

*Guia do Usuário do IBM SPSS Modeler  
Entity Analytics 17.1*

**IBM**

**Nota**

Antes de utilizar estas informações e o produto suportado por elas, leia as informações em “Avisos” na página 57.

**Informações sobre o Produto**

Esta edição se aplica à versão 17, liberação 1, modificação 0 do IBM(r) SPSS(r) Modeler e a todas as liberações e modificações subsequentes até que seja indicado de outra forma em novas edições.

# Índice

<b>Prefácio</b> . . . . .	<b>v</b>
---------------------------	----------

<b>Capítulo 1. Análise de entidade</b> . . . . .	<b>1</b>
Sobre a análise de entidade . . . . .	1
Análise de entidade e análise preditiva . . . . .	2

<b>Capítulo 2. Análise de entidade com o IBM SPSS Modeler</b> . . . . .	<b>3</b>
---	----------

Usando análise de entidade com o IBM SPSS Modeler . . . . .	3
Estágio 1: Ler os dados de origem para o SPSS Modeler . . . . .	4
Estágio 2: Criar o repositório . . . . .	4
Estágio 3: Conectar o SPSS Modeler ao repositório . . . . .	5
Estágio 4: Mapear os campos de entrada para variáveis de repositório . . . . .	5
Estágio 5: Exportar dados para o repositório e resolver identidades . . . . .	5
Estágio 6: Analisando as identidades resolvidas . . . . .	7
Estágio 7: Resolver novos casos com relação ao repositório . . . . .	8
Estágio 8: Gerar alertas . . . . .	9

<b>Capítulo 3. Tarefas de análise de entidade</b> . . . . .	<b>11</b>
---	-----------

Sobre as tarefas . . . . .	11
Configurando um repositório de entidade (nó Exportação do EA) . . . . .	11
O repositório de entidade . . . . .	11
Conectando-se a uma origem de dados . . . . .	12
Criando o repositório . . . . .	12
Mapeando campos de entrada para variáveis (nó Exportação do EA) . . . . .	14
Exibindo os mapeamentos de campo (nó Exportação do EA) . . . . .	16
Configurando um repositório de entidade . . . . .	16
Visualizando os mapeamentos de origem de dados . . . . .	17
Mantendo as variáveis de repositório . . . . .	17
Incluindo ou editando uma variável . . . . .	19
Anonimizando as variáveis de repositório . . . . .	20
Mantendo os tipos de entidade . . . . .	21
Configurando o limite para correspondência de entidade . . . . .	23
Reutilizando uma configuração de repositório . . . . .	23
Salvando as mudanças na configuração . . . . .	23
Fechando a janela de configuração . . . . .	24
Analisando as identidades resolvidas [nó de origem Entity Analytics (EA)] . . . . .	24
Selecionando uma origem de dados . . . . .	24
Renomeando campos de dados . . . . .	25
Configurando informações de tipo para campos de dados . . . . .	25

Incluindo nós no fluxo . . . . .	26
Comparando novos casos com relação ao repositório (nó EA de fluxo) . . . . .	26
Mapeando campos de entrada para variáveis (nó EA de fluxo) . . . . .	27
Exibindo os mapeamentos de campo e as origens de dados (nó EA de fluxo) . . . . .	28
Saída do nó EA de fluxo . . . . .	29
Usando o IBM SPSS Modeler Entity Analytics com outros produtos IBM SPSS . . . . .	30
Tarefas administrativas . . . . .	30
Configurando designações de porta . . . . .	30
Gerenciando credenciais do administrador para a base de dados do repositório . . . . .	31
Movendo o repositório para um diretório de armazenamento . . . . .	32
Configurando propriedades de fluxo para os campos data/hora e registro de data e hora . . . . .	32
Ajustando as configurações de tempo limite . . . . .	32
Executando o IBM SPSS Modeler Entity Analytics com o cliente SPSS Modeler e o SPSS Modeler Server no mesmo sistema Windows . . . . .	33
Limpando um repositório de entidade . . . . .	33
Excluindo origens de dados não usadas de um repositório . . . . .	33
Excluindo um repositório de entidade . . . . .	34
Excluindo um repositório quando não é possível se conectar a ele . . . . .	34

<b>Capítulo 4. Análise de entidade em ação</b> . . . . .	<b>37</b>
--	-----------

Sobre este exemplo . . . . .	37
O modelo original . . . . .	37
Incluindo análise de entidade . . . . .	40
Colocando os dados de origem no repositório . . . . .	40
Lendo as identidades resolvidas . . . . .	41
Comparando a saída de análise de entidade ao modelo original . . . . .	47
Resumo . . . . .	51

<b>Apêndice. Propriedades de script do IBM SPSS Modeler Entity Analytics</b> . . . . .	<b>53</b>
--	-----------

Script com o IBM SPSS Modeler Entity Analytics . . . . .	53
Propriedades comuns . . . . .	53
Propriedades entityanalytics_exporthnode . . . . .	53
Propriedades entityanalytics_sourcenode . . . . .	54
Propriedades entityanalytics_processnode . . . . .	54

<b>Avisos</b> . . . . .	<b>57</b>
Marcas comerciais . . . . .	58

<b>Índice Remissivo</b> . . . . .	<b>61</b>
-----------------------------------	-----------



---

## Prefácio

IBM® SPSS Modeler é o ambiente de trabalho de mineração de dados de força corporativa do IBM Corp.. O SPSS Modeler ajuda as organizações a melhorarem as relações com o cliente e com o cidadão por meio de um entendimento profundo dos dados. As organizações utilizam o insight adquirido do SPSS Modeler para reter clientes rentáveis, identificar oportunidades de venda cruzada, atrair novos clientes, detectar fraude, reduzir o risco e melhorar a entrega de serviço de governo.

A interface visual do SPSS Modeler convida os usuários a aplicarem seus conhecimentos de negócios específicos, levando a modelos preditivos mais poderosos e reduzindo o tempo para a solução. O SPSS Modeler oferece muitas técnicas de modelagem, como previsão, classificação, segmentação e algoritmos de detecção de associação. Quando os modelos são criados, o IBM SPSS Modeler Solution Publisher permite entregá-los aos tomadores de decisão na empresa ou a um banco de dados.

## Sobre o IBM Business Analytics

O software IBM Business Analytics fornece informações completas, consistentes e exatas nas quais os tomadores de decisão confiam para melhorar o desempenho de negócios. Um portfólio abrangente de inteligência de negócios, análise preditiva, gerenciamento de desempenho financeiro e estratégias aplicativos analíticos fornecem insight claro, imediato e prático sobre o desempenho atual e a capacidade de prever resultados futuros. Combinado com soluções para segmentos do mercado, práticas comprovadas e serviços profissionais completos, organizações de qualquer tamanho poderão conduzir maior produtividade, automatizar as decisões de modo confiável e entregar melhores resultados.

Como parte deste portfólio, o software IBM SPSS Predictive Analytics ajuda as organizações a preverem eventos futuros e agirem proativamente nesse insight para conduzir os melhores resultados de negócios. Clientes comerciais, governamentais e acadêmicos no mundo todo baseiam-se na tecnologia IBM SPSS como uma vantagem competitiva para atrair, reter e ampliar clientes, reduzindo a fraude e minimizando o risco. Ao incorporar o software IBM SPSS em suas operações diárias, as organizações se tornam empresas preditivas – capazes de dirigir e automatizar decisões para atender às metas de negócio e obter vantagem competitiva mensurável. Para obter mais informações ou entrar em contato com um representante, visite <http://www.ibm.com/spss>.

## Suporte técnico

O suporte técnico está disponível para manutenção dos clientes. Os clientes podem entrar em contato com o Suporte Técnico para obter assistência no uso de produtos IBM Corp. ou para obter ajuda na instalação em um dos ambientes de hardware suportados. Para entrar em contato com o Suporte técnico, consulte o website da IBM Corp. em <http://www.ibm.com/support>. Esteja preparado para se identificar, identificar sua organização e sua concordância de suporte ao solicitar assistência.



---

# Capítulo 1. Análise de entidade

---

## Sobre a análise de entidade

O **IBM SPSS Modeler Entity Analytics** inclui uma dimensão extra à análise preditiva do **IBM SPSS Modeler**. Considerando que a análise preditiva tenta prever comportamento futuro de dados passados, a análise de entidade foca melhorar a coerência e a consistência dos dados atuais resolvendo conflitos de identidade nos registros em si. Uma identidade pode ser de um indivíduo, de uma organização, de um objeto ou de qualquer outra entidade para a qual possa existir ambiguidade. A resolução de identidade pode ser vital em diversos campos, incluindo gerenciamento de relacionamento com o cliente, detecção de fraude, medidas contra lavagem de dinheiro e segurança nacional e internacional.

Suponha que haja os seguintes registros de cliente a partir de duas origens diferentes e você não tenha certeza se eles se referem à mesma pessoa ou a pessoas diferentes.

### Origem 1

Record no.: 70001  
Name: Jon Smith  
Address: 123 Main Street  
Tax Reference: 555-00-1111  
Driv. License: 0001133107  
Cred. Card: 10229127

### Origem 2

Record no.: 9103  
Name: JOHNATHAN Smith  
Date of Birth: 06/17/1934  
Telephone: 555-1212  
Cred. Card: 10229128  
Email: jls@mail.com  
IP address: 9.50.18.77

Não há correspondências exatas nos dados entre os dois registros. No entanto, se introduzirmos uma terceira origem, encontraremos alguns atributos comuns.

### Origem 3

Record no.: 6251  
Name: Jon Smith  
Telephone: 555-1212  
Driv. License: 0001133107  
Cred. Card: 10229132

O número da carteira de motorista liga aos registros na Origem 1 e na Origem 3, enquanto o número de telefone liga às Origens 2 e 3. Portanto é possível ter uma certeza razoável de que as três origens se referem à mesma pessoa.

Mas e se a distinção não for tão fácil de fazer? É possível que haja muito poucos dados para basear a decisão. Considere os dois registros a seguir.

### Origem 4

Record no.: S45286  
Name: John T Smith Jr  
Address: 456 Main Street  
Telephone: 703-555-2000  
Date of birth: 03/12/1984

Record no.: S45287  
Name: John T Smith  
Address: 456 Main Street  
Telephone: 703-555-2000  
Driv. License: 009900991

É evidente que esse não é o mesmo Sr. Smith dos registros anteriores, as diferenças são suficientes para definir isso como regra. No entanto, ainda existe um problema. Dois registros diferentes, a partir da mesma origem de dados, parecem estar relacionados à mesma pessoa. São registros duplicados? Não podemos ter certeza, a não ser que possamos encontrar outro registro relacionado que forneça mais informações, talvez a partir de uma origem diferente.

## Origem 5

Record no.: 769582-2  
 Name: John T Smith Sr  
 Address: 456 Main Street  
 Telephone: 703-555-2000  
 Driv. License: 009900991  
 Date of birth: 06/25/1959

Isso resolve o problema. Os dois registros na Origem 4 não são duplicados, mas são, de fato, um pai e filho com o mesmo nome, que moram no mesmo endereço e usam o mesmo número de telefone. Em um sistema manual, levaria semanas de procura para localizar o registro que resolveu as identidades. Com um sistema de análise de entidade automatizado, o tempo de resolução é radicalmente reduzido

---

## Análise de entidade e análise preditiva

Se todos os dados consistissem em uma única origem de registros completos e inequívocos, seria relativamente simples para o IBM SPSS Modeler resolver quaisquer conflitos de identidade. Usando somente a análise preditiva, é possível ler os dados para o IBM SPSS Modeler, executar o processamento e obter resultados confiáveis.

No entanto, na realidade, o cenário é geralmente muito diferente. Geralmente, os dados estão longe de serem completos, são frequentemente ambíguos e geralmente estão espalhados em várias origens de dados diferentes, registrando muitos atributos com poucos campos sobrepostos. Parte do valor da análise de entidade está na coleta de dados a partir de todas as origens diferentes em uma única área de armazenamento central, conhecida como **repositório**. Em seguida, o sistema de análise de entidade examina os dados nos mínimos detalhes para resolver conflitos, incluindo um identificador exclusivo para os registros que se originam da mesma pessoa ou organização.

A tabela a seguir ilustra as diferenças entre os dois tipos de análise.

*Tabela 1. Diferenças entre análise preditiva e de entidade.*

Característica	Análise preditiva	Análise de entidade
Tipos de dados de treinamento	Baseados em conjuntos relativamente pequenos e intervalos numéricos	Não é possível explorar grandes conjuntos (campos sem tipo) como nomes e endereços
Tamanho dos dados de treinamento	Geralmente ignora conjuntos grandes (campos sem tipo)	Todos os dados usados
Generalização	O algoritmo generaliza entre os dados de treinamento para formar um modelo conciso	Dados persistentes em estruturas adequadas para correspondência de entidade e detecção de relacionamento
Detecção de fraude	Registros sinalizados como potencialmente fraudulentos se tiverem as características típicas de um aplicativo fraudulento	Registros sinalizados como potencialmente fraudulentos se estiverem relacionados a registros fraudulentos ou se forem originados a partir dos mesmos indivíduos, mas com identidades diferentes

---

## Capítulo 2. Análise de entidade com o IBM SPSS Modeler

---

### Usando análise de entidade com o IBM SPSS Modeler

Você suspeita que talvez haja problemas de identidade com os dados. Por exemplo, indivíduos podem aparecer mais de uma vez ou indivíduos distintos podem parecer estarem mesclados ou ausentes. Como o IBM SPSS Modeler Entity Analytics pode ajudar a resolver isso? A seguir há um procedimento sugerido, embora possa ser necessário variá-lo para adequar aos seus requisitos específicos.

- Leia os dados de origem no IBM SPSS Modeler
- Crie um repositório pronto para armazenar os dados
- Conecte o IBM SPSS Modeler ao repositório
- Mapeie os campos de dados para as variáveis do repositório
- Exporte os dados no repositório e resolva as identidades
- Analise as identidades resolvidas
- Resolva os novos casos com relação ao repositório
- Gere quaisquer alertas necessários (em lote ou em tempo real)

Nesse momento, é necessário saber um pouco sobre como o IBM SPSS Modeler funciona. O IBM SPSS Modeler é uma ferramenta muito fácil e simples, baseada na representação gráfica de um fluxo de dados que flui por vários nós. Cada nó representa um determinado estágio do fluxo de trabalho.

O próprio IBM SPSS Modeler fornece uma ampla variedade de nós, cobrindo todas as funções de mineração de dados padrão. O IBM SPSS Modeler Entity Analytics inclui nós para serem usados especificamente na análise de entidade. São eles o nó Exportação do EA, o nó de origem Entity Analytics (EA) e o nó de processamento EA de fluxo.

A figura a seguir ilustra o processo.

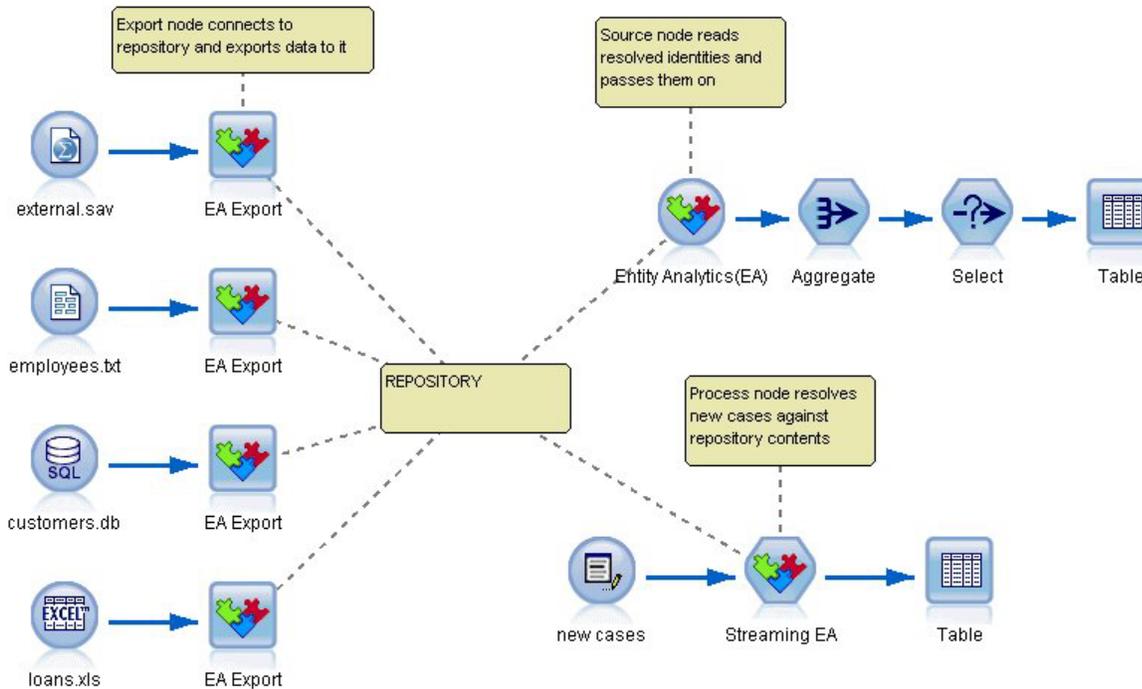


Figura 1. O processo de análise de entidade

## Estágio 1: Ler os dados de origem para o SPSS Modeler

Sua primeira tarefa é ler seus dados em SPSS Modeler por meio de um ou mais nós de origem, denotados em SPSS Modeler por um ícone redondo.

Os dados podem estar em qualquer formato suportado pelo SPSS Modeler, como arquivos de texto, tabelas de banco de dados, planilhas, arquivos XML e assim por diante, mas cada formato diferente requer um nó de origem do SPSS Modeler correspondente. Na ilustração, trata-se de um nó de origem Banco de dados.

Cada arquivo de origem de dados deve ter um campo que identifique exclusivamente cada registro. Se alguma origem de dados não tiver um campo desse, será possível incluir um facilmente no SPSS Modeler. Consulte o tópico “Incluindo um identificador de registro exclusivo” na página 12 para obter mais informações.

Consulte o tópico “Conectando-se a uma origem de dados” na página 12 para obter mais informações.

**Nota:** Dados de caracteres não latinos não são suportados. Quando os dados consistirem em uma combinação de registros de conjuntos de caracteres latinos (Europa Ocidental, por exemplo) e não latinos, somente as entradas de dados latinos serão resolvidas.

## Estágio 2: Criar o repositório

O ponto focal de todos os esforços de análise de entidade é o repositório - a área de armazenamento central onde são coletados todos os registros de dados.

Para criar um repositório, comece conectando a origem de dados a um nó Exportação do EA, representado pelo ícone quadrado.

A partir do nó de exportação, é possível criar um novo repositório (ou selecionar um existente), pronto para receber os dados exportados.

O processo de criação de um repositório é completamente descrito mais adiante. Consulte o tópico “Configurando um repositório de entidade (nó Exportação do EA)” na página 11 para obter mais informações.

**Nota:** Se estiver executando no modo de servidor remoto, você deverá criar o repositório na máquina do Modeler Server (ou seja, você deverá estar conectado ao Modeler Server a partir do Modeler Client ao criar o repositório, para que o repositório do EA seja criado na máquina servidor).

Depois de configurar um repositório, será possível manter seus conteúdos de várias formas. Consulte o tópico “Configurando um repositório de entidade” na página 16 para obter mais informações.

### **Estágio 3: Conectar o SPSS Modeler ao repositório**

Depois de criar o repositório, conecte-o ao fluxo do SPSS Modeler.

Consulte o tópico “Opções de repositório de entidade” na página 14 para obter mais informações.

### **Estágio 4: Mapear os campos de entrada para variáveis de repositório**

As origens de dados podem conter vários tipos diferentes de informações da entidade. Alguns tipos de informações são comuns para a maioria das origens de dados de entidade, enquanto outras podem ser específicas para uma determinada origem de dados. Em um repositório da entidade, esses tipos de informações diferentes são conhecidos como **variáveis**. O repositório fornece várias variáveis como padrão, e também é possível criar as suas próprias variáveis.

Uma variável de repositório é um tipo de informação individual que pode ser usado com uma origem de dados de entidade. Algumas variáveis (por exemplo, Nome, Sobrenome, Data de nascimento e assim por diante) têm a possibilidade de serem usadas com muitas origens de dados diferentes, enquanto outras variáveis são específicas de uma determinada origem de dados. Uma variável é, geralmente, o equivalente a um campo em um registro de dados ou uma coluna em uma tabela de base de dados.

Após criar um repositório e conectar-se a ele, designe um campo dos dados de entrada como o campo **chave exclusiva**, que será usado nas análises subsequentes. Além disso, mapeie os campos de dados de entrada para suas variáveis correspondentes no repositório. O mapeamento para variáveis predefinidas informa ao repositório de entidade quais campos devem ser comparados e, principalmente, como compará-los. O nó Exportação do EA fornece uma tabela de mapeamento na qual é possível criar os mapeamentos.

Consulte o tópico “Mapeando campos de entrada para variáveis (nó Exportação do EA)” na página 14 para obter mais informações.

### **Estágio 5: Exportar dados para o repositório e resolver identidades**

Cada nó de origem de dados precisa de seu próprio nó Exportação do EA, portanto, se os dados estiverem espalhados em várias origens diferentes, poderá haver um fluxo com várias origens de dados, cada uma conectada a um nó Exportação do EA separado. Consulte o tópico “Usando análise de entidade com o IBM SPSS Modeler” na página 3 para obter mais informações.

É possível escolher ler os registros de uma, de algumas ou de todas as origens de dados, caso haja mais de uma. O sistema Entity Analytics analisa os registros selecionados e inclui um campo identificador nomeado como \$EA\_ID em cada um. Quando dois ou mais registros relacionados a identidades anteriormente ambíguas puderem ser resolvidos, os identificadores incluídos nesses registros serão exclusivos em todo o repositório. O sistema também inclui um campo para mostrar a origem de dados a partir da qual o registro se originou.

Você conecta cada nó de origem de dados ao seu próprio nó Exportação do EA, mapeia os campos de entrada para as variáveis de repositório e, em seguida, executa o fluxo para exportar os dados a partir do SPSS Modeler para o repositório e resolver quaisquer conflitos de identidade, em uma única operação. Para ilustrar como isso funciona, suponha que os seguintes registros existam em quatro origens de dados diferentes.

#### Dados externos

*Tabela 2. Dados externos*

Nome	Telefone	Risco de crédito
Mike	555-1234	560
Joe	555-4567	780

#### Funcionários

*Tabela 3. Funcionários*

Nome	Address	Telefone
Michael	Rua 5, 1234	555-1234
Fred	Avenida 1, 543	555-9876

#### Clientes

*Tabela 4. Clientes*

Nome	Address	Economia
Susan	Rua 5, 1234	\$ 1234
Joe	Rua Oak, 777	\$ 5

#### Empréstimos

*Tabela 5. Empréstimos*

Nome	Address	Telefone	Empréstimo
Sue	Rua 5, 1234	555-1234	\$ 10.000
Joseph	Rua Oak, 777	555-4567	\$ 50.000

Como vimos, você exporta uma origem de dados de cada vez. Ao fazer isso, o repositório atualiza a resolução de cada registro. No repositório, cada registro é precedido por um campo identificador (nomeado como \$EA-ID) e um campo indicador (nomeado como \$EA-SRC), que mostram a origem de dados na qual o registro se originou. Portanto, nesse exemplo, após a exportação das quatro origens de dados, os conteúdos do repositório serão semelhantes a estes.

*Tabela 6. Exemplo de conteúdos do repositório após o estágio de exportação.*

\$EA-ID	\$EA-SRC	Nome	Telefone	Address	Risco de crédito	Economia	Empréstimo
1	Funcionários	Michael	555-1234	Rua 5, 1234			
1	Externo	Mike	555-1234		560		
2	Clientes	Joe		777 Oak St		\$ 5	
2	Externo	Joe	555-4567		780		
2	Empréstimos	Joseph	555-4567	777 Oak St			\$ 50.000

Tabela 6. Exemplo de conteúdos do repositório após o estágio de exportação (continuação).

\$EA-ID	\$EA-SRC	Nome	Telefone	Address	Risco de crédito	Economia	Empréstimo
3	Funcionários	Fred	555-9876	Avenida 1, 543			
4	Clientes	Susan		Rua 5, 1234		\$ 1234	
4	Empréstimos	Sue	555-1234	Rua 5, 1234			\$ 10.000

O sistema Entity Analytics determinou que *Mike* no conjunto de dados *Externo* é a mesma pessoa que *Michael* no conjunto de dados *Funcionários*, com base no número de telefone em comum e designou a ele o ID 1.

O caso de *Joe* no conjunto de dados *Externo* é um pouco mais difícil. Ele é a mesma pessoa que o *Joe* em *Clientes*? É impossível saber considerando somente essas duas origens de dados, mas temos uma terceira origem, *Empréstimos*, que contém um *Joseph*. Agora, temos uma correspondência: o número de telefone do *Joseph* é igual ao do *Joe* no conjunto de dados *Externo*. Com essa base, o sistema determina que eles são a mesma pessoa e fornece a eles o identificador 2.

Não há diversos registros para *Fred*, portanto, ele recebe o ID 3. *Susan* de *Clientes* é identificada como a mesma pessoa que *Sue* de *Empréstimos* pois elas têm o mesmo endereço, portanto o ID 4 é designado a ela.

*Observação:* Esse é um exemplo de correspondência otimista para propósitos ilustrativos. É possível escolher um conjunto de regras mais pessimista, em que um simples nome e um número de telefone ou endereço sozinhos não constituiriam uma correspondência exata e designar a ambos os registros o mesmo identificador.

## Estágio 6: Analisando as identidades resolvidas

Já tendo resolvido os conflitos de identidade no repositório, agora é possível executar análise e processamento adicionais nos resultados. Por exemplo, se você suspeitar de uma possível atividade fraudulenta com a existência de registros duplicados para a mesma identidade, será interessante produzir um relatório listando os duplicados.

Comece criando um nó de origem Entity Analytics (EA) e ligando-o ao repositório.

A saída comum do nó consiste nos seguintes campos.

- O campo identificador incluído pelo sistema (*\$EA-ID* no exemplo do Estágio 5)
- O campo indicador de origem incluído pelo sistema (*\$EA-SRC* no exemplo do Estágio 5)
- O campo de chave exclusiva designado no Estágio 4

Além disso, se estiver verificando relacionamentos, a seguinte saída será produzida. Consulte o tópico “Selecionando uma origem de dados” na página 24 para obter mais informações.

- O grau de separação entre as entidades (*\$EA-DEGREE*)
- O campo-pai (*\$EA-PARENT*)
- O campo-filho (*\$EA-CHILD*)
- A regra que identifica o relacionamento (*\$EA-RULE*)

Para visualizar a saída no SPSS Modeler, é possível anexar um nó de saída do SPSS Modeler, como um nó de tabela ou um nó de relatório e executar essa parte do fluxo. Caso seja necessário sumarizar a saída, que pode ser muito grande, será possível incluir nós de operação de registro, como os nós de agregado ou de seleção.

O nó de origem Entity Analytics (EA) é completamente descrito mais adiante. Consulte o tópico “Analisando as identidades resolvidas [nó de origem Entity Analytics (EA)]” na página 24 para obter mais informações.

## Estágio 7: Resolver novos casos com relação ao repositório

Você resolveu as identidades de todos os registros em todas as origens de dados. Mas o que acontece quando você deseja comparar um conjunto de registros novos para ver como eles se relacionam com o que você conhece, para obter uma escoragem melhor? É aí que entra o nó EA de fluxo.

Primeiro, inclua um novo nó de origem de dados do SPSS Modeler para ler os novos dados para o fluxo. Em seguida, conecte esse nó de origem a um nó EA de fluxo. Para visualizar a saída, inclua um nó de tabela, como antes.

Ao executar essa parte do fluxo, o nó EA de fluxo lê cada registro novo e compara-o com relação aos conteúdos do repositório. Se ele localizar registros correspondentes no repositório, o nó EA de fluxo emitirá todos os registros correspondentes juntamente com o novo registro, ao qual incluirá os campos de ID e indicador de origem. Se nenhuma correspondência for localizada, o nó de processamento emitirá somente o novo registro com os campos de ID e indicador de origem incluídos.

Para ilustrar isso, suponha que o repositório consista atualmente nos conteúdos que foram emitidos pelo nó de origem Entity Analytics (EA). Consulte Tabela 6 na página 6.

Agora, recebemos os seguintes registros novos. Eles estão relacionados a alguém que conhecemos?

Tabela 7. Novos registros a serem escorados.

Nome	Address	Telefone	Empréstimo
Suzan	Rua 5, 1234	555-1234	\$ 100.000
Mark	Avenida 9, 888	555-9999	\$ 60.000

Comparando os novos dados com os conteúdos existentes do repositório, o nó EA de fluxo corresponde o primeiro novo registro com a pessoa que tem o identificador 4 nos registros existentes. No entanto, nenhuma correspondência pode ser localizada para o segundo novo registro, portanto, um novo identificador exclusivo, 5, é designado a ele.

O nó EA de fluxo inclui os campos identificador e indicador de origem e emite os novos registros juntamente com todos os seus registros correspondentes. Assim, a saída será a seguinte.

Tabela 8. Saída do nó EA de fluxo.

\$EA-ID	\$EA-SRC	Nome	Telefone	Address	Risco de crédito	Economia	Empréstimo
4	Custo	Susan		Rua 5, 1234		\$ 1234	
4	Empréstimo	Sue	555-1234	Rua 5, 1234			\$ 10.000
4	Novo empréstimo	Suzan	555-1234	Rua 5, 1234			\$ 100.000
5	Novo empréstimo	Mark	555-9999	Avenida 9, 888			\$ 60.000

Em seguida, essa saída pode ser agregada usando o identificador de análise de entidade como a chave de agregação e passado aos outros nós para processamento adicional.

O nó EA de fluxo é completamente descrito mais adiante.

## Estágio 8: Gerar alertas

Mais uma vez, a atividade potencialmente suspeita pode tornar-se aparente. Nesse caso, a pessoa com o identificador 4 já tem um empréstimo de \$ 10,000 e agora está aplicando, com um nome um pouco diferente, para um outro dez vezes esse tamanho. Naturalmente, isso pode ser perfeitamente aceitável e não ter intenção fraudulenta. No entanto, se essa atividade contar como suspeita de acordo com as regras de negócios, talvez seja válido verificá-la.

É possível, por exemplo, anexar e executar um nó de tabela do SPSS Modeler ou um nó de relatório, imprimir os conteúdos de sua janela de saída e solicitar que alguém os leia e gere alertas manualmente. Como alternativa, é possível passar a saída do nó EA de fluxo para um modelo de avaliação de risco criado anteriormente no IBM SPSS Modeler, produzindo um conjunto de escores que reflitam as regras de negócios com mais proximidade. Outra possibilidade é exportar a saída para um banco de dados ou algum outro meio para processamento adicional. Com o IBM SPSS Modeler, há uma ampla variedade de ações para atender aos seus requisitos específicos.



---

## Capítulo 3. Tarefas de análise de entidade

---

### Sobre as tarefas

Esta seção descreve as seguintes tarefas de análise de entidade.

- Instalando um repositório de entidade
- Configurando um repositório de entidade
- Analisando as identidades resolvidas
- Resolvendo novos casos com relação ao repositório de entidade
- Limpando um repositório de entidade
- Excluindo um repositório de entidade
- Usando a análise de entidade com outros produtos IBM SPSS
- Administrando a análise de entidade

---

### Configurando um repositório de entidade (nó Exportação do EA)

O processo de configuração de um repositório de entidade consiste nas seguintes tarefas.

1. Conecte-se a uma origem de dados. Consulte o tópico “Conectando-se a uma origem de dados” na página 12 para obter mais informações.
2. Crie o repositório. Consulte o tópico “Criando o repositório” na página 12 para obter mais informações.
3. Mapeie os campos de entrada na origem de dados para as variáveis no repositório. Consulte o tópico “Mapeando campos de entrada para variáveis (nó Exportação do EA)” na página 14 para obter mais informações.

Depois de configurar os mapeamentos, será possível exibi-los para a origem de dados atual ou para todas as origens de dados conhecidas do repositório. Consulte o tópico “Exibindo os mapeamentos de campo (nó Exportação do EA)” na página 16 para obter mais informações.

**Nota:** A partir da versão 16, o SPSS Entity Analytics suporta repositórios no produto IBM DB2. Como um repositório é específico de uma versão do SPSS Modeler e não pode ser importado de uma versão anterior, se houver um repositório existente e você estiver fazendo upgrade para o SPSS Entity Analytics versão 16, será necessário recriar esse repositório no novo banco de dados DB2.

### O repositório de entidade

O repositório fornece uma área de armazenamento central, agindo como um cache de dados para todas as informações de entidade. Como o repositório é ativo, ele tem um único estado, portanto, não existe o conceito de versionamento com um repositório de entidade. O repositório mantém o estado atual de todos os dados de entrada e pode ficar muito grande.

É possível manter os conteúdos do repositório por meio de uma interface gráfica fácil de usar. Consulte o tópico “Configurando um repositório de entidade” na página 16 para obter mais informações.

*Importante:* A partir da versão 16, o IBM SPSS Modeler Entity Analytics suporta os repositórios no produto IBM DB2. As versões anteriores do SPSS Entity Analytics suportavam os repositórios hospedados no IBM solidDB. Se houver algum repositório solidDB existente, será necessário recriá-lo no novo banco de dados DB2 no caso de upgrade para o SPSS Entity Analytics versão 16 ou posterior.

*Observação:* A versão do IBM SPSS Modeler Entity Analytics fornecida com o IBM SPSS Modeler Premium somente suporta um único repositório hospedado no produto IBM DB2 que seja incluído em pacote configurável com o SPSS Entity Analytics. Com essa versão, deve-se excluir o repositório existente antes de criar um novo. Está disponível um upgrade licenciado separadamente para o SPSS Entity Analytics (conhecido como IBM SPSS Modeler Entity Analytics Unleashed) que permite a coexistência de mais de um repositório no mesmo sistema. Cada repositório pode conter mais de 10 milhões de linhas e usar mais de quatro núcleos do processador. Entre em contato com o representante de suporte IBM para obter detalhes.

## Conectando-se a uma origem de dados

Comece lendo os dados de origem no SPSS Modeler por meio do nó de origem.

Para conectar-se a uma origem de dados

1. Na guia **Origens**, na paleta de nós, na parte inferior da janela principal do SPSS Modeler, clique duas vezes em um ícone que corresponda ao tipo de dados de origem. Essa ação inclui um nó de origem na tela.
2. Na tela, clique duas vezes no ícone para abrir sua caixa de diálogo.
3. No campo **Arquivo**, insira o local e o nome do arquivo de dados de origem.
4. Preencha o restante da caixa de diálogo conforme o necessário (clique no botão **Ajuda** para obter mais informações) e, em seguida, clique em **OK**.
5. Se o arquivo de dados de origem não tiver um campo que identifique exclusivamente cada registro, inclua um por meio de um nó de deriva. Consulte o tópico “Incluindo um identificador de registro exclusivo” para obter mais informações.

**Nota:** Dados de caracteres não latinos não são suportados. Quando os dados consistirem em uma combinação de registros de conjuntos de caracteres latinos (Europa Ocidental, por exemplo) e não latinos, somente as entradas de dados latinos serão resolvidas.

## Incluindo um identificador de registro exclusivo

Cada arquivo de origem de dados que for uma entrada para o repositório de entidade deverá ter um campo que o identifique exclusivamente. Se algum arquivo de origem de dados não tiver um campo desse, será possível incluir um por meio do nó de deriva do SPSS Modeler.

Para incluir um identificador de registro exclusivo em um arquivo de origem de dados

1. Na tela, clique no nó de origem incluído na tarefa anterior.
2. Na guia **Ops. de campo** na paleta de nós, clique duas vezes no ícone **Derivar** para anexar um nó de deriva ao nó de origem.
3. Na tela, clique duas vezes no nó de deriva para abrir sua caixa de diálogo.
4. No campo **Derivar**, substitua o nome padrão por um nome significativo (como o **ID**) para o campo identificador que está sendo incluído.
5. Assegure-se de que o campo **Derivar como** esteja configurado como **Fórmula**.
6. Configure **Tipo de campo** para **Contínuo**.
7. Na caixa de texto **Fórmula**, digite **@INDEX** e clique em **OK**.

## Criando o repositório

É necessário criar um repositório para armazenar todos os dados de entrada.

**Nota:** Se estiver executando no modo de servidor remoto, você deverá criar o repositório na máquina do Modeler Server (ou seja, você deverá estar conectado ao Modeler Server a partir do Modeler Client ao criar o repositório, para que o repositório do EA seja criado na máquina servidor).

Para criar um repositório

1. Na guia Exportar da paleta de nós do SPSS Modeler, coloque um nó Exportação do EA na tela do fluxo.  
*Observação:* Se estiver criando um repositório pela primeira vez, use um nó Exportação do EA e conecte-o ao nó de origem do SPSS Modeler que contém os dados que deseja inserir no repositório (ou no nó de deriva, caso precise de um para obter um campo identificador). Para conectar os nós, faça o seguinte.
  - a. Clique com o botão direito no nó de origem do SPSS Modeler.
  - b. Escolha Conectar.
  - c. Clique no nó Exportação do EA.
2. Clique duas vezes no nó Exportação do EA para abrir sua caixa de diálogo.
3. Clique na lista **Repositório de entidade**.
4. Clique em <Procurar...> para exibir a caixa de diálogo Repositórios de entidade.
5. Na caixa de diálogo Repositórios de entidade, clique no campo Nome do repositório.
6. Escolha <Criar novo repositório...> para exibir o assistente Criar repositório.

## Assistente Criar repositório

### Etapa 1

Aqui você escolhe criar um repositório local, usando o produto IBM DB2 incluído em pacote configurável com o IBM SPSS Modeler Entity Analytics, ou usar um banco de dados externo para o repositório.

**Criar repositório local.** Especifique um nome do usuário administrador e os detalhes de senha para o banco de dados IBM DB2 que deve hospedar o repositório que está sendo criado. Confirme a senha e clique em **Avançar**.

**Nota:** Não é possível usar os símbolos traço ou sublinhado no nome do usuário.

As credenciais que devem ser usadas para o banco de dados IBM DB2 dependem do sistema operacional. Os usuários UNIX devem usar o nome do usuário g2user e a senha G2password.

As tarefas de administração de repositório nos nós Entity Analytics, como criar ou destruir um repositório, requerem permissões adicionais. No UNIX, o usuário conectado ao IBM SPSS Modeler Server deve ser o usuário raiz ou o g2user e deve ser membro do grupo db2iadm1. No Windows, o usuário que efetua login no IBM SPSS Modeler Server precisa ser membro do grupo DB2ADMNS para realizar administração de repositório.

Mais tarde, caso seja necessário alterar as credenciais do administrador, use editor de linha de comando do banco de dados. Consulte o tópico “Gerenciando credenciais do administrador para a base de dados do repositório” na página 31 para obter mais informações.

**Nota:** Somente uma combinação de nome do usuário e senha é possível. Todos os usuários que efetuam login no repositório compartilham o mesmo nome do usuário e a mesma senha.

**Incluir repositório externo.** Use essa opção se desejar usar um banco de dados externo para hospedar o repositório. Digite o local do arquivo .ini do banco de dados no campo **Selecionar arquivo .ini de repositório** e clique em **Avançar**.

### Etapa 2

**Nome do novo repositório.** Digite um nome exclusivo para o novo repositório.

**Importar configuração de.** (Somente repositório local) Se você desejar basear a configuração naquela de um repositório existente, escolha o repositório aqui, caso contrário, escolha **Padrão**. Consulte o tópico “Configurando um repositório de entidade” na página 16 para obter mais informações.

Se você escolher um repositório existente, insira os detalhes da conexão caso sejam diferentes dos inseridos na tela anterior.

Clique em **OK** para criar o novo repositório e exibir a caixa de diálogo Instâncias de resolução da entidade, que pode ser usada para se conectar ao repositório.

## Opções de repositório de entidade

A caixa de diálogo Repositórios de entidade contém várias opções para criar, conectar, configurar e manter um repositório de entidade.

**Conectar ao repositório.** Use essas opções para criar um novo repositório de entidade ou para conectar-se a um existente.

- **Nome do repositório.** Mostra o repositório de entidade atual, caso exista algum. Para escolher um repositório diferente, caso exista mais que um, selecione um na lista.  
Para criar um novo repositório, selecione <**Criar novo repositório...**>. Essa ação inicia um assistente que orienta o processo de criação.
- **Nome do usuário.** Insira um nome do usuário válido para o repositório selecionado.
- **Senha.** A senha para esse nome do usuário.
- **Conectar.** Clique para conectar-se ao repositório atual.

**Gerenciar repositório.** A tabela lista as origens de dados que foram carregadas no repositório atual (aquele ao qual você está conectado), mostrando o número de registros em cada origem de dados.

- **Atualizar.** Atualiza a origem de dados e as informações de tamanho na tabela, por exemplo, se você tiver incluído uma nova origem de dados ou alterado o tamanho de uma existente.
- **Limpar todos.** Remove todos os dados de origem do repositório, mas mantém todos os detalhes de configuração. Será possível usar essa opção se as informações de configuração ainda forem úteis, mas você desejar remover todos os registros de dados do repositório. Consulte o tópico “Limpar um repositório de entidade” na página 33 para obter mais informações.
- **Excluir não usado.** Remove todos os dados de origem destacados do repositório, mas mantém todos os detalhes de configuração. Consulte o tópico “Excluindo origens de dados não usadas de um repositório” na página 33 para obter mais informações.
- **Renomear origem.** Abre uma caixa de diálogo na qual é possível alterar o nome da origem de dados destacada.

**Nota:** Essa ação renomeia o origem de dados dentro do repositório. Ainda será necessário selecionar novamente esse novo nome da origem de dados em qualquer exportação ou nó de fluxo existentes a partir do qual ele é referenciada.

**Destruir repositório inteiro.** Destroi completamente os conteúdos do repositório atual e os detalhes de configuração. Consulte o tópico “Excluindo um repositório de entidade” na página 34 para obter mais informações.

**Configurar repositório.** Exibe uma janela na qual é possível configurar o repositório atual. Consulte o tópico “Configurando um repositório de entidade” na página 16 para obter mais informações.

## Mapeando campos de entrada para variáveis (nó Exportação do EA)

O repositório fornece várias variáveis predefinidas como padrão. Diferentes origens de dados podem usar diferentes nomes de campo (por exemplo, **Endereço01** ou **Linha de endereço 1**) para tipos de informação que correspondam à mesma variável. Para evitar duplicação, é necessário mapear os campos de origem de dados de entrada para variáveis de repositório específicas. Não é necessário mapear todos os campos no conjunto de dados, somente aqueles que possam corresponder à mesma variável em outros conjuntos de dados.

Quando uma origem de dados usar campos correspondentes a outros tipos de informação que não sejam predefinidos no repositório, será possível criar novas variáveis a partir da janela Configuração de repositório. Consulte o tópico “Configurando um repositório de entidade” na página 16 para obter mais informações.

Para mapear campos para variáveis

1. Anexe um nó Exportação do EA a um nó de origem de dados na tela de fluxo. Cada nó de origem de dados usado deve ser anexado ao seu próprio nó Exportação do EA.
2. Abra o nó Exportação do EA para exibir a guia Entradas, que contém opções para mapear os campos de entrada. Consulte o tópico “Opções de entrada de repositório para mapeamento” para obter mais informações.
3. No nó Exportação do EA, selecione a guia Repositório para visualizar as designações de mapeamento para a origem de dados atual ou para todas as origens de dados, se estiver usando mais de uma.
4. Para salvar um conjunto de designações de mapeamento (por exemplo, para usar com um nó de exportação para uma origem de dados diferente) clique em **Exportar mapeamento**.

Ao concluir o mapeamento do primeiro nó de origem de dados, repita o processo para todos os outros nós de origem de dados que desejar usar.

### Opções de entrada de repositório para mapeamento

A guia Entradas contém as opções para o mapeamento de campos de origem de dados para variáveis de repositório, prontas para serem exportadas para o repositório. Configure as designações de mapeamento nessa guia e, opcionalmente, clique na guia Repositório para ver o mapeamento de outras origens de dados e, em seguida, clique em **Executar** para exportar os dados para o repositório.

Se você já tiver armazenado um conjunto de mapeamentos em um arquivo XML, será possível usá-lo clicando em **Importar mapeamento**.

**Modo.** Deixe a seleção padrão, **Incluir no repositório**, se desejar incluir os registros do arquivo de origem nos conteúdos existentes do repositório. Se desejar limpar os conteúdos do repositório, mas manter as informações de configuração antes de incluir os registros de origem, escolha **Limpar repositório antes de exportar**.

**Repositório de entidade.** Mostra o repositório de entidade atual, caso exista algum. Para escolher um repositório diferente, caso exista mais que um, selecione um na lista. Para criar um novo repositório, escolha <Procurar...> para exibir uma caixa de diálogo que pode ser usada para criar o repositório. Consulte o tópico “Opções de repositório de entidade” na página 14 para obter mais informações.

**Mapear para tipo de entidade.** Uma lista de tipos de entidade (ou seja, conjuntos de variáveis) definidos no repositório. Escolha um na lista ou escolha <Incluir novo tipo de entidade...> para exibir a janela de configuração do repositório, na qual é possível definir um novo tipo de entidade. Consulte o tópico “Configurando um repositório de entidade” na página 16 para obter mais informações.

**Tag de origem.** Uma lista de tags indicando as origens de dados atualmente conhecidas para o repositório. Escolha uma na lista ou escolha <Incluir nova tag de origem...> para criar uma tag para uma nova origem de dados.

**Chave exclusiva.** (obrigatório) O campo de entrada a ser usado para os identificadores exclusivos para os registros de dados.

**Tabela de mapeamento.** Nessa tabela, é possível mapear cada campo de entrada para uma variável correspondente no repositório. Se não existir uma variável no tipo de entidade selecionada, será possível criar uma nova variável aqui.

- **Campo.** O conjunto de campos de entrada na origem de dados selecionada. Cada campo tem um ícone que indica o nível de medição (ou seja, o tipo de dados) para o campo.

- **Mapeado para variável.** Para mapear um campo para uma variável, clique duas vezes nessa coluna (ou pressione a barra de espaço) na linha de campo e escolha uma variável na lista. Se uma variável adequada não estiver disponível, escolha <Incluir nova variável...> para exibir a janela de configuração de repositório, na qual é possível definir uma nova variável para esse tipo de entidade. Consulte o tópico “Configurando um repositório de entidade” para obter mais informações.
- **Uso.** Indica o contexto de um determinado campo no qual mais de um contexto é possível, por exemplo, os números de telefone de trabalho e residenciais. Existem tipos de uso predefinidos disponíveis para as variáveis ADDRESS e PHONE, e é possível criar seus próprios tipos de uso para todas as variáveis. Para configurar um uso que não seja o padrão (**Automático**), clique nessa coluna na linha desejada e escolha um dos tipos de uso existentes (caso haja algum) ou clique em <Incluir uso...> para criar um novo. Consulte o tópico “Mantendo os tipos de entidade” na página 21 para obter mais informações.

**Importar mapeamento.** Importa um conjunto exportado anteriormente de mapeamentos de campo-para-variável a partir de um arquivo XML externo. Isso pode ser útil no caso de diferentes origens de dados com os mesmos requisitos de mapeamento, pois evita a necessidade de redefinir os mesmos mapeamentos para diferentes origens.

**Exportar mapeamento.** Exporta para um arquivo XML externo o conjunto de mapeamentos de campo-para-variável mostrado na tabela de mapeamento.

## Exibindo os mapeamentos de campo (nó Exportação do EA)

Na guia Repositório, clique no botão **Atualizar** para ver quais variáveis de repositório têm campos de entrada mapeados para eles. Isso pode ser visto para a origem de dados atual (a controlada pelo nó de origem anexado a esse nó de exportação) ou para todas as origens de dados.

**Mostrar entradas para.** Escolha uma opção para exibir os mapeamentos da origem de dados atual ou de todas as origens de dados conhecidas do repositório.

**Atualizar.** Atualiza a exibição para a opção de entrada selecionada.

**Variáveis.** Uma lista de todas as variáveis que têm mapeamentos nas origens de dados exibidas. As variáveis não mapeadas não são mostradas.

<Origem de dados>. Cada coluna lista os campos mapeados em uma determinada origem de dados para cada variável que tenha um mapeamento definido.

---

## Configurando um repositório de entidade

É possível manter os conteúdos do repositório da janela Configuração de repositório, que fornece uma interface visual fácil de usar para o repositório inteiro.

Se você for usar mais de um repositório com configurações iguais ou semelhantes, será possível definir uma configuração básica e exportá-la para um arquivo que, em seguida, poderá ser importado para outros repositórios. Consulte o tópico “Reutilizando uma configuração de repositório” na página 23 para obter mais informações.

**Nota:** A partir da versão 16, o SPSS Entity Analytics suporta repositórios no produto IBM DB2. Como um repositório é específico de uma versão do SPSS Modeler e não pode ser importado de uma versão anterior, se houver um repositório existente e você estiver fazendo upgrade para o SPSS Entity Analytics versão 16, será necessário recriar esse repositório no novo banco de dados DB2.

### **CUIDADO:**

**Se você modificar e salvar a configuração de um repositório que já contenha dados, poderá ser solicitado que você limpe os conteúdos do repositório e recarregue os dados. Essa ação evita que o repositório fique em um estado inconsistente.**

Para definir uma configuração de repositório

1. Abra qualquer nó Entity Analytics.
2. Clique na lista **Repositório de entidade**.
3. Clique em **<Procurar...>** para exibir a caixa de diálogo Instâncias de resolução da entidade.
4. Na caixa de diálogo Instâncias de resolução da entidade, clique na lista **Nome do repositório**.
5. Selecione o repositório para o qual deseja definir a configuração.
6. Se você ainda não estiver conectado, insira o nome do usuário administrador e a senha e clique em **Conectar**.
7. Quando o botão **Configurar repositório** for ativado, clique nele para exibir a janela Configuração de repositório.
8. Crie os detalhes de configuração conforme é explicado nas seções a seguir.

A área de janela de navegação no lado esquerdo da janela Configuração de repositório contém uma estrutura em árvore a partir da qual é possível gerenciar as diferentes características do repositório.

*Tabela 9. Elementos principais da janela Configuração de repositório.*

Seção	Descrição	
Origens de dados	Exibir os mapeamentos de todas as origens de dados para as várias variáveis do repositório.	Consulte o tópico “Visualizando os mapeamentos de origem de dados” para obter mais informações.
Variáveis	Criar uma nova variável, ou duplicar, editar ou excluir um existente.	Consulte o tópico “Mantendo as variáveis de repositório” para obter mais informações.
Tipos de entidade	Criar um novo tipo de entidade ou gerenciar os existentes (duplicar, renomear, anexar ou remover variáveis, excluir).	Consulte o tópico “Mantendo os tipos de entidade” na página 21 para obter mais informações.
Regras de resolução	Configurar o limite para a correspondência de entidade.	Consulte o tópico “Configurando o limite para correspondência de entidade” na página 23 para obter mais informações.

## Visualizando os mapeamentos de origem de dados

Na seção Origens de dados da janela Configuração de repositório, a entrada Todas as origens fornece uma exibição somente leitura dos mapeamentos de todas as origens de dados para as várias variáveis do repositório.

Clique em **Atualizar** para atualizar a lista se novas origens de dados forem incluídas no repositório.

*Observação:* Não é possível incluir uma origem de dados no repositório aqui. As origens de dados somente podem ser incluídas ao criar um nó de origem do SPSS Modeler e conectá-lo a um nó Exportação do Entity Analytics. Consulte o tópico “Conectando-se a uma origem de dados” na página 12 para obter mais informações.

## Mantendo as variáveis de repositório

Uma variável de repositório é um tipo de informação individual que pode ser usado com uma origem de dados de entidade. Algumas variáveis (por exemplo, Nome, Sobrenome, Data de nascimento e assim por diante) têm a possibilidade de serem usadas com muitas origens de dados diferentes, enquanto outras variáveis são específicas de uma determinada origem de dados. Uma variável pode conter um ou mais elementos. Cada elemento é, geralmente, o equivalente a um campo em um registro de dados ou uma coluna em uma tabela de base de dados.

Na seção Variáveis da janela Configuração de repositório, a entrada Todos as variáveis fornece os meios para manter todas as variáveis de repositório. É possível fazer o seguinte.

- Criar uma nova variável
- Duplicar uma variável existente (por exemplo, para criar uma nova variável com base em uma existente)
- Editar uma variável existente
- Excluir uma variável existente

Instruções para essas tarefas são fornecidas posteriormente nesta seção.

A lista de variáveis mostra todas as variáveis que foram definidas nesse repositório. As colunas na lista mostram as diversas propriedades que uma variável pode ter.

**Variável.** O nome da variável. Um símbolo de cadeado próximo a um nome de variável indica que a variável está bloqueada. As variáveis bloqueadas não podem ser excluídas ou duplicadas, a única mudança nelas que pode ser salva é alterar o atributo de anonimização.

**Frequência.** Indica quantas entidades podem ter o mesmo valor para essa variável. Os valores válidos são **Um** (por exemplo, para um número de passaporte), **Poucos** (por exemplo, para um endereço) ou **Muitos** (por exemplo, para uma data de nascimento).

**Exclusividade.** Indica que, geralmente, uma entidade deve ter somente um desse tipo de variável. Por exemplo, uma data de nascimento ou um número de identidade nacional teriam o valor **Sim** aqui, enquanto um endereço ou um número do cartão de crédito teriam o valor **Não** (pois uma entidade pode ter vários endereços ou cartões de crédito).

**Estabilidade.** Indica o valor de estabilidade dessa variável (ou seja, se ela *provavelmente não* se altera durante o tempo de vida de uma entidade). Por exemplo, uma variável de data de nascimento teria o valor **Sim**, pois ela nunca se altera, mas uma variável de endereço teria **Não** pois é muito provável que ela se altere e ela é menos estável. *Observação:* O sexo é geralmente estável durante um tempo de vida, mas como ele é frequentemente especificado de forma incorreta devido a dados inválidos, a configuração padrão fornece o valor **Não**.

**Anonimar.** Indica se a variável foi anonimizada. As entradas são **Sim** ou **Não**. Consulte o tópico “Anonimizando as variáveis de repositório” na página 20 para obter mais informações.

Para criar uma nova variável

1. Execute um dos procedimentos a seguir.
  - Clique no botão Criar nova variável (o botão superior no lado direito da tela).
  - Clique com o botão direito em **Todas as variáveis** na área de janela de navegação no lado esquerdo da tela e escolha **Nova variável**.
2. Preencha a caixa de diálogo Incluir/editar variável. Consulte o tópico “Incluindo ou editando uma variável” na página 19 para obter mais informações.

Para duplicar uma variável existente

1. Na coluna **Variável** da tabela no lado direito da tela, selecione a variável que deseja duplicar.
2. Clique no botão Duplicar variável selecionada (o segundo botão no lado direito da tela).
3. Preencha a caixa de diálogo Incluir/editar variável. Consulte o tópico “Incluindo ou editando uma variável” na página 19 para obter mais informações.

Para editar uma variável existente

**CUIDADO** Se você editar, excluir ou anonimizar uma variável ou elemento de variável quando o repositório já contiver dados, será necessário limpar o repositório e recarregar os dados. Essa ação evita deixar o repositório em um estado inconsistente.

1. Na coluna **Variável** da tabela no lado direito da tela, selecione a variável que deseja editar. *Observação:* Somente é possível editar as variáveis que você criou, não as variáveis fornecidas pelo sistema.
2. Clique no botão Editar variável selecionada (o terceiro botão no lado direito da tela).
3. Preencha a caixa de diálogo Incluir/editar variável. Consulte o tópico “Incluindo ou editando uma variável” para obter mais informações.

Para excluir uma variável existente

**CUIDADO** Se você editar, excluir ou anonimizar uma variável ou elemento de variável quando o repositório já contiver dados, será necessário limpar o repositório e recarregar os dados. Essa ação evita deixar o repositório em um estado inconsistente.

1. Na coluna **Variável** da tabela no lado direito da tela, selecione a variável que deseja excluir. *Observação:* Somente é possível excluir as variáveis que você criou, não as variáveis fornecidas pelo sistema.
2. Execute um dos procedimentos a seguir.
  - Clique no botão Excluir variável selecionada (o botão inferior no lado direito da tela).
  - Clique com o botão direito em **Todas as variáveis** na área de janela de navegação no lado esquerdo da tela e escolha **Excluir**.
3. Clique em **Continuar** para confirmar a exclusão da variável.

**CUIDADO:**

**Não é possível desfazer a exclusão de uma variável.**

## Incluindo ou editando uma variável

**CUIDADO** Se você editar, excluir ou anonimizar uma variável ou elemento de variável quando o repositório já contiver dados, será necessário limpar o repositório e recarregar os dados. Essa ação evita deixar o repositório em um estado inconsistente.

Na caixa de diálogo Incluir/editar variável, é possível criar uma nova variável de repositório ou duplicar ou editar uma variável existente.

*Observação:* Se uma variável existente estiver bloqueada, não será possível editar seus detalhes nesta caixa de diálogo.

**Tipo de variável.** Um rótulo que indica o tipo de informação ao qual a variável está relacionada. Esse rótulo forma a primeira parte do identificador de variável.

**Descrição.** Uma breve descrição de texto do tipo de variável, somente para informação.

**Frequência.** Indica quantas entidades podem ter o mesmo valor para essa variável. Os valores válidos são **Um** (por exemplo, para um número de passaporte), **Poucos** (por exemplo, para um endereço) ou **Muitos** (por exemplo, para uma data de nascimento).

**Exclusividade.** Indica que, geralmente, uma entidade deve ter somente um desse tipo de variável. Por exemplo, uma data de nascimento ou um número de identidade nacional teriam o valor **Sim** aqui, enquanto um endereço ou um número do cartão de crédito teriam o valor **Não** (pois uma entidade pode ter vários endereços ou cartões de crédito).

**Estabilidade.** Indica o valor de estabilidade dessa variável (ou seja, se ela *provavelmente não* se altera durante o tempo de vida de uma entidade). Por exemplo, uma variável de data de nascimento teria o valor **Sim**, pois ela nunca se altera, mas uma variável de endereço teria **Não** pois é muito provável que

ela se altere e ela é menos estável. *Observação:* O sexo é geralmente estável durante um tempo de vida, mas como ele é frequentemente especificado de forma incorreta devido a dados inválidos, a configuração padrão fornece o valor **Não**.

**Tabela de elementos.** Uma lista dos elementos que essa variável abrange.

- **Elemento.** O nome do elemento.
- **Descrição.** Uma breve descrição do que o elemento fornece.
- **Tipo de dados.** O tipo de dados que pode ser usado para esse elemento. Os tipos disponíveis são: Sequência, Número inteiro, Real e Data.

**Botão Incluir novo elemento.** Inclui uma nova linha na tabela de elementos, para que seja possível definir um novo elemento.

**Botão Excluir elemento.** Exclui uma linha selecionada da tabela de elementos. Não é possível desfazer essa operação.

**CUIDADO** Se você editar, excluir ou anonimizar uma variável ou elemento de variável quando o repositório já contiver dados, será necessário limpar o repositório e recarregar os dados. Essa ação evita deixar o repositório em um estado inconsistente.

**Anonimar.** Para proteção de dados, é possível escolher anonimizar os dados conforme eles são incluídos em um repositório. Para ativar isso para uma variável, selecione **Sim**. Consulte o tópico “Anonimizando as variáveis de repositório” para obter mais informações.

## Anonimizando as variáveis de repositório

Como parte da segurança de dados, é interessante anonimizar os dados quando eles são incluídos no repositório para reduzir o risco de que informações de identificação pessoal sejam divulgadas inadvertidamente.

Quando dados anonimizados são exportados para um repositório, é necessário um método de anonimização que ainda permita que a resolução da entidade seja realizada com os dados anonimizados. Por exemplo, se dois registros de dados de detalhes do cartão de crédito de uma pessoa forem anonimizados como "anon\_s21" e "anon\_s9271" eles perderão seu relacionamento. No entanto, usando uma ligação interna de segundo plano entre os registros, o sistema ainda poderá entender que um nome é uma forma abreviada do outro.

As ligações e os identificadores de segundo plano que permitem que os dados anonimizados sejam vinculados são gerados quando o repositório é criado e são exclusivos para o repositório. Os dados criptografados são armazenados internamente e, posteriormente, lidos quando um fluxo se conecta a um repositório.

Ao configurar o repositório, é possível especificar se cada variável individual será anonimizada ou não. Se uma variável for anonimizada, todos os seus elementos serão anonimizados e eles serão sempre anonimizados, independentemente de seu tipo de uso. Consulte o tópico “Incluindo ou editando uma variável” na página 19 para obter mais informações.

**Nota:** Assegure-se de não anonimizar todos os campos do SPSS Entity Analytics ou não será possível identificar quais dados retornam. É recomendado deixar pelo menos um campo (mesmo que seja somente um número de linha) sem o anonimato, para que seja possível controlar a nova mesclagem com os dados originais posteriormente.

Uma coluna na lista de variável na janela Configuração de repositório mostra quais variáveis foram configuradas para anonimizar. As entradas são **Sim** ou **Não**.

**Nota:** Se um repositório existente contiver quaisquer dados antes que as variáveis sejam anonimizadas, você deverá limpar todos os dados primeiro, caso contrário não ocorrerão correspondências entre as variáveis anonimizadas e não anonimizadas.

## Mantendo os tipos de entidade

Um **tipo de entidade** é um conjunto nomeado de variáveis de repositório que se pertencem logicamente, em conjunto. Por exemplo, um tipo de entidade para ser usado com um conjunto de dados de cliente pode consistir em variáveis como Nome, Data de nascimento, Gênero, Endereço, Número de telefone e assim por diante.

O repositório do IBM SPSS Modeler Entity Analytics é fornecido com um conjunto padrão de tipos de entidade, e é possível incluir seus próprios tipos.

A seção Tipos de entidade da janela Configuração de repositório lista os diferentes tipos de entidade que foram criados. É possível fazer o seguinte.

- Criar um novo tipo de entidade
- Duplicar um tipo de entidade existente (por exemplo, para criar um novo tipo de entidade com base em um existente)
- Anexar variáveis a um tipo de entidade
- Remover variáveis de um tipo de entidade
- Renomear um tipo de entidade
- Excluir um tipo de entidade

**Tipo de entidade.** O nome do tipo de entidade selecionado.

**Variável.** A lista de variáveis válidas que esse tipo de entidade engloba.

**Tipo de uso.** (Opcional) Indica diferentes contextos nos quais essa variável pode ser usada. Clique duas vezes nessa coluna para incluir ou editar um tipo de uso, separando os tipos de uso com uma vírgula e um espaço. Os valores especificados aqui definem os valores exibidos nos nós Exportação do EA ou EA de fluxo quando um usuário clica na coluna Uso, de uma variável na guia Entradas. Consulte o tópico “Opções de entrada de repositório para mapeamento” na página 15 para obter mais informações.

Informações gerais sobre os tipos de uso:

- Os tipos de uso são rótulos arbitrários.
- É possível criar um tipo de uso a partir de quase qualquer entrada de texto, no entanto, não é permitido inserir espaços e caracteres inválidos.
- O conteúdo digitado é automaticamente alterado para letras maiúsculas ao ser inserido.
- É possível ter quantos tipos de uso desejar.
- Os tipos de uso não precisam ser significativos, mas ajudará, ao mapeá-los posteriormente, o uso de uma convenção de nomenclatura que faça sentido para você e outros usuários.
- Durante o mapeamento, um aviso será exibido em fonte vermelha se você tiver mapeado alguns elementos com um tipo de uso e outros com um diferente.

Geralmente, um erro será exibido se você tentar mapear dois campos para o mesmo `feature.element`. Os tipos de uso são uma forma de mapear dois ou mais campos para o mesmo `feature.element` e poder corresponder entre eles.

Por exemplo, se você tiver definido duas variáveis separadas: *HOMEADDRESS* e *WORKADDRESS*, não haverá correspondência entre eles. Se uma entidade tiver um *HOMEADDRESS* que seja igual ao

*WORKADDRESS* de outra entidade, não haverá correspondência por eles serem variáveis diferentes. No entanto, se você reutilizar uma única variável com tipos de uso diferentes, a resolução entenderá que *ADDRESS.WORK* é igual a *ADDRESS.HOME*.

É possível reutilizar os tipos de uso para variáveis diferentes ou ter tipos diferentes. Por exemplo, *HM* e *WK* para telefone e *HOME* e *WORK* para endereço. Isso não importa porque telefones não são correspondidos com relação a endereços. No entanto, se for possível ser consistente, isso ajudará a identificar e agrupar campos posteriormente.

Quando vários tipos de entidade são alimentados em um único repositório, desde que você esteja usando a mesma variável, não importa sejam os tipos de uso. Por exemplo, se você definir *WK* e *HM* com tipos de uso de *ADDRESS* para o tipo de entidade *COMPANY* que ainda será correspondido com relação a *WORK* e *HOME* como tipos de uso de *ADDRESS* para *PERSON*.

Para criar um novo tipo de entidade

1. Clique com o botão direito em **Tipos de entidade** na área de janela de navegação no lado esquerdo da tela.
2. Escolha **Novo tipo de entidade**.
3. Insira um nome exclusivo para o tipo de entidade e clique em OK.
4. Anexe variáveis ao tipo de entidade (consulte a seção a seguir).

Para anexar variáveis a um tipo de entidade

1. Selecione o tipo de entidade na área de janela de navegação no lado esquerdo da tela.
2. Clique no botão Anexar variável (o botão superior no lado direito da tela).
3. Na lista de variáveis disponíveis, escolha uma ou mais (use Ctrl-clique para escolher várias variáveis) e clique em OK.

Para remover variáveis de um tipo de entidade

1. Selecione o tipo de entidade na área de janela de navegação no lado esquerdo da tela.
2. Selecione uma ou mais variáveis na tabela de variáveis anexadas no lado direito da tela. Use Ctrl-clique para escolher várias variáveis.
3. Clique no botão Remover variável (o botão inferior no lado direito da tela).

Para duplicar um tipo de entidade existente

1. Na área de janela de navegação no lado esquerdo da tela, clique com o botão direito no tipo de entidade que deseja duplicar.
2. Escolha **Duplicar tipo de entidade**.
3. Insira um nome exclusivo para o novo tipo de entidade e clique em OK.
4. Anexe variáveis ou remova variáveis do tipo de entidade, conforme o necessário (consulte as instruções anteriores).

Para renomear um tipo de entidade

**CUIDADO** Se você editar, excluir ou anonimizar uma variável ou elemento de variável quando o repositório já contiver dados, será necessário limpar o repositório e recarregar os dados. Essa ação evita deixar o repositório em um estado inconsistente.

1. Na área de janela de navegação no lado esquerdo da tela, clique com o botão direito no tipo de entidade que deseja renomear.
2. Escolha **Renomear**.
3. Insira o novo nome para o tipo de entidade e clique em OK.

Para excluir um tipo de entidade

**CUIDADO** Se você editar, excluir ou anonimizar uma variável ou elemento de variável quando o repositório já contiver dados, será necessário limpar o repositório e recarregar os dados. Essa ação evita deixar o repositório em um estado inconsistente.

1. Na área de janela de navegação no lado esquerdo da tela, clique com o botão direito no tipo de entidade que deseja excluir.
2. Escolha **Excluir**.
3. Clique em **OK** para confirmar a exclusão do tipo de entidade.

**CUIDADO:**

**Não é possível desfazer a exclusão de um tipo de entidade.**

## Configurando o limite para correspondência de entidade

Na seção Regras de resolução da janela Configuração de repositório, escolha o limite no qual a correspondência de entidade ocorrerá.

Ao criar o repositório, a correspondência é pré-configurada para o limite padrão.

Escolha **Configurar para resolução severa** se você não estiver localizando correspondências suficientes nos registros para executar a resolução da entidade.

Escolha **Definir para resolução padrão** para retornar para o limite padrão a partir de uma das outras configurações.

Escolha **Definir para resolução conservadora** se você estiver localizando muitas correspondências.

Para construir um repositório para as entidades e os relacionamentos, selecione **Incluir relacionamentos**. Observe que essa opção somente estará disponível se você tiver o upgrade licenciado separadamente conhecido como IBM SPSS Modeler Entity Analytics Unleashed.

## Reutilizando uma configuração de repositório

Se você já tiver definido uma configuração e desejar usá-la para outro repositório, será possível exportar a configuração existente para um arquivo XML e importar o arquivo no outro repositório (destino). Isso somente é possível em uma configuração existente. Por exemplo, não é possível migrar uma configuração de repositório de uma versão do IBM SPSS Modeler para outra ou de um tipo de banco de dados para outro.

Para reutilizar uma configuração existente

1. Exiba a janela Configuração de repositório para o repositório cuja configuração deseja usar. Consulte o tópico “Configurando um repositório de entidade” na página 16 para obter mais informações.
2. No menu nessa janela, escolha **Configuração > Exportar configuração**.
3. Na caixa de diálogo Salvar como, escolha o nome e o local do arquivo XML de exportação.
4. Exiba a janela Configuração de repositório para o repositório de destino.
5. No menu nessa janela, escolha **Configuração > Importar configuração**.
6. Na caixa de diálogo Salvar como, escolha o nome e o local do arquivo XML exportado anteriormente e clique em **Abrir**.

## Salvando as mudanças na configuração

Para salvar as mudanças na configuração

No menu na janela Configuração de repositório, escolha

Arquivo > Salvar.

## Fechando a janela de configuração

Para sair da janela de configuração

No menu na janela Configuração de repositório, escolha

Arquivo > Sair.

Se houver mudanças não salvas na configuração, clique em **OK** para salvar as mudanças e sair ou em **Cancelar** para sair sem salvar.

---

## Analizando as identidades resolvidas [nó de origem Entity Analytics (EA)]

Depois que os dados forem exportados para o repositório, será possível usar o nó de origem Entity Analytics (EA) para passar as identidades resolvidas a outros nós do IBM SPSS Modeler para análise ou processamento adicional, como criar um relatório listando as identidades resolvidas.

Para analisar as identidades resolvidas

1. Inclua um nó de origem Entity Analytics (EA) em um fluxo.
2. Abra o nó Entity Analytics (EA).
3. Na guia Dados, selecione o repositório de entidade e uma ou mais de suas origens de dados de entrada (clique em **Atualizar** para atualizar a contagem de registro). Consulte o tópico “Selecionando uma origem de dados” para obter mais informações.
4. Inclua nós adicionais no fluxo para executar o processamento desejado. Consulte o tópico “Incluindo nós no fluxo” na página 26 para obter mais informações.

## Selecionando uma origem de dados

Na guia Dados, selecione pelo menos um origem de dados no repositório em que deseja executar algum processamento adicional. Para atualizar a contagem de registros para a origem de dados listada, clique em **Atualizar**.

**Repositório de entidade.** Mostra o repositório de entidade atual, caso exista algum. Para escolher um repositório diferente, se mais de um existir, selecione um na lista. Para criar um novo repositório, escolha <Procurar...> para exibir uma caixa de diálogo que pode ser usada para criar o repositório. Consulte o tópico “Opções de repositório de entidade” na página 14 para obter mais informações.

**Incluir registros a partir de origens de dados.** Essa tabela lista as diferentes origens de dados que foram inseridas no repositório, juntamente com o número de registros em cada origem. Marque a caixa de seleção **Incluir** das origens de dados que deseja usar para executar análise e processamento adicionais. Para selecionar ou cancelar a seleção de todas as origens de dados, clique em **Incluir todos** ou em **Excluir todos** respectivamente.

**Relacionamentos.** Selecione o tipo de relacionamento a ser incluído no repositório. Observe que essa opção somente estará disponível se você tiver o upgrade licenciado separadamente, conhecido como IBM SPSS Modeler Entity Analytics Unleashed, e se o repositório tiver sido configurado para incluir relacionamentos.

- **Sem relacionamento.** Os detalhes de relacionamento não são usados.

- **Relacionamentos próximos.** Seleciona somente as entidades proximaamente relacionadas. A proximidade de um relacionamento depende de muitas variáveis, como as propriedades das variáveis mapeadas, quais variáveis são compartilhadas e se a resolução está configurada como conservadora ou severa.
- **Todos os relacionamentos.** Seleciona todas as entidades relacionadas.

**Grau máximo de separação.** Somente estará disponível se **Relacionamentos próximos** ou **Todos os relacionamentos** estiver selecionado. Selecione o número de graus de separação a ser usado para identificar um relacionamento. Por exemplo, se Ann e Bob não se conhecem, mas John conhece Ann e Bob, Ann e Bob estão relacionados com dois graus de separação.

**Tipo de entidade de saída.** Por padrão, se o repositório contiver detalhes, essa opção mostrará o primeiro tipo de entidade no repositório. Se o repositório tiver mais de um, a seleção de um tipo de entidade aqui alterará as variáveis mostradas na guia Filtrar, para listar as variáveis desse tipo. É possível selecionar qualquer um dos tipos de entidade usados no repositório.

## Renomeando campos de dados

É possível usar a guia Filtrar para renomear qualquer um dos campos de identidade resolvida que são enviados para processamento adicional. É interessante renomear um campo de identidade resolvida, por exemplo, para manter a compatibilidade de nome do campo ao mesclar o recebimento de dados com outro conjunto de dados.

Os campos com seus nomes originais são os seguintes.

*Tabela 10. Campos de identidade resolvida*

Campo	Descrição
\$EA-ID	Identificador de entidade
\$EA-SRC	Tag de origem que identifica a origem de dados na qual os registros se originaram
\$EA-KEY	Campo designado como chave exclusiva no arquivo de origem de dados

*Observação:* Embora também seja possível usar a guia Filtrar para filtrar campos, isso não deve ser feito aqui, pois os campos de identidade são o mínimo necessário para o processamento de análise de entidade.

## Configurando informações de tipo para campos de dados

Na guia Tipos, é possível visualizar ou alterar as diversas propriedades dos campos de identidade resolvida que são enviados para processamento adicional.

As propriedades que podem ser alteradas são as mesmas que as da guia Tipos de um nó Tipo do SPSS Modeler regular e são as seguintes.

*Tabela 11. Propriedades de tipo por campos*

Propriedade	Descrição
<b>Medição</b>	O nível de medição (ou seja, o tipo de dados), usado para descrever as características dos dados no campo.
<b>Valores</b>	Fornece opções para ler os valores de dados a partir do conjunto de dados.
<b>Omisso</b>	Usado para especificar como os valores ausentes do campo serão manipulados.
<b>Verificar</b>	As opções de validação para assegurar que os valores desse campo estejam adequados aos valores ou intervalos especificados.
<b>Função</b>	Especifica como o campo será usado se os dados forem passados a um nó de modelagem ou a um nugget do modelo.

## Incluindo nós no fluxo

É possível incluir vários nós do SPSS Modeler no fluxo para executar análises ou operações de processamento na saída do nó de origem Entity Analytics (EA). Por exemplo, é possível incluir um ou mais dos seguintes.

- Nó de agregado ou de distinção para sumarizar a saída, que pode ser muito grande
- Nó de seleção para selecionar um subconjunto da saída
- Nó de tabela para visualizar a saída do nó de origem Entity Analytics (EA)
- Nó de relatório para imprimir a saída em um relatório
- Nó de exportação do SPSS Modeler para exportar a saída para um formato diferente, como planilha ou banco de dados

Para obter mais informações, consulte as seções sobre os nós Operações de registro, Saída e Exportação no *IBM SPSS Modeler Source, Process and Output Nodes Guide*.

---

## Comparando novos casos com relação ao repositório (nó EA de fluxo)

Depois de já ter executado algumas resoluções de identidade no repositório, será possível usar o nó EA de fluxo para comparar os novos casos encontrados posteriormente com os conteúdos do repositório. Esse nó processa os registros a partir de uma nova origem de dados, compara-os com as entidades já resolvidas no repositório e envia os registros correspondentes para processamento adicional. As correspondências podem ser configuradas para serem exatas ou relacionadas com menos proximidade às entidades existentes.

Como o nó Exportação do EA, o nó EA de fluxo usa um único nó de origem do SPSS Modeler como entrada. No entanto, o nó EA de fluxo difere da seguinte forma. Enquanto o nó de exportação emite registros para todas as entidades relacionadas aos seus registros de entrada, o nó EA de fluxo somente emite registros para as entidades relacionadas às entidades já resolvidas no repositório. Consulte o tópico “Saída do nó EA de fluxo” na página 29 para obter mais informações.

Para comparar novos casos com relação ao repositório

1. Conecte a origem de dados que contém os novos registros que deseja comparar com as entidades existentes. Consulte o tópico “Conectando-se a uma origem de dados” na página 12 para obter mais informações.
2. Na guia Ops. de registro, anexe um nó EA de fluxo ao nó de origem de dados.
3. Clique duas vezes no nó Exportação do Entity Analytics para abrir sua caixa de diálogo.
4. Clique na lista **Repositório de entidade**.
5. Clique em <Procurar...> para exibir a caixa de diálogo Repositórios de entidade.
6. Na caixa de diálogo Repositórios de entidade, clique no campo Nome do repositório.
7. Clique no nome do repositório que deseja usar.
8. Insira o nome do usuário e a senha para esse repositório e clique em **Conectar**. Clique em **OK** quando o repositório for conectado.
9. Na caixa de diálogo EA de fluxo, selecione o Tipo de entidade para o qual deseja mapear. Consulte o tópico “Mantendo os tipos de entidade” na página 21 para obter mais informações.
10. Mapeie os campos de entrada na origem de dados para as variáveis no repositório. Consulte o tópico “Mapeando campos de entrada para variáveis (nó EA de fluxo)” na página 27 para obter mais informações.
11. Opcionalmente, é possível atualizar os registros no repositório em tempo real, ao escorar os dados. Consulte o tópico “Mapeando campos de entrada para variáveis (nó EA de fluxo)” na página 27 para obter mais informações.

12. Clique na guia **Saídas** para ver detalhes das várias origens de dados que foram inseridas no repositório e configure os critérios de seleção para recuperar as entidades existentes. Consulte o tópico “Exibindo os mapeamentos de campo e as origens de dados (nó EA de fluxo)” na página 28 para obter mais informações.
13. Clique na guia **Filtrar** para ver os detalhes dos campos de entrada e das variáveis armazenadas no repositório. As variáveis que não tiverem sido mapeadas no nó serão filtradas, por padrão, no entanto, será possível alterar isso, caso seja necessário.
14. Clique em **OK** quando o nó estiver configurado corretamente.
15. Anexe um nó de tabela ao nó EA de fluxo e execute o fluxo.

A janela de saída do nó de tabela lista todas as entidades recuperadas que correspondem aos novos registros na origem de dados. Os campos de saída tem o prefixo **\$EA-** incluído. Consulte o tópico “Saída do nó EA de fluxo” na página 29 para obter mais informações.

**Nota:** É possível encontrar um erro do tipo **Número incorreto de campos detectado no modelo de dados do servidor** ao executar o nó EA de fluxo. Isso poderá ocorrer se você editar a configuração de repositório após a criação do nó EA de fluxo. A edição da configuração nessas circunstâncias pode ter o efeito de alterar o número e os nomes da saída dos campos a partir do nó. Para resolver o problema, abra o nó EA de fluxo e clique no botão **Atualizar**. Essa ação faz com que o número e os nomes dos campos de saída sejam recalculados.

## Mapeando campos de entrada para variáveis (nó EA de fluxo)

A guia Entradas contém as opções para o mapeamento de campos na entrada para esse nó para as variáveis de repositório. Configure as designações de mapeamento nessa guia ou selecione a guia **Visualizar** para ver os detalhes de todas as origens de dados no repositório e, em seguida, clique em **OK**.

Se você já tiver armazenado um conjunto de mapeamentos em um arquivo XML, será possível usá-lo clicando em **Importar mapeamento**.

**Repositório de entidade.** Mostra o repositório de entidade atual, caso exista algum. Para escolher um repositório diferente, caso exista mais que um, selecione um na lista. Para criar um novo repositório, escolha **<Procurar...>** para exibir uma caixa de diálogo que pode ser usada para criar o repositório. Consulte o tópico “Opções de repositório de entidade” na página 14 para obter mais informações.

**Mapear para tipo de entidade.** Uma lista de tipos de entidade (ou seja, conjuntos de variáveis) definidos no repositório. Escolha um na lista ou escolha **<Incluir novo tipo de entidade...>** para exibir a janela de configuração do repositório, na qual é possível definir um novo tipo de entidade. Consulte o tópico “Configurando um repositório de entidade” na página 16 para obter mais informações.

**Persistir procuras.** Se desejar atualizar os registros no repositório em tempo real durante o escopo de dados, selecione essa opção.

**Tag de origem.** Somente está disponível ao selecionar **Persistir procuras**. Uma lista de tags indicando as origens de dados atualmente conhecidas para o repositório. Escolha uma na lista ou escolha **<Incluir nova tag de origem...>** para criar uma tag para uma nova origem de dados.

**Chave exclusiva.** Somente está disponível ao selecionar **Persistir procuras**. O campo de entrada a ser usado para identificadores exclusivos para os registros de dados.

**Tabela de mapeamento.** Nessa tabela, é possível mapear cada campo de entrada para uma variável correspondente no repositório. Se não existir uma variável no tipo de entidade selecionada, será possível criar uma nova variável aqui.

- **Campo.** O conjunto de campos de entrada na origem de dados selecionada. Cada campo tem um ícone que indica o nível de medição (ou seja, o tipo de dados) para o campo.

- **Mapeado para variável.** Para mapear um campo para uma variável, clique duas vezes nessa coluna (ou pressione a barra de espaço) na linha de campo e escolha uma variável na lista. Se uma variável adequada não estiver disponível, escolha <Incluir nova variável...> para exibir a janela de configuração de repositório, na qual é possível definir uma nova variável para esse tipo de entidade. Consulte o tópico “Configurando um repositório de entidade” na página 16 para obter mais informações.
- **Uso.** Indica o contexto de um determinado campo no qual mais de um contexto é possível, por exemplo, os números de telefone de trabalho e residenciais. Consulte o tópico “Mantendo os tipos de entidade” na página 21 para obter mais informações.

**Importar mapeamento.** Importa um conjunto exportado anteriormente de mapeamentos de campo-para-variável a partir de um arquivo XML externo. Isso pode ser útil no caso de diferentes origens de dados com os mesmos requisitos de mapeamento, pois evita a necessidade de redefinir os mesmos mapeamentos para diferentes origens.

**Exportar mapeamento.** Exporta para um arquivo XML externo o conjunto de mapeamentos de campo-para-variável mostrado na tabela de mapeamento.

## Exibindo os mapeamentos de campo e as origens de dados (nó EA de fluxo)

Na guia Saída, é possível ver os detalhes das várias origens de dados que foram inseridas no repositório. Essas são as origens de dados com relação às quais a entrada para este nó é processada para procurar e recuperar entidades correspondentes. Clique em **Atualizar** para atualizar a contagem de registro.

**Incluir correspondências das origens de dados.** Essa tabela lista as diferentes origens de dados disponíveis no repositório, juntamente com o número de registros em cada origem.

**Correspondências.** Essas opções especificam como exatamente as informações de mapeamento de campo-para-variável especificadas na guia Entradas devem ser correspondidas com relação aos registros candidatos (ou seja, aos conteúdos de todo o repositório). Quanto mais aproximado for o critério de correspondência, menos entidades serão recuperadas.

*Observação:* Se mais de 20 correspondências forem localizadas, somente as primeiras 20 serão retornadas.

- **Somente incluir correspondências exatas.** Esse é o critério de correspondência mais aproximado e resulta na seleção de menos registros. Use essa opção quando desejar retornar somente as entidades consideradas como correspondências exatas.
- **Incluir correspondências possíveis.** Use essa configuração quando desejar retornar entidades correspondentes e entidade que compartilhem os mesmos identificadores (aquelas com variáveis que tenham sido configurados com o valor de frequência Um, por exemplo, números de cartão de crédito, números de ID imposto e assim por diante).
- **Incluir todas as correspondências.** Use essa opção quando desejar ver o maior número possível de entidades no repositório que tenham variáveis compartilhadas. Esse é o critério de correspondência menos aproximado e resulta na seleção do maior número de registros possíveis. Essa opção retorna correspondências exatas e entidades que compartilham quase qualquer variável (geralmente as que têm o valor de frequência Um ou Poucos). Por exemplo, ambas as entidades com o mesmo número de ID e as entidades com endereços semelhantes seriam incluídas.

**Relacionamentos.** Somente disponível se o repositório tiver sido configurado para incluir relacionamentos. Para configurar o repositório para incluir relacionamentos, deve-se ter o upgrade licenciado separadamente conhecido como IBM SPSS Modeler Entity Analytics Unleashed. Selecione o tipo de relacionamento a ser incluído na saída.

- **Sem relacionamento.** Os detalhes de relacionamento não são usados.

- **Relacionamentos próximos.** Seleciona somente as entidades proxivamente relacionadas. A proximidade de um relacionamento depende de muitas variáveis, como as propriedades das variáveis mapeadas, quais variáveis são compartilhadas e se a resolução está configurada como conservadora ou severa.
- **Todos os relacionamentos.** Seleciona todas as entidades relacionadas.

**Grau máximo de separação.** Somente estará disponível se **Relacionamentos próximos** ou **Todos os relacionamentos** estiver selecionado. Selecione o número de graus de separação a ser usado para identificar um relacionamento. Por exemplo, se Ann e Bob não se conhecem, mas John conhece Ann e Bob, Ann e Bob estão relacionados com dois graus de separação.

**Tipo de entidade de saída.** Por padrão, se o repositório contiver detalhes, essa opção mostrará o primeiro tipo de entidade no repositório. Se o repositório tiver mais de um, a seleção de um tipo de entidade aqui alterará as variáveis mostradas na guia Filtrar, para listar as variáveis desse tipo. É possível selecionar qualquer um dos tipos de entidade usados no repositório.

## Saída do nó EA de fluxo

A saída do nó EA de fluxo consiste nos seguintes campos para cada registro recuperado.

Campo	Descrição
<i>Field1</i> [, <i>Field2</i> [, ... <i>FieldN</i> ]]	Campos da origem de dados que contêm os novos registros.
\$EA-ID	Identificador de entidade para esse registro no repositório.
\$EA-SRC	Tag de origem que identifica a origem de dados da qual esse registro se originou.
\$EA-KEY	Valor da chave exclusiva para esse registro no arquivo de origem de dados.
\$EA-SC	Campo que indica a proximidade da correspondência entre esse registro e uma entidade observada no repositório, um valor de 1,0 (correspondência fraca) a 10,0 (boa correspondência).
\$EA-Feature1[, \$EA-Feature2[, ... \$EA-FeatureN]]	Valores das variáveis mapeadas para esse registro no repositório.

Se os campos de relacionamento forem ativados no repositório e o grau de separação for maior que zero na guia Saídas, a saída do nó EA de fluxo também conterá os seguintes campos para cada registro recuperado.

Campo	Descrição
\$EA-DEGREE	Grau de separação.
\$EA-PARENT	Identificador do registro a partir do qual a separação é calculada.
\$EA-CHILD	Identificador do registro para o qual a separação é calculada.
\$EA-RULE	

---

## Usando o IBM SPSS Modeler Entity Analytics com outros produtos IBM SPSS

Há instaladores disponíveis para permitir o uso do IBM SPSS Modeler Entity Analytics com os seguintes produtos:

- IBM SPSS Collaboration and Deployment Services
- IBM SPSS Modeler Batch para Windows
- IBM SPSS Modeler Solution Publisher

Será necessário executar esses instaladores antes que seja possível usar as variáveis do IBM SPSS Modeler Entity Analytics com esses produtos. Para obter mais informações, consulte o guia *IBM SPSS Modeler Premium Installation*.

Após a instalação, você deverá usar o cliente IBM SPSS Collaboration and Deployment Services Deployment Manager para criar uma definição de servidor de repositório Entity Analytics. Isso é necessário para usar um fluxo do IBM SPSS Modeler que contenha um nó do Entity Analytics em uma tarefa do IBM SPSS Collaboration and Deployment Services (em outras palavras, para executar fluxos do Entity Analytics no IBM SPSS Collaboration and Deployment Services). A definição de servidor deve corresponder ao nome do repositório no fluxo. Essa definição é usada para informar o fluxo onde localizar o repositório e fornecê-lo as informações de conexão necessárias.

---

### Tarefas administrativas

Para os repositórios criados no Entity Analytics, um novo serviço de banco de dados é criado usando o produto IBM DB2. Há algumas tarefas administrativas associadas ao DB2. Geralmente, essas tarefas são executadas pelo administrador de banco de dados ou o administrador do sistema e são:

- Configurando designações de porta
- Gerenciando credenciais do administrador para a base de dados do repositório

Outras tarefas administrativas que podem precisar ser executadas se aplicam a todos os repositórios e são as seguintes.

- Movendo o repositório para um diretório de armazenamento
- Configurando propriedades de fluxo para os campos data/hora e registro de data e hora
- Ajustando as configurações de tempo limite
- Executando o IBM SPSS Modeler Entity Analytics com o cliente SPSS Modeler e o SPSS Modeler Server no mesmo sistema Windows
- Limpando um repositório de entidade
- Excluindo um repositório de entidade
- Excluindo um repositório quando não é possível se conectar a ele

### Configurando designações de porta

Cada serviço do banco de dados DB2 deve ter uma porta alocada que não pode ser alocada a outros serviços em execução na máquina. Os serviços de banco de dados residem na mesma máquina que executa o IBM SPSS Modeler Server (ou, quando o IBM SPSS Modeler é usado sem uma conexão com o IBM SPSS Modeler Server, a máquina que executa o IBM SPSS Modeler).

Por padrão, o Entity Analytics designa portas no intervalo de 1320 a 1520, começando na porta 1320 para o primeiro repositório criado. No caso de conflito, é possível configurar a designação de portas, editando o arquivo: `<modeler server installation path>/ext/bin/pasw.entityanalytics/ea.cfg` e definindo os valores apropriados para a configuração `min_port` e `max_port`. Os conteúdos padrão desse arquivo são mostrados abaixo:

```
# configuração de intervalo de portas para análise de entidade
#
#     esse intervalo de portas controla quais portas os bancos de dados DB2
#     (criado para armazenar repositórios do Entity Analytics)
#     podem usar. Configure se o intervalo de portas padrão for
#     introduzir um conflito no sistema.
#
# min_port padrão = 1320
# max_port padrão = 1520
min_port, 1320
max_port, 1520
```

## Gerenciando credenciais do administrador para a base de dados do repositório

O nome do usuário administrador e a senha para o banco de dados DB2 que hospeda um repositório de entidade são definidos quando o repositório é criado. Se você souber as credenciais atuais, será possível alterar esses detalhes por meio do editor de SQL do DB2.

Para iniciar o editor de SQL do DB2

1. Em uma máquina cliente, abra uma janela de prompt de comandos.
2. Insira:

```
cd modeler_install_dir\ext\bin\pasw.entityanalytics\DB2\bin
```

em que *modeler\_install\_dir* é o diretório no qual o SPSS Modeler está instalado.

3. Insira:

```
sqlsql -c "C:\Documents and Settings\All Users\Application Data\IBM\SPSS\Modeler\version\EA\
repositories\repos_name
```

em que *version* é o número da versão da instalação do SPSS Modeler e *repos\_name* é o nome do repositório.

4. Nos prompts, insira o nome do usuário administrador do banco de dados atual e a senha para exibir o prompt `sqlsql>`.

Para alterar a senha do administrador do banco de dados

1. No prompt `sqlsql>`, insira:

```
alter user username identified by password;
commit work;
```

em que *username* é o nome do usuário atual do administrador do banco de dados e *password* é a nova senha.

2. Insira sair para fechar o editor.
3. Reinicie o cliente SPSS Modeler.

Para obter informações sobre outras tarefas administrativas relacionadas ao banco de dados DB2, consulte a documentação da versão apropriada do IBM DB2 em <http://publib.boulder.ibm.com/>.

## Movendo o repositório para um diretório de armazenamento

Por padrão, os arquivos de repositório são armazenados em um diretório nomeado *EA* nos seguintes locais:

- C:\Documents and Settings\All Users\ApplicationData\IBM\SPSS\Modeler\version\EA (sistema Windows)
- *modeler\_install\_directory*/ext/bin/pasw.entityanalytics/EA (sistema UNIX)

Como os arquivos usados para armazenar o repositório podem ficar muito grandes, poderá ser necessário movê-los para um disco ou uma partição diferente para disponibilizar mais espaço.

Para mover o repositório para um diretório diferente

1. Saia do SPSS Modeler.
2. Mova o diretório *EA* do local original (listado anteriormente) para um novo local. Por exemplo, no Windows é interessante movê-lo para um local como *F:\data\EA*.
3. Edite o arquivo *<modeler\_server\_installation\_path>/ext/bin/pasw.entityanalytics/ea.cfg* para incluir a opção a seguir:

```
repository_data_directory, new_location
```

em que *new\_location* é o diretório para o qual o diretório *EA* foi movido, por exemplo *F:\data\EA*.

## Configurando propriedades de fluxo para os campos data/hora e registro de data e hora

Se os dados de origem incluírem campos contendo dados de data/hora ou de registro de data e hora, assegure-se de que as propriedades de fluxo correspondentes estejam configuradas para o formato reconhecido pelo IBM SPSS Modeler Entity Analytics.

Para configurar o formato de propriedade de fluxo

1. No menu principal do SPSS Modeler, escolha:  
**Ferramentas > Propriedades de fluxo > Opções.**
2. Selecione **Data/hora**.
3. Configure o **Formato de data** para **AAAA-MM-DD**.
4. Configure o **Formato de horário** para **HH:MM:SS**.
5. Clique em **OK**.

## Ajustando as configurações de tempo limite

Em sistemas lentos ou fortemente carregados, se você encontrar erros ao criar ou acessar repositórios, poderá ser necessário aumentar as configurações de tempo limite para iniciar e parar o mecanismo de análise de entidade ou o servidor de banco de dados de análise de entidade.

Para ajustar o tempo limite do mecanismo de análise de entidade

1. Saia do SPSS Modeler.
2. Edite o arquivo *<modeler\_server\_installation\_path>/ext/bin/pasw.entityanalytics/ea.cfg* para aumentar o valor da opção a seguir:

```
timeout, value
```

em que *value* é o valor do tempo limite, em segundos, do mecanismo de análise de entidade (o padrão é 60).

Para ajustar o tempo limite do servidor de banco de dados de análise de entidade (somente DB2)

1. Saia do SPSS Modeler.

2. Edite o arquivo `<modeler server installation path>/ext/bin/pasw.entityanalytics/ea.cfg` para aumentar o valor da opção a seguir:

`timeout, value`

em que *value* é o valor do tempo limite, em segundos, do servidor de banco de dados DB2 de análise de entidade (o padrão é 100).

## Executando o IBM SPSS Modeler Entity Analytics com o cliente SPSS Modeler e o SPSS Modeler Server no mesmo sistema Windows

Se o IBM SPSS Modeler Entity Analytics for instalado no cliente SPSS Modeler e no SPSS Modeler Server no mesmo sistema Windows, por padrão, o cliente e o servidor compartilharão o mesmo repositório. Se você desejar usar repositórios separados, será necessário editar o arquivo de configuração *ea.cfg* em **um** dos sistemas para configurá-lo para usar um intervalo de portas e uma pasta de repositório diferentes.

*Observação:* Em especial, se você usar um cliente SPSS Modeler de 32 bits e um SPSS Modeler Server de 64 bits (ou vice-versa), será necessário realizar este procedimento.

1. Abra o arquivo `<modeler [server] installation path>/ext/bin/pasw.entityanalytics/ea.cfg` para edição.
2. Altere as configurações `min_port` e `max_port` para usar portas diferentes do que o outro sistema. Consulte o tópico “Configurando designações de porta” na página 30 para obter mais informações.
3. Altere a configuração `repository_data_directory` para usar um diretório diferente do que o outro sistema.
4. Salve e feche o arquivo *ea.cfg*.

## Limpando um repositório de entidade

Se desejar limpar os registros de dados de um repositório de entidade, mas manter as informações de configuração, será possível limpar os dados do repositório.

Para limpar todos os dados de um repositório:

1. Abra um nó Entity Analytics.
2. Clique na lista **Repositório de entidade**.
3. Clique em **<Procurar...>** para exibir a caixa de diálogo Instâncias de resolução da entidade.
4. Na caixa de diálogo Instâncias de resolução da entidade, clique na lista **Nome do repositório**.
5. Selecione o repositório que deseja limpar.
6. Se você ainda não estiver conectado, insira o nome do usuário administrador e a senha e clique em **Conectar**.
7. Quando o botão **Limpar todos** for ativado, clique nele.
8. Na caixa de diálogo Limpar todas as origens de dados, clique em **Limpar** para confirmar a limpeza do repositório.

## Excluindo origens de dados não usadas de um repositório

Se houver uma origem de dados que não seja mais usada ou necessária em um repositório de entidade, será possível excluí-la do repositório. É possível selecionar uma ou mais origens de dados para exclusão.

Para excluir uma origem de dados selecionada de um repositório:

1. Abra um nó Entity Analytics.
2. Clique na lista **Repositório de entidade**.
3. Clique em **<Procurar...>** para exibir a caixa de diálogo Instâncias de resolução da entidade.
4. Na caixa de diálogo Instâncias de resolução da entidade, clique na lista **Nome do repositório**.
5. Selecione o repositório a partir do qual deseja excluir uma origem de dados.

6. Se você ainda não estiver conectado, insira o nome do usuário administrador e a senha e clique em **Conectar**.
7. Na lista **Gerenciar repositório**, selecione a origem de dados a ser excluída. Caso seja necessário, use Ctrl-clique para selecionar origens de dados adicionais.
8. Quando o botão **Excluir não usados** for ativado, clique nele.
9. Na caixa de diálogo Excluir dados não usados, clique em **Excluir** para confirmar a limpeza do repositório.

## Excluindo um repositório de entidade

Quando um repositório não for mais necessário, será possível excluí-lo completamente.

*Cuidado:* Isso faz exatamente o que diz. **Não é possível desfazer essa operação.** Se você não tiver certeza, use o botão **Limpar** para remover todos os dados de origem. Essa ação não remove a configuração de repositório. Consulte o tópico “Limpando um repositório de entidade” na página 33 para obter mais informações.

*Observação:* O procedimento a seguir assume que seja possível se conectar ao repositório a partir do SPSS Modeler e que você saiba o nome do usuário administrador e a senha do banco de dados que hospeda o repositório. Se esse não for o caso, siga o procedimento para excluir um repositório quando não é possível se conectar a ele. Consulte o tópico “Excluindo um repositório quando não é possível se conectar a ele” para obter mais informações.

Para excluir um repositório

1. Abra um nó Entity Analytics.
2. Clique na lista **Repositório de entidade**.
3. Clique em <Procurar...> para exibir a caixa de diálogo Instâncias de resolução da entidade.
4. Na caixa de diálogo Instâncias de resolução da entidade, clique na lista **Nome do repositório**.
5. Selecione o repositório que deseja excluir.
6. Se você ainda não estiver conectado, insira o nome do usuário administrador e a senha e clique em **Conectar**.
7. Quando o botão **Excluir repositório inteiro** for ativado, clique nele.
8. Clique em **Excluir** para confirmar a exclusão do repositório.
9. Clique em **OK** para reconhecer a exclusão bem-sucedida.

## Excluindo um repositório quando não é possível se conectar a ele

Use o procedimento a seguir se desejar excluir um repositório de entidade, mas não for possível se conectar a ele, devido a problemas de conectividade com o SPSS Modeler ou porque você esqueceu o nome do usuário ou a senha.

Execute este procedimento na máquina que hospeda a base de dados do repositório.

### Sistemas Windows

1. Abra uma janela de Prompt de comandos.
2. Insira:

```
cd modeler_install_dir
cd ext\bin\pasw.entityanalytics
delete_repository.bat repos_name
```

em que *modeler\_install\_dir* é o diretório no qual o SPSS Modeler está instalado e *repos\_name* é o nome do repositório.

**Nota:** O nome de repositório faz distinção entre maiúsculas e minúsculas.

3. Continue a partir de "Concluindo o procedimento" posteriormente nesta seção.

## Sistemas UNIX

1. Abra um shell.
2. Insira:

```
cd modeler_server_install_dir  
cd ext/bin/pasw.entityanalytics  
./delete_repository.sh repos_name
```

em que *modeler\_server\_install\_dir* é o diretório no qual o SPSS Modeler Server está instalado e *repos\_name* é o nome do repositório.

**Nota:** O nome de repositório faz distinção entre maiúsculas e minúsculas.

### Concluindo o procedimento (todos os sistemas)

1. No prompt, confirme a exclusão do repositório, inserindo S.
2. Exclua o diretório que tem o mesmo nome que o repositório excluído. Caso não seja possível excluir esse diretório, reinicie a máquina e tente novamente.



---

## Capítulo 4. Análise de entidade em ação

---

### Sobre este exemplo

Neste exemplo, veremos como a inclusão de análise de entidade pode melhorar ainda mais os resultados já impressionantes que podem ser obtidos usando o IBM SPSS Modeler.

Este exemplo usa o fluxo *loan\_entity\_analytics.str*, que faz referência ao arquivo de dados *loan\_applications.csv*. Esses arquivos estão disponíveis no diretório *Demos* de qualquer instalação do IBM SPSS Modeler que também tenha o IBM SPSS Modeler Entity Analytics instalado. É possível acessar o diretório *Demos* do grupo de programa IBM SPSS Modeler no menu Iniciar do Windows. O arquivo *loan\_entity\_analytics.str* está no diretório *Entity\_Analytics*.

*Observação:* Antes que seja possível executar este fluxo de exemplo, será necessário criar um repositório no sistema. Faça isso antes de continuar com o exemplo. Consulte o tópico “Criando o repositório” na página 12 para obter mais informações.

Vamos começar com uma situação familiar: executivos de um banco estão preocupados com a probabilidade de os clientes ficarem inadimplentes em empréstimos para os quais existem aplicativos pendentes. O departamento de TI do banco é um usuário antigo do SPSS Modeler, portanto, sua equipe já criou um fluxo e construiu um modelo preditivo a partir de seus dados existentes sobre 700 empréstimos que o banco já fez. Esses empréstimos foram reembolsados ou os clientes ficaram inadimplentes em seus reembolsos.

---

### O modelo original

Aqui está como a equipe do banco construiu seu modelo e o que eles aprenderam com isso.

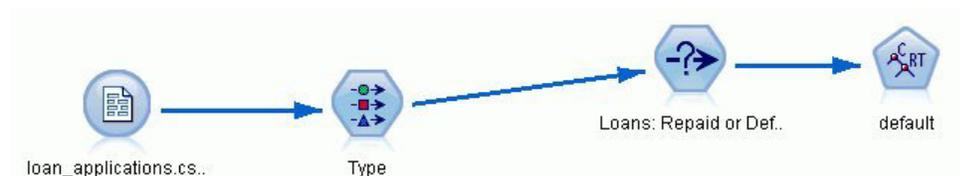


Figura 2. Fluxo inicial com nó de modelagem

Além de detalhes de empréstimos passados, o conjunto de dados *loan\_applications.csv* inclui detalhes de 150 clientes cujos aplicativos de empréstimos ainda estão pendentes, fornecendo um total de 850 registros.

Nem todos os campos do conjunto de dados são úteis para fazer a predição. Por exemplo, os campos de nome podem ser ignorados. O nó Tipo filtra os campos a serem ignorados, configurando o papel deles como **Nenhum**. Os campos a serem usados para fazer a predição têm seus papéis configurados como **Entrada** e o campo cujo valor o modelo está tentando prever tem seu papel configurado como **Destino**.

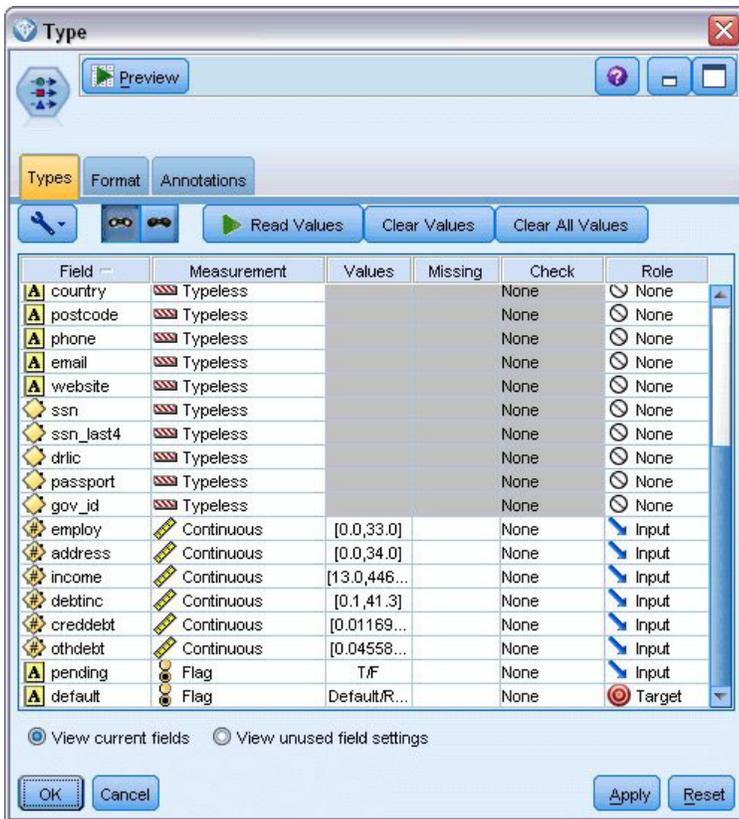


Figura 3. Papéis de campo configurados no nó Tipo

Como o modelo deve fazer suas previsões unicamente com base nos dados passados, o fluxo inclui um nó de seleção, que somente inclui os empréstimos que *não* estão marcados como Pendentes, descartando os 150 empréstimos pendentes.

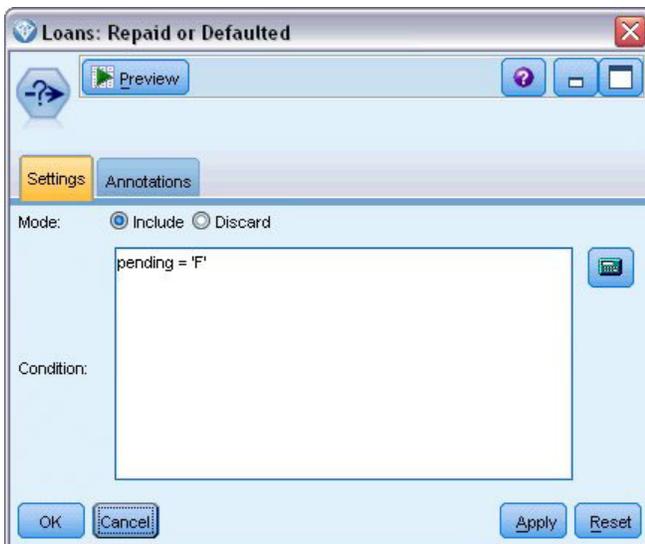


Figura 4. Descartando os aplicativos de empréstimos pendentes

Com os empréstimos pendentes descartados, somente os detalhes dos 700 empréstimos restantes que foram reembolsados ou ficaram inadimplentes são passados para o nó de modelagem. O banco poderia ter usado um entre inúmeros algoritmos do SPSS Modeler para produzir um bom modelo. Nesse caso,

eles usam um nó da árvore C&R, que será usado para construir um modelo para prever prováveis inadimplentes com base no desempenho anterior dos clientes do banco.

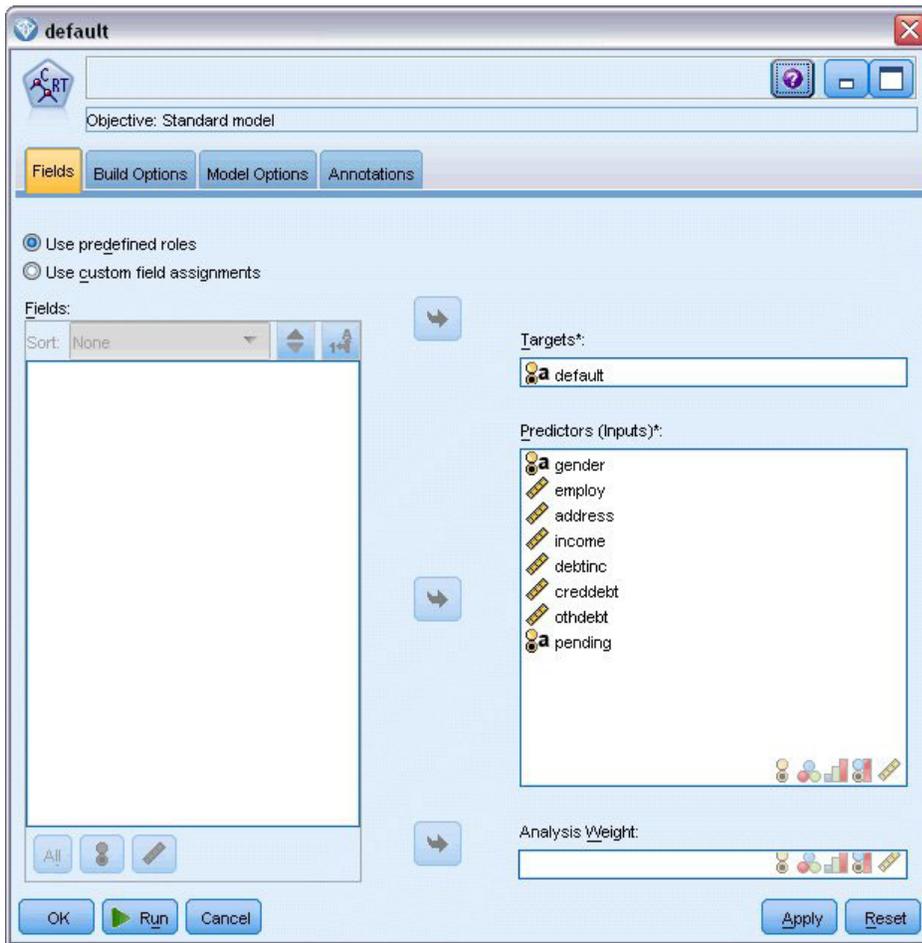


Figura 5. Designando campos preditores e de destino

Os campos usados para fazer a predição são designados como os campos preditores, e o campo cujo valor o modelo está tentando prever (**padrão** nesse caso) é configurado como o campo de destino, conforme definido anteriormente pelo nó Tipo.

A execução desse fluxo produz um nugget do modelo, contendo o modelo que foi construído a partir dos campos preditores.

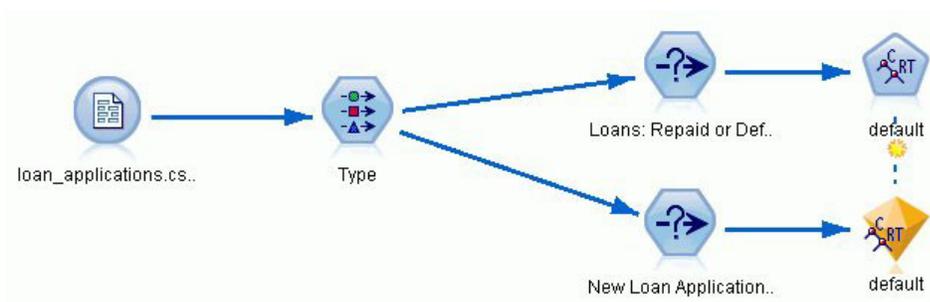


Figura 6. Fluxo com nugget do modelo incluído

Agora, o analista do banco pode usar o modelo para começar a prever se os clientes com reembolsos pendentes provavelmente ficarão inadimplentes. Usando o conjunto de dados original, o analista insere um nó de seleção que, desta vez, inclui somente os 150 registros de empréstimo marcados como Pendentes, em vez de descartá-los. O analista passa esses registros diretamente para o modelo, incluindo um nó de distribuição para uma representação visual das previsões do modelo.



Figura 7. Fluxo selecionando novos aplicativos de empréstimos e com nó de distribuição incluído

O nó de distribuição mostra a distribuição dos valores do campo *\$R-default* no modelo. Esse campo é incluído no modelo de dados pelo nó da árvore C&R quando ele é executado. O campo contém a previsão da chance de cada novo requerente reembolsar ou ficar inadimplente e ele será usado mais tarde para comparar o efeito da inclusão da análise de entidade.

Executando essa parte do fluxo, o analista verifica na saída do nó de distribuição que 137 dos 150 novos requerentes estão preditos para reembolsarem seus empréstimos. Os 13 restantes estão preditos para ficarem inadimplentes, portanto, o analista provavelmente recomendará ao banco que rejeite seus aplicativos.



Figura 8. Saída do nó de distribuição sem análise de entidade

## Incluindo análise de entidade

Agora veremos se a situação pode ser melhorada ao incluir a análise de entidade na equação. Imagine que você seja um especialista em análise de entidade, chamado pelo banco para investigar possíveis entradas fraudulentas nos registros de cliente, nos dados de origem. É possível que haja registros duplicados causados por erros de entrada de dados, mas também é possível que um requerente de empréstimo possa estar tentando disfarçar sua identidade. Em qualquer caso, o banco precisa saber qual é o verdadeiro contexto.

Para esse exemplo, assumiremos que um repositório de entidade já tenha sido criado. Consulte o tópico “Criando o repositório” na página 12 para obter mais informações.

## Colocando os dados de origem no repositório

Primeiro, é necessário incluir um nó Exportação do EA no nó de origem de dados para que seja possível exportar os dados de origem para o repositório de entidade.

Antes que seja possível exportar os dados, é necessário mapear os campos na origem de dados para as variáveis no repositório de entidade. Isso é necessário porque as diferentes origens de dados podem usar

nomes de campos diferentes para o mesmo tipo de informação. O repositório de entidade fornece um conjunto padrão de tipos de informação (conhecidos como "variáveis") para evitar duplicação.

No nó Exportação do EA, configure os detalhes sobre o repositório: os detalhes de conexão, a tag de origem (para identificar a origem de dados, nesse caso, **TEST**), o tipo de entidade (o conjunto de variáveis que estamos usando, nomeado como **PERSON**), e o campo-chave exclusivo (para identificar exclusivamente cada registro). Nesse caso, use o campo-**chave** como a chave exclusiva.

Agora é possível configurar os mapeamentos. No conjunto de variáveis usado, existem variáveis que correspondem aos campos *fname*, *mname*, *lname*, *geração*, *dob*, *sexo*, *addr1*, *cidade*, *país*, *CEP*, *telefone*, *email*, *ssn*, *drlic* e *passaporte*.

Comece configurando o mapeamento para *fname*. Clique duas vezes na coluna **Mapeado para variável** na tabela, na linha *fname*, role para baixo até a entrada **NAME.GIVEN\_NAME** e clique nela para criar o mapeamento.

Agora, mapeie os campos que têm variáveis correspondentes, para que o conjunto integral de mapeamentos fique semelhante a este.

Tabela 12. Campos mapeados para variáveis de repositório.

Campo	Mapeado para variável
<i>fname</i>	NAME.GIVEN_NAME
<i>mname</i>	NAME.MIDDLE_NAME
<i>lname</i>	NAME.SUR_NAME
<i>geração</i>	NAME.NAME_GEN
<i>dob</i>	DOB.DOB
<i>sexo</i>	GENDER.GENDER
<i>addr1</i>	ADDRESS.ADDR1
<i>cidade</i>	ADDRESS.CITY
<i>país</i>	ADDRESS.COUNTRY
<i>CEP</i>	ADDRESS.POSTAL_CODE
<i>telefone</i>	PHONE.PHONE_NUM
<i>email</i>	EMAIL_ADDR.ADDR
<i>ssn</i>	SSN.ID_NUM
<i>drlic</i>	DRLIC.ID_NUM
<i>passaporte</i>	PASSPORT.ID_NUM

Clique em **Executar** para exportar os dados no repositório. Isso demora um pouco, mas quando a caixa de diálogo Feedback de execução fechar, a exportação estará concluída.

## Lendo as identidades resolvidas

Ao exportar dados para o repositório, o sistema de análise de entidade começa a resolver possíveis conflitos de identidade, designando um identificador de entidade exclusivo, que você verá mais tarde como o campo *\$EA-ID*. (*Observação:* Esse não é o mesmo que o campo Chave exclusiva no nó Exportação do EA, ele é usado para identificar exclusivamente os registros de origem de dados.)

A primeira etapa para ler as identidades resolvidas é incluir um nó de origem Entity Analytics (EA) no fluxo. Esse nó de origem não deve ser conectado a nada neste estágio.

Abra o nó de origem Entity Analytics (EA) e configure os detalhes do repositório de entidade. É exibida uma lista das origens de dados que foram exportadas para o repositório, nesse caso, há somente uma.

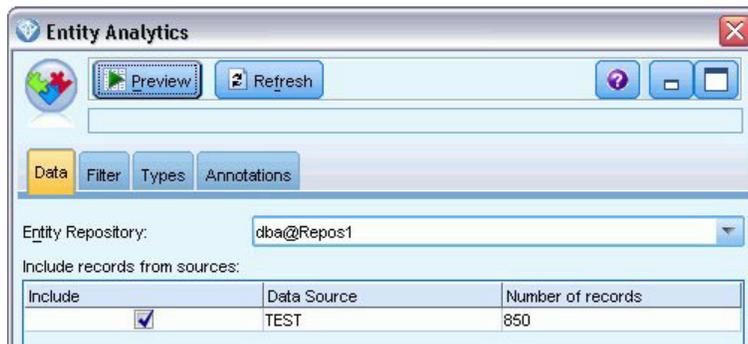


Figura 9. Selecionando uma origem de dados no repositório

Marque a caixa de seleção da origem de dados **TEST** e clique em OK.

Vamos verificar o que o sistema de análise de entidade fez com os dados. Anexe um nó de tabela ao nó de origem Entity Analytics (EA), abra o nó de tabela e clique em **Executar** para exibir a janela de saída do nó de tabela.

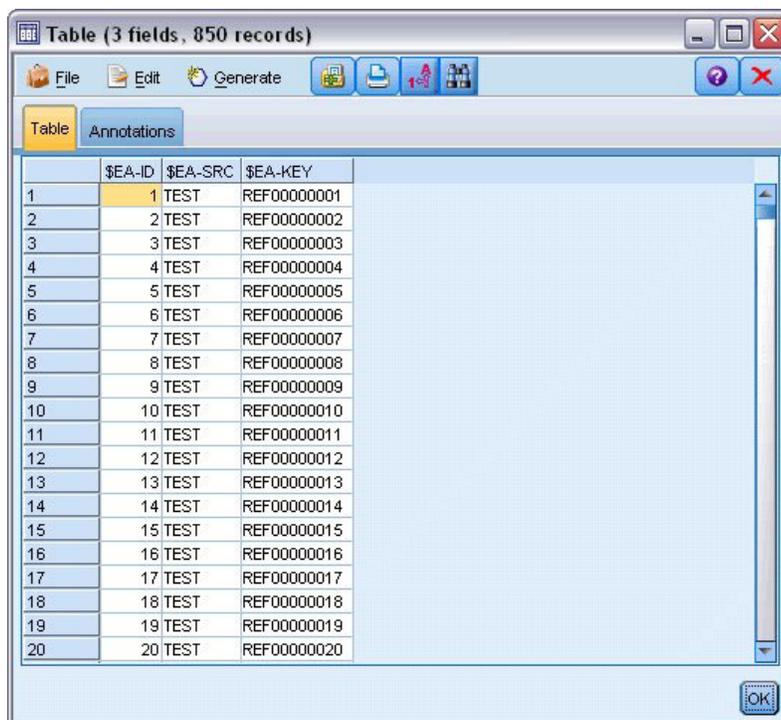


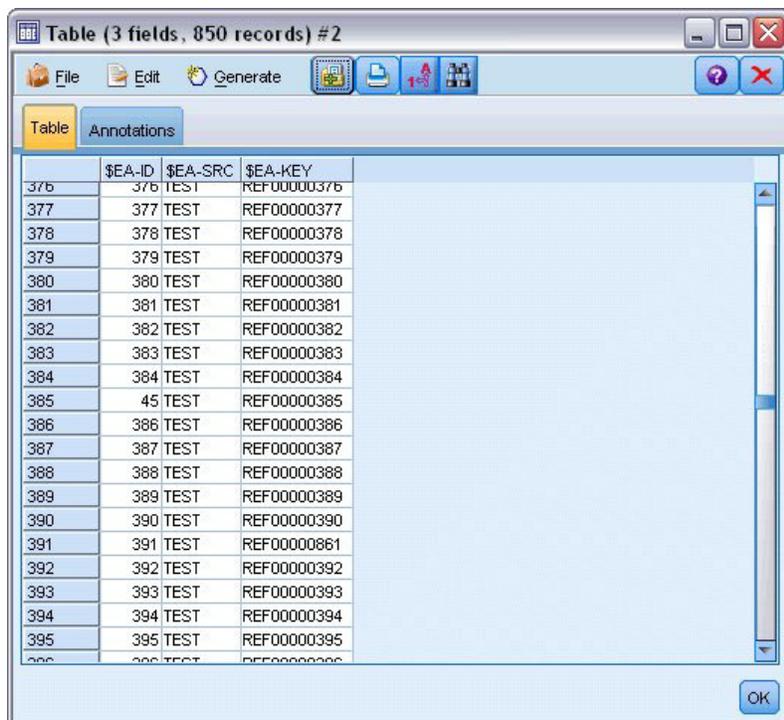
Figura 10. Saída do nó de tabela

Somente um campo parece familiar, aquele que tem o rótulo **\$EA-KEY**. Esse é, na verdade, o campo-chave dos dados de origem e ele está aqui porque você o escolheu como o campo Chave exclusiva no nó Exportação do EA.

No entanto, o sistema incluiu dois outros campos. O campo *\$EA-ID* é o identificador exclusivo, não dos registros de origem, mas das identidades resolvidas. Veremos as diferenças em breve. O campo *\$EA-SRC* identifica de onde os dados vêm, ele informa **TEST** aqui porque essa é a tag da origem designada no nó Exportação do EA.

O que ocorreu com todos os outros campos nos dados de origem? Não se preocupe, eles ainda estão no repositório, ocorre que, por motivo de desempenho, o nó de origem Entity Analytics (EA) somente passa o conjunto mínimo de campos para processamento adicional.

Agora, role a saída do nó de tabela para baixo até a linha 385.



The screenshot shows a window titled "Table (3 fields, 850 records) #2". The window has a menu bar with "File", "Edit", and "Generate". Below the menu bar are several icons. The window is divided into two tabs: "Table" and "Annotations". The "Table" tab is active, showing a table with three columns: "\$EA-ID", "\$EA-SRC", and "\$EA-KEY". The table contains 850 records. The first 384 records have \$EA-ID values from 376 to 395. The 385th record has an \$EA-ID of 45, which is out of sequence. The following records have \$EA-ID values from 386 to 395. The \$EA-SRC column contains the value "TEST" for all records. The \$EA-KEY column contains values like "REF00000376", "REF00000377", etc., with a jump at the 385th record to "REF00000385".

	\$EA-ID	\$EA-SRC	\$EA-KEY
376	376	TEST	REF00000376
377	377	TEST	REF00000377
378	378	TEST	REF00000378
379	379	TEST	REF00000379
380	380	TEST	REF00000380
381	381	TEST	REF00000381
382	382	TEST	REF00000382
383	383	TEST	REF00000383
384	384	TEST	REF00000384
385	45	TEST	REF00000385
386	386	TEST	REF00000386
387	387	TEST	REF00000387
388	388	TEST	REF00000388
389	389	TEST	REF00000389
390	390	TEST	REF00000390
391	391	TEST	REF00000861
392	392	TEST	REF00000392
393	393	TEST	REF00000393
394	394	TEST	REF00000394
395	395	TEST	REF00000395

Figura 11. Diferenças entre as linhas de saída de tabela e os números de *\$EA-ID*

Observe como o número *\$EA-ID* parece estar fora de sequência aqui. O sistema de análise inteiro determinou que o registro REF00000385 faz referência à pessoa identificada como entidade 45, que também tem o registro REF00000045. Rolando mais para baixo na saída, há mais números fora de sequência, por exemplos, nas linhas 485, 517, 520 e assim por diante. É melhor verificarmos mais de perto.

Primeiro, vamos destacar o fato de que o campo *\$EA-KEY* contém os dados do campo-chave nos dados de origem, renomeando-o como *chave*. Anexe um nó de filtro ao nó de origem Entity Analytics (EA) e abra o nó de filtro. Clique duas vezes na sequência de caracteres *\$EA-KEY* na segunda coluna **Campo** e digite *chave*.

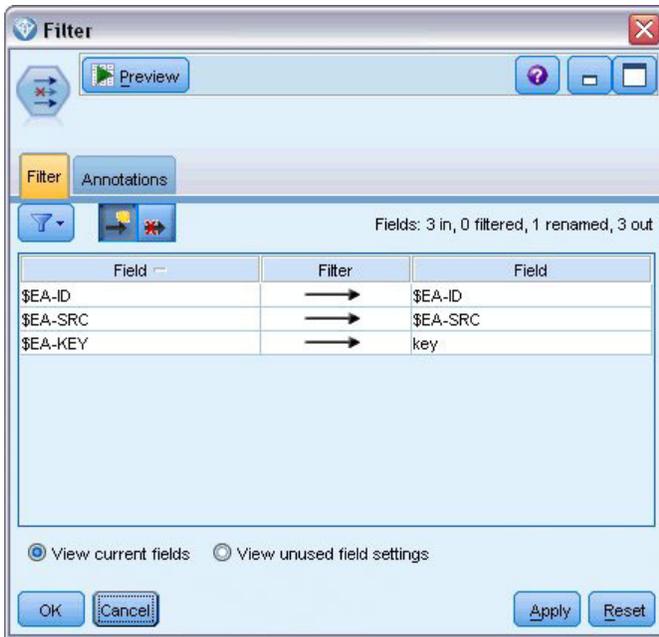


Figura 12. Renomeando o campo \$EA-KEY

Clique em **OK** para fechar o nó de filtro.

Agora, é necessário ordenar os IDs de entidade \$EA-ID em ordem ascendente. Anexe um nó de ordem ao nó de filtro. Abra o nó de ordem, clique no botão superior ao lado da tabela **Ordenar por**, selecione \$EA-ID e clique em **OK**.



Figura 13. Ordenando os IDs de entidade em ordem ascendente

Deixe a ordenação como **Ascendente** e clique em **OK**.

Agora, é necessário criar um campo extra que indique se um registro é exclusivo ou duplicado. Anexe um nó de deriva ao nó de ordem. Abra o nó de deriva e configure o nome **Campo de deriva** como IsDuplicate. Na lista **Derivar como**, escolha **Flag**, que também configura o **Tipo de campo** como **Flag**. Configure o campo **Valor real** como **Duplicado** e o **Valor falso** como **Exclusivo**.

Para localizar registros duplicados, você usará uma função de sequência especial chamada @OFFSET, que é fornecida com o SPSS Modeler.

Digite o seguinte no campo **Se**:

```
'$EA-ID' = @OFFSET('$EA-ID',1) or '$EA-ID' = @OFFSET('$EA-ID',-1)
```



Figura 14. Configurando a condição no nó de deriva

Com os IDs de entidade ordenados em ordem ascendente, a função @OFFSET verifica se há IDs de entidade adjacentes idênticos, representando registros duplicados. Caso haja, seus valores *IsDuplicate* serão configurados como *Duplicados*, caso contrário como *Exclusivos*.

Clique em **OK** para fechar o nó.

Para ver o efeito do nó de deriva, anexe um nó de tabela ao nó de deriva, abra o nó de tabela de clique em **Executar**. Role para baixo na janela de saída do nó de tabela, até a linha 45.

	\$EA-ID	\$EA-SRC	key	IsDuplicate
39	39	TEST	REF00000039	Unique
40	40	TEST	REF00000040	Unique
41	41	TEST	REF00000041	Unique
42	42	TEST	REF00000042	Unique
43	43	TEST	REF00000043	Unique
44	44	TEST	REF00000044	Unique
45	45	TEST	REF00000045	Duplicate
46	45	TEST	REF00000385	Duplicate
47	46	TEST	REF00000046	Unique
48	47	TEST	REF00000047	Unique
49	48	TEST	REF00000048	Unique
50	49	TEST	REF00000049	Unique
51	50	TEST	REF00000050	Unique
52	51	TEST	REF00000051	Unique
53	52	TEST	REF00000052	Unique
54	53	TEST	REF00000053	Unique
55	54	TEST	REF00000054	Unique
56	55	TEST	REF00000055	Unique
57	56	TEST	REF00000056	Unique
58	57	TEST	REF00000057	Unique

Figura 15. Saída do nó de deriva

Lembra-se de quando verificamos a saída diretamente a partir do nó de origem Entity Analytics (EA)? O sistema já tinha identificado que o registro REF00000385 fazia referência à mesma pessoa que a entidade 45. Agora, nós avançamos ainda mais com relação a esse estágio e sinalizamos o fato de que os registros REF00000045 e REF00000385 são duplicados, pois ambos fazem referência à entidade 45.

Role mais para baixo na janela de saída e veja os outros registros sinalizados como duplicados.

Para obter um relatório listando os registros duplicados, anexe um nó de relatório (da guia Saída da paleta de nós) ao nó de deriva *IsDuplicate*. Abra o nó de relatório, copie o texto a seguir no campo de entrada da guia Modelo e clique em **Executar**.

```
<html>
<h1>List of duplicate customer records.

<h2>This report was generated: [:@TODAY]

<h2>Duplicate records
<table>
  <tr>
    <td>Entity ID</td>
    <td>Key</td>
  </tr>

#WHERE IsDuplicate = "Duplicate"

  <tr>
    <td>['$EA-ID']</td>
    <td>[key]</td>
  </tr>
#
</table>
</html>
```

Isso fornece a saída a seguir.

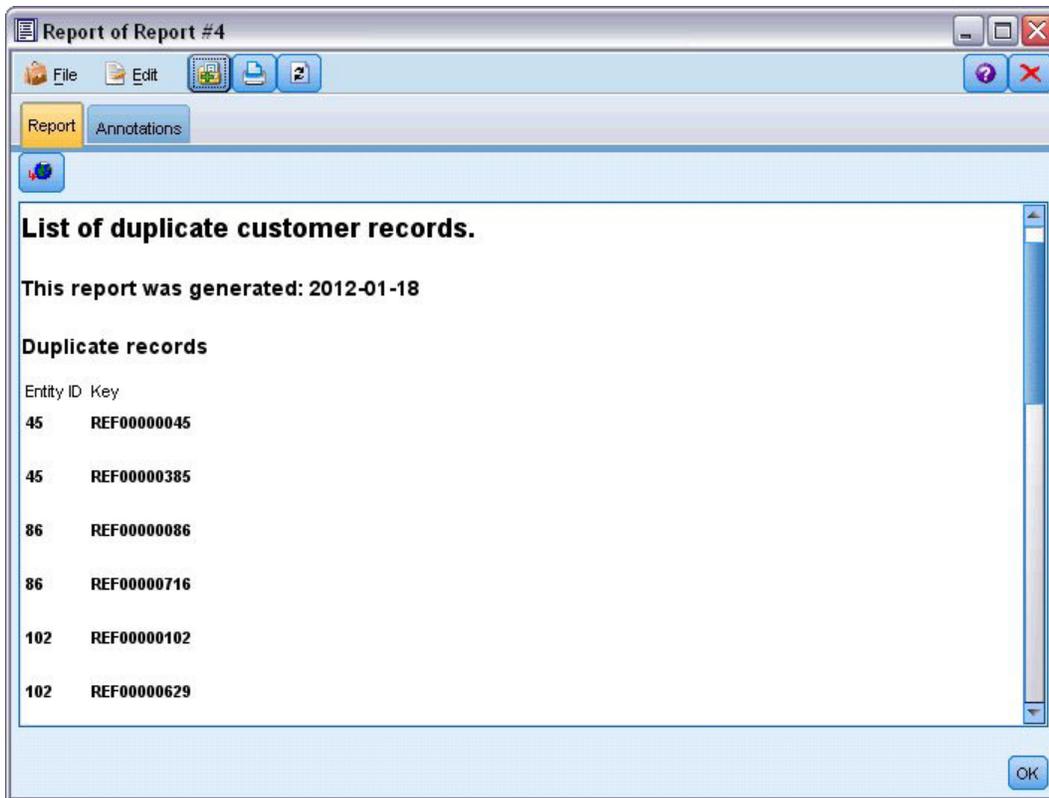


Figura 16. Saída do nó de relatório

O relatório usa o formato HTML nesse caso, embora também seja possível usar a formatação XML ou ASCII.

---

## Comparando a saída de análise de entidade ao modelo original

O estágio final desse exemplo é ver se a inclusão da análise de entidade faz alguma diferença para a predição original do banco. Lembre-se de que o modelo original previa 13 inadimplentes dos 150 aplicativos pendentes. Você usará um nó Mesclagem para mesclar a saída do modelo com as informações sobre os registros duplicados da análise de entidade para ver se essa ação altera a predição.

Primeiro, é necessário assegurar-se de que os novos campos incluídos por análise de entidade tenham os tipos de dados ou *níveis de medição* corretos, como são conhecidos no SPSS Modeler. Anexe um nó Tipo ao nó de deriva **IsDuplicate**, abra o nó Tipo e clique no botão **Ler valores**.

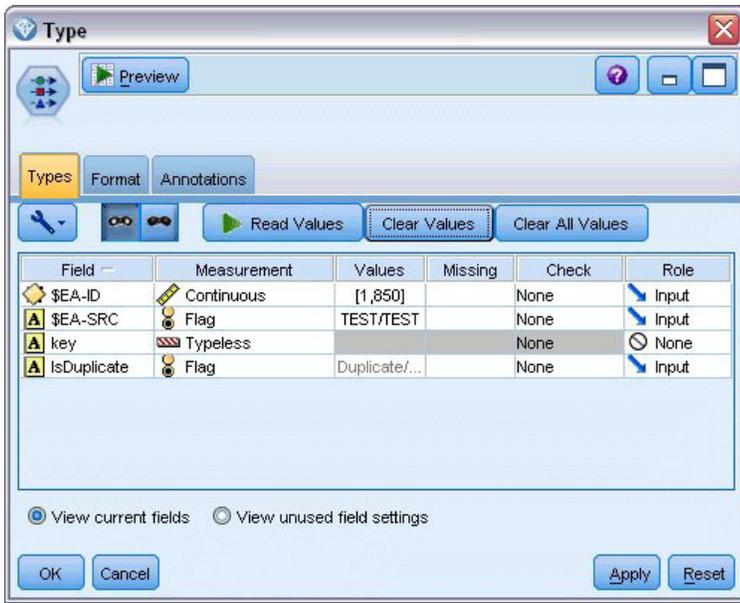


Figura 17. Configurações para o nó Tipo

Agora é possível incluir o nó Mesclagem. Anexe-o ao nó Tipo e também conecte-o ao nugget ouro contendo o modelo original. Para isso, clique com o botão direito no nugget ouro, escolha **Conectar** e, em seguida, clique no nó Mesclagem, que agora deve ter duas setas de entrada.

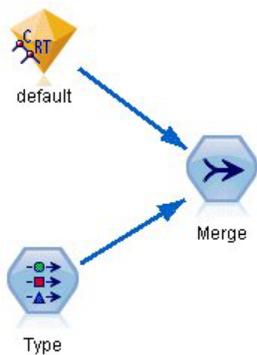


Figura 18. Entradas para o nó Mesclagem

Abra o nó Mesclagem, configure o **Método de mesclagem** como **Chaves** e clique no botão Seta à Direita para mover o campo **chave** de **Chaves possíveis** para **Chaves para mesclar**, em seguida, clique em **OK**.

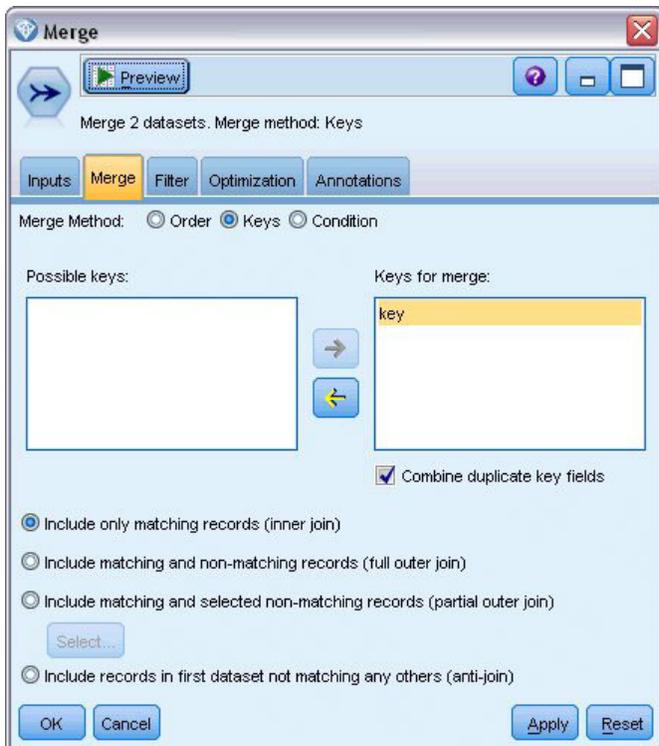


Figura 19. Especificando o campo de chave para a operação de mesclagem

Agora já está quase tudo pronto para fazer a comparação. No entanto, se você anexar um nó de distribuição e executá-lo nesse momento, não haverá nenhuma mudança com relação à predição original. Embora agora o fluxo mescle a saída do nugget do modelo original com os novos campos criados pela análise de entidade, o campo de predição em si (*\$R-default*) no modelo de dados não é atualizado com as novas informações.

Para fazer isso, use um nó de preenchimento, que pode substituir os valores do campo. Anexe um nó de preenchimento a um nó de mesclagem e abra o nó de preenchimento.

Clique no botão superior à direita de **Preencha os campos**, role para a parte inferior da lista, escolha **\$R-default** e clique em **OK**. Esse é o campo cujos valores deverão ser alterados se a condição especificada no restante da caixa de diálogo for cumprida.

Para especificar a condição, assegure-se de que **Substituir** esteja configurado para **Baseado na condição** e, em seguida, no campo **Condição** insira:

```
default != "default" and IsDuplicate = "Duplicate"
```

No campo **Substituir por**, insira:

```
"default"
```



Figura 20. Especificando a condição para substituir valores de campo

Essas definições precisam de algumas explicações. A condição afirma que, para cada registro no qual o valor do campo *padrão* do conjunto de dados original não for igual a **Padrão** e o registro tiver sido sinalizado como um duplicado, o valor do campo *\$R-default* no modelo será configurado para **Padrão**.

O campo *\$R-default* é o campo no modelo que contém a predição da probabilidade de que um cliente fique inadimplente no empréstimo. Dessa forma, os clientes com registros duplicados são incluídos no modelo como inadimplentes potenciais.

Clique em **OK** para fechar o nó de preenchimento.

Finalmente, está tudo pronto para ver a diferença que a análise de entidade fez. A partir da paleta Gráficos, anexe um nó de Distribuição ao nó de Preenchimento e abra o nó de Distribuição. Clique na lista **Campo** e escolha **\$R-default**.

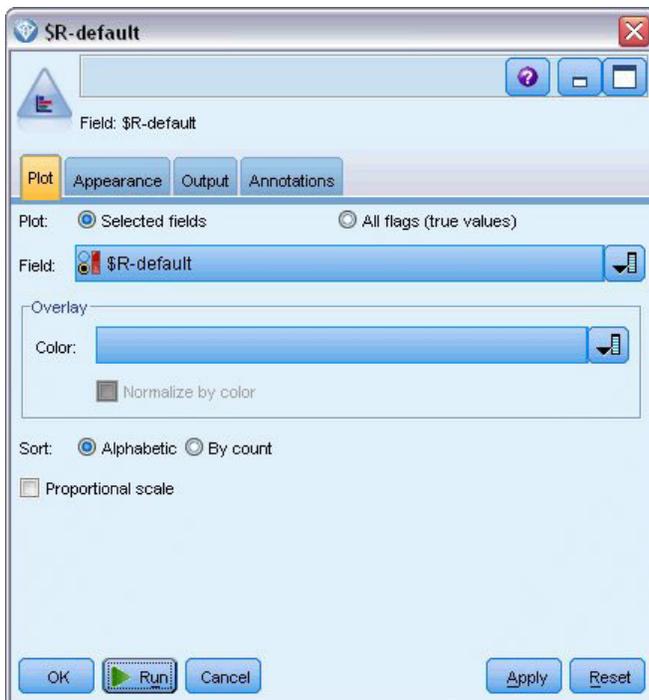


Figura 21. Configurações para o nó de distribuição

Clique em **Executar** para produzir o gráfico da nova predição.

Value	Proportion	%	Count
Default		10.67	16
Repaid		89.33	134

Figura 22. Saída do nó de distribuição após a análise de entidade

Agora, existem 16 aplicativos arriscados em vez de 13. Esses extras poderiam custar muito caro se fossem assumidos como padrão, portanto, é possível demonstrar graficamente ao banco o benefício de incluir a análise de entidade em suas operações de avaliação de risco.

## Resumo

Este exemplo mostrou como, usando a análise de entidade, é possível eliminar a duplicação de registros em dados sobre pessoas ou organizações, levando a uma melhoria na qualidade da predição.

*Observação:* O ideal seria eliminar os registros duplicados antes de executar qualquer outro processamento. Isso poderia ser realizado usando um nó de preparação de dados automatizada (ADP) para analisar os dados e identificar correções, não aceitar no sistema campos que sejam problemáticos ou provavelmente inúteis, derivar novos atributos quando apropriado e melhorar o desempenho por meio de técnicas de triagem inteligentes.

A combinação de análise de entidade e preparação de dados automatizada pode ajudar a assegurar que você trabalhe com dados que sejam os mais limpos possíveis.



---

## Apêndice. Propriedades de script do IBM SPSS Modeler Entity Analytics

---

### Script com o IBM SPSS Modeler Entity Analytics

A criação de script no IBM SPSS Modeler Entity Analytics é uma ferramenta poderosa para automatizar processos na interface com o usuário. Os scripts podem executar os mesmos tipos de ações que podem ser executadas com o mouse ou teclado e podem ser utilizados para automatizar tarefas que seriam altamente repetitivas ou que demorariam muito tempo para serem executadas manualmente. Para obter uma explicação do uso de script, consulte o guia *ScriptingAutomation.pdf*, disponível com o IBM SPSS Modeler.

---

### Propriedades comuns

As propriedades comuns aos nós do IBM SPSS Modeler Entity Analytics são listadas na tabela a seguir. Informações sobre os nós específicos são fornecidas nas seções a seguir.

Tabela 13. Propriedades comuns

Nome da propriedade	Tipo de dados	Descrição da propriedade
entity_repository	[ <i>'field'</i> , <i>'field'</i> , ... , <i>'field'</i> ]	A sequência de conexões do repositório. Formato: [ <i>'reposname'</i> , <i>'username'</i> , <i>'password'</i> ]  Exemplo: entity_repository = [ <i>'repos1'</i> , <i>'dba'</i> , <i>'psw1'</i> ]
entity_type	<i>string</i>	O tipo de entidade (conjunto de variáveis) a ser usado.  Exemplo: entity_type = <i>'PERSON'</i>

---

### Propriedades entityanalytics\_exportnode



O nó Exportação do EA é um nó terminal que lê os dados de entidade a partir de uma origem de dados e exporta-os para um repositório para o propósito da resolução da entidade.

Tabela 14. Propriedades entityanalytics\_exportnode

Propriedades entityanalytics_exportnode	Tipo de dados	Descrição da propriedade
mode	Add PurgeFirst	Modo de exportação. Add inclui registros de arquivo nos conteúdos existentes do repositório; PurgeFirst remove os conteúdos existentes antes de exportar.
source_tag	<i>string</i>	O identificador da origem de dados.  Exemplo: source_tag = <i>'CUST'</i>

Tabela 14. Propriedades *entityanalytics\_exportnode* (continuação)

Propriedades <i>entityanalytics_exportnode</i>	Tipo de dados	Descrição da propriedade
unique_key_field	<i>string</i>	Campo de entrada a ser usado para identificadores exclusivos para registros de dados.  Exemplo: unique_key_field = 'ID'
field_mapping	[ [ 'field_name' 'feature.element' 'usage_type' ]... ]	Mapeia os campos de entrada à variável correspondente no repositório.  Exemplo: field_mapping = [ [ 'fname' 'NAME.GIVEN_NAME' ' ' ] [ 'addr1' 'ADDRESS.ADDR1' 'PRIMARY' ] ] <i>Observação:</i> Para configurar <i>usage_type</i> para o equivalente de "(Auto)", use ' ' como no primeiro exemplo acima.

## Propriedades *entityanalytics\_sourcenode*



O nó de origem Entity Analytics (EA) lê as entidades resolvidas a partir do repositório e passa esses dados ao fluxo para um processamento adicional, como a formatação em um relatório.

Tabela 15. Propriedades *entityanalytics\_sourcenode*

Propriedades <i>entityanalytics_sourcenode</i>	Tipo de dados	Descrição da propriedade
source_tags	<i>lista</i>	Lista de tags das origens de dados que devem ser recuperadas do repositório.  Exemplo: source_tags=[ 'LOANS', 'CUSTOMERS' ]
relationships	None Fechar All	Critério de correspondência para recuperar detalhes de relacionamento do repositório. None não retorna relacionamentos. Close retorna correspondências próximas, dependendo de detalhes como o grau de separação. All retorna todos os relacionamentos possíveis.
max_degree_separation	<i>integer</i>	Mínimo 0, máximo 3.
output_entity_type	<i>string</i>	Lista de tipos de entidade que são usados no repositório.

## Propriedades *entityanalytics\_processnode*



O nó EA de fluxo compara os casos novos com relação aos dados de entidade no repositório.

Tabela 16. Propriedades *entityanalytics\_processnode*

Propriedades <i>entityanalytics_processnode</i>	Tipo de dados	Descrição da propriedade
<i>match</i>	Exata ByIdentifier All	Critério de correspondência para recuperar entidades do repositório. Exact retorna somente correspondências exatas. ByIdentifier retorna correspondências e entidades exatas que compartilham os mesmos identificadores. All retorna todas as correspondências possíveis.
<i>save_search_records</i>	<i>boolean</i>	
<i>relationships</i>	None Fechar All	Critério de correspondência para recuperar detalhes de relacionamento do repositório. None não retorna relacionamentos. Close retorna correspondências próximas, dependendo de detalhes como o grau de separação. All retorna todos os relacionamentos possíveis.
<i>max_degree_separation</i>	<i>integer</i>	Mínimo 0, máximo 3.
<i>output_entity_type</i>	<i>string</i>	Lista de tipos de entidade que são usados no repositório.



---

## Avisos

Estas informações foram desenvolvidas para produtos e serviços oferecidos no mundo todo.

É possível que a IBM não ofereça os produtos, serviços ou recursos discutidos nesta publicação em outros países. Consulte um representante IBM local para obter informações sobre produtos e serviços disponíveis atualmente em sua área. Qualquer referência a produtos, programas ou serviços IBM não significa que apenas produtos, programas ou serviços IBM possam ser usados. Qualquer produto, programa ou serviço funcionalmente equivalente, que não infrinja nenhum direito de propriedade intelectual da IBM poderá ser usado em substituição. Entretanto, a avaliação e verificação da operação de qualquer produto, programa ou serviço não IBM são de responsabilidade do Cliente.

A IBM pode ter patentes ou solicitações de patentes relativas a assuntos tratados nesta publicação. O fornecimento desta publicação não lhe garante direito algum sobre tais patentes. Pedidos de licença devem ser enviados, por escrito, para:

Gerência de Relações Comerciais e Industriais da IBM Brasil  
IBM Corporation  
Av. Pasteur, 138-146, Botafogo  
Rio de Janeiro. RJ  
CEP 22290-240  
Brasil

Para pedidos de licença relacionados a informações de byte duplo (DBCS), entre em contato com o Departamento de Propriedade Intelectual da IBM em seu país ou envie pedidos de licença, por escrito, para:

Licença de Propriedade Intelectual  
Lei de Propriedade Intelectual  
IBM Japan Ltd.  
1623-14, Shimotsuruma, Yamato-shi  
Kanagawa 242-8502 Japan

O parágrafo a seguir não se aplica a nenhum país em que tais disposições não estejam de acordo com a legislação local: A INTERNATIONAL BUSINESS MACHINES CORPORATION FORNECE ESTA PUBLICAÇÃO "NO ESTADO EM QUE SE ENCONTRA", SEM GARANTIA DE NENHUM TIPO, SEJA EXPRESSA OU IMPLÍCITA, INCLUINDO, MAS A ELAS NÃO SE LIMITANDO, AS GARANTIAS IMPLÍCITAS DE NÃO INFRAÇÃO, COMERCIALIZAÇÃO OU ADEQUAÇÃO A UM DETERMINADO PROPÓSITO. Alguns países não permitem a exclusão de garantias expressas ou implícitas em determinadas transações, portanto, essa disposição pode não se aplicar ao Cliente.

Essas informações podem conter imprecisões técnicas ou erros tipográficos. São feitas alterações periódicas nas informações aqui contidas; tais alterações serão incorporadas em futuras edições desta publicação. A IBM pode, a qualquer momento, aperfeiçoar e/ou alterar os produtos e/ou programas descritos nesta publicação, sem aviso prévio.

As referências nestas informações a websites não IBM são fornecidas apenas por conveniência e não representam de forma alguma um endosso a estes websites. Os materiais contidos nesses Web sites não fazem parte dos materiais deste produto IBM e a utilização desses Web sites é de inteira responsabilidade do Cliente.

A IBM pode usar ou distribuir as informações fornecidas da forma que julgar apropriada sem incorrer em qualquer obrigação para com o Cliente.

Licenciados deste programa que desejam obter informações sobre este assunto com objetivo de permitir: (i) a troca de informações entre programas criados independentemente e outros programas (incluindo este) e (ii) a utilização mútua das informações trocadas, devem entrar em contato com:

IBM Software Group  
ATTN: Licensing  
200 W. Madison St.  
Chicago, IL; 60606  
CEP 22290-240

Tais informações podem estar disponíveis, sujeitas a termos e condições apropriadas, incluindo em alguns casos o pagamento de uma taxa.

O programa licenciado descrito nesta publicação e todo o material licenciado disponível são fornecidos pela IBM sob os termos do IBM Customer Agreement, Contrato de Licença do Programa Internacional IBM ou qualquer outro contrato equivalente.

Quaisquer dados de desempenho aqui contidos foram determinados em um ambiente controlado. Portanto, os resultados obtidos em outros ambientes operacionais podem variar significativamente. Algumas medidas podem ter sido tomadas em sistemas em nível de desenvolvimento e não há garantia de que estas medidas serão as mesmas em sistemas geralmente disponíveis. Além disso, algumas medidas podem ter sido estimadas através de extrapolação. Os resultados reais podem variar. Os usuários desta publicação devem verificar os dados aplicáveis para seu ambiente específico.

As informações relacionadas a produtos não IBM foram obtidas junto aos fornecedores destes produtos, de seus anúncios publicados ou de outras fontes disponíveis publicamente. A IBM não testou estes produtos e não pode confirmar a precisão de desempenho, compatibilidade nem qualquer outra reivindicação relacionada a produtos não IBM. Perguntas sobre os recursos de produtos não IBM devem ser encaminhadas aos fornecedores desses produtos.

Todas as instruções relativas aos objetivos ou intenção futura da IBM estão sujeitas a alterações ou cancelamento sem aviso prévio e representam apenas metas e objetivos.

Essas informações contêm exemplos de dados e relatórios usados em operações diárias de negócios. Para ilustrá-los da forma mais completa possível, os exemplos incluem nomes de indivíduos, empresas, marcas e produtos. Todos estes nomes são fictícios e qualquer semelhança com nomes e endereços usados por uma empresa real é mera coincidência.

Se essas informações estiverem sendo exibidas em formato eletrônico, as fotografias e ilustrações coloridas poderão não aparecer.

---

## **Marcas comerciais**

IBM, o logotipo IBM e [ibm.com](http://ibm.com) são marcas comerciais ou marcas registradas da International Business Machines Corp., registradas em muitas jurisdições em todo o mundo. Outros nomes de produtos e serviços podem ser marcas comerciais da IBM ou de outras empresas. Uma lista atual de marcas comerciais da IBM trademarks está disponível na web em "Informações de Copyright e marcas comerciais" em [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Intel, logotipo Intel, Intel Inside, logotipo Intel Inside, Intel Centrino, logotipo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium e Pentium são marcas registradas ou marcas comerciais da Intel Corporation ou suas subsidiárias nos Estados Unidos e outros países.

Linux é uma marca registrada da Linus Torvalds nos Estados Unidos e/ou em outros países.

Microsoft, Windows, Windows NT e o logotipo Windows são marcas comerciais da Microsoft Corporation nos Estados Unidos e/ou em outros países.

UNIX é uma marca registrada do The Open Group nos Estados Unidos e em outros países.

Java e todas as marcas e logotipos baseados em Java são marcas comerciais ou marcas registradas da Oracle e/ou suas afiliadas.

Outros nomes de produtos e serviços podem ser marcas comerciais da IBM ou de outras empresa.



---

# Índice Remissivo

## A

- análise de entidade
  - comparada com a análise preditiva 2
  - definidos 1
  - usando com o IBM SPSS Modeler 3
  - usando com outros produtos IBM SPSS 30
- anonimização de variáveis
  - repositório de entidade 20
- anonimizando variáveis
  - repositório de entidade 20

## C

- chaves exclusivas
  - análise de entidade 5
  - repositório de entidade 15
- configuração
  - repositório de entidade 16, 23, 24
- correspondência de entidade,
  - configurando limite 23
- credenciais do administrador
  - gerenciando para análises de entidade 31

## D

- designações de porta
  - configurando para análise de entidade 30

## E

- excluindo
  - repositório de entidade 34
- excluindo dados não usados
  - repositório de entidade 33
- exportando
  - dados para um repositório de entidade 5

## I

- identidades resolvidas, analisando com
  - análise de entidade 24
- informações de tipo, configurando para
  - análise de entidade 25

## L

- limite para correspondência de entidade,
  - análise de entidade 23
- limpando
  - repositório de entidade 33

## M

- mapeando campos
  - para variáveis de repositório de entidade 5, 14, 15, 16, 17, 27, 28

## N

- nó de origem Entity Analytics (EA) 24
- Nó EA de fluxo, análise de entidade 26
- Nó Exportação do EA, análise de entidade 11
- nós
  - incluindo em um fluxo de análise de entidade 26
- nós de exportação
  - análise de entidade 5, 11
- nós de origem
  - análise de entidade 7, 24
- nós de processamento
  - análise de entidade 8, 26
- novos casos, comparando com relação ao
  - repositório de análise de entidade 26

## O

- origem de dados, selecionando para
  - análise de entidade 24
- origens de dados
  - conectando com a análise de entidade 4, 12
  - visualizando para análise de entidade 14, 28

## P

- propriedades
  - script 53
- propriedades de fluxo
  - configurando análise de entidade 32

## R

- regras de resolução, análise de entidade 23
- renomeando
  - campos de dados para análise de entidade 25
- repositório
  - administrando a análise de entidade 30
  - análise de entidade 4, 5, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 23, 24, 26, 27, 28, 33, 34
  - diretório de armazenamento para análise de entidade, alterando 32
- repositório de entidade 11
  - anonimizando 20
  - comparando novos casos com relação 26

- repositório de entidade (*continuação*)
  - conectando ao IBM SPSS Modeler 5
  - configurando 11, 16, 23, 24
  - configurando designações de porta 30
  - configurando propriedades de fluxo 32
  - criação 4, 12, 13
  - excluindo 34
  - excluindo dados não usados 33
  - gerenciando credenciais do administrador 31
  - limpando 33
  - mantendo 17
  - Movendo para um diretório de armazenamento diferente 32
  - opções 14
  - tarefas administrativas 30
  - variáveis 19
- resolução de identidade, análise de entidade 5
- resolvendo identidades, análise de entidade 5

## S

- saída
  - da análise de entidade 29
- script
  - propriedades 53

## T

- tags de origem
  - repositório de entidade 15
- tipos de entidade
  - análise de entidade 21
  - repositório de entidade 15
- tipos de uso, análise de entidade 21

## V

- variáveis
  - repositório de entidade 5, 14, 15, 16, 17, 19, 27, 28







Impresso no Brasil