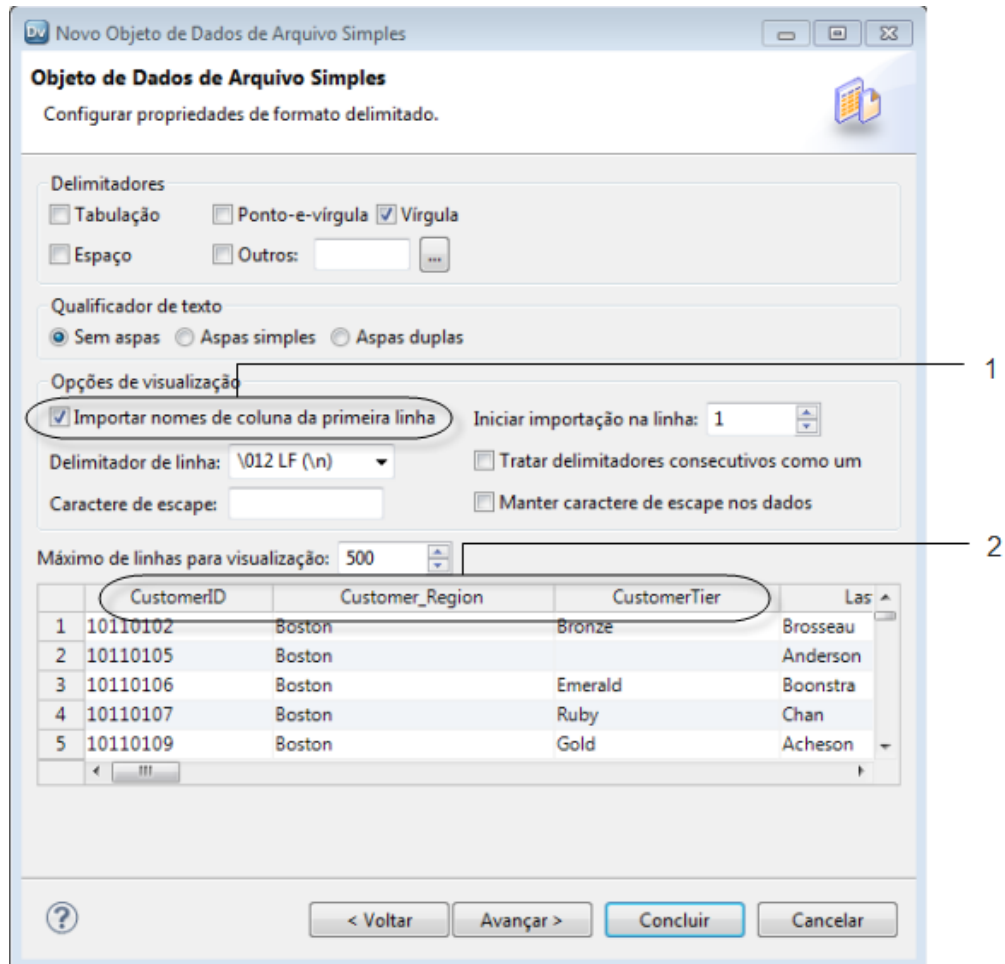
Máximo de linhas para visualização: 500

	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 ☒ Vírgula

☐ 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:

Delimitador de linha: \012 LF (\n)

☐ Tratar delimitadores consecutivos como um

Caractere de escape:

☐ Manter caractere de escape nos dados

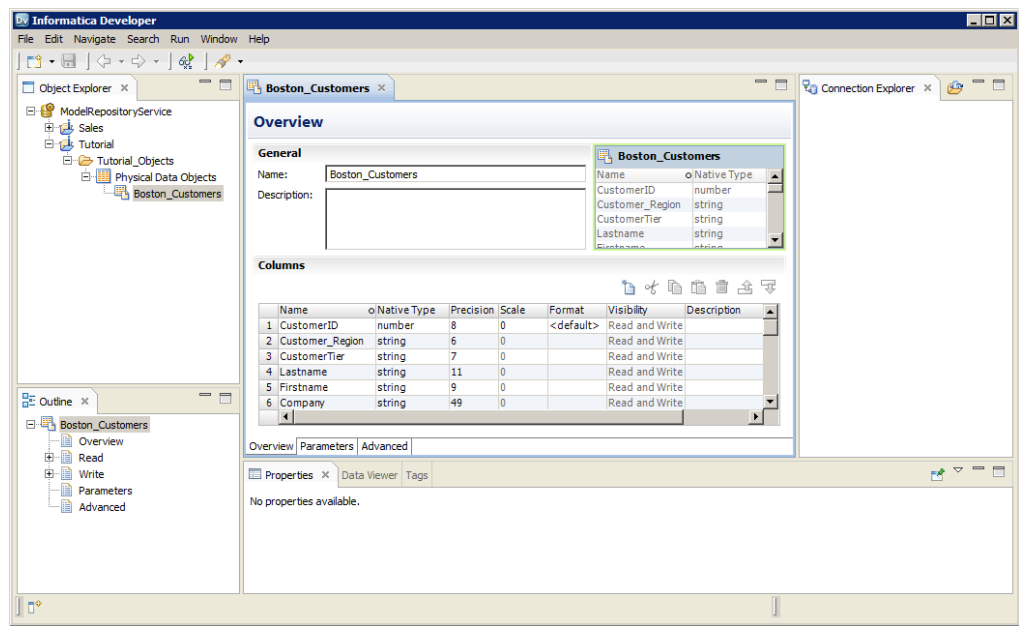
Máximo de linhas para visualização:

	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

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.

ÍNDICE

C

- configurando a ferramenta Analyst
 - visão geral [18](#)
- configurando a Developer tool
 - visão geral [46](#)
- criação de perfil
 - visão geral [8](#)
- criando objetos de dados
 - visão geral [21](#)
- criando perfis padrão
 - visão geral [24](#)
- 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](#)