

IBM SPSS Modeler Entity Analytics
15-Benutzerhandbuch



Hinweis: Lesen Sie zunächst die allgemeinen Informationen unter Hinweise auf S. 67, bevor Sie dieses Informationsmaterial sowie das zugehörige Produkt verwenden.

Diese Ausgabe bezieht sich auf und alle nachfolgenden Versionen sowie Anpassungen, sofern dies in neuen Ausgaben nicht anders angegeben ist.

Screenshots von Adobe-Produkten werden mit Genehmigung von Adobe Systems Incorporated abgedruckt.

Screenshots von Microsoft-Produkten werden mit Genehmigung der Microsoft Corporation abgedruckt.

Lizenziertes Material - Eigentum von IBM

Eingeschränkte Rechte für Benutzer der US-Regierung: Verwendung, Vervielfältigung und Veröffentlichung eingeschränkt durch GSA ADP Schedule Contract mit der IBM Corp.

Vorwort

IBM® SPSS® Modeler ist die auf Unternehmensebene einsetzbare Data-Mining-Workbench von IBM Corp.. Mit SPSS Modeler können Unternehmen und Organisationen die Beziehungen zu ihren Kunden bzw. zu den Bürgern durch ein tief greifendes Verständnis der Daten verbessern. Organisationen benutzen die mithilfe von SPSS Modeler gewonnenen Erkenntnisse zur Bindung profitabler Kunden, zur Ermittlung von Cross-Selling-Möglichkeiten, zur Gewinnung neuer Kunden, zur Ermittlung von Betrugsfällen, zur Reduzierung von Risiken und zur Verbesserung der Verfügbarkeit öffentlicher Dienstleistungen.

Die visuelle Benutzeroberfläche von SPSS Modeler erleichtert die Anwendung des spezifischen Geschäftswissens der Benutzer, was zu leistungsstärkeren Vorhersagemodellen führt und die Zeit bis zur Lösungserstellung verkürzt. SPSS Modeler bietet zahlreiche Modellierungsverfahren, beispielsweise Algorithmen für Vorhersage, Klassifizierung, Segmentierung und Assoziationserkennung. Nach der Modellerstellung ermöglicht IBM® SPSS® Modeler Solution Publisher die unternehmensweite Bereitstellung für Entscheidungsträger oder in einer Datenbank.

Über IBM Business Analytics

IBM Business Analytics-Software bietet vollständige, einheitliche und genaue Informationen, auf die Entscheidungsträger vertrauen, um die Unternehmensleistung zu steigern. Ein umfassendes Portfolio von Anwendungen für [Unternehmensinformationen](#), [Vorhersageanalysen](#), [Verwaltung der Finanzleistung und Strategie](#) sowie [Analysen](#) bietet sofort klare und umsetzbare Einblicke in die aktuelle Leistung und ermöglicht die Vorhersage zukünftiger Ergebnisse. In Kombination mit umfassenden Branchenlösungen, bewährten Vorgehensweisen und professionellen Dienstleistungen können Unternehmen jeder Größe optimale Produktivität erreichen, die Entscheidungsfindung zuverlässig automatisieren und bessere Ergebnisse erzielen.

Als Teil dieses Portfolios unterstützt die IBM SPSS Predictive Analytics-Software Unternehmen dabei, zukünftige Ereignisse vorherzusagen und aktiv auf diese Erkenntnisse zu reagieren, um bessere Geschäftsergebnisse zu erzielen. Kunden aus den Bereichen Wirtschaft, Behörden und Bildung aus aller Welt verlassen sich auf die IBM SPSS-Technologie. Sie bringt Ihnen beim Gewinnen, Halten und Ausbauen neuer Kundenbeziehungen einen Wettbewerbsvorteil und verringert gleichzeitig das Betrugs- sowie andere Risiken. Durch Integration der IBM SPSS-Software in den täglichen Betrieb können diese Unternehmen qualifizierte Vorhersagen treffen und dadurch die Entscheidungsfindung so ausrichten und automatisieren, dass Geschäftsziele erreicht werden und ein messbarer Wettbewerbsvorteil entsteht. Wenn Sie weitere Informationen wünschen oder einen Mitarbeiter kontaktieren möchten, ist dies unter <http://www.ibm.com/spss> möglich.

Technischer Support

Kunden mit Wartungsvertrag können den technischen Support in Anspruch nehmen. Kunden können sich an den technischen Support wenden, wenn sie Hilfe bei der Arbeit mit IBM Corp.-Produkten oder bei der Installation in einer der unterstützten Hardware-Umgebungen benötigen. Die Kontaktdaten des Technischen Supports finden Sie auf der IBM Corp.-Website

unter <http://www.ibm.com/support>. Sie müssen bei der Kontaktaufnahme Ihren Namen, Ihre Organisation und Ihre Supportvereinbarung angeben.

1	Entitätsanalyse	1
	Informationen zur Entitätsanalyse	1
	Entitätsanalyse und Vorhersageanalyse	3
2	Entitätsanalyse mit IBM SPSS Modeler	4
	Verwendung der Entitätsanalyse mit IBM SPSS Modeler	4
	Schritt 1: Datenquelle in SPSS Modeler einlesen	5
	Schritt 2: Repository erstellen	6
	Schritt 3: SPSS Modeler mit dem Repository verbinden	7
	Schritt 4: Eingabefelder den Repository-Merkmalen zuordnen	7
	Schritt 5: Daten in das Repository exportieren und Konflikte auflösen	9
	Schritt 6: Analyse der aufgelösten Identitäten	11
	Schritt 7: Neue Fälle anhand des Repositorys auflösen	12
	Schritt 8: Warnungen generieren	13
3	Entitätsanalyseaufgaben	14
	Informationen zu den Aufgaben	14
	Einrichten eines Entitäts-Repositorys (EA-Exportknotens)	14
	Das Entitäts-Repository	14
	Verbindung mit einer Datenquelle	15
	Erstellen des Repository	16
	Erstellen einer Zuordnung zwischen Eingabefeldern und Merkmalen (EA-Exportknoten)	19
	Anzeigen der Feldzuordnungen (EA-Exportknoten)	22
	Konfigurieren eines Entitäts-Repository	23
	Anzeigen der Datenquellenzuordnungen	24
	Verwaltung der Repository-Merkmale	25
	Hinzufügen bzw. Bearbeiten von Merkmalen	28
	Verwaltung der Elementtypen	30
	Festlegen des Schwellenwerts für den Entitätsabgleich	32
	Wiederverwenden von Repository-Konfigurationen	33
	Speichern der Konfigurationsänderungen	33
	Schließen des Konfigurationsfensters	33
	Analyse der aufgelösten Identitäten (Entity Analytics-Quellenknoten)	33
	Auswählen einer Datenquelle	34
	Umbenennen von Datendateien	34

Festlegen der Typinformationen für Datenfelder	35
Hinzufügen von Knoten zum Stream	35
Vergleich neuer Fälle mit dem Repository (Knoten "Streaming von EA")	36
Erstellen einer Zuordnung zwischen Eingabefeldern und Merkmalen (Knoten "Streaming von EA")	37
Anzeigen der Feldzuordnungen (Knoten "Streaming von EA")	39
Anzeigen der Datenzuordnungen (Knoten "Streaming von EA")	40
Ausgabe aus dem Knoten "Streaming von EA"	41
Verwendung von IBM SPSS Modeler Entity Analytics mit anderen IBM SPSS-Produkten	41
Verwaltungsaufgaben	42
Konfigurieren von Port-Zuweisungen	42
Verwalten der Administratorberechtigungen für die Repository-Datenbank	43
Verschieben des Speicherverzeichnisses für das Repository	44
Festlegen von Stream-Optionen für Datums-/Zeit- und Zeitstempelfelder	44
Ausführen von IBM SPSS Modeler Entity Analytics mit SPSS Modeler-Client und SPSS Modeler Server auf demselben Windows-System	45
Löschen des Inhalts eines Entitäts-Repository	45
Löschen eines Entitäts-Repository	46
Löschen eines Repository, wenn keine Verbindung damit hergestellt werden kann	46

4 Entitätsanalyse in Aktion 48

Zu diesem Beispiel	48
Das ursprüngliche Modell	48
Hinzunahme der Entitätsanalyse	52
Übertragen der Quelldaten in das Repository	53
Lesen der aufgelösten Identitäten	55
Vergleich der Entitätsanalyseausgabe mit dem ursprünglichen Modell	62
Zusammenfassung	66

Anhang

A Hinweise 67

Index 70

Entitätsanalyse

Informationen zur Entitätsanalyse

IBM® SPSS® Modeler Entity Analytics fügt eine völlig neue Dimension zu den IBM® SPSS® Modeler-Vorhersageanalysen hinzu. Während bei Vorhersageanalysen versucht wird, zukünftiges Verhalten aus früheren Daten vorherzusagen, liegt der Schwerpunkt bei der Entitätsanalyse auf der Verbesserung von Kohärenz und Konsistenz der aktuellen Daten, indem Identitätskonflikte innerhalb der Datensätze selbst aufgelöst werden. Bei der Identität kann es sich um die Identität einer Person, einer Organisation, eines Objekts oder einer anderen Entität handeln, bei der Unklarheiten bestehen könnten. Die Identitätsauflösung kann in einer Reihe von Bereichen entscheidend sein, darunter Customer Relationship Management, Betrugserkennung, Bekämpfung der Geldwäsche sowie nationale und internationale Sicherheit.

Angenommen, Sie besitzen die folgenden Kundendatensätze aus zwei verschiedenen Quellen und sind sich nicht sicher, ob sie sich auf dieselbe Person oder auf zwei verschiedene Personen beziehen.

Quelle 1

Datensatznr.: 70001
Name: Hans Schmidt
Adresse: Hauptstraße 123
Steuernummer: 555001111
Führerschein: 0001133107
Kreditkarte: 10229127

Quelle 2

Datensatznr.: 9103
Name: JOHANN Schmidt
Geburtsdatum: 17.06.1934
Telefon: 555-1212
Kreditkarte: 10229128
E-Mail: jls@mail.com
IP-Adresse: 9.50.18.77

Es gibt keine genauen Übereinstimmungen in den Daten zwischen den beiden Datensätzen. Wenn wir jedoch noch eine dritte Quelle hinzunehmen, finden wir gemeinsame Attribute.

Quelle 3

Datensatznr.: 6251
Name: Hans Schmidt
Telefon: 555-1212
Führerschein: 0001133107

Kreditkarte: 10229132

Die Führerscheinnummer verknüpft die Datensätze in Quelle 1 und Quelle 3 und die Telefonnummer verknüpft die Quellen 2 und 3. Wir können also hinreichend sicher sein, dass sich alle drei Quellen auf dieselbe Person beziehen.

Aber was tun wir, wenn der Sachverhalt weniger klar ist? Möglicherweise haben wir nur sehr wenige Daten, die wir unserer Entscheidung zugrunde legen können. Betrachten Sie die beiden folgenden Datensätze.

Quelle 4

Datensatznr.: S45286
Name: Hans T. Schmidt Jr.
Adresse: Hauptstraße 456
Telefon: 703-555-2000
Geburtsdatum: 12.03.1984

Datensatznr.: S45287
Name: Hans T. Schmidt
Adresse: Hauptstraße 456
Telefon: 703-555-2000
Führerschein: 009900991

Offensichtlich ist dies nicht der Herr Schmidt aus den vorangegangenen Datensätzen: die Unterschiede sind groß genug, dass wir dies ausschließen können. Wir haben jedoch trotzdem ein Problem. Zwei verschiedene Datensätze aus derselben Datenquelle scheinen sich auf dieselbe Person zu beziehen. Handelt es sich um doppelte Datensätze? Wir können nicht sicher sein, solange wir keinen anderen zugehörigen Datensatz finden, der uns weitere Informationen liefert, möglicherweise einen Datensatz aus einer anderen Quelle.

Quelle 5

Datensatznr.: 769582-2
Name: Hans T. Schmidt Sr.
Adresse: Hauptstraße 456
Telefon: 703-555-2000
Führerschein: 009900991
Geburtsdatum: 25.06.1959

Dies löst das Problem. Bei den beiden Datensätzen in Quelle 4 handelt es sich nicht um Duplikate, sondern tatsächlich um einen Vater und einen Sohn mit demselben Namen und unter derselben Adresse, die dieselbe Telefonnummer verwenden. Bei einem manuellen System könnte eine wochenlange Suche erforderlich sein, um den einen Datensatz zu finden, der die Identitäten auflöst. Mit einem automatisierten Entitätsanalysesystem wird der Zeitaufwand für die Auflösung erheblich reduziert.

Entitätsanalyse und Vorhersageanalyse

Wenn alle Ihre Daten aus einer einzigen, vollständigen und unzweideutigen Datenquelle bestehen würden, wäre es für IBM® SPSS® Modeler relativ einfach, etwaige Identitätskonflikte aufzulösen. Sie bräuchten lediglich die Vorhersageanalyse, um Ihre Daten in SPSS Modeler einzulesen, die Verarbeitung durchzuführen und zuverlässige Ergebnisse zu erhalten.

In der realen Welt sieht die Sache jedoch normalerweise deutlich anders aus. Die Daten sind typischerweise alles andere als vollständig, häufig mehrdeutig und oftmals über viele verschiedene Datenquellen verstreut, in denen viele verschiedene Attribute mit wenigen sich überlappenden Feldern dokumentiert sind. Der Nutzen der Entitätsanalyse liegt zum Teil darin, dass Daten aus all den verschiedenen Quellen in einem einzigen, zentralen Bereich, dem so genannten **Repository**, gesammelt werden. Das Entitätsanalyzesystem untersucht die Daten dann bis ins kleinste Detail, um Konflikte aufzulösen. Dabei erhalten Datensätze, die von derselben Person oder Organisation stammen, eine eindeutige ID.

In der folgenden Tabelle werden die Unterschiede zwischen den beiden Analysetypen veranschaulicht.

Tabelle 1-1
Unterschiede zwischen Vorhersage- und Entitätsanalyse

Eigenschaft	Vorhersageanalyse	Entitätsanalyse
Training	Beruh auf relativ kleinen Sets und numerischen Bereichen	Beruh auf großen Sets, z. B. Namen, Adressen
Für das Training verwendete Daten	Teilmenge (Trainingspartition)	Alle Daten werden verwendet
Verallgemeinerung	Algorithmus verallgemeinert über die gesamten Trainingsdaten hinweg, um ein prägnantes Modell zu erstellen.	Die Daten werden in Strukturen persistiert, die sich für Entitätsabgleich und Beziehungserkennung eignen.
Betrugserkennung	Datensätze werden als potenziell betrügerisch gekennzeichnet, wenn sie typische Eigenschaften einer betrügerischen Anwendung aufweisen.	Datensätze werden als potenziell betrügerisch gekennzeichnet, wenn sie in Bezug zu bekannten betrügerischen Datensätzen stehen oder von denselben Personen, jedoch unter verschiedenen Identitäten, stammen.

Entitätsanalyse mit IBM SPSS Modeler

Verwendung der Entitätsanalyse mit IBM SPSS Modeler

Sie stellen fest, dass bei Ihren Daten Identitätsprobleme vorliegen könnten. Wie kann IBM® SPSS® Modeler Entity Analytics Sie bei der Lösung dieser Probleme unterstützen? Im Folgenden wird eine Vorgehensweise vorgeschlagen, die Sie jedoch eventuell abwandeln müssen, um sie an Ihre speziellen Anforderungen anzupassen.

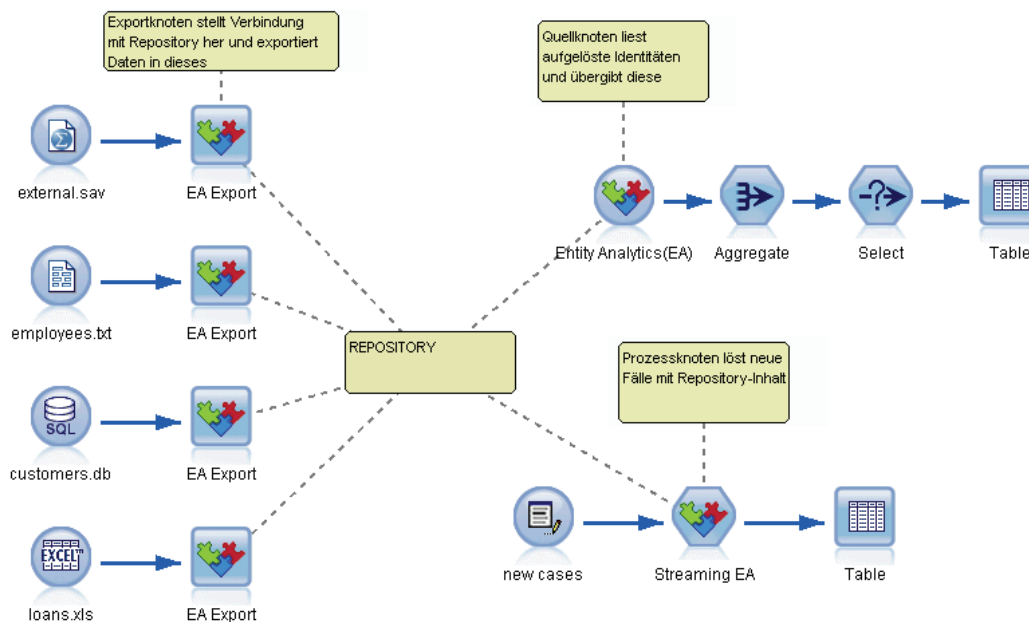
- Datenquelle in IBM® SPSS® Modeler einlesen
- Repository erstellen, in dem die Daten gespeichert werden können
- SPSS Modeler mit dem Repository verbinden
- Datenfelder den Repository-Merkmalen zuordnen
- Daten in das Repository exportieren und Identitäten auflösen
- Aufgelöste Identitäten analysieren
- Neue Fälle anhand des Repositorys auflösen
- Etwaige notwendige Warnungen generieren (als Stapelvorgang oder in Echtzeit)

Sie benötigen nun Kenntnisse darüber, wie SPSS Modeler funktioniert. SPSS Modeler ist ein sehr benutzerfreundliches Tool, das auf der grafischen Darstellung eines Datenstreams beruht, der eine Reihe von Knoten durchläuft. Jeder Knoten steht für einen bestimmten Schritt des Arbeitsablaufs.

SPSS Modeler selbst enthält eine große Palette an Knoten, die alle standardmäßigen Data Mining-Funktionen abdecken. IBM SPSS Modeler Entity Analytics fügt Knoten speziell für die Verwendung in Entitätsanalysen hinzu. Hierbei handelt es sich um den EA-Exportknoten, den Entity Analytics-(EA-)Quellenknoten und den Prozessknoten "Streaming von EA".

Der Prozess wird durch die folgende Abbildung dargestellt.

Abbildung 2-1
Der Entitätsanalyseprozess



Schritt 1: Datenquelle in SPSS Modeler einlesen

Ihre erste Aufgabe besteht darin, Ihre Daten mithilfe eines Quellenknoten in SPSS Modeler einzulesen. Quellenknoten werden in SPSS Modeler durch ein rundes Symbol gekennzeichnet.

Abbildung 2-2
IBM SPSS Modeler-Quellenknoten



Die Daten können in jedem beliebigen Format vorliegen, das von SPSS Modeler unterstützt wird, beispielsweise Textdateien, Datenbanktabellen, Kalkulationstabellen, XML-Dateien usw. Für jedes verschiedene Format ist jedoch jeweils ein entsprechender SPSS Modeler-Quellenknoten erforderlich. In der Abbildung handelt es sich um einen Datenbank-Quellenknoten.

Jede Datenquellendatei benötigt ein Feld, das die einzelnen Datensätze eindeutig kennzeichnet. Wenn eine Datenquelle kein derartiges Feld aufweist, können Sie problemlos eines in SPSS Modeler hinzufügen. Für weitere Informationen siehe Thema [Hinzufügen einer eindeutigen Datensatz-ID](#) in Kapitel 3 auf S. 15.

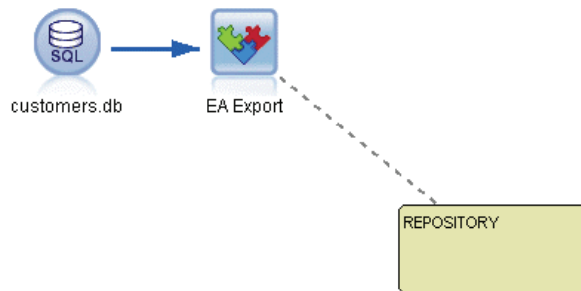
Für weitere Informationen siehe Thema [Verbindung mit einer Datenquelle](#) in Kapitel 3 auf S. 15.

Schritt 2: Repository erstellen

Der Schwerpunkt all Ihrer Bemühungen im Bereich der Entitätsanalyse liegt auf dem Repository, dem zentralen Speicherbereich, in dem Sie alle Datensätze sammeln können.

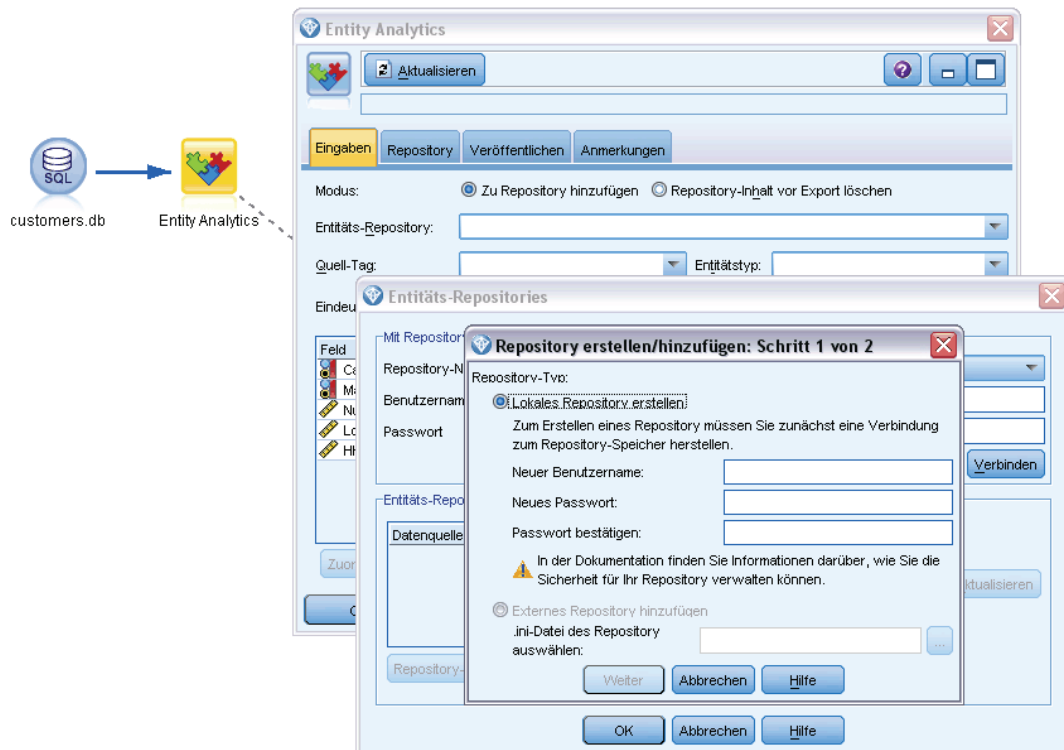
Zur Erstellung eines Repositorys verbinden Sie zunächst die Datenquelle mit einem EA-Exportknoten, der durch das quadratische Symbol repräsentiert wird.

Abbildung 2-3
Verbinden mit einem Repository



Aus dem Exportknoten können Sie ein neues Repository erstellen (oder ein bestehendes auswählen), das zur Aufnahme der exportierten Daten bereit ist.

Abbildung 2-4
Erstellen eines Repository



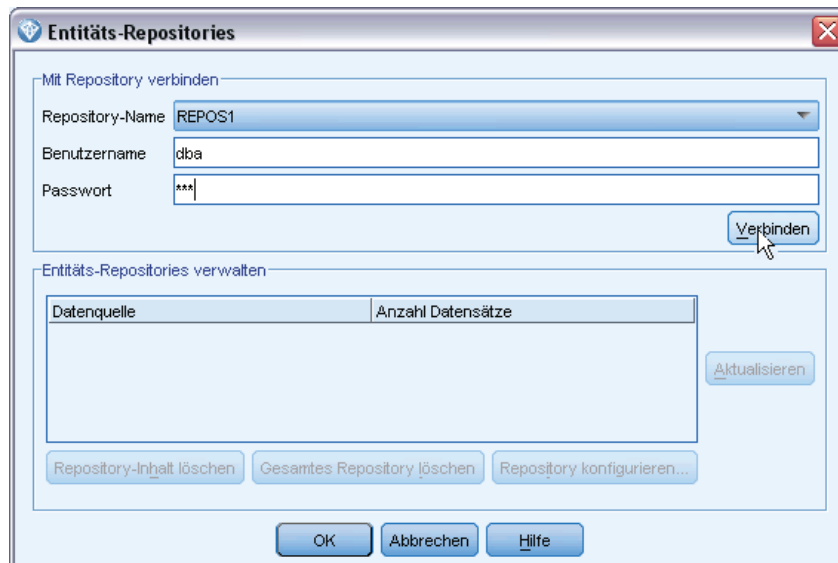
Der Vorgang der Repository-Erstellung wird weiter unten ausführlich beschrieben. Für weitere Informationen siehe Thema [Einrichten eines Entitäts-Repositorys \(EA-Exportknotens\)](#) in Kapitel 3 auf S. 14.

Nachdem Sie ein Repository eingerichtet haben, können Sie seine Inhalte auf verschiedene Weisen verwalten. Für weitere Informationen siehe Thema [Konfigurieren eines Entitäts-Repository in](#) Kapitel 3 auf S. 23.

Schritt 3: SPSS Modeler mit dem Repository verbinden

Nach der Erstellung des Repositorys müssen Sie es nun mit dem SPSS Modeler-Stream verbinden.

Abbildung 2-5
Verbinden mit dem Repository



Für weitere Informationen siehe Thema [Optionen für das Entitäts-Repository](#) in Kapitel 3 auf S. 18.

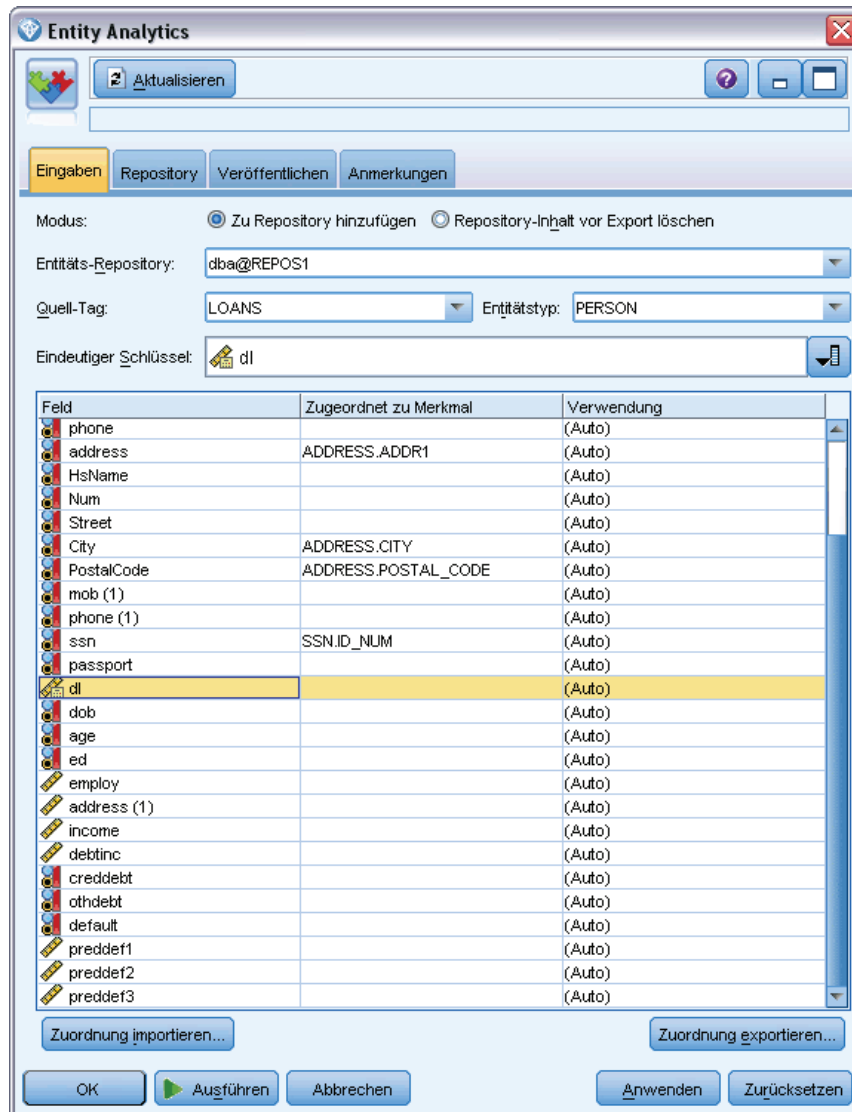
Schritt 4: Eingabefelder den Repository-Merkmalen zuordnen

Datenquellen können mehrere Arten von Entitätsinformationen enthalten. Einige Informationstypen kommen bei den meisten Entitätsdatenquellen vor, während andere möglicherweise nur bei einer bestimmten Datenquelle vorkommen. Bei einem Entitäts-Repository werden diese verschiedenen Informationstypen als **Merkmale** bezeichnet. Das Repository stellt eine Reihe von Merkmalen als Standard bereit, und Sie können auch eigene Merkmale erstellen.

Ein Repository-Merkmal ist ein einzelner Informationstyp, der zusammen mit einer Entitätsdatenquelle verwendet werden kann. Einige Merkmale (z. B. Vorname, Nachname, Geburtsdatum usw.) können mit vielen verschiedenen Datenquellen verwendet werden, während andere Merkmale nur bei einer bestimmten Datenquelle vorkommen. Ein Merkmal entspricht üblicherweise einem Feld in einem Datensatz oder einer Spalte in einer Datenbanktabelle.

Wenn Sie ein Repository erstellt und eine Verbindung damit hergestellt haben, kennzeichnen Sie ein Feld Ihrer Eingabedaten als Feld mit dem **eindeutigen Schlüssel**, das in der nachfolgenden Analyse verwendet wird. Außerdem ordnen Sie die Eingabedatenfeldern den entsprechenden Merkmalen im Repository zu. Damit sollen Dopplungen im Repository vermieden werden, für den Fall, dass verschiedene Datenquellen unterschiedliche Namen für Felder verwenden, die dieselbe Art von Informationen enthalten. Der EA-Exportknoten enthält eine Zuordnungstabelle, in der Sie die Zuordnungen erstellen können.

Abbildung 2-6
Erstellen einer Zuordnung zwischen Eingabefeldern und Merkmalen

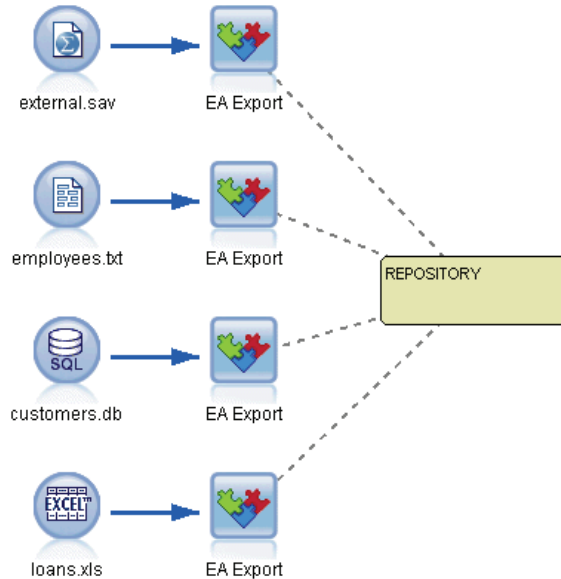


Für weitere Informationen siehe Thema [Erstellen einer Zuordnung zwischen Eingabefeldern und Merkmalen \(EA-Exportknoten\)](#) in Kapitel 3 auf S. 19.

Schritt 5: Daten in das Repository exportieren und Konflikte auflösen

Für die einzelnen Datenquellen ist jeweils ein eigener EA-Exportknoten erforderlich, wenn Ihre Daten also über mehrere verschiedene Quellen verstreut sind, sieht Ihr Stream wohl in etwa wie folgt aus.

Abbildung 2-7
Exportieren von Daten aus mehreren Datenquellen in das Repository



Wenn Sie über mehrere Datenquellen verfügen, können Sie auswählen, ob die Datensätze aus einer, einigen oder allen Datenquellen eingelesen werden sollen. Das Entitätsanalyzesystem analysiert die von Ihnen ausgewählten Datensätze und fügt zu jedem ein ID-Feld mit der Bezeichnung “\$EA_ID” hinzu. Wenn nun zwei oder mehr Datensätze, die sich auf zuvor mehrdeutige Identitäten bezogen, aufgelöst werden können, sind die zu diesen Datensätzen hinzugefügten IDs im gesamten Repository eindeutig. Das System fügt auch ein Feld hinzu, das die Datenquelle angibt, aus der der Datensatz stammt.

Sie verbinden die einzelnen Datenquellenknoten jeweils mit ihrem eigenen EA-Exportknoten, ordnen die Eingabefelder den Repository-Funktionen zu und führen dann den Stream aus, um die Daten aus SPSS Modeler in das Repository zu exportieren und etwaige Identitätskonflikte in einem einzigen Vorgang aufzulösen. Stellen Sie sich zur Veranschaulichung der Funktionsweise vor, dass Sie die folgenden Datensätze in vier verschiedenen Datenquellen besitzen.

Externe Daten

Name	Telefon	Kreditrisiko
Mike	555-1234	560
Joe	555-4567	780

Mitarbeiter

Name	Adresse	Telefon
Michael	1234 5th Street	555-1234
Fred	543 1st Avenue	555-9876

Kunden

Name	Adresse	Sparkonto
Susan	1234 5th Street	\$1234
Joe	777 Oak Street	\$5

Kredite

Name	Adresse	Telefon	Kredit
Sue	1234 5th Street	555-1234	\$10,000
Joseph	777 Oak Street	555-4567	\$50,000

Wie wir gesehen haben, exportieren Sie die einzelnen Datenquellen nacheinander in das Repository. Dabei aktualisiert das Repository die Auflösung der einzelnen Datensätze. Im Repository wird jedem Datensatz ein ID-Feld (mit dem Namen *SEA-ID*) und ein Feld zur Quellenangabe (mit dem Namen *SEA-SRC*) vorangestellt, das die Datenquelle anzeigt, aus der der Datensatz stammt. In unserem Beispiel sieht der Repository-Inhalt nach dem Export aller vier Datenquellen wie folgt aus:

Tabelle 2-1

Beispiel für Repository-Inhalte nach Exportschritt

SEA-ID	SEA-SRC	Name	Telefon	Adresse	Kreditrisiko	Sparkonto	Kredit
1	Mitarbeiter	Michael	555-1234	1234 5th St			
1	External	Mike	555-1234		560		
2	Kunden	Joe		777 Oak St		\$5	
2	External	Joe	555-4567		780		
2	Kredite	Joseph	555-4567	777 Oak St			\$50,000
3	Mitarbeiter	Fred	555-9876	543 1st Ave			
4	Kunden	Susan		1234 5th St		\$1234	
4	Kredite	Sue	555-1234	1234 5th St			\$10,000

Das Entitätsanalysesystem hat anhand der gemeinsamen Telefonnummer ermittelt, dass *Mike* im Daten-Set *External* dieselbe Person ist wie *Michael* im Datenset *Mitarbeiter*, und ihm die ID 1 zugewiesen.

Der Fall von *Joe* im Daten-Set *Extern* ist ein wenig komplizierter. Handelt es sich um dieselbe Person wie bei *Joe* in *Kunden*? Dies lässt sich aus diesen beiden Datenquellen allein unmöglich sagen. Wir verfügen jedoch über eine dritte Quelle, *Kredite*, die einen *Joseph* enthält. Nun haben wir eine Übereinstimmung: Die Telefonnummer von *Joseph* stimmt mit der von *Joe* im Daten-Set *Extern* überein. Auf dieser Grundlage ermittelt das System, dass es sich bei allen um dieselbe Person handelt, und weist ihnen die ID 2 zu.

Es gibt nicht mehrere Datensätze für *Fred*; er erhält also ID 3. *Susan* aus *Kunden* wird als dieselbe Person identifiziert wie *Sue* aus *Kredite*, da sie dieselbe Adresse haben; ihr wird also ID 4 zugewiesen.

Hinweis: Dies ist ein Beispiel für einen optimistischen Abgleich zum Zwecke der Illustration. Sie könnten eine pessimistischere Regelmenge auswählen, sodass ein einfacher Name und eine Telefonnummer bzw. Adresse allein noch nicht ausreicht, um eine exakte Übereinstimmung festzustellen und beiden Datensätzen dieselbe ID zuzuweisen.

Schritt 6: Analyse der aufgelösten Identitäten

Nach der Auflösung der Identitätskonflikte im Repository können Sie nun weitere Analysen und Verarbeitungsschritte an den Ergebnissen durchführen. Wenn Sie beispielsweise beim Vorliegen doppelter Datensätze für dieselbe Identität mögliche betrügerische Aktivitäten vermuten, kann es sinnvoll sein, einen Bericht zu erstellen, in dem die Duplikate aufgelistet werden.

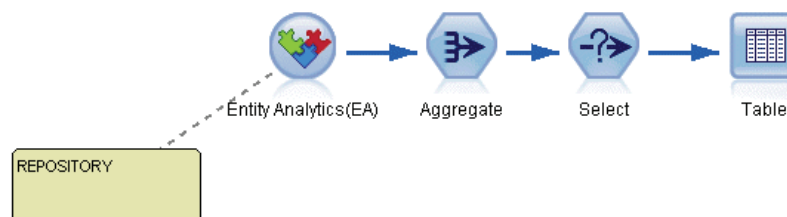
Dazu erstellen Sie zuerst einen Entity Analytics-(EA-)Quellenknoten und verknüpfen ihn mit dem Repository.

Die Ausgabe aus dem Knoten besteht aus den folgenden Feldern:

- Vom System hinzugefügtes ID-Feld (*\$EA-ID* im Beispiel für Schritt 5)
- Vom System hinzugefügtes Feld zur Quellenangabe (*\$EA-SRC* im Beispiel für Schritt 5)
- In Schritt 4 hinzugefügtes Feld mit eindeutigem Schlüssel

Um die Ausgabe in SPSS Modeler anzuzeigen, können Sie einen SPSS Modeler-Ausgabeknoten, beispielsweise einen Tabellenknoten oder einen Berichtsknoten, anfügen und diesen Teil des Streams ausführen. Um die Ausgabe zusammenzufassen, die sehr groß werden könnte, können Sie Knoten für Datensatzoperationen mit aufnehmen, beispielsweise Aggregatknoten und Auswahlknoten, wie in der folgenden Darstellung.

Abbildung 2-8
Anzeigen der Ausgabe



Der Entity Analytics-(EA-)Quellenknoten wird weiter unten ausführlich beschrieben. Für weitere Informationen siehe Thema [Analyse der aufgelösten Identitäten \(Entity Analytics-Quellenknoten\)](#) in Kapitel 3 auf S. 33.

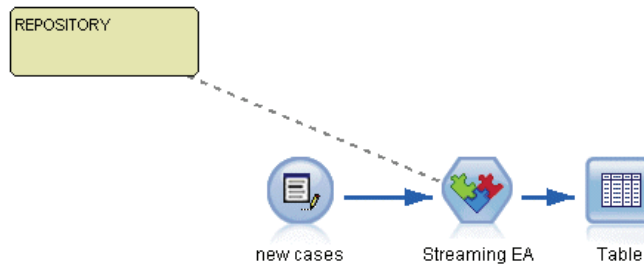
Schritt 7: Neue Fälle anhand des Repositorys auflösen

Sie haben nun also die Identitäten aller Datensätze in allen Datenquellen aufgelöst. Doch was geschieht, wenn Sie eine Menge an neuen Datensätzen vergleichen möchten, um zu ermitteln, in welchem Zusammenhang sie mit Ihren bisherigen Erkenntnissen stehen, und so das Scoring zu verbessern? Hier kommt der Knoten “Streaming von EA” ins Spiel.

Zunächst fügen Sie einen neuen SPSS Modeler-Datenquellenknoten hinzu, um Ihre neuen Daten in den Stream einzulesen. Anschließend verbinden Sie diesen Quellenknoten mit einem Knoten des Typs “Streaming von EA”. Zur Anzeige der Ausgabe fügen Sie wie zuvor einen Tabellenknoten hinzu, sodass dieser Teil des Streams wie folgt aussieht.

Abbildung 2-9

Auflösen neuer Fälle



Wenn Sie diesen Teil des Streams ausführen, liest der Knoten “Streaming von EA” neue Datensätze ein und vergleicht sie mit dem Inhalt des Repositorys. Wenn übereinstimmende Datensätze im Repository gefunden werden, gibt der Knoten “Streaming von EA” alle übereinstimmenden Datensätze zusammen mit dem neuen Datensatz aus und fügt die Felder für die ID und die Quellenangabe hinzu. Wenn keine Übereinstimmung gefunden wird, gibt der Prozessknoten nur den neuen Datensatz mit den hinzugefügten Feldern für die ID und die Quellenangabe aus.

Stellen Sie sich zur Veranschaulichung vor, dass das Repository derzeit aus den Inhalten besteht, die durch den Entity Analytics-(EA-)Quellenknoten ausgegeben wurden. Unter [Tabelle 2-1](#) auf S. 10.

Nun erhalten wir folgende neuen Datensätze. Beziehen Sie sich auf Personen, die wir bereits kennen?

Tabelle 2-2

Mit Scoring zu bewertende neue Datensätze

Name	Adresse	Telefon	Kredit
Suzan	1234 5th Street	555-1234	\$100,000
Mark	888 9th Ave	555-9999	\$60,000

Beim Vergleich der neuen Daten mit den bestehenden Repository-Inhalten findet der Knoten “Streaming von EA” eine Übereinstimmung zwischen dem ersten neuen Datensatz und der Person, die in den bestehenden Datensätzen die ID 4 trägt. Für den zweiten neuen Datensatz wird jedoch keine Übereinstimmung gefunden, weshalb ihr die neue, eindeutige ID 5 zugewiesen wird.

Der Knoten "Streaming von EA" fügt die Felder für die ID und die Quellenangabe hinzu und gibt die neuen Datensätze zusammen mit allen übereinstimmenden Datensätzen aus. Die Ausgabe sieht also wie folgt aus:

Tabelle 2-3
Ausgabe aus dem Knoten "Streaming von EA"

SEA-ID	SEA-SRC	Name	Telefon	Adresse	Kreditrisiko	Sparkonto	Kredit
4	Cust	Susan		1234 5th St		\$1234	
4	Kredit	Sue	555-1234	1234 5th St			\$10,000
4	Neuer Kredit	Suzan	555-1234	1234 5th Street			\$100,000
5	Neuer Kredit	Mark	555-9999	888 9th Ave			\$60,000

Diese Ausgabe kann dann mit einer Entitätsanalyse-ID aggregiert und zur weiteren Verarbeitung an andere Knoten weiter unten im Stream weitergegeben werden.

Der Knoten "Streaming von EA" wird weiter unten ausführlich beschrieben.

Schritt 8: Warnungen generieren

Auch hier können eventuell wieder potenziell verdächtige Aktivitäten aufgedeckt werden. In diesem Fall hat die Person mit ID 4 bereits einen Kredit über 10.000 Dollar aufgenommen und beantragt nun, unter einem leicht abweichenden Namen, einen weiteren Kredit in zehnfacher Höhe. Dies kann natürlich vollkommen in Ordnung sein und ohne betrügerische Absicht geschehen. Wenn eine solche Handlungsweise jedoch gemäß Ihren Geschäftsregeln als verdächtig gilt, kann es sinnvoll sein, einen genaueren Blick darauf zu werfen.

Sie könnten beispielsweise einen SPSS Modeler-Tabellenknoten oder -Berichtsknoten anfügen und ausführen, den Inhalt des zugehörigen Ausgabefensters ausdrucken und ihn von jemandem lesen und manuell Warnungen erstellen lassen. Alternativ könnten Sie die Ausgabe des Knotens "Streaming von EA" an ein Risikobewertungsmodell weiterleiten, das Sie zuvor in IBM® SPSS® Modeler erstellt haben, wodurch ein Score-Set ausgegeben wird, das Ihren Geschäftsregeln besser entspricht. Eine weitere Möglichkeit besteht darin, die Ausgabe des Entity Analytics-Prozessknotens zur weiteren Verarbeitung in eine Datenbank oder ein anderes Medium auszugeben. Mit SPSS Modeler steht Ihnen eine Vielzahl an Handlungsmöglichkeiten für Ihre speziellen Anforderungen zur Verfügung.

Entitätsanalyseaufgaben

Informationen zu den Aufgaben

In diesem Abschnitt werden die folgenden Aufgaben aus dem Bereich der Entitätsanalyse beschrieben.

- Einrichten eines Entitäts-Repositorys
- Konfigurieren eines Entitäts-Repository
- Analyse der aufgelösten Identitäten
- Auflösen neuer Fälle anhand des Entitäts-Repository
- Löschen des Inhalts eines Entitäts-Repository
- Löschen eines Entitäts-Repository
- Verwalten der Entitätsanalyse

Einrichten eines Entitäts-Repositorys (EA-Exportknotens)

Die Einrichtung eines Entitäts-Repositorys besteht aus folgenden Aufgaben:

1. Verbindung mit einer Datenquelle. Für weitere Informationen siehe Thema [Verbindung mit einer Datenquelle](#) auf S. 15.
2. Erstellen des Repositorys. Für weitere Informationen siehe Thema [Erstellen des Repository](#) auf S. 16.
3. Zuordnung der Eingabefelder in der Datenquelle zu Merkmalen im Repository. Für weitere Informationen siehe Thema [Erstellen einer Zuordnung zwischen Eingabefeldern und Merkmalen \(EA-Exportknoten\)](#) auf S. 19.

Nachdem Sie die Zuordnungen eingerichtet haben, können Sie sie entweder für die aktuelle Datenquelle anzeigen oder für alle Datenquellen, die dem Repository bekannt sind. Für weitere Informationen siehe Thema [Anzeigen der Feldzuordnungen \(EA-Exportknoten\)](#) auf S. 22.

Das Entitäts-Repository

Das Repository bietet einen zentralen Speicherbereich, der als Daten-Cache für alle Entitätsinformationen dient. Da es sich um ein Live-Repository handelt, hat es nur einen einzigen Zustand. Es gibt also bei einem Entitäts-Repository keine Versionserstellung und -verwaltung. Das Repository enthält den aktuellen Status sämtlicher Eingabedateien. Es kann somit sehr groß werden.

Sie können die Repository-Inhalte mithilfe einer benutzerfreundlichen grafischen Oberfläche verwalten. Für weitere Informationen siehe Thema [Konfigurieren eines Entitäts-Repository](#) auf S. 23.

Hinweis: Die im Lieferumfang von IBM® SPSS® Modeler Premium enthaltene Version von IBM® SPSS® Modeler Entity Analytics unterstützt ein einzelnes Repository mit dem IBM solidDB-Produkt als Host, das mit IBM SPSS Modeler Entity Analytics gebündelt ist. Bei dieser Version müssen Sie ein bestehendes Repository löschen, bevor Sie ein neues erstellen. Es ist ein separat lizenziertes Upgrade für IBM SPSS Modeler Entity Analytics (IBM SPSS Modeler Entity Analytics Unleashed) verfügbar, mit dem mehrere Repositories auf demselben System vorhanden sein können. Jedes Repository kann mehr als 10 Millionen Zeilen enthalten und mehr als vier Prozessorkerne verwenden. Details hierzu erhalten Sie bei Ihrem lokalen IBM Support-Mitarbeiter.

Verbindung mit einer Datenquelle

Sie beginnen, indem Sie Ihre Quelldaten mit einem Quellenknoten in SPSS Modeler einlesen.

So stellen Sie die Verbindung mit einer Datenquelle her

1. Doppelklicken Sie auf der Registerkarte “Quellen” auf der Knotenpalette unten im SPSS Modeler-Hauptfenster auf das Symbol, das dem Typ der Quelldaten entspricht. Dadurch wird ein Quellenknoten zum Bildschirm-Zeichenbereich hinzugefügt.
2. Doppelklicken Sie im Bildschirm-Zeichenbereich auf das Symbol, um das zugehörige Dialogfeld zu öffnen.
3. Geben Sie im Feld “Datei” den Standort und den Namen der Quelldatendatei ein.
4. Füllen Sie den Rest des Dialogfelds nach Bedarf aus (klicken Sie auf “Hilfe”, wenn Sie weitere Informationen benötigen) und klicken Sie dann auf “OK”.
5. Wenn die Quelldatendatei kein Feld enthält, mit dem die einzelnen Datensätze eindeutig identifiziert werden, fügen Sie mithilfe eines Ableitungsknotens ein solches Feld hinzu. Für weitere Informationen siehe Thema [Hinzufügen einer eindeutigen Datensatz-ID](#) auf S. 15.

Hinzufügen einer eindeutigen Datensatz-ID

Jede Datenquellendatei, die in das Entitäts-Repository eingegeben wird, benötigt ein Feld, das die einzelnen Datensätze eindeutig identifiziert. Falls eine Datenquelle kein derartiges Feld enthält, können Sie eines mithilfe eines SPSS Modeler-Ableitungsknotens hinzufügen.

So fügen Sie eine eindeutige Datensatz-ID zu einer Datenquellendatei hinzu:

1. Klicken Sie im Bildschirm-Zeichenbereich auf den Quellenknoten, den Sie in der vorangegangenen Aufgabe hinzugefügt haben.
2. Doppelklicken Sie auf der Registerkarte Feldoperationen der Knotenpalette auf das Symbol Ableiten, um einen Ableitungsknoten zum Quellenknoten hinzuzufügen.

3. Doppelklicken Sie im Bildschirm-Zeichenbereich auf den Ableitungsknoten, um das zugehörige Dialogfeld zu öffnen.
4. Ersetzen Sie im Feld Ableiten den Standardnamen durch einen aussagekräftigen Namen (z. B. ID) für das hinzuzufügende ID-Feld.
5. Vergewissern Sie sich, dass das Feld Ableitungstyp auf Formel gesetzt ist.
6. Setzen Sie Feldtyp auf Stetig.
7. Geben Sie im Textfeld Formel@INDEX ein und klicken Sie auf OK.

Erstellen des Repository

Sie müssen ein Repository erstellen, in dem alle Eingabedaten gespeichert werden.

So erstellen Sie ein Repository:

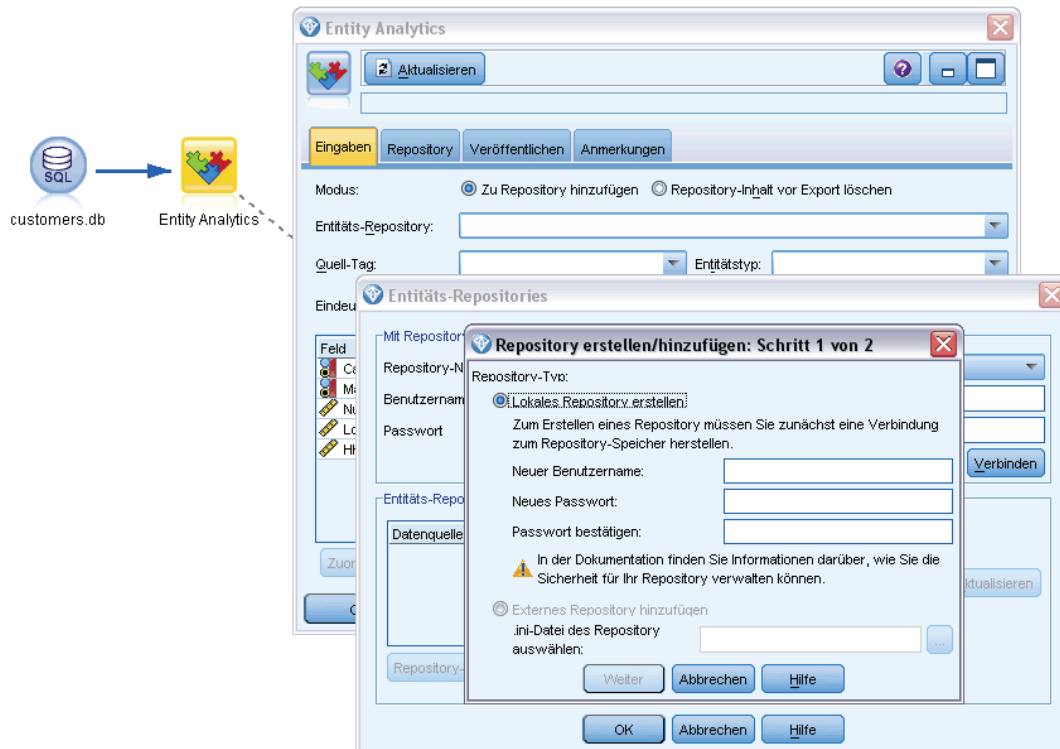
1. Platzieren Sie auf der Registerkarte "Export" der SPSS Modeler-Knotenpalette einen EA-Exportknoten im Stream-Zeichenbereich.

Hinweis: Wenn Sie zum ersten Mal ein Repository erstellen, verwenden Sie einen EA-Exportknoten und verbinden Sie ihn mit dem SPSS Modeler-Quellenknoten, der die Daten enthält, die in das Repository eingegeben werden sollen (bzw. mit dem Ableitungsknoten, wenn Sie einen hinzugefügt haben, um ein Feld mit einer eindeutigen ID zu erhalten). Gehen Sie zur Verbindung der Knoten wie folgt vor.

- Klicken Sie mit der rechten Maustaste auf den SPSS Modeler-Quellenknoten.
 - Wählen Sie die Option "Verbinden".
 - Klicken Sie auf den EA-Exportknoten.
2. Doppelklicken Sie dann auf den EA-Exportknoten, um das zugehörige Dialogfeld zu öffnen.
 3. Klicken Sie auf die Liste Entitäts-Repository.
 4. Klicken Sie auf <Durchsuchen...>, um das Dialogfeld "Entitäts-Repositories" anzuzeigen.
 5. Klicken Sie im Dialogfeld "Entitäts-Repositories" auf das Feld "Repository-Name".
 6. Wählen Sie die Option <Neues Repository erstellen...>, um den Assistenten zum Erstellen von Repositories anzuzeigen.

Assistent zum Erstellen von Repositories

Abbildung 3-1
Assistent zum Erstellen von Repositories



Schritt 1

Hier können Sie auswählen, ob ein lokales Repository mit dem IBM solidDB-Produkt erstellt werden soll, das mit IBM SPSS Modeler Entity Analytics gebündelt ist, oder ob eine externe Datenbank für das Repository verwendet werden soll.

Lokales Repository erstellen. Geben Sie den Benutzername und das Passwort des Administrators für die IBM solidDB-Datenbank an, die als Host für das zu erstellende Repository fungieren soll. Bestätigen Sie das Passwort und klicken Sie auf Weiter.

Hinweis: Wenn Sie anschließend die Administratorberechtigungen ändern müssen, verwenden Sie dazu den Befehlszeileneditor für die Datenbank. Für weitere Informationen siehe Thema [Verwalten der Administratorberechtigungen für die Repository-Datenbank](#) auf S. 43.

Externes Repository hinzufügen. Verwenden Sie diese Option, wenn Sie eine externe Datenbank als Host für das Repository verwenden möchten. Geben Sie den Speicherort der .ini-Datei der Datenbank in das Feld .ini-Datei des Repositories auswählen ein und klicken Sie auf Weiter.

Schritt 2

Neuer Repository-Name. Geben Sie einen eindeutigen Namen für das neue Repository ein.

Konfiguration importieren aus. (Nur lokales Repository) Wenn die Konfiguration auf der eines bestehenden Repositorys beruhen soll, wählen Sie hier das Repository aus. Falls nicht, wählen Sie Standard aus. Für weitere Informationen siehe Thema [Konfigurieren eines Entitäts-Repository](#) auf S. 23.

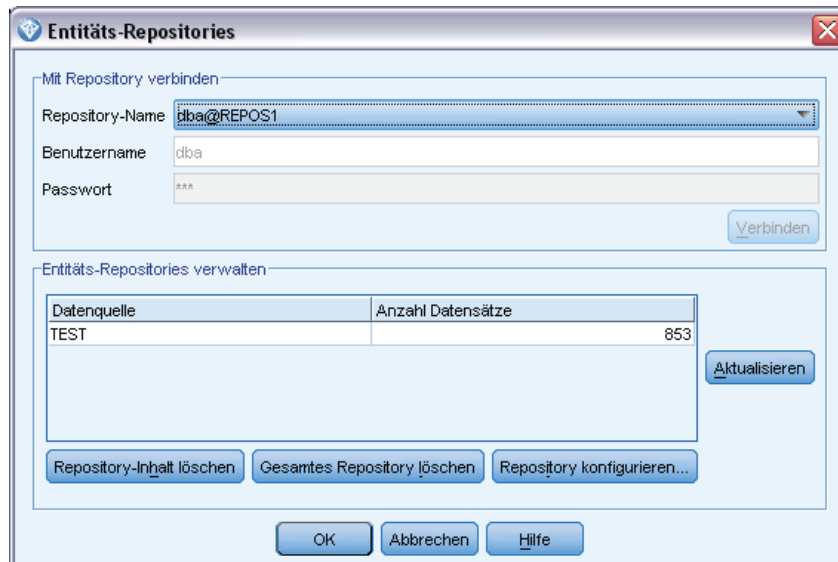
Wenn Sie ein bestehendes Repositorys auswählen, geben Sie die Verbindungsdetails ein, sofern diese von den im vorherigen Bildschirm eingegebenen abweichen.

Klicken Sie auf OK, um das neue Repository zu erstellen und das Dialogfeld “Entitätsauflösungsinstanzen” anzuzeigen, von dem aus Sie eine Verbindung zum Repository herstellen können.

Optionen für das Entitäts-Repository

Das Dialogfeld “Entitäts-Repositories” enthält eine Reihe von Optionen, mit denen Sie ein Entitäts-Repository erstellen, eine Verbindung damit herstellen und es konfigurieren und verwalten können.

Abbildung 3-2
Optionen für das Entitäts-Repository



Mit Repository verbinden. Mit diesen Optionen können Sie ein neues Entitäts-Repository erstellen bzw. eine Verbindung mit einem bestehenden herstellen.

- **Repository-Name.** Zeigt das aktuelle Entitäts-Repository an, sofern eines vorhanden ist. Wenn mehrere Repositories vorhanden sind, können Sie in der Liste ein anderes Repository auswählen.

Um ein neues Repository zu erstellen, wählen Sie die Option <Neues Repository erstellen...>. Dadurch wird ein Assistent gestartet, der Sie durch den Erstellungsvorgang führt.

- **Benutzername.** Geben Sie Ihren Benutzernamen für das ausgewählte Repository ein.
- **Passwort.** Das Passwort für diesen Benutzernamen.
- **Verbinden.** Klicken Sie auf diese Option, um eine Verbindung mit dem aktuellen Repository herzustellen.

Entitäts-Repositories verwalten. In der Tabelle werden die Datenquellen aufgelistet, die in das aktuelle Repository (dasjenige, mit dem Sie verbunden sind) geladen wurden. Dabei wird auch jeweils die Anzahl der Datensätze in den einzelnen Datenquellen angezeigt.

- **Aktualisieren.** Aktualisiert die Tabellendaten zu Datenquellen und Datengröße, beispielsweise, wenn Sie eine neue Datenquelle hinzugefügt oder die Größe einer bestehenden Datenquelle geändert haben.
- **Repository-Inhalt löschen.** Entfernt alle Quelldaten aus dem Repository, behält jedoch alle Konfigurationsdetails bei. Sie können diese Option verwenden, wenn Sie die Konfigurationsinformationen noch brauchen können, jedoch alle Datensätze aus dem Repository entfernen möchten. Für weitere Informationen siehe Thema [Löschen des Inhalts eines Entitäts-Repository](#) auf S. 45.

Gesamtes Repository löschen. Löscht die aktuellen Repository-Inhalte sowie die Konfigurationsdetails vollständig. Für weitere Informationen siehe Thema [Löschen eines Entitäts-Repository](#) auf S. 46.

Repository konfigurieren. Zeigt ein Fenster an, in dem Sie das aktuelle Repository konfigurieren können. Für weitere Informationen siehe Thema [Konfigurieren eines Entitäts-Repository](#) auf S. 23.

Erstellen einer Zuordnung zwischen Eingabefeldern und Merkmalen (EA-Exportknoten)

Das Repository bietet standardmäßig eine Reihe vordefinierter Merkmale. Verschiedene Datenquellen können für Informationstypen, die sich auf dasselbe Merkmal beziehen, unterschiedliche Feldnamen verwenden (z. B. Adresse1 oder Adresszeile 1). Um Dopplungen zu vermeiden, müssen die Felder der Eingabedatenquellen bestimmten Repository-Merkmalen zugeordnet werden. Sie brauchen nicht jedes Feld im Daten-Set zuzuordnen, nur diejenigen, die vermutlich demselben Merkmal in anderen Daten-Sets entsprechen.

Wenn eine Datenquelle Felder verwendet, die anderen Datentypen entsprechen, die nicht im Repository vordefiniert sind, können Sie über das Fenster "Repository-Konfiguration" neue Merkmale erstellen. Für weitere Informationen siehe Thema [Konfigurieren eines Entitäts-Repository](#) auf S. 23.

So erstellen Sie eine Zuordnung zwischen Eingabefeldern und Merkmalen:

1. Verbinden Sie einen EA-Exportknoten mit einem Datenquellenknoten im Stream-Zeichenbereich. Jeder verwendete Datenquellenknoten muss mit seinem eigenen EA-Exportknoten verbunden sein.

2. Öffnen Sie den EA-Exportknoten, um die Registerkarte “Eingaben” anzuzeigen, die Optionen für die Zuordnung der Eingabefelder enthält. Für weitere Informationen siehe Thema [Repository-Eingabeoptionen für die Zuordnung](#) auf S. 20.
3. Wählen Sie im EA-Exportknoten die Registerkarte “Repository” aus, um entweder die Zuordnungszuweisungen für die aktuelle Datenquelle oder für alle Datenquellen anzuzeigen, wenn Sie mehrere verwenden.
4. Um eine Gruppe von Zuordnungszuweisungen zu speichern (z. B. zur Verwendung mit einem anderen Datenquellenknoten), klicken Sie auf Zuordnung exportieren.

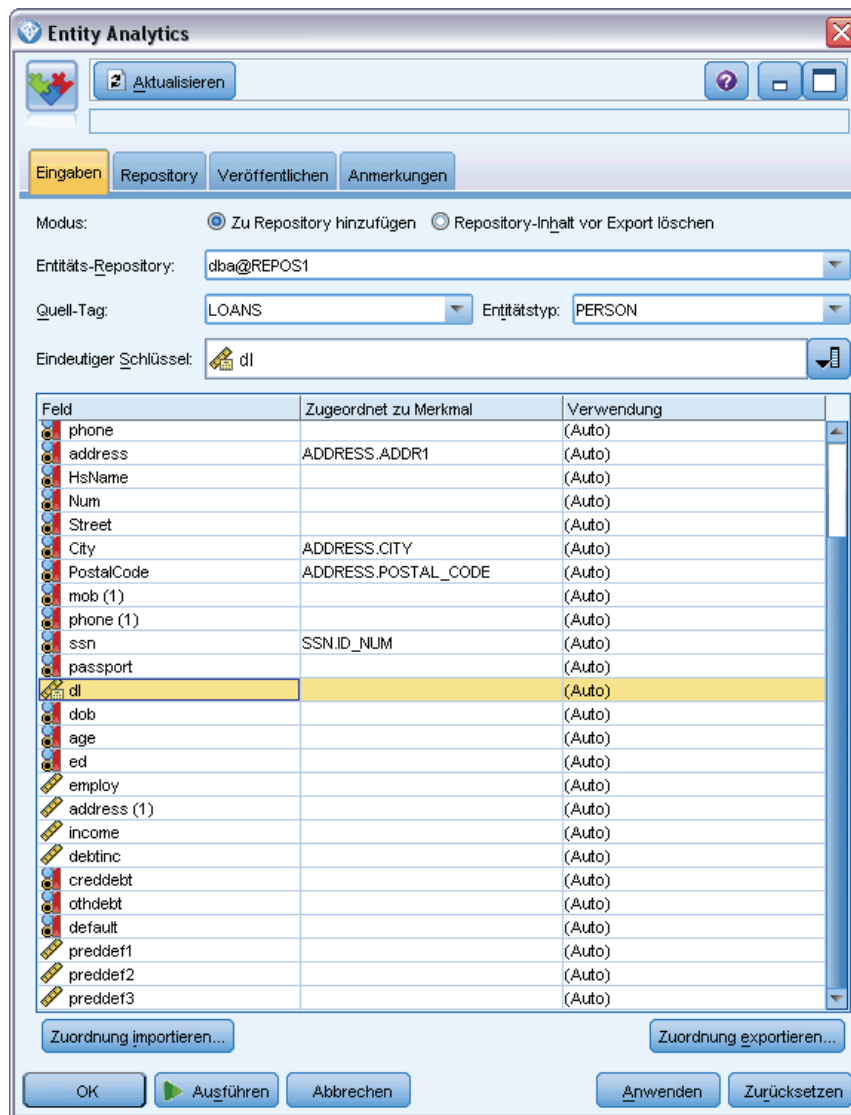
Wenn Sie mit der Zuordnung des ersten Datenquellenknotens fertig sind, wiederholen Sie den Vorgang für alle anderen Datenquellenknoten, die verwendet werden sollen.

Repository-Eingabeoptionen für die Zuordnung

Die Registerkarte “Eingaben” enthält die Optionen zur Zuordnung von Datenquellenfeldern zu Repository-Merkmalen, die für den Export in das Repository bereit sind. Richten Sie die Zuordnungszuweisungen auf dieser Registerkarte ein. Klicken Sie optional auf die Registerkarte “Repository”, um die Zuordnung für andere Datenquellen anzuzeigen, und klicken Sie dann auf Ausführen, um die Daten in das Repository zu exportieren.

Wenn Sie bereits eine Gruppe von Zuordnungen in einer XML-Datei gespeichert haben, können Sie diese durch Klicken auf Zuordnung importieren verwenden.

Abbildung 3-3
Erstellen einer Zuordnung zwischen Eingabefeldern und Merkmalen



Modus. Behalten Sie die Standardauswahl Zu Repository hinzufügen bei, wenn Sie die Datensätze der Quellendatei zu den bestehenden Inhalten des Repositorys hinzufügen möchten. Wenn Sie die Repository-Inhalte löschen, jedoch die Konfigurationsdateien beibehalten möchten, bevor Sie die Quelldatensätze hinzufügen, wählen Sie die Datei Repository-Inhalt vor Export löschen aus.

Entitäts-Repository. Zeigt das aktuelle Entitäts-Repository an, sofern eines vorhanden ist. Wenn mehrere Repositories vorhanden sind, können Sie in der Liste ein anderes Repository auswählen. Wählen Sie zur Erstellung eines neuen Repository die Option <Durchsuchen...>, um ein Dialogfeld anzuzeigen, in dem Sie das Repository erstellen können. Für weitere Informationen siehe Thema [Optionen für das Entitäts-Repository](#) auf S. 18.

Quell-Tag. Eine Liste von Tags, die Datenquellen angeben, die dem Repository derzeit bekannt sind. Wählen Sie ein Tag aus der Liste aus oder erstellen Sie mit <Neues Quell-Tag hinzufügen...> ein Tag für eine neue Datenquelle.

Entitätstyp. Eine Liste der im Repository definierten Entitätstypen (d. h. Merkmals-Sets). Wählen Sie einen aus der Liste aus oder zeigen Sie mithilfe der Option <Neuen Entitätstyp hinzufügen...> das Fenster Repository-Konfiguration an, in dem Sie einen neuen Entitätstyp definieren können. Für weitere Informationen siehe Thema [Konfigurieren eines Entitäts-Repository](#) auf S. 23.

Eindeutiger Schlüssel. (erforderlich) Das für die eindeutigen IDs für die Datensätze zu verwendende Eingabefeld.

Zuordnungstabelle. In dieser Tabelle können Sie jedes Eingabefeld einem entsprechenden Merkmal im Repository zuordnen. Wenn im ausgewählten Entitätstyp kein geeignetes Merkmal vorhanden ist, können Sie hier ein neues Merkmal erstellen.

- **Feld.** Die Menge an Eingabefeldern in der ausgewählten Datenquelle. Zu jedem Feld gehört ein Symbol, das das Messniveau (also den Datentyp) für das Feld angibt.
- **Zugeordnet zu Merkmal.** Um ein Feld einem Merkmal zuzuordnen, doppelklicken Sie auf diese Spalte (oder drücken Sie die Leertaste) in der Zeile des betreffenden Felds und wählen Sie ein Merkmal aus der Liste aus. Wenn kein geeignetes Merkmal verfügbar ist, können Sie mit der Option <Neues Merkmal hinzufügen...> das Fenster Repository-Konfiguration anzeigen, in dem Sie ein neues Merkmal für diesen Entitätstyp definieren können. Für weitere Informationen siehe Thema [Konfigurieren eines Entitäts-Repository](#) auf S. 23.
- **Verwendung.** Gibt den Kontext eines bestimmten Felds an, wenn mehrere Kontexte möglich sind, beispielsweise die private und die dienstliche Telefonnummer. Für die Merkmale “Adresse” und “Telefon” sind voreingestellte Verwendungstypen verfügbar. Des Weiteren können Sie für alle Merkmale Ihre eigenen Verwendungstypen erstellen. Zur Festlegung eines anderen Verwendungstyps als dem standardmäßig eingestellten Typ (Auto) klicken Sie in dieser Spalte auf die gewünschte Zeile und wählen Sie entweder einen der bestehenden Verwendungstypen aus (falls vorhanden) oder klicken Sie auf die Option Verwendung hinzufügen..., um einen neuen Verwendungstyp zu erstellen. Für weitere Informationen siehe Thema [Verwaltung der Elementtypen](#) auf S. 30.

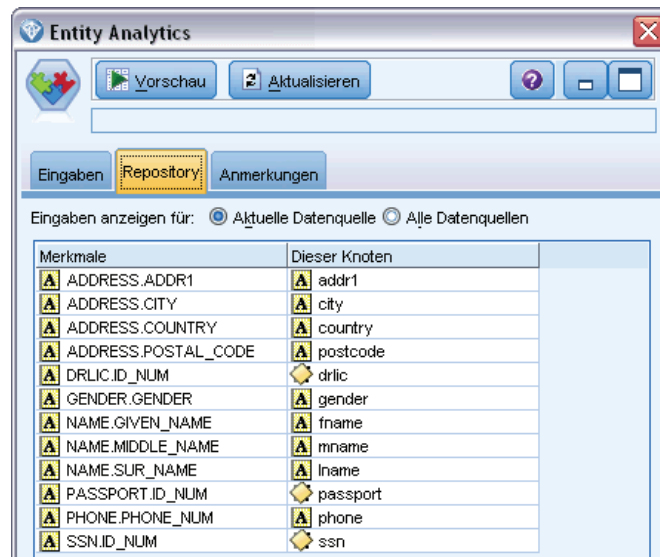
Zuordnung importieren. Importiert eine zuvor exportierte Menge an Zuordnungen zwischen Feldern und Merkmalen aus einer externen XML-Datei. Diese kann nützlich sein, wenn Sie verschiedene Datenquellen mit denselben Zuordnungsanforderungen verwenden, da Sie dadurch nicht dieselben Zuordnungen für die verschiedenen Quellen erneut zu definieren brauchen.

Zuordnung exportieren. Exportiert die Menge der in der Zuordnungstabelle angegebenen Zuordnungen zwischen Feldern und Merkmalen in eine externe XML-Datei.

Anzeigen der Feldzuordnungen (EA-Exportknoten)

Klicken Sie auf der Registerkarte “Repository” auf die Schaltfläche Aktualisieren, um anzuzeigen, welchen Repository-Merkmalen Eingabefelder zugeordnet sind. Sie können dies entweder für die aktuelle Datenquelle (die Datenquelle, die durch den mit diesem Exportknoten verknüpften Quellenknoten gesteuert wird) oder für alle Datenquellen anzeigen.

Abbildung 3-4
Anzeigen der Feldzuordnungen



Eingaben anzeigen für. Wählen Sie eine Option, um die Zuordnungen entweder für die aktuelle Datenquelle anzuzeigen oder für alle Datenquellen, die dem Repository bekannt sind.

Aktualisieren. Aktualisiert die Anzeige für die ausgewählte Eingabeoption.

Strukturen. Eine Liste aller Merkmale, die über Zuordnungen in den angezeigten Datenquellen verfügen. Merkmale ohne Zuordnung werden nicht angezeigt.

<Datenquelle>. In den einzelnen Spalten sind zugeordnete Felder in einer bestimmten Datenquelle für jedes Merkmal aufgeführt, für das eine Zuordnung definiert wurde.

Konfigurieren eines Entitäts-Repository

Sie können die Repository-Inhalte über das Fenster "Repository-Konfiguration" verwalten. Dieses bietet eine benutzerfreundliche visuelle Benutzeroberfläche für das gesamte Repository.

Wenn Sie vorhaben, mehrere Repositories mit denselben oder ähnlichen Konfigurationen zu verwenden, können Sie eine Grundkonfiguration einrichten und in eine Datei exportieren, die Sie später in andere Repositories importieren können. Für weitere Informationen siehe Thema [Wiederverwenden von Repository-Konfigurationen](#) auf S. 33.

Vorsicht: Wenn Sie die Konfiguration eines Repositories ändern und speichern, das bereits Daten enthält, wird dringend empfohlen, die Repository-Inhalte zu löschen und die Daten erneut zu laden. Dadurch wird ein inkonsistenter Zustand des Repositories vermieden.

So richten Sie eine Repository-Konfiguration ein:

1. Öffnen Sie einen beliebigen Entity Analytics-Knoten.
2. Klicken Sie auf die Liste Entitäts-Repository.

3. Klicken Sie auf <Durchsuchen...>, um das Dialogfeld “Entitätsauflösungsinstanzen” anzuzeigen.
4. Klicken Sie im Dialogfeld “Entitätsauflösungsinstanzen” auf die Liste Repository-Name.
5. Wählen Sie das Repository aus, für das Sie die Konfiguration einrichten möchten.
6. Wenn noch keine Verbindung besteht, geben Sie den Administrator-Benutzernamen und das zugehörige Passwort ein und klicken Sie auf Verbinden.
7. Wenn die Schaltfläche Repository konfigurieren aktiviert ist, klicken Sie darauf, um das Fenster “Repository-Konfiguration” anzuzeigen.
8. Erstellen Sie die Konfigurationsdetails, wie in den folgenden Abschnitten beschrieben.

Der Navigationsbereich auf der linken Seite des Fensters “Repository-Konfiguration” enthält eine Baumstruktur, über die Sie die verschiedenen Eigenschaften des Repositorys verwalten können.

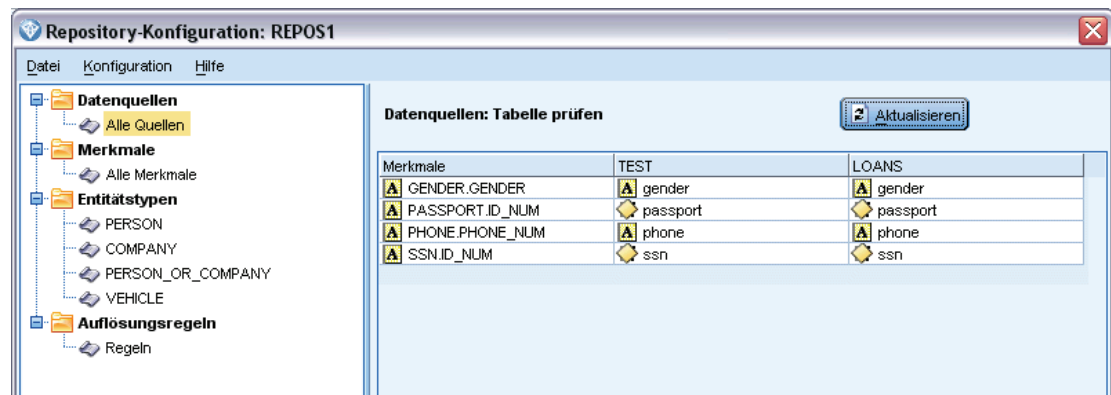
Tabelle 3-1
Hauptelemente des Fensters “Repository-Konfiguration”

Abschnitt	Beschreibung	
Datenquellen	Zeigt die Zuordnungen aller Datenquellen zu den verschiedenen Repository-Merkmalen an.	Für weitere Informationen siehe Thema Anzeigen der Datenquellenzuordnungen auf S. 24.
Merkmale	Dient zum Erstellen eines neuen Merkmals bzw. zum Duplizieren, Bearbeiten oder Löschen eines bestehenden Merkmals.	Für weitere Informationen siehe Thema Verwaltung der Repository-Merkmale auf S. 25.
Entitätstypen	Dient zum Erstellen eines neuen Merkmals bzw. zum Verwalten bestehender Merkmale (Duplizieren, Umbenennen, Anfügen oder Entfernen von Merkmalen, Löschen).	Für weitere Informationen siehe Thema Verwaltung der Elementtypen auf S. 30.
Auflösungsregeln	Dient zur Festlegung des Schwellenwerts für den Entitätsabgleich.	Für weitere Informationen siehe Thema Festlegen des Schwellenwerts für den Entitätsabgleich auf S. 32.

Anzeigen der Datenquellenzuordnungen

Im Abschnitt “Datenquellen” des Fensters “Repository-Konfiguration” bietet der Eintrag “Alle Quellen” eine schreibgeschützte Anzeige der Zuordnungen aller Datenquellen zu den verschiedenen Repository-Merkmalen.

Abbildung 3-5
Datenquellenzuordnungen im Konfigurationsfenster



Klicken Sie auf Aktualisieren, um die Liste zu aktualisieren, wenn neue Datenquellen in das Repository aufgenommen wurden.

Hinweis: Hier kann keine Datenquelle zum Repository hinzugefügt werden. Die einzige Möglichkeit zum Hinzufügen von Datenquellen besteht darin, einen SPSS Modeler-Quellenknoten zu erstellen und ihn mit einem Entity Analytics-Exportknoten zu verbinden. Für weitere Informationen siehe Thema [Verbindung mit einer Datenquelle](#) auf S. 15.

Verwaltung der Repository-Merkmale

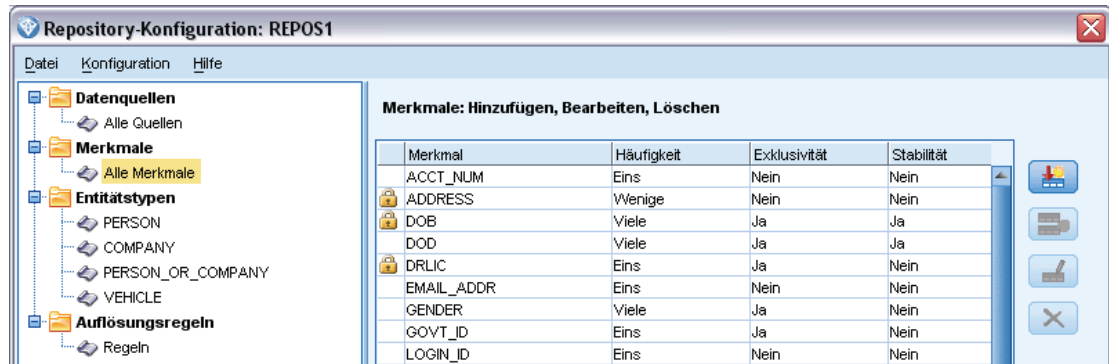
Ein Repository-Merkmal ist ein einzelner Informationstyp, der zusammen mit einer Entitätsdatenquelle verwendet werden kann. Einige Merkmale (z. B. Vorname, Nachname, Geburtsdatum usw.) können mit vielen verschiedenen Datenquellen verwendet werden, während andere Merkmale nur bei einer bestimmten Datenquelle vorkommen. Ein Merkmal kann ein oder mehrere Elemente enthalten; jedes Element entspricht üblicherweise einem Feld in einem Datensatz oder einer Spalte in einer Datenbanktabelle.

Im Abschnitt “Merkmale” des Fensters “Repository-Konfiguration” bietet der Eintrag “Alle Merkmale” die Möglichkeit zur Verwaltung aller Repository-Merkmale. Sie haben folgende Möglichkeiten:

- Neues Merkmal erstellen
- Bestehendes Merkmal duplizieren (beispielsweise, um ein neues Merkmal auf der Grundlage eines bestehenden Merkmals zu erstellen)
- Bestehendes Merkmal bearbeiten
- Bestehendes Merkmal löschen

Anweisungen für diese Aufgaben erhalten Sie später in diesem Abschnitt.

Abbildung 3-6
Im Konfigurationsfenster aufgeführte Merkmale



Die Merkmalsliste zeigt alle Merkmale, die in diesem Repository definiert wurden. In den Spalten in der Liste werden die verschiedenen Eigenschaften angezeigt, die ein Merkmal aufweisen kann.

Merkmal. Der Name des Merkmals. Ein Schlosssymbol neben dem Namen eines Merkmals zeigt an, dass das Merkmal gesperrt ist. Gesperrte Merkmale können nicht gelöscht oder dupliziert werden. Auch ist es nicht möglich, Änderungen daran zu speichern.

Häufigkeit. Gibt an, wie viele Identitäten denselben Wert für dieses Merkmal aufweisen können. Gültige Werte sind Keine, Name, Eins (z. B. für eine Passnummer), Wenige (z. B. für eine Adresse) und Viele (z. B. für ein Geburtsdatum).

Exklusivität. Gibt an, dass eine Entität diesen Merkmalstyp typischerweise nur einmal aufweisen sollte. So hätten hier z. B. die Merkmale "Geburtsdatum" oder "Steuernummer" den Wert Ja, während die Merkmale "Adresse" oder "Kreditkartennummer" den Wert Nein hätten, da eine Entität mehrere Adressen oder Kreditkartennummern haben kann.

Stabilität. Gibt den Stabilitätswert des betreffenden Merkmals an (d. h. ob es *unwahrscheinlich* ist, dass es sich während der Lebensdauer einer Entität ändert). Das Merkmal "Geburtsdatum" beispielsweise hätte den Wert Ja, da es sich niemals ändert, das Merkmal "Adresse" dagegen hätte den Wert Nein, da hierfür die Wahrscheinlichkeit einer Änderung recht hoch und das Merkmal somit weniger stabil ist. *Hinweis:* Das Merkmal "Geschlecht" ist in der Regel lebenslang stabil. Da es aufgrund von fehlerhaften Daten aber häufig falsch angegeben wird, erhält das Merkmal bei der Standardkonfiguration den Wert Nein.

So erstellen Sie ein neues Merkmal:

- Führen Sie eine der folgenden Aktionen aus.
 - Klicken Sie auf die Schaltfläche "Neues Merkmal erstellen" (die oberste Schaltfläche rechts im Bildschirm).
 - Klicken Sie mit der rechten Maustaste im Navigationsbereich links im Bildschirm auf Alle Merkmale und wählen Sie die Option Neues Merkmal aus.
- Arbeiten Sie das Dialogfeld "Merkmal hinzufügen/bearbeiten" ab. Für weitere Informationen siehe Thema [Hinzufügen bzw. Bearbeiten von Merkmalen](#) auf S. 28.

So duplizieren Sie ein bestehendes Merkmal:

1. Wählen Sie in der Spalte Merkmal auf der rechten Bildschirmseite das zu duplizierende Merkmal aus.
2. Klicken Sie auf die Schaltfläche “Ausgewähltes Merkmal duplizieren” (die zweite Schaltfläche rechts im Bildschirm).
3. Arbeiten Sie das Dialogfeld “Merkmal hinzufügen/bearbeiten” ab. Für weitere Informationen siehe Thema [Hinzufügen bzw. Bearbeiten von Merkmalen](#) auf S. 28.

So bearbeiten Sie ein bestehendes Merkmal:

Vorsicht: Wenn Sie ein Merkmal oder ein Element eines Merkmals bearbeiten oder löschen und das Repository bereits Daten enthält, sollten Sie den Repository-Inhalt anschließend löschen und die Daten neu laden. Dadurch wird ein inkonsistenter Zustand des Repositorys vermieden.

1. Wählen Sie in der Spalte Merkmal auf der rechten Bildschirmseite das zu bearbeitende Merkmal aus. *Hinweis:* Sie können nur Merkmale auswählen, die Sie selbst erstellt haben, keine vom System bereitgestellten Merkmale.
2. Klicken Sie auf die Schaltfläche “Ausgewähltes Merkmal bearbeiten” (die dritte Schaltfläche rechts im Bildschirm).
3. Arbeiten Sie das Dialogfeld “Merkmal hinzufügen/bearbeiten” ab. Für weitere Informationen siehe Thema [Hinzufügen bzw. Bearbeiten von Merkmalen](#) auf S. 28.

So löschen Sie ein bestehendes Merkmal:

Vorsicht: Wenn Sie ein Merkmal oder ein Element eines Merkmals bearbeiten oder löschen und das Repository bereits Daten enthält, sollten Sie den Repository-Inhalt anschließend löschen und die Daten neu laden. Dadurch wird ein inkonsistenter Zustand des Repositorys vermieden.

1. Wählen Sie in der Spalte Merkmal auf der rechten Bildschirmseite das zu löschende Merkmal aus. *Hinweis:* Sie können nur Merkmale löschen, die Sie selbst erstellt haben, keine vom System bereitgestellten Merkmale.
2. Führen Sie eine der folgenden Aktionen aus.
 - Klicken Sie auf die Schaltfläche “Ausgewähltes Merkmal löschen” (die unterste Schaltfläche rechts im Bildschirm).
 - Klicken Sie mit der rechten Maustaste im Navigationsbereich links im Bildschirm auf Alle Merkmale und wählen Sie die Option Löschen aus.
3. Klicken Sie auf Weiter, um das Löschen des Merkmals zu bestätigen. *Hinweis:* Das Löschen von Merkmalen kann nicht rückgängig gemacht werden.

Hinzufügen bzw. Bearbeiten von Merkmalen

Vorsicht: Wenn Sie ein Merkmal oder ein Element eines Merkmals bearbeiten oder löschen und das Repository bereits Daten enthält, sollten Sie den Repository-Inhalt anschließend löschen und die Daten neu laden. Dadurch wird ein inkonsistenter Zustand des Repositories vermieden.

Im Dialogfeld “Merkmal hinzufügen/bearbeiten” können Sie ein neues Repository-Merkmal erstellen oder ein bestehendes Merkmal duplizieren bzw. bearbeiten.

Abbildung 3-7
Bearbeiten von Merkmalen

	Element	Beschreib...	Datentyp	Suchen	Vergleichen	Deskriptiv
1	ADDR_TYPE	Address ty...	Zeichenkette	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
2	FULL_ADDR	Full address	Zeichenkette	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	ADDR1	Addr line 1	Zeichenkette	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
4	ADDR2	Addr line 2	Zeichenkette	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
5	ADDR3	Addr line 3	Zeichenkette	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
6	CITY	City	Zeichenkette	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
7	STATE	State	Zeichenkette	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
8	POSTAL_C...	Postal code	Zeichenkette	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
9	COUNTRY	Country	Zeichenkette	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
10	STR_NUM	Street num...	Zeichenkette	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
11	STR_NAME	Street name	Zeichenkette	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
12	UNIT_TYPE	Unit type	Zeichenkette	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
13	UNIT_NUM	Unit number	Zeichenkette	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Merkmalstyp. Eine Beschriftung, die den Typ der Informationen angibt, auf die sich das Merkmal bezieht. Diese Beschriftung bildet den ersten Teil der Merkmals-ID.

Beschreibung. Eine kurze Textbeschreibung des Merkmalstyps, nur zu Informationszwecken.

Häufigkeit. Gibt an, wie viele Identitäten denselben Wert für dieses Merkmal aufweisen können. Gültige Werte sind Keine, Name, Eins (z. B. für eine Passnummer), Wenige (z. B. für eine Adresse) und Viele (z. B. für ein Geburtsdatum).

Exklusivität. Gibt an, dass eine Entität diesen Merkmalstyp typischerweise nur einmal aufweisen sollte. So hätten hier z. B. die Merkmale “Geburtsdatum” oder “Steuernummer” den Wert Ja, während die Merkmale “Adresse” oder “Kreditkartennummer” den Wert Nein hätten, da eine Entität mehrere Adressen oder Kreditkartennummern haben kann.

Stabilität. Gibt den Stabilitätswert des betreffenden Merkmals an (d. h. ob es *unwahrscheinlich* ist, dass es sich während der Lebensdauer einer Entität ändert). Das Merkmal “Geburtsdatum” beispielsweise hätte den Wert Ja, da es sich niemals ändert, das Merkmal “Adresse” dagegen hätte den Wert Nein, da hierfür die Wahrscheinlichkeit einer Änderung recht hoch und das Merkmal somit weniger stabil ist. *Hinweis:* Das Merkmal “Geschlecht” ist in der Regel lebenslang stabil. Da es aufgrund von fehlerhaften Daten aber häufig falsch angegeben wird, erhält das Merkmal bei der Standardkonfiguration den Wert Nein.

Historie persistieren. Wenn diese Option auf “Ja” (Standardvorgabe) gesetzt ist, wird der Änderungsverlauf des Feldwerts beibehalten. Dies ist beispielsweise bei einem Adressfeld für einen Kunden nützlich. In diesem Fall kann das Beibehalten des Änderungsverlaufs den Abgleich verbessern, wenn ein Kunde mehrmals die Adresse gewechselt hat. Bei Feldern wie dem Kontostand dagegen sollte diese Option auf “Nein” gesetzt werden, um zu vermeiden, dass übermäßig große Datenmengen beibehalten werden.

Elementetabelle. Eine Liste der Elemente, aus denen dieses Merkmal besteht.

- **Element.** Der Name des Elements.
- **Beschreibung.** Eine kurze Beschreibung dessen, was das Element angibt.
- **Datentyp.** Typ der Daten, die für dieses Element verwendet werden können.
- **Suchen.** Wenn diese Option ausgewählt ist, kann nach diesem Element des Merkmals alleine oder in Verbindung mit anderen Elementen der Merkmalsbeschriftungssuche gesucht werden. Nehmen Sie beispielsweise an, das Merkmal PASSPORT habe zwei Elemente, ID_NUM und COUNTRY. Wenn es möglich sein soll, die Suche für ID_NUM alleine durchzuführen, aktivieren Sie das zugehörige Kontrollkästchen Suchen und deaktivieren Sie das Kontrollkästchen Suchen für COUNTRY. Alle Modifizierfaktoren für das Merkmal, wie Land oder Bundesland bzw. Ausgabe- und Ablaufdatum, sollten nicht Teil der Suche sein.
- **Vergleichen.** Gibt an, welche Elemente des Merkmals zur Entitätsauflösung an die Vergleichsroutine gesendet werden. In dem Beispiel mit dem Merkmal PASSPORT, das die beiden Elemente ID_NUM und COUNTRY aufweist, sollte nur ID_NUM an die Vergleichsroutine gesendet werden, da COUNTRY möglicherweise nicht immer angegeben ist.
- **Deskriptiv.** Gibt an, welche Elemente des Merkmals in der internen Beschreibung vorkommen sollen. Wenn im betrachteten Beispiel mit dem Merkmal PASSPORT und den beiden Elementen ID_NUM und COUNTRY beide Elemente in der Beschreibung des Merkmals vorkommen sollen, muss das Feld Deskriptiv für beide ausgewählt werden. Nehmen wir jedoch an, dass das Merkmal noch ein weiteres Element, ISSUE_DT, aufweist. Das Kontrollkästchen Deskriptiv für dieses Element müsste gelöscht werden, um sicherzustellen, dass es nicht in der Merkmalsbeschreibung vorkommt.

Schaltfläche “Neues Element hinzufügen”. Fügt eine neue Zeile zur Elementtabelle hinzu, sodass Sie ein neues Element definieren können.

Schaltfläche “Element löschen”. Löscht eine ausgewählte Zeile aus der Elementetabelle. Dieser Vorgang kann nicht rückgängig gemacht werden.

Vorsicht: Wenn Sie ein Merkmal oder ein Element eines Merkmals bearbeiten oder löschen und das Repository bereits Daten enthält, sollten Sie den Repository-Inhalt anschließend löschen und die Daten neu laden. Dadurch wird ein inkonsistenter Zustand des Repositorys vermieden.

Verwaltung der Elementtypen

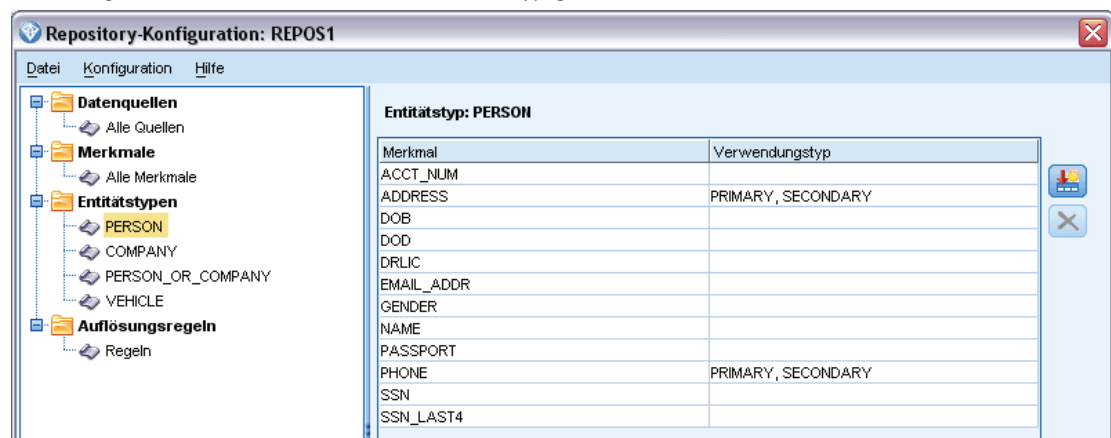
Ein **Entitätstyp** ist eine benannte Menge an Repository-Merkmalen, die logisch zusammengehören. Beispielsweise könnte ein Entitätstyp, der für die Verwendung mit einem Kunden-Daten-Set vorgesehen ist, aus Merkmalen wie “Name”, “Geburtsdatum”, “Geschlecht”, “Adresse”, “Telefonnummer” usw. bestehen.

Das IBM SPSS Modeler Entity Analytics-Repository wird mit einer Standardauswahl an Entitätstypen bereitgestellt und Sie können Ihre eigenen hinzufügen.

Im Abschnitt “Entitätstypen” des Fensters “Repository-Konfiguration” werden die verschiedenen Entitätstypen aufgeführt, die erstellt wurden. Sie haben folgende Möglichkeiten:

- Neuen Entitätstyp erstellen
- Bestehenden Entitätstyp duplizieren (beispielsweise, um einen neuen Entitätstyp auf der Grundlage eines bestehenden Entitätstyps zu erstellen)
- Merkmale zu einem Entitätstyp hinzufügen
- Merkmale aus einem Entitätstyp entfernen
- Entitätstyp umbenennen
- Entitätstyp löschen

Abbildung 3-8
Auflistung der Merkmale, die zu einem Entitätstyp gehören.



Entitätstyp. Der Name des ausgewählten Entitätstyps.

Merkmal. Eine Liste gültiger Merkmale, aus denen dieser Entitätstyp besteht.

Verwendungstyp. (Optional) Gibt verschiedene Kontexte an, in denen dieses Merkmal verwendet werden könnte. Doppelklicken Sie auf diese Spalte, um einen Verwendungstyp hinzuzufügen bzw. zu bearbeiten. Trennen Sie dabei die einzelnen Verwendungstypen durch Komma und Leerzeichen.

voneinander. Die Werte, die Sie hier angeben, definieren die Werte, die im EA-Export-Knoten bzw. im Knoten "Streaming von EA" angezeigt werden, wenn ein Benutzer auf der Registerkarte "Eingaben" auf die Spalte "Verwendung" für ein Merkmal klickt. Für weitere Informationen siehe Thema [Repository-Eingabeoptionen für die Zuordnung](#) auf S. 20.

So erstellen Sie einen neuen Entitätstyp:

1. Klicken Sie mit der rechten Maustaste im Navigationsbereich links im Bildschirm auf Entitätstypen.
2. Wählen Sie die Option Neuer Entitätstyp aus.
3. Geben Sie einen eindeutigen Namen für den Entitätstyp ein und klicken Sie auf "OK".
4. Fügen Sie Merkmale zu dem Entitätstyp hinzu (siehe nächster Abschnitt).

So fügen Sie Merkmale zu einem Entitätstyp hinzu:

1. Wählen Sie den Entitätstyp im Navigationsbereich links im Bildschirm aus.
2. Klicken Sie auf die Schaltfläche "Merkmal anfügen" (die obere Schaltfläche rechts im Bildschirm).
3. Wählen Sie ein oder mehrere Merkmale aus der Liste der verfügbaren Merkmale (zur Auswahl mehrerer Merkmale halten Sie beim Klicken die Taste "Strg" gedrückt) und klicken Sie auf OK.

So entfernen Sie Merkmale aus einem Entitätstyp:

1. Wählen Sie den Entitätstyp im Navigationsbereich links im Bildschirm aus.
2. Wählen Sie ein oder mehrere Merkmale aus der Tabelle der angefügten Merkmale auf der rechten Bildschirmseite aus. Halten Sie beim Klicken die Taste "Strg" gedrückt, um mehrere Merkmale auszuwählen.
3. Klicken Sie auf die Schaltfläche "Merkmal lösen" (die untere Schaltfläche rechts im Bildschirm).

So duplizieren Sie einen bestehenden Entitätstyp:

1. Klicken Sie mit der rechten Maustaste im Navigationsbereich links im Bildschirm auf den zu duplizierenden Entitätstyp.
2. Wählen Sie die Option Entitätstyp duplizieren aus.
3. Geben Sie einen eindeutigen Namen für den neuen Entitätstyp ein und klicken Sie auf "OK".
4. Fügen Sie nach Bedarf Merkmale zum Entitätstyp hinzu bzw. entfernen Sie sie daraus (siehe Anweisungen weiter oben).

So benennen Sie einen Entitätstyp um:

Vorsicht: Wenn Sie ein Merkmal oder ein Element eines Merkmals bearbeiten oder löschen und das Repository bereits Daten enthält, sollten Sie den Repository-Inhalt anschließend löschen und die Daten neu laden. Dadurch wird ein inkonsistenter Zustand des Repositorys vermieden.

1. Klicken Sie mit der rechten Maustaste im Navigationsbereich links im Bildschirm auf den umzubenennenden Entitätstyp.
2. Wählen Sie die Option Umbenennen.
3. Geben Sie den neuen Namen für den Entitätstyp ein und klicken Sie auf “OK”.

So löschen Sie einen Entitätstyp:

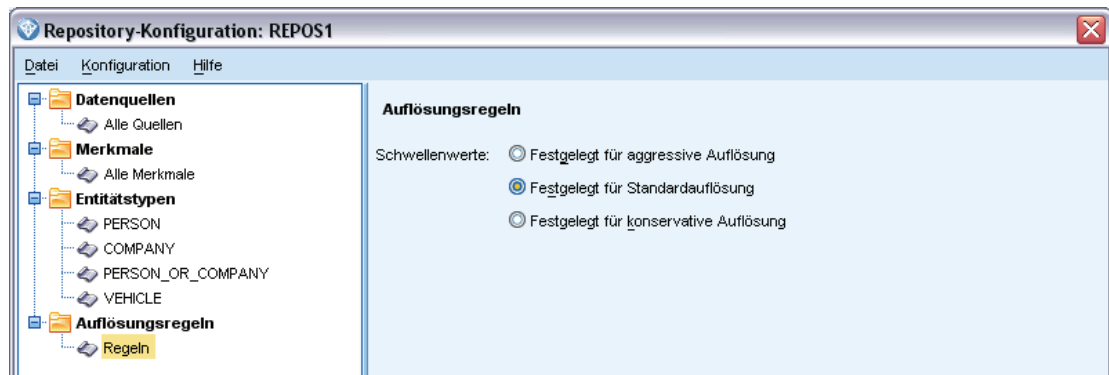
Vorsicht: Wenn Sie ein Merkmal oder ein Element eines Merkmals bearbeiten oder löschen und das Repository bereits Daten enthält, sollten Sie den Repository-Inhalt anschließend löschen und die Daten neu laden. Dadurch wird ein inkonsistenter Zustand des Repositorys vermieden.

1. Klicken Sie mit der rechten Maustaste im Navigationsbereich links im Bildschirm auf den zu löschenden Entitätstyp.
2. Wählen Sie Löschen.
3. Klicken Sie auf OK, um das Löschen des Entitätstyps zu bestätigen. *Vorsicht:* Das Löschen von Entitätstypen kann nicht rückgängig gemacht werden.

Festlegen des Schwellenwerts für den Entitätsabgleich

Im Abschnitt “Auflösungsregeln” des Fensters “Repository-Konfiguration” wählen Sie den Schwellenwert aus, bei dem der Entitätsabgleich erfolgen soll.

Abbildung 3-9
Festlegen der Entitätsauflösungsregeln



Beim Erstellen des Repositorys ist der Abgleich auf den Standard-Schwellenwert voreingestellt.

Wählen Sie die Option Festgelegt für aggressive Auflösung, wenn Sie in Ihren Datensätzen nicht genug Treffer finden, um eine Entitätsauflösung durchzuführen.

Wählen Sie Festgelegt für Standardauflösung, um von einer der anderen Einstellungen zum Standard-Schwellenwert zurückzukehren.

Wählen Sie die Option Festgelegt für konservative Auflösung, wenn Sie zu viele Treffer finden.

Wiederverwenden von Repository-Konfigurationen

Wenn Sie bereits eine Konfiguration eingerichtet haben und sie für ein anderes Repository verwenden möchten, können Sie die bestehende Konfiguration in eine XML-Datei exportieren und die Datei in das andere Repository (das Ziel-Repository) importieren.

So können Sie eine bestehende Konfiguration wiederverwenden:

1. Zeigen Sie das Fenster “Repository-Konfiguration” für das Repository an, dessen Konfiguration Sie verwenden möchten. Für weitere Informationen siehe Thema [Konfigurieren eines Entitäts-Repository](#) auf S. 23.
2. Wählen Sie aus dem Menü in diesem Fenster folgende Optionsfolge:
Konfiguration > Konfiguration exportieren.
3. Wählen Sie im Dialogfeld “Speichern unter” den Namen und den Speicherort der XML-Exportdatei aus.
4. Zeigen Sie das Fenster “Repository-Konfiguration” für das Ziel-Repository an.
5. Wählen Sie aus dem Menü in diesem Fenster folgende Optionsfolge:
Konfiguration > Konfiguration importieren.
6. Wählen Sie im Dialogfeld “Öffnen” den Namen und den Speicherort der zuvor exportierten XML-Exportdatei aus und klicken Sie auf Öffnen.

Speichern der Konfigurationsänderungen

So speichern Sie die Änderungen an der Konfiguration:

Wählen Sie aus dem Menü im Fenster “Repository-Konfiguration” folgende Optionsfolge:
Datei > Speichern.

Schließen des Konfigurationsfensters

So verlassen Sie das Konfigurationsfenster:

Wählen Sie aus dem Menü im Fenster “Repository-Konfiguration” folgende Optionsfolge:
Datei > Beenden.

Klicken Sie bei nicht gespeicherten Änderungen an der Konfiguration auf OK, um die Änderungen zu speichern und das Dialogfeld zu verlassen, bzw. auf Abbrechen, um das Dialogfeld ohne Speichern zu verlassen.

Analyse der aufgelösten Identitäten (Entity Analytics-Quellenknoten)

Wenn mindestens eine Datenquelle in das Repository eingegeben wurde, können Sie den Entity Analytics-(EA-)Quellenknoten verwenden, um die aufgelösten Identitäten zur weiteren Analyse bzw. Verarbeitung an andere IBM® SPSS® Modeler-Knoten weiterzuleiten, beispielsweise, um einen Bericht zu erstellen, in dem die aufgelösten Identitäten aufgeführt sind.

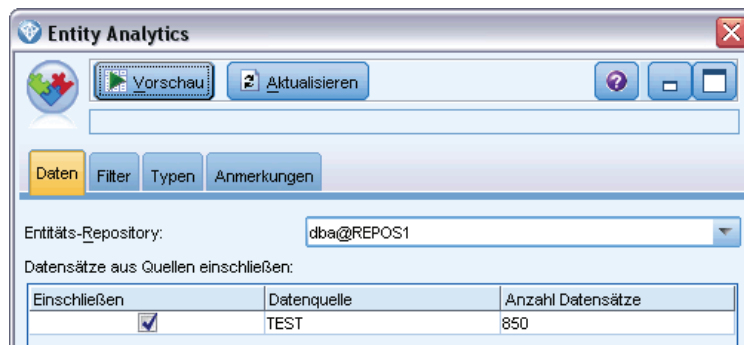
So analysieren Sie die aufgelösten Identitäten:

1. Fügen Sie einen Entity Analytics-(EA-)Quellenknoten zu einem Stream hinzu.
2. Öffnen Sie den Entity Analytics-(EA-)Quellenknoten.
3. Wählen Sie auf der Registerkarte “Daten” das Entitäts-Repository und mindestens eine seiner Eingabedatenquellen aus (klicken Sie auf Aktualisieren, um die Datensatzanzahl zu aktualisieren). Für weitere Informationen siehe Thema [Auswählen einer Datenquelle](#) auf S. 34.
4. Fügen Sie weitere Knoten zum Stream hinzu, um die gewünschte Verarbeitung durchzuführen. Für weitere Informationen siehe Thema [Hinzufügen von Knoten zum Stream](#) auf S. 35.

Auswählen einer Datenquelle

Auf der Registerkarte “Daten” wählen Sie mindestens eine Datenquelle im Repository aus, für die Sie weitere Verarbeitungsschritte durchführen. Klicken Sie auf Aktualisieren, um die Datensatzanzahl für die aufgelisteten Datenquellen zu aktualisieren.

Abbildung 3-10
Auswählen einer Datenquelle im Quellenknoten



Entitäts-Repository. Zeigt das aktuelle Entitäts-Repository an, sofern eines vorhanden ist. Wenn mehrere Repositories vorhanden sind, können Sie in der Liste ein anderes Repository auswählen. Wählen Sie zur Erstellung eines neuen Repository die Option <Durchsuchen...>, um ein Dialogfeld anzuzeigen, in dem Sie das Repository erstellen können. Für weitere Informationen siehe Thema [Optionen für das Entitäts-Repository](#) auf S. 18.

Datensätze aus Quellen einschließen. In dieser Tabelle werden die verschiedenen Datenquellen aufgelistet, die in das Repository eingegeben wurden, zusammen mit der Anzahl der Datensätze in den einzelnen Datenquellen. Aktivieren Sie das Kontrollkästchen Einschließen für diejenigen Datenquellen, die Sie für die weitere Analyse und Verarbeitung verwenden möchten.

Umbenennen von Datendateien

Sie können die Registerkarte “Filter” verwenden, um eines der aufgelösten Identitätsfelder umbenennen, die zur weiteren Verarbeitung nach unten im Stream weitergegeben werden. Es kann sinnvoll sein, ein aufgelöstes Identitätsfeld umbenennen, beispielsweise um die

Kompatibilität der Feldnamen beim Zusammenführen mit einem anderen Daten-Set weiter unten im Stream zu gewährleisten.

Die Felder mit ihren ursprünglichen Namen lauten wie folgt.

Tabelle 3-2
Aufgelöste Identitätsfelder

Feld	Beschreibung
\$EA-ID	Entitäts-ID
\$EA-SRC	Quell-Tag zur Angabe der Datenquelle, aus der die Datensätze stammen
\$EA-KEY	Als eindeutiger Schlüssel in der Datenquellendatei gekennzeichnetes Feld

Hinweis: Sie können zwar auch die Registerkarte “Filter” verwenden, um Felder herauszufiltern, sollten dies hier jedoch nicht tun, da die aufgelösten Identitätsfelder das absolute Minimum sind, das für die Entity Analytics-Verarbeitung benötigt wird.

Festlegen der Typinformationen für Datenfelder

Auf der Registerkarte “Typen” können Sie verschiedene Eigenschaften der aufgelösten Identitätsfelder anzeigen bzw. ändern, die zur weiteren Verarbeitung nach unten im Stream weitergegeben werden.

Sie können dieselben Eigenschaften ändern wie auf der Registerkarte “Typen” eines regulären SPSS Modeler-Typknoten, nämlich:

Tabelle 3-3
Typeigenschaften für Felder

Eigenschaft	Beschreibung
Messung	Das Messniveau (d. h. der Datentyp), das zur Beschreibung der Eigenschaften der Daten im Feld verwendet wird.
Werte	Bietet Optionen zum Lesen von Datenwerten aus dem Daten-Set.
Fehlende Werte	Wird verwendet, um anzugeben, wie fehlende Werte für das Feld behandelt werden.
Überprüfen	Validierungsoptionen, die sicherstellen, dass Feldwerte den angegebenen Werten bzw. Bereichen entsprechen.
Rolle	Gibt an, wie das Feld verwendet wird, wenn die Daten an Modellierungsknoten oder ein Modell-Nugget weitergegeben werden.

Hinzufügen von Knoten zum Stream

Sie können verschiedene SPSS Modeler-Knoten zum Stream hinzufügen, um Analyse- oder Verarbeitungsvorgänge an der Ausgabe aus dem Entity Analytics-(EA-)Quellenknoten vorzunehmen. Beispielsweise könnten Sie eines oder mehrere der folgenden Elemente hinzufügen.

- Aggregatknoten, zur Zusammenfassung der Ausgabe, die sehr groß sein kann
- Auswahlknoten zur Auswahl einer Teilmenge der Ausgabe
- Tabellenknoten zur Anzeige der Ausgabe aus dem Entity Analytics-(EA-)Quellenknoten

- Berichtknoten zum Drucken der Ausgabe in einem Bericht
- SPSS Modeler-Exportknoten zum Exportieren der Ausgabe in ein anderes Format, beispielsweise eine Kalkulationstabelle oder eine Datenbank

Weitere Informationen finden Sie in den Abschnitten zu Datensatzoperationsknoten, Ausgabeknoten und Exportknoten im Handbuch *IBM SPSS Modeler 15 Quellen-, Prozess- und Ausgabeknoten*.

Vergleich neuer Fälle mit dem Repository (Knoten "Streaming von EA")

Wenn Sie bereits eine Identitätsauflösung im Repository durchgeführt haben, können Sie den Knoten "Streaming von EA" verwenden, um neue Fälle, auf die Sie danach stoßen, mit dem Repository-Inhalt zu vergleichen. Dieser Knoten verarbeitet Datensätze aus einer neuen Datenquelle, vergleicht sie mit den bereits im Repository befindlichen Entitäten und leitet alle übereinstimmenden Datensätze für die weitere Verarbeitung weiter. Es kann festgelegt werden, dass die Übereinstimmungen exakt sein müssen, oder es kann auch eine lockerere Beziehung mit den bestehenden Entitäten zugelassen werden.

Ebenso wie der EA-Exportknoten verwendet auch der Knoten "Streaming von EA" einen einzelnen SPSS Modeler-Quellenknoten als Eingabe. Der Knoten "Streaming von EA" weist jedoch folgende Abweichungen auf: Während der Exportknoten Datensätze zu allen Entitäten ausgibt, die mit seinen Eingabedatensätzen in Beziehung stehen, gibt der Knoten "Streaming von EA" Datensätze nur zu denjenigen Entitäten aus, die in Beziehung mit Entitäten stehen, die bereits im Repository aufgelöst wurden. Für weitere Informationen siehe Thema [Ausgabe aus dem Knoten "Streaming von EA"](#) auf S. 41.

So können Sie neue Fälle mit dem Repository vergleichen:

1. Stellen Sie eine Verbindung zu der Datenquelle her, die die neuen Datensätze enthält, die mit den bestehenden Entitäten verglichen werden sollen. Für weitere Informationen siehe Thema [Verbindung mit einer Datenquelle](#) auf S. 15.
2. Verbinden Sie auf der Registerkarte "Datensatzoperationen" einen Knoten vom Typ "Streaming von EA" mit dem Datenquellenknoten.
3. Doppelklicken Sie dann auf den Entity Analytics-Exportknoten, um das zugehörige Dialogfeld zu öffnen.
4. Klicken Sie auf die Liste Entitäts-Repository.
5. Klicken Sie auf <Durchsuchen...>, um das Dialogfeld "Entitäts-Repositories" anzuzeigen.
6. Klicken Sie im Dialogfeld "Entitäts-Repositories" auf das Feld "Repository-Name".
7. Klicken Sie auf den Namen des zu verwendenden Repository.
8. Geben Sie Benutzernamen und Passwort für dieses Repository ein und klicken Sie auf Verbinden. Klicken Sie auf OK, wenn die Verbindung mit dem Repository hergestellt wurde.

9. Wählen Sie im Dialogfeld “Streaming von EA” den zu verwendenden Entitätstyp aus. Für weitere Informationen siehe Thema [Verwaltung der Elementtypen](#) auf S. 30.
10. Ordnen Sie die Eingabefelder in der Datenquelle Merkmalen im Repository zu. Für weitere Informationen siehe Thema [Erstellen einer Zuordnung zwischen Eingabefeldern und Merkmalen \(Knoten “Streaming von EA”\)](#) auf S. 37.
11. Optional können Sie auf die Registerkarte Repository klicken, um eine Tabelle mit der Zuordnung für diese Datenquelle (oder alle zugeordneten Datenquellen) anzuzeigen. Für weitere Informationen siehe Thema [Anzeigen der Feldzuordnungen \(Knoten “Streaming von EA”\)](#) auf S. 39.
12. Klicken Sie auf die Registerkarte Ansicht, um Details der verschiedenen Datenquellen anzuzeigen, die in das Repository eingegeben wurden, und die Auswahlkriterien für das Abrufen bestehender Entitäten festzulegen. Für weitere Informationen siehe Thema [Anzeigen der Datenzuordnungen \(Knoten “Streaming von EA”\)](#) auf S. 40.
13. Klicken Sie auf OK, wenn der Knoten korrekt eingerichtet wurde.
14. Verbinden Sie einen Tabellenknoten mit dem Knoten “Streaming von EA” und führen Sie den Stream aus.

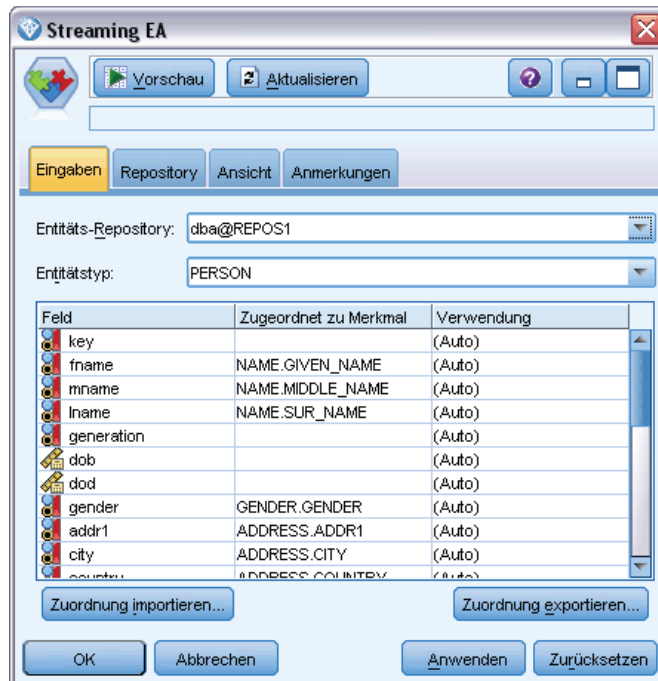
Im Ausgabefenster des Tabellenknotens werden alle abgerufenen Entitäten aufgeführt, die mit den neuen Datensätzen in der Datenquelle übereinstimmen. Die Ausgabefelder wurden mit dem Präfix \$EA- versehen. Für weitere Informationen siehe Thema [Ausgabe aus dem Knoten “Streaming von EA”](#) auf S. 41.

Erstellen einer Zuordnung zwischen Eingabefeldern und Merkmalen (Knoten “Streaming von EA”)

Die Registerkarte “Eingaben” enthält die Optionen zur Zuordnung von Feldern in der Eingabe für diesen Knoten zu Repository-Merkmalen. Richten Sie die Zuordnungszuweisungen auf dieser Registerkarte ein oder alternativ auf der Registerkarte Ansicht, um Details aller Datenquellen im Repository anzuzeigen, und klicken Sie anschließend auf OK.

Wenn Sie bereits eine Gruppe von Zuordnungen in einer XML-Datei gespeichert haben, können Sie diese durch Klicken auf Zuordnung importieren verwenden.

Abbildung 3-11
Erstellen einer Zuordnung zwischen Feldern von neuen Datensätzen und Merkmalen im Repository



Entitäts-Repository. Zeigt das aktuelle Entitäts-Repository an, sofern eines vorhanden ist. Wenn mehrere Repositories vorhanden sind, können Sie in der Liste ein anderes Repository auswählen. Wählen Sie zur Erstellung eines neuen Repository die Option <Durchsuchen...>, um ein Dialogfeld anzuzeigen, in dem Sie das Repository erstellen können. Für weitere Informationen siehe Thema [Optionen für das Entitäts-Repository](#) auf S. 18.

Entitätstyp. Eine Liste der im Repository definierten Entitätstypen (d. h. Merkmals-Sets). Wählen Sie einen aus der Liste aus oder zeigen Sie mithilfe der Option <Neuen Entitätstyp hinzufügen...> das Fenster Repository-Konfiguration an, in dem Sie einen neuen Entitätstyp definieren können. Für weitere Informationen siehe Thema [Konfigurieren eines Entitäts-Repository](#) auf S. 23.

Zuordnungstabelle. In dieser Tabelle können Sie jedes Eingabefeld einem entsprechenden Merkmal im Repository zuordnen. Wenn im ausgewählten Entitätstyp kein geeignetes Merkmal vorhanden ist, können Sie hier ein neues Merkmal erstellen.

- **Feld.** Die Menge an Eingabefeldern in der ausgewählten Datenquelle. Zu jedem Feld gehört ein Symbol, das das Messniveau (also den Datentyp) für das Feld angibt.
- **Zugeordnet zu Merkmal.** Um ein Feld einem Merkmal zuzuordnen, doppelklicken Sie auf diese Spalte (oder drücken Sie die Leertaste) in der Zeile des betreffenden Felds und wählen Sie ein Merkmal aus der Liste aus. Wenn kein geeignetes Merkmal verfügbar ist, können Sie mit der Option <Neues Merkmal hinzufügen...> das Fenster Repository-Konfiguration anzeigen, in dem

Sie ein neues Merkmal für diesen Entitätstyp definieren können. Für weitere Informationen siehe Thema [Konfigurieren eines Entitäts-Repository](#) auf S. 23.

- **Verwendung.** Gibt den Kontext eines bestimmten Felds an, wenn mehrere Kontexte möglich sind, beispielsweise die private und die dienstliche Telefonnummer. Für weitere Informationen siehe Thema [Verwaltung der Elementtypen](#) auf S. 30.

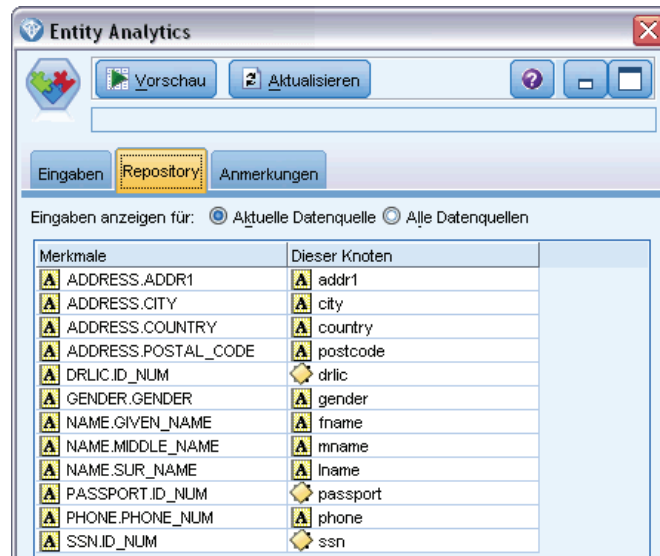
Zuordnung importieren. Importiert eine zuvor exportierte Menge an Zuordnungen zwischen Feldern und Merkmalen aus einer externen XML-Datei. Diese kann nützlich sein, wenn Sie verschiedene Datenquellen mit denselben Zuordnungsanforderungen verwenden, da Sie dadurch nicht dieselben Zuordnungen für die verschiedenen Quellen erneut zu definieren brauchen.

Zuordnung exportieren. Exportiert die Menge der in der Zuordnungstabelle angegebenen Zuordnungen zwischen Feldern und Merkmalen in eine externe XML-Datei.

Anzeigen der Feldzuordnungen (Knoten "Streaming von EA")

Klicken Sie auf der Registerkarte "Repository" auf Aktualisieren, um anzuzeigen, welchen Repository-Merkmalen Eingabefelder zugeordnet sind. Sie können dies entweder für die aktuelle Datenquelle (die Datenquelle, die durch den mit diesem Exportknoten verknüpften Quellenknoten gesteuert wird) oder für alle Datenquellen anzeigen.

Abbildung 3-12
Anzeigen der Feldzuordnungen



Eingaben anzeigen für. Wählen Sie eine Option, um die Zuordnungen entweder für die aktuelle Datenquelle anzuzeigen oder für alle Datenquellen, die dem Repository bekannt sind.

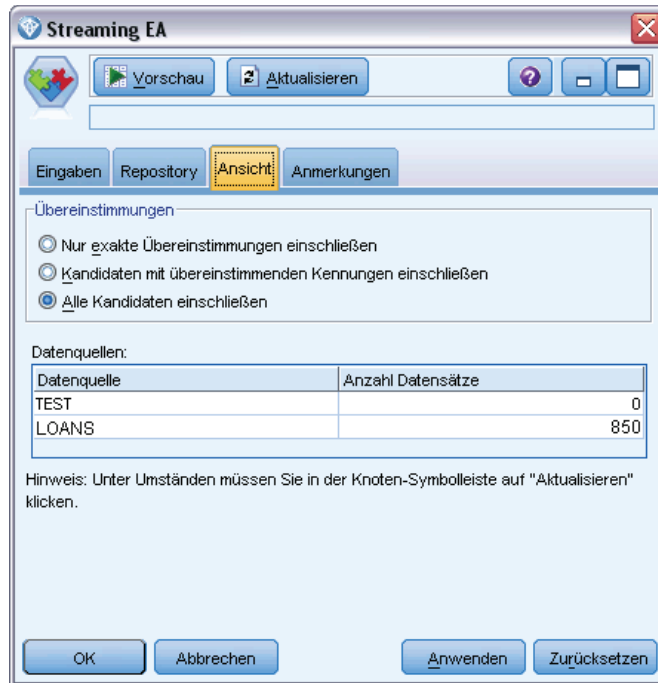
Strukturen. Eine Liste aller Merkmale, die über Zuordnungen in den angezeigten Datenquellen verfügen. Merkmale ohne Zuordnung werden nicht angezeigt.

Dieser Knoten. In den einzelnen Spalten sind zugeordnete Felder in einer bestimmten Datenquelle für jedes Merkmal aufgeführt, für das eine Zuordnung definiert wurde.

Anzeigen der Datenzuordnungen (Knoten "Streaming von EA")

Auf der Registerkarte "Ansicht" können Sie Details der verschiedenen Datenquellen anzeigen, die in das Repository eingegeben wurden. Dabei handelt es sich um Datenquellen, anhand derer die Eingaben für diesen Knoten verarbeitet werden, um nach übereinstimmenden Entitäten zu suchen und diese abzurufen. Klicken Sie zum Aktualisieren der Datensatzanzahl auf Aktualisieren.

Abbildung 3-13
Anzeigen der Datenquellendetails



Übereinstimmungen. Mit den folgenden Optionen wird angegeben, wie eng die Informationen für Zuordnungen zwischen Feldern und Merkmalen, die Sie auf der Registerkarte "Eingaben" angeben, mit den in Frage kommenden Datensätzen (also den gesamten Repository-Inhalten) abgeglichen werden sollen. Je enger das Übereinstimmungskriterium, desto weniger Entitäten werden abgerufen.

- **Nur exakte Übereinstimmungen einschließen.** Dies ist das engste Übereinstimmungskriterium und führt zur Auswahl der geringsten Anzahl an Datensätzen. Verwenden Sie diese Option, wenn Sie nur die Entitäten abrufen möchten, die als exakte Übereinstimmung eingestuft werden.
- **Kandidaten mit übereinstimmenden Kennungen einschließen.** Verwenden Sie diese Einstellung, wenn Sie sowohl übereinstimmende Entitäten als auch Entitäten mit den gleichen Kennungen einbeziehen möchten (Entitäten mit Merkmalen, für die eine Häufigkeit von Eins konfiguriert wurde, z. B. übereinstimmende Kreditkartennummern, Steuernummern usw.).
- **Alle Kandidaten einschließen.** Verwenden Sie diese Option, wenn Sie die größtmögliche Anzahl an Entitäten im Repository anzeigen möchten, die gemeinsame Merkmale haben. Dies ist das unschärfste Übereinstimmungskriterium und führt zur Auswahl der größten Anzahl an Datensätzen. Bei dieser Option werden exakte Übereinstimmungen sowie Entitäten mit

fast beliebigen gemeinsamen Merkmalen zurückgegeben (typischerweise Entitäten mit einer Häufigkeit von Eins oder Wenige). So würden z. B. Entitäten mit der gleichen Steuernummer und Entitäten mit ähnlichen Adressen einbezogen werden.

Datenquellen. In dieser Tabelle werden die Quell-Tags der verschiedenen Datenquellen aufgelistet, die in das Repository geladen wurden, zusammen mit der Anzahl der Datensätze in den einzelnen Datenquellen.

Ausgabe aus dem Knoten “Streaming von EA”

Die Ausgabe aus dem Knoten “Streaming von EA” besteht aus folgenden Feldern für jeden abgerufenen Datensatz.

Feld	Beschreibung
<i>Feld1[, Feld2[, ... FeldN]]</i>	Felder aus der Datenquelle, die die neuen Datensätze enthält.
\$EA-ID	Entitäts-ID für diesen Datensatz im Repository.
\$EA-SC	Feld, das die Genauigkeit der Übereinstimmung zwischen diesem Datensatz und einer beobachteten Entität im Repository angibt; Wert von 1,0 (schlechte Übereinstimmung) bis 10,0 (gute Übereinstimmung).
\$EA-SRC	Quell-Tag zur Angabe der Datenquelle, aus der dieser Datensatz stammt.
\$EA-KEY	Wert des eindeutigen Schlüssels für diesen Datensatz in der Datenquellendatei.
<i>\$EA-Merkmal1[, \$EA-Merkmal2[, ... \$EA-MerkmalN]]</i>	Werte der zugeordneten Merkmale für diesen Datensatz im Repository.

Verwendung von IBM SPSS Modeler Entity Analytics mit anderen IBM SPSS-Produkten

Es stehen Installationsprogramme zur Verfügung, die die Verwendung von IBM SPSS Modeler Entity Analytics in Verbindung mit folgenden Produkten ermöglichen:

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

Sie müssen diese Installationsprogramme ausführen, bevor Sie die Funktionen von IBM SPSS Modeler Entity Analytics in Verbindung mit diesen Produkten verwenden können. Weitere Informationen finden Sie im Handbuch *IBM SPSS Modeler Premium-Installation*.

Verwaltungsaufgaben

Für diejenigen Repositories, die in Entity Analytics erstellt werden, wird unter Verwendung des Produkts IBM solidDB ein neuer Datenbankdienst erstellt. Zu solidDB gehören einige Verwaltungsaufgaben. Diese Aufgaben werden typischerweise vom Datenbankadministrator oder vom Systemadministrator ausgeführt. Es handelt sich dabei um folgende Aufgaben:

- Konfigurieren von Port-Zuweisungen
- Verwalten der Administratorberechtigungen für die Repository-Datenbank

Weitere Verwaltungsaufgaben, die möglicherweise durchgeführt werden müssen, gelten für alle Repositories. Es handelt sich dabei um folgende Aufgaben:

- Verschieben des Speicherverzeichnisses für das Repository
- Festlegen von Stream-Optionen für Datums-/Zeit- und Zeitstempelfelder
- Ausführen von IBM SPSS Modeler Entity Analytics mit SPSS Modeler-Client und SPSS Modeler Server auf demselben Windows-System
- Löschen des Inhalts eines Entitäts-Repository
- Löschen eines Entitäts-Repository
- Löschen eines Repository, wenn keine Verbindung damit hergestellt werden kann

Konfigurieren von Port-Zuweisungen

Jedem solidDB-Datenbankdienst muss ein Port zugeteilt sein, der keinen anderen Diensten, die auf dem Computer ausgeführt werden, zugeteilt sein darf. Die Datenbankdienste befinden sich auf demselben Computer, auf dem auch IBM® SPSS® Modeler Server ausgeführt wird (oder, wenn IBM® SPSS® Modeler ohne Verbindung mit SPSS Modeler Server verwendet wird, auf dem Computer, auf dem IBM® SPSS® Modeler ausgeführt wird).

Standardmäßig weist Entity Analytics Ports im Bereich 1320 bis 1520 zu, beginnend mit Port 1320 für das erste erstellte Repository. Im Falle eines Konflikts können Sie die Zuweisung von Ports durch Bearbeiten der Datei *<Modeler Server-Installationspfad>/ext/bin/pasw.entityanalytics/ea.cfg* und Festlegung geeigneter Werte für die Einstellungen *min_port* und *max_port* konfigurieren. Im Folgenden finden Sie die Standardinhalte dieser Datei:

```
# port range configuration for entity analytics

#

# this port range controls which ports SolidDB databases

# (created to store Entity Analytics Repositories in)

# may use. Configure this if the default port range will

# introduce a conflict on your system.
```



```
#
# default min_port = 1320
# default max_port = 1520
min_port, 1320
max_port, 1520
```

Verwalten der Administratorberechtigungen für die Repository-Datenbank

Der Benutzername und das Passwort für die solidDB-Datenbank, die als Host für ein Entitäts-Repository fungiert, werden bei der Erstellung des Repositorys definiert. Wenn Ihnen die aktuellen Anmeldeberechtigungen bekannt sind, können Sie diese Angaben mithilfe des solidDB-SQL-Editors ändern.

So starten Sie den solidDB-SQL-Editor:

1. Öffnen Sie an einem Client-Computer ein Eingabeaufforderungsfenster.
2. Geben Sie Folgendes ein:

```
cd modeler-installationsverzeichnis\ext\bin\pasw.entityanalytics\solidDB\bin
```

dabei ist *modeler-installationsverzeichnis* das Verzeichnis, in dem SPSS Modeler installiert wurde.

3. Geben Sie Folgendes ein:

```
solsql -c "C:\Dokumente und Einstellungen\All
Users\Anwendungsdaten\IBM\SPSS\Modeler\version\EA\repositories\repos-name
```

Dabei ist *version* die Versionsnummer der SPSS Modeler-Installation und *repos-name* ist der Name des Repositorys.

4. Geben Sie an der Eingabeaufforderung den aktuellen Benutzernamen des Datenbankadministrators und das zugehörige Passwort ein, um die Eingabeaufforderung `solsql>` anzuzeigen.

So ändern Sie das Passwort für den Datenbankadministrator:

1. Geben Sie an der Eingabeaufforderung `solsql>` Folgendes ein:

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

Dabei ist *username* der aktuelle Benutzername des Datenbankadministrators und *password* ist das neue Passwort.

2. Geben Sie "exit" ein, um den Editor zu schließen.
3. Starten Sie den SPSS Modeler-Client neu.

So führen Sie andere Verwaltungsaufgaben durch:

Informationen zu anderen Verwaltungsaufgaben, die mit der solidDB-Datenbank ausgeführt werden können, finden Sie in der Dokumentation zu der entsprechenden Version von IBM solidDB unter <http://publib.boulder.ibm.com/>.

Verschieben des Speicherverzeichnisses für das Repository

Standardmäßig gelten für die Repository-Dateien im Verzeichnis *EA* folgende Speicherorte:

- C:\Dokumente und Einstellungen\All Users\Anwendungsdaten\IBM\SPSS\Modeler\Version\EA (Windows-Systeme)
- *Modeler-Installationsverzeichnis*/ext/bin/pasw.entityanalytics/EA (UNIX-Systeme)

Da die zum Speichern des Repositorys verwendeten Dateien sehr groß werden können, müssen Sie sie möglicherweise auf einen anderen Datenträger bzw. in eine andere Partition verlagern, um mehr Speicherplatz verfügbar zu machen. Gehen Sie dazu wie folgt vor.

1. Beenden Sie SPSS Modeler.
2. Verschieben Sie das Verzeichnis *EA* vom ursprünglichen Speicherort (siehe oben) an einen neuen Speicherort. Unter Windows könnten Sie es beispielsweise in ein Verzeichnis wie *F:\Daten\EA* verschieben.
3. Bearbeiten Sie die Datei *<Modeler-Server-Installationspfad>/ext/bin/pasw.entityanalytics/ea.cfg*, indem Sie folgende Option hinzufügen:

```
repository_data_directory, neuer_Speicherort
```

Dabei ist *neuer_Speicherort* das Verzeichnis, in das Sie das EA-Verzeichnis verlagert haben, z. B. *F:\data\EA*.

Festlegen von Stream-Optionen für Datums-/Zeit- und Zeitstempelfelder

Wenn die Quelldaten Felder mit Datums-/Zeit- bzw. Zeitstempeldaten enthalten, müssen Sie sicherstellen, dass die entsprechenden Stream-Optionen auf das von IBM SPSS Modeler Entity Analytics erkannte Format gesetzt sind.

So legen Sie das Format für die Stream-Optionen fest:

1. Klicken Sie im SPSS Modeler-Hauptmenü auf:
Werkzeuge > Stream-Eigenschaften > Optionen.
2. Wählen Sie Datum/Uhrzeit.
3. Setzen Sie Datumsformat auf JJJJ-MM-TT.
4. Setzen Sie Zeitformat auf HH:MM:SS.
5. Klicken Sie auf OK.

Ausführen von IBM SPSS Modeler Entity Analytics mit SPSS Modeler-Client und SPSS Modeler Server auf demselben Windows-System

Wenn Sie IBM SPSS Modeler Entity Analytics sowohl auf SPSS Modeler-Client als auch auf SPSS Modeler Server auf demselben Windows-System installiert haben, verwenden Client und Server standardmäßig dasselbe Repository. Wenn getrennte Repositories verwendet werden sollen, müssen Sie die Konfigurationsdatei *ea.cfg* auf **einem** der Systeme so bearbeiten, dass es einen anderen Port-Bereich und einen anderen Repository-Ordner verwendet.

Hinweis: Dieses Verfahren muss insbesondere dann ausgeführt werden, wenn Sie eine 32-Bit-Version des SPSS Modeler-Clients und eine 64-Bit-Version von SPSS Modeler Server (oder umgekehrt) verwenden.

1. Öffnen Sie die Datei `<Modeler-[Server-]Installationspfad>/ext/bin/pasw.entityanalytics/ea.cfg` für die Bearbeitung.
2. Ändern Sie die Einstellungen `min_port` und `max_port` so, dass andere Ports verwendet werden als beim anderen System. Für weitere Informationen siehe Thema [Konfigurieren von Port-Zuweisungen](#) auf S. 42.
3. Ändern Sie die Einstellung `repository_data_directory` so, dass ein anderes Verzeichnis verwendet wird als beim anderen System.
4. Speichern und schließen Sie die Datei *ea.cfg*.

Löschen des Inhalts eines Entitäts-Repository

Wenn Sie die Datensätze aus einem Entitäts-Repository entfernen, die Konfigurationsinformationen jedoch beibehalten möchten, verwenden Sie die Option zur Löschung des Repository-Inhalts.

So löschen Sie den Inhalt eines Repository:

1. Öffnen Sie einen Entity Analytics-Knoten.
2. Klicken Sie auf die Liste Entitäts-Repository.
3. Klicken Sie auf `<Durchsuchen...>`, um das Dialogfeld "Entitätsauflösungsinstanzen" anzuzeigen.
4. Klicken Sie im Dialogfeld "Entitätsauflösungsinstanzen" auf die Liste Repository-Name.
5. Wählen Sie das Repository aus, dessen Inhalt Sie löschen möchten.
6. Wenn noch keine Verbindung besteht, geben Sie den Administrator-Benutzernamen und das zugehörige Passwort ein und klicken Sie auf Verbinden.
7. Wenn die Schaltfläche Repository-Inhalt löschen aktiviert ist, klicken Sie darauf.
8. Klicken Sie im Dialogfeld "Datenquelleninhalt löschen" auf Repository-Inhalt löschen, um das Löschen des Repository-Inhalts zu bestätigen.

Löschen eines Entitäts-Repository

Wenn Sie ein Repository nicht mehr benötigen, können Sie es vollständig löschen.

Vorsicht: Dabei geschieht genau dies. **Dieser Vorgang kann nicht rückgängig gemacht werden.** Wenn Sie sich nicht sicher sind, verwenden Sie die Schaltfläche Inhalt löschen, um alle Quelldaten zu entfernen. Dabei bleibt die Repository-Konfiguration erhalten. Für weitere Informationen siehe Thema [Löschen des Inhalts eines Entitäts-Repository](#) auf S. 45.

Hinweis: Beim folgenden Verfahren wird davon ausgegangen, dass Sie von SPSS Modeler aus eine Verbindung mit dem Repository herstellen können und dass Ihnen der Benutzername und das Passwort des Administrators für die Datenbank bekannt ist, die als Host für das Repository fungiert. Wenn dies nicht der Fall ist, verwenden Sie das Verfahren zum Löschen eines Repository, wenn keine Verbindung damit hergestellt werden kann. Für weitere Informationen siehe Thema [Löschen eines Repository, wenn keine Verbindung damit hergestellt werden kann](#) auf S. 46.

So löschen Sie ein Repository:

1. Öffnen Sie einen Entity Analytics-Knoten.
2. Klicken Sie auf die Liste Entitäts-Repository.
3. Klicken Sie auf <Durchsuchen...>, um das Dialogfeld “Entitätsauflösungsinstanzen” anzuzeigen.
4. Klicken Sie im Dialogfeld “Entitätsauflösungsinstanzen” auf die Liste Repository-Name.
5. Wählen Sie das Repository aus, das Sie löschen möchten.
6. Wenn noch keine Verbindung besteht, geben Sie den Administrator-Benutzernamen und das zugehörige Passwort ein und klicken Sie auf Verbinden.
7. Wenn die Schaltfläche Gesamtes Repository löschen aktiviert ist, klicken Sie darauf.
8. Klicken Sie auf Löschen, um das Löschen des Repositorys zu bestätigen.
9. Klicken Sie auf OK, um die erfolgreiche Löschung zu bestätigen.

Löschen eines Repository, wenn keine Verbindung damit hergestellt werden kann

Gehen Sie wie folgt vor, wenn Sie ein Entitäts-Repository löschen möchten, aber keine Verbindung damit herstellen können, entweder aufgrund von Konnektivitätsproblemen mit SPSS Modeler oder weil Sie den Benutzernamen bzw. das Passwort vergessen haben.

Führen Sie das folgende Verfahren an dem Computer durch, der als Host für die Repository-Datenbank fungiert.

Windows-Systeme

1. Öffnen Sie ein Eingabeaufforderungsfenster.

2. Geben Sie Folgendes ein:

```
cd modeler-installationsverzeichnis  
cd ext\bin\pasw.entityanalytics\tools  
delete_repository.bat repos-name
```

Dabei ist *modeler-installationsverzeichnis* das Verzeichnis, in dem SPSS Modeler installiert ist, und *repos-name* ist der Name des Repositorys.

3. Fahren Sie mit “Abschluss des Verfahrens” weiter unten in diesem Abschnitt fort.

UNIX-Systeme

1. Öffnen Sie eine Shell.
2. Geben Sie Folgendes ein:

```
cd modeler-server-installationsverzeichnis  
cd ext/bin/pasw.entityanalytics/tools  
./delete_repository.sh repos-name
```

Dabei ist *modeler-server-installationsverzeichnis* das Verzeichnis, in dem SPSS Modeler Server installiert ist, und *repos-name* ist der Name des Repositorys.

Abschluss des Verfahrens (alle Systeme)

1. Bestätigen Sie an der Eingabeaufforderung die Löschung des Repositorys durch Eingabe von Y.
2. Nachdem das Repository gelöscht wurde, wird folgende Meldung angezeigt:

```
Info - please remove repository files from the following directory:  
  verzeichnis-pfad
```

(Hinweis: Möglicherweise müssen Sie zuerst einen Neustart des Systems durchführen, bevor Sie die Repository-Dateien entfernen können.)

3. Löschen Sie das Verzeichnis, das denselben Namen trägt wie das gelöschte Repository. Wenn Sie das Verzeichnis nicht löschen können, führen Sie einen Neustart des Computers durch und versuchen Sie es noch einmal.

Entitätsanalyse in Aktion

Zu diesem Beispiel

In diesem Beispiel wird gezeigt, wie Sie durch die Hinzunahme der Entitätsanalyse die bereits beeindruckenden Ergebnisse, die Sie mit IBM® SPSS® Modeler erzielen können, weiter verbessern können.

In diesem Beispiel wird der Stream *loan_entity_analytics.str* verwendet, der auf die Datendatei *loan_applications.csv* verweist. Die Dateien stehen im Verzeichnis *Demos* jeder SPSS Modeler-Installation zur Verfügung, bei der auch IBM® SPSS® Modeler Entity Analytics installiert ist. Das Verzeichnis *Demos* kann über die Programmgruppe “SPSS Modeler” im Windows-Startmenü aufgerufen werden. Die Datei *loan_entity_analytics.str* befindet sich im Verzeichnis *Entity_Analytics*.

Hinweis: Damit Sie diesen Beispiel-Stream ausführen können, müssen Sie ein Repository in Ihrem System erstellen. Erledigen Sie dies, bevor Sie mit diesem Beispiel fortfahren. Für weitere Informationen siehe Thema [Erstellen des Repository](#) in Kapitel 3 auf S. 16.

Beginnen wir mit einer vertrauten Situation: Die leitenden Angestellten einer Bank fragen sich, ob es bei bestimmten Kunden wahrscheinlich ist, dass es bei Krediten, die sie beantragt haben, zu einem Zahlungsausfall kommt. Die IT-Abteilung der Bank verwendet bereits seit langer Zeit SPSS Modeler, weshalb die Mitarbeiter der Abteilung bereits einen Stream erstellt und aus den bestehenden Daten zu ca. 700 früheren Krediten der Bank ein Vorhersagemodell erstellt haben. Diese Kredite wurden entweder zurückgezahlt, oder die Kunden kamen ihren Rückzahlungsverpflichtungen nicht nach.

Das ursprüngliche Modell

Hier sehen Sie, wie die Bankmitarbeiter ihr Modell erstellt und welche Erkenntnisse sie daraus gewonnen haben.

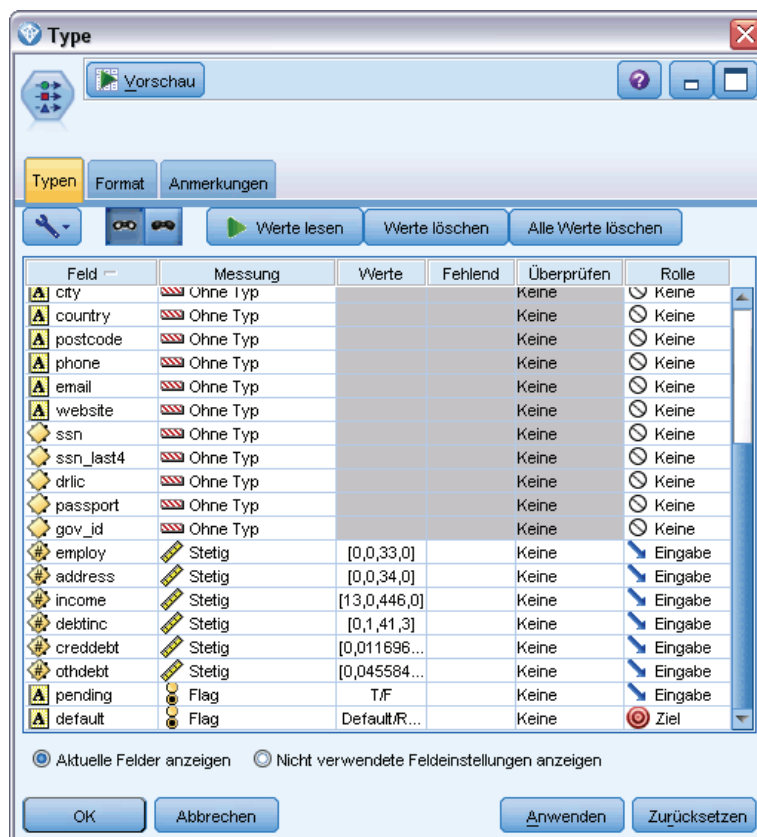
Abbildung 4-1
Ursprünglicher Stream mit Modellierungsknoten



Neben den Details zu früheren Krediten enthält das Daten-Set *loan_applications.csv* Details zu 150 Kunden, über deren Kreditanträge noch nicht entschieden wurde. Dies ergibt insgesamt 850 Datensätze.

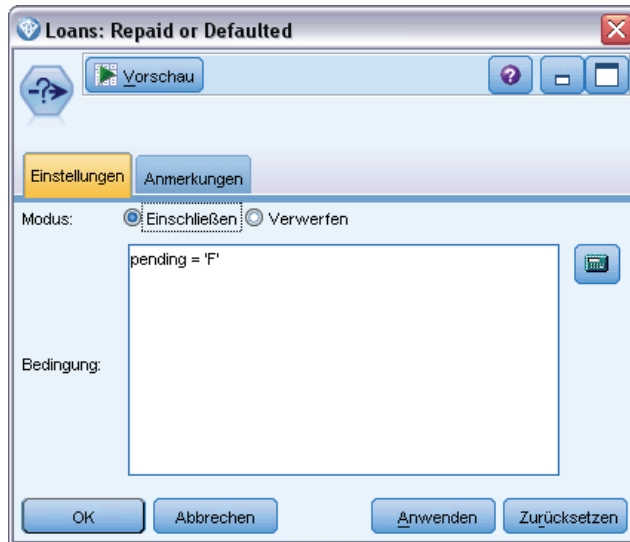
Nicht alle Felder aus dem Daten-Set sind für die Erstellung der Vorhersage von Nutzen. So können beispielsweise Namensfelder ignoriert werden. Der Typknoten filtert die zu ignorierenden Felder heraus, indem er deren Rolle auf Keine setzt. Bei den für die Vorhersage zu verwendenden Feldern wird die Rolle auf Eingabe gesetzt, und bei dem Feld, dessen Wert das Modell vorherzusagen versucht, wird die Rolle auf Ziel gesetzt.

Abbildung 4-2
Im Typknoten festgelegte Feldrollen



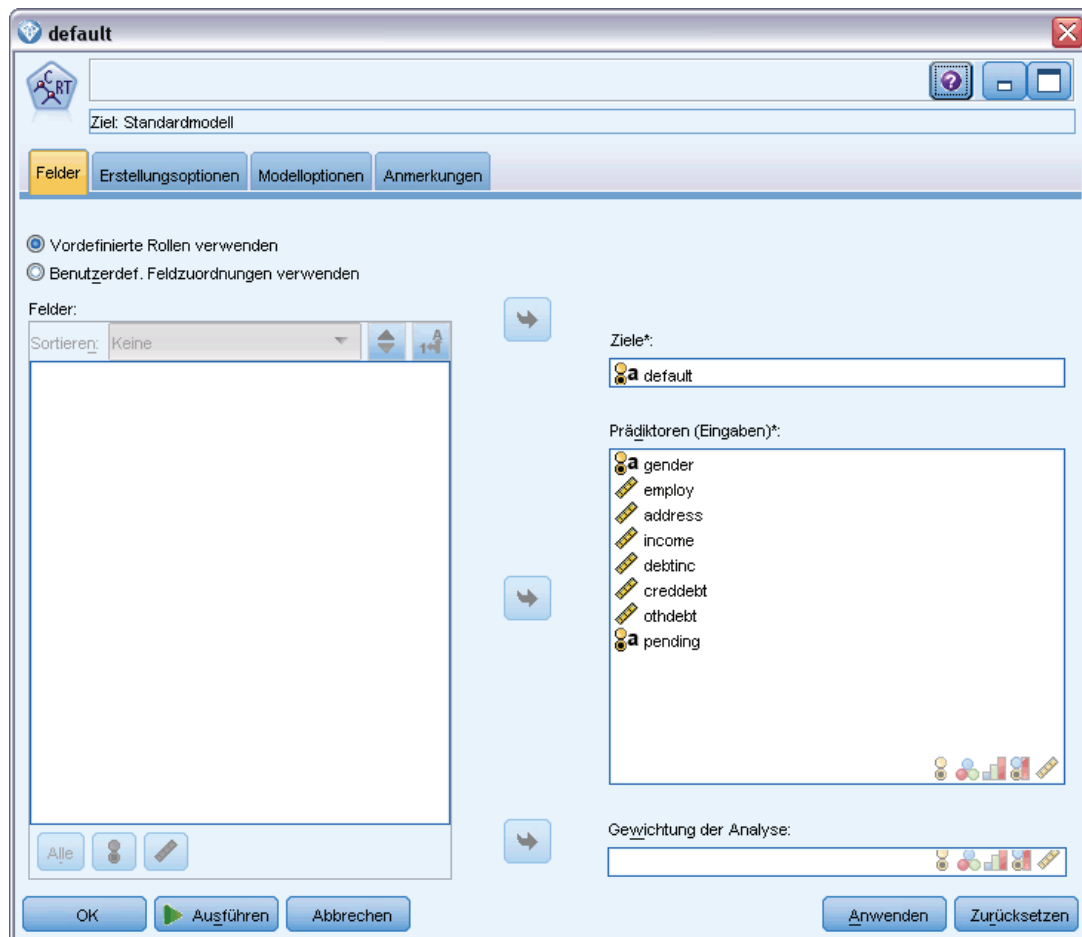
Da das Modell seine Vorhersagen ausschließlich auf der Grundlage der Daten aus der Vergangenheit erstellen darf, enthält der Stream einen Auswahlknoten, der nur diejenigen Kredite mit aufnimmt, die *nicht* als “Pending” (Ausstehend) gekennzeichnet sind. Die 150 ausstehenden Kredite werden somit verworfen.

Abbildung 4-3
Verwerfen der ausstehenden Kreditanträge



Nachdem die ausstehenden Kredite verworfen wurden, werden nur die Details der verbleibenden 700 Kredite, die entweder zurückgezahlt oder nicht zurückgezahlt wurden, an den Modellierungsknoten weitergegeben. Die Bank könnte einen von mehreren SPSS Modeler-Algorithmen verwendet haben, um ein gutes Modell zu erstellen. In diesem Fall wurde ein C&R-Baumknoten verwendet, mit dem ein Modell erstellt wird, das die wahrscheinlichen Kreditausfälle auf der Grundlage der früheren Leistung der Bankkunden vorhersagt.

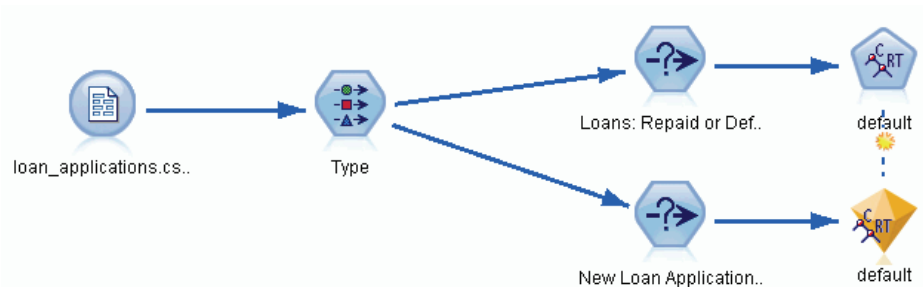
Abbildung 4-4
Zuweisen von Prädiktor- und Zielfeldern



Die zur Erstellung der Vorhersage verwendeten Felder werden als Prädiktorfelder gekennzeichnet und das Feld, dessen Wert mit dem Modell vorhergesagt werden soll — in diesem Fall default (Zahlungsausfall) —, wird als Zielfeld festgelegt, wie zuvor vom Typknoten definiert.

Bei der Ausführung dieses Streams wird ein Modell-Nugget erstellt, das das aus den Prädiktorfeldern erstellte Modell enthält.

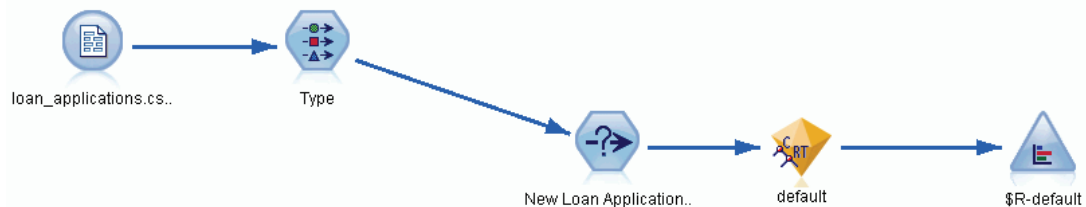
Abbildung 4-5
Stream mit hinzugefügtem Modell-Nugget



Nun kann der Analyst der Bank das Modell verwenden, um mit der Vorhersage zu beginnen, ob es bei Kunden mit ausstehenden Kreditanträgen mit hoher Wahrscheinlichkeit zu einem Zahlungsausfall kommt. Der Analyst verwendet das ursprüngliche Daten-Set und fügt einen Auswahlknoten ein, der dieses Mal nur die 150 als “Pending” (Ausstehend) gekennzeichneten Datensätze enthält, die zuvor verworfen worden waren. Der Analyst leitet diese Datensätze direkt in das Modell weiter und fügt dabei einen Verteilungsknoten für eine visuelle Darstellung der Vorhersagen des Modells hinzu.

Abbildung 4-6

Stream zur Auswahl neuer Kreditanträge mit hinzugefügten Verteilungsknoten



Der Verteilungsknoten zeigt die Verteilung der Werte des Felds *\$R-default* im Modell. Dieses Feld wird bei der Ausführung vom C&R-Baumknoten zum Datenmodell hinzugefügt. Das Feld enthält die Vorhersage, ob die einzelnen neuen Bewerber ihren Kredit zurückzahlen oder nicht und wir verwenden dieses Feld später, um zu vergleichen, welche Wirkung die Hinzunahme der Entitätsanalyse hat.

Bei der Ausführung dieses Teils des Streams erfährt der Analyst aus der Ausgabe des Verteilungsknotens, dass 137 der 150 neuen Antragsteller ihre Kredite vermutlich zurückzahlen werden. Für die verbleibenden 13 wird ein Zahlungsausfall vorhergesagt, weshalb der Analyst der Bank vermutlich empfehlen wird, diese Anträge abzulehnen.

Abbildung 4-7

Ausgabe aus dem Verteilungsknoten ohne Entitätsanalyse

Wert	Anteil	%	Anzahl
Default		8.67	13
Repaid		91.33	137

Hinzunahme der Entitätsanalyse

Betrachten wir nun, ob die Situation verbessert werden kann, wenn eine Entitätsanalyse mit in die Gleichung aufgenommen wird. Stellen Sie sich vor, Sie sind ein Experte auf dem Gebiet der Entitätsanalyse, der von der Bank hinzugezogen wurde, um mögliche betrügerische Einträge in den Kundendatensätzen in den Quelldaten zu untersuchen. Es kann möglicherweise doppelte Datensätze aufgrund von Fehlern beim Dateneintrag geben, es ist jedoch ebenfalls möglich, dass ein Antragsteller für einen Kredit versucht, seine wahre Identität zu verschleiern. In jedem Fall muss die Bank den tatsächlichen Sachverhalt aufdecken.

Im Rahmen dieses Beispiels gehen wir davon aus, dass bereits ein Entitäts-Repository erstellt wurde. Für weitere Informationen siehe Thema [Erstellen des Repository](#) in Kapitel 3 auf S. 16.

Übertragen der Quelldaten in das Repository

Zunächst müssen Sie einen EA-Exportknoten zum Datenquellenknoten hinzufügen, damit Sie die Quelldaten in das Entitäts-Repository exportieren können.

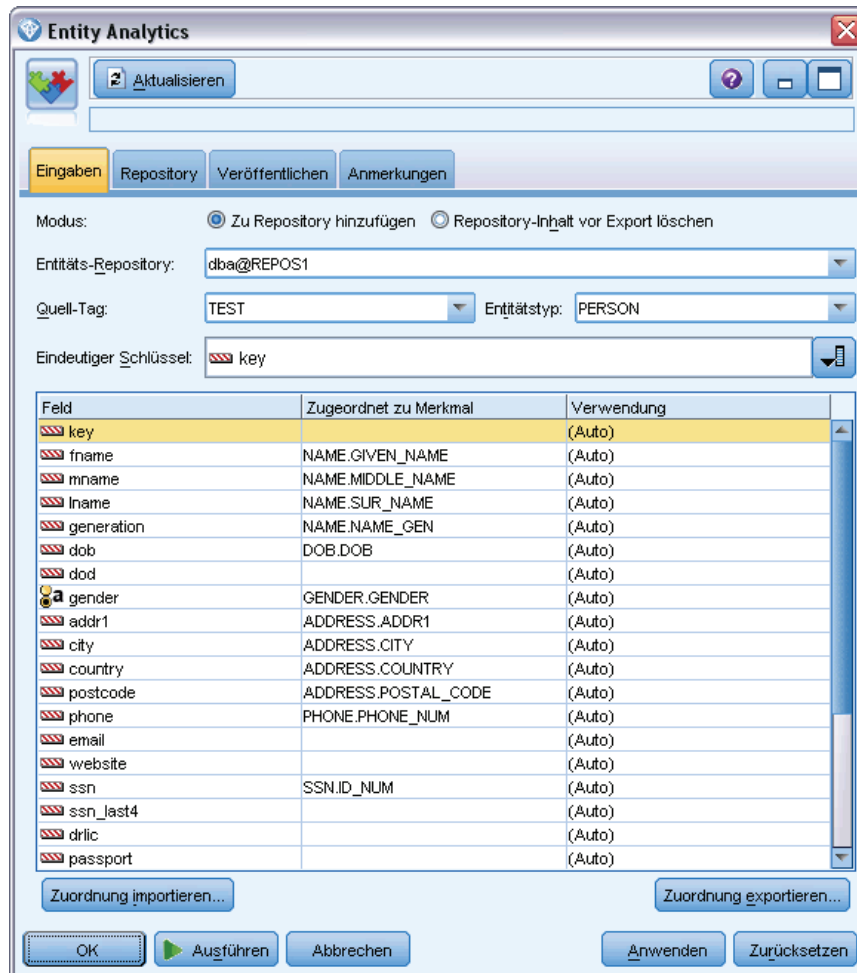
Abbildung 4-8

Verbinden eines EA-Exportknotens mit dem Datenquellenknoten



Bevor Sie die Daten exportieren können, müssen Sie Felder in der Datenquelle Merkmalen im Entitäts-Repository zuordnen. Dies ist notwendig, da verschiedene Datenquellen möglicherweise verschiedene Feldnamen für dieselbe Art von Informationen verwenden. Das Entitätsrepository stellt eine Standardmenge an Informationstypen (so genannte "Merkmale") bereit, um Dopplungen zu vermeiden.

Abbildung 4-9
Erstellen einer Zuordnung zwischen Feldern und Merkmalen



Richten Sie im EA-Exportknoten die Details zum Repository ein: die Verbindungsdetails, das Quell-Tag (in diesem Fall zur Angabe der Datenquelle – TEST), den Entitätstyp (das verwendete Merkmals-Set – dasjenige mit dem Namen PERSON), und das Feld mit dem eindeutigen Schlüssel (zur eindeutigen Angabe der einzelnen Datensätze). Verwenden Sie in diesem Fall das Feld key (Schlüssel) als eindeutigen Schlüssel.

Nun können Sie die Zuordnungen einrichten. In dem verwendeten Merkmals-Set gibt es Merkmale, die den Feldern *fname*, *mname*, *lname*, *generation*, *dob*, *gender*, *addr1*, *city*, *country*, *postcode*, *phone*, *email*, *ssn*, *drlic* und *passport* entsprechen.

Beginnen Sie, indem Sie die Zuordnung für *fname* festlegen. Doppelklicken Sie in der Spalte Zugeordnet zu Merkmal der Tabelle auf die Zeile *fname*, blättern Sie nach unten bis zum Eintrag NAME.GIVEN_NAME und klicken Sie darauf, um die Zuordnung zu erstellen.

Ordnen Sie nun die restlichen Felder, die entsprechende Merkmale aufweisen zu, sodass die vollständige Menge der Zuordnungen wie folgt aussieht.

Tabelle 4-1
Repository-Merkmalen zugeordnete Felder

Feld	Zugeordnet zu Merkmal
<i>fname</i>	NAME.GIVEN_NAME
<i>mname</i>	NAME.MIDDLE_NAME
<i>lname</i>	NAME.SUR_NAME
<i>generation</i>	NAME.NAME_GEN
<i>dob</i>	DOB.DOB
<i>gender</i>	GENDER.GENDER
<i>addr1</i>	ADDRESS.ADDR1
<i>city</i>	ADDRESS.CITY
<i>country</i>	ADDRESS.COUNTRY
<i>postcode</i>	ADDRESS.POSTAL_CODE
<i>phone</i>	PHONE.PHONE_NUM
<i>email</i>	EMAIL_ADDR.ADDR
<i>ssn</i>	SSN.ID_NUM
<i>drlic</i>	DRLIC.ID_NUM
<i>passport</i>	PASSPORT.ID_NUM

Klicken Sie auf **Ausführen**, um die Daten in das Repository zu exportieren. Dies nimmt eine gewisse Zeit in Anspruch. Wenn das Dialogfeld "Ausführungs-Feedback" automatisch geschlossen wird, ist der Exportvorgang abgeschlossen.

Lesen der aufgelösten Identitäten

Beim Export der Daten in das Repository beginnt das Entitätsanalyzesystem mit der Auflösung möglicher Identitätskonflikte, indem es eine eindeutige Entitäts-ID zuweist, die später als Feld *\$EA-ID* angezeigt wird. (*Hinweis*: Dies ist nicht dasselbe Feld wie das Feld "Eindeutiger Schlüssel" im EA-Exportknoten, das zur eindeutigen Kennzeichnung von Datenquellen-Datensätzen dient.)

Der erste Schritt beim Lesen der aufgelösten Identitäten besteht darin, einen Entity Analytics-(EA-)Quellenknoten zum Stream hinzuzufügen. Dieser Quellenknoten sollte in dieser Phase noch völlig unverbunden sein.

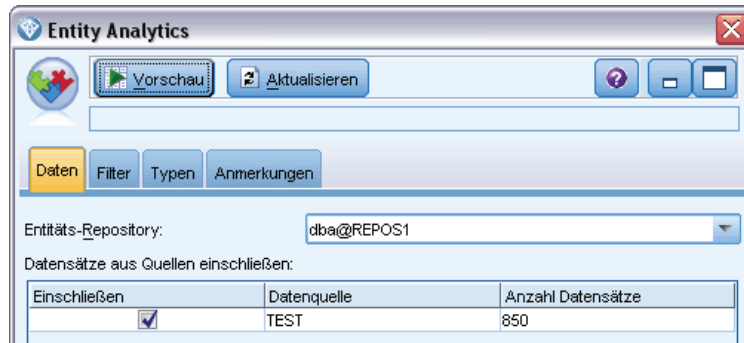
Abbildung 4-10
Entity Analytics-(EA-)Quellenknoten



Entity Analytics(EA)

Öffnen Sie den Entity Analytics-(EA-)Quellenknoten und legen Sie die Details für das Entity-Repository fest. Es wird dann eine Liste der Datenquellen angezeigt, die in das Repository exportiert wurden. – Im vorliegenden Fall gibt es nur eine einzige.

Abbildung 4-11
Auswählen einer Datenquelle im Repository



Aktivieren Sie das Kontrollkästchen für die Datenquelle TEST und klicken Sie auf “OK”.

Betrachten wir nun, wie das Entitätsanalyzesystem die Daten bearbeitet hat. Verbinden Sie einen Tabellenknoten mit dem Entity Analytics-(EA-)Quellenknoten, öffnen Sie den Tabellenknoten und klicken Sie auf Ausführen, um das Ausgabefenster des Tabellenknotens anzuzeigen.

Abbildung 4-12
Ausgabe aus dem Tabellenknoten

The screenshot shows the 'Table (3 Felder, 850 Datensätze)' window. It displays a table with the following columns: '\$EA-ID', '\$EA-SRC', and '\$EA-KEY'. The table contains 20 rows of data:

	\$EA-ID	\$EA-SRC	\$EA-KEY
1	1	TEST	REF00000001
2	2	TEST	REF00000002
3	3	TEST	REF00000003
4	4	TEST	REF00000004
5	5	TEST	REF00000005
6	6	TEST	REF00000006
7	7	TEST	REF00000007
8	8	TEST	REF00000008
9	9	TEST	REF00000009
10	10	TEST	REF00000010
11	11	TEST	REF00000011
12	12	TEST	REF00000012
13	13	TEST	REF00000013
14	14	TEST	REF00000014
15	15	TEST	REF00000015
16	16	TEST	REF00000016
17	17	TEST	REF00000017
18	18	TEST	REF00000018
19	19	TEST	REF00000019
20	20	TEST	REF00000020

Nur ein einziges Feld sieht bekannt aus – das mit der Beschriftung *\$EA-KEY*. Dabei handelt es sich tatsächlich um das Feld *key* (Schlüssel) aus den Quelldaten und es ist enthalten, weil Sie es im EA-Exportknoten als Feld “Eindeutiger Schlüssel” ausgewählt hatten.

Das System hat jedoch noch zwei weitere Felder hinzugefügt. Das Feld *\$EA-ID* ist nicht die eindeutige ID der Quelldatensätze, sondern die eindeutige ID der aufgelösten Identitäten. Der Unterschied wird in Kürze deutlich. Das Feld *\$EA-SRC* gibt an, woher die Daten stammen. – Hier steht TEST, da dies das Quell-Tag ist, das Sie im EA-Exportknoten zugewiesen haben.

Was ist mit den ganzen anderen Feldern in den Quelldaten geschehen? Keine Sorge, sie befinden sich noch immer im Repository. – Es ist nur so, dass der Entity Analytics-(EA-)Quellenknoten aus Leistungsgründen nur das Mindest-Set an Feldern zur weiteren Verarbeitung nach unten in den Stream weitergibt.

Blättern Sie nun in der Ausgabe des Tabellenknotens nach unten bis zur Zeile 385.

Abbildung 4-13

Unterschied zwischen Tabellen-Ausgabezeilen und *\$EA-ID*-Nummern

	<i>\$EA-ID</i>	<i>\$EA-SRC</i>	<i>\$EA-KEY</i>
375	375	TEST	REF00000375
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

Beachten Sie, dass die *\$EA-ID*-Nummer hier nicht in der richtigen Reihenfolge zu stehen scheint. Das Entitätsanalysessystem hat ermittelt, dass Datensatz REF00000385 auf die Person verweist, die als Entität 45 ermittelt wurde und die außerdem Datensatz REF00000045 aufweist. Wenn wir in der Ausgabe weiter nach unten blättern, gibt es weitere Nummern mit falscher Reihenfolge, beispielsweise in den Zeilen 485, 517, 520 usw. Wir sehen uns das besser ein wenig genauer an.

Heben wir zunächst die Tatsache hervor, dass das Feld *\$EA-KEY* die Daten aus dem Feld *key* (Schlüssel) in den Quelldaten enthält, indem wir es in *key* umbenennen. Verbinden Sie einen Filterknoten mit dem Entity Analytics-(EA-)Quellenknoten und öffnen Sie den Filterknoten. Doppelklicken Sie auf die Zeichenfolge *\$EA-KEY* in der zweiten Spalte Feld und geben Sie *key* ein.

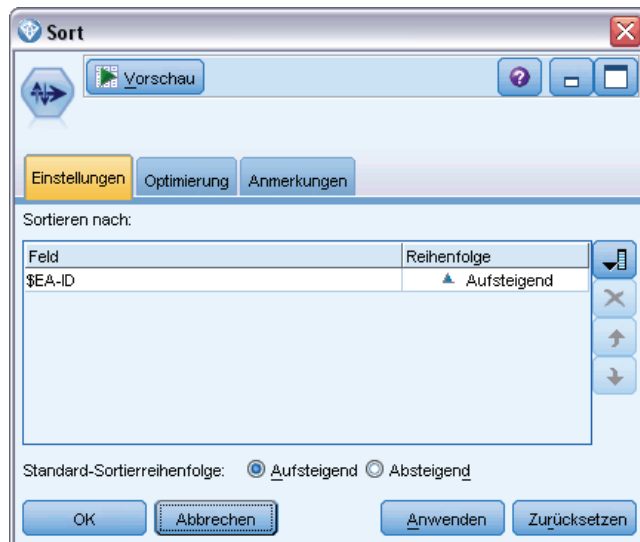
Abbildung 4-14
Umbenennen des Felds "\$EA-KEY"



Klicken Sie auf OK, um den Filterknoten zu schließen.

Nun müssen wir die *\$EA-ID*-Entitäts-IDs in aufsteigender Reihenfolge sortieren. Verbinden Sie einen Sortierknoten mit dem Filterknoten. Öffnen Sie den Sortierknoten, klicken Sie auf die oberste Schaltfläche neben der Tabelle Sortieren nach, wählen Sie *\$EA-ID* und klicken Sie auf OK.

Abbildung 4-15
Sortieren der Entitäts-IDs in aufsteigender Reihenfolge



Behalten Sie die Sortierreihenfolge Aufsteigend bei und klicken Sie auf OK.

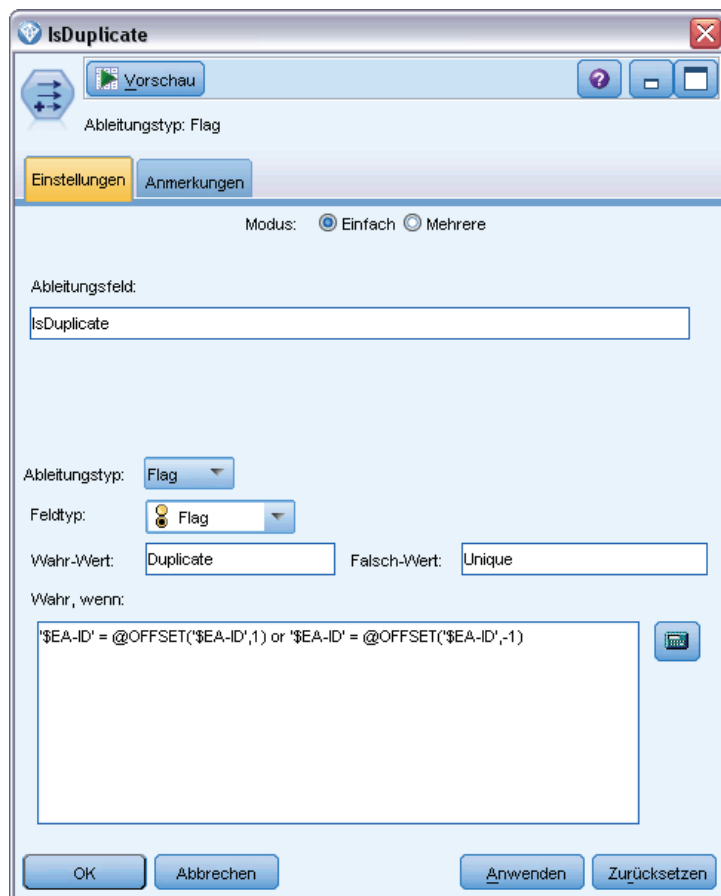
Nun müssen Sie ein zusätzliches Feld erstellen, das angibt, ob ein Datensatz eindeutig oder ein Duplikat ist. Verbinden Sie einen Ableitungsknoten mit dem Sortierknoten. Öffnen Sie den Ableitungsknoten und setzen Sie den Wert für Ableitungsfeld auf `IsDuplicate`. Wählen Sie in der Liste Ableitungstyp die Option `Flag`. Dadurch wird auch Feldtyp auf `Flag` gesetzt. Setzen Sie das Feld Wahr-Wert auf `Duplicate` und das Feld Falsch-Wert auf `Unique`.

Um doppelte Datensätze zu finden, verwenden Sie eine spezielle Sequenzfunktion, `@OFFSET`, die im Lieferumfang von SPSS Modeler enthalten ist.

Geben Sie Folgendes in das Feld Wenn ein:

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

Abbildung 4-16
Festlegen der Bedingung im Ableitungsknoten



Wenn die Entitäts-IDs in aufsteigender Reihenfolge sortiert sind, können Sie mit der Funktion `@OFFSET` überprüfen, ob untereinander stehende Entitäts-IDs identisch sind, was bedeuten würde, dass es sich bei den Datensätzen um Duplikate handelt. In diesem Fall wird der zugehörige Wert für `IsDuplicate` auf `Duplicate` gesetzt; andernfalls wird er auf `Unique` gesetzt.

Klicken Sie auf `OK`, um den Knoten zu schließen.

Um die Wirkung des Ableitungsknotens anzuzeigen, verbinden Sie einen Tabellenknoten mit dem Ableitungsknoten, öffnen Sie den Tabellenknoten und klicken Sie auf Ausführen. Blättern Sie im Ausgabefenster des Tabellenknoten nach unten bis zur Zeile 45.

Abbildung 4-17
Ausgabe aus dem Ableitungsknoten

	\$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

Erinnern Sie sich an die Anzeige der Ausgabe direkt aus dem Entity Analytics-(EA-)Quellenknoten. Das System hatte bereits ermittelt, dass Datensatz REF00000385 dieselbe Person referenziert wie Entität 45. Nun sind wir dabei einen Schritt weitergegangen und haben die Tatsache, dass es sich bei den Datensätzen REF00000045 und REF00000385 um Duplikate handelt, da sie beide Entität 45 referenzieren, durch ein Flag gekennzeichnet.

Blättern Sie im Ausgabefenster weiter nach unten, um die anderen Datensätze zu sehen, die als Duplikate gekennzeichnet sind.

Um einen Bericht zu erhalten, in dem die doppelten Datensätze aufgeführt sind, verbinden Sie einen Berichtsknoten (von der Registerkarte "Ausgabe" der Knotenpalette) mit dem Ableitungsknoten *IsDuplicate*. Öffnen Sie den Berichtsknoten, kopieren Sie den folgenden Text in das Eingabefeld der Registerkarte "Vorlage" und klicken Sie auf Ausführen.

```
<html>
<h1>Liste der doppelten Kundendatensätze.

<h2>Dieser Bericht wurde generiert am: [@TODAY]

<h2>Duplicate records
<table>
```

```
<tr>
  <td>Entity ID</td>
  <td>Key</td>
</tr>

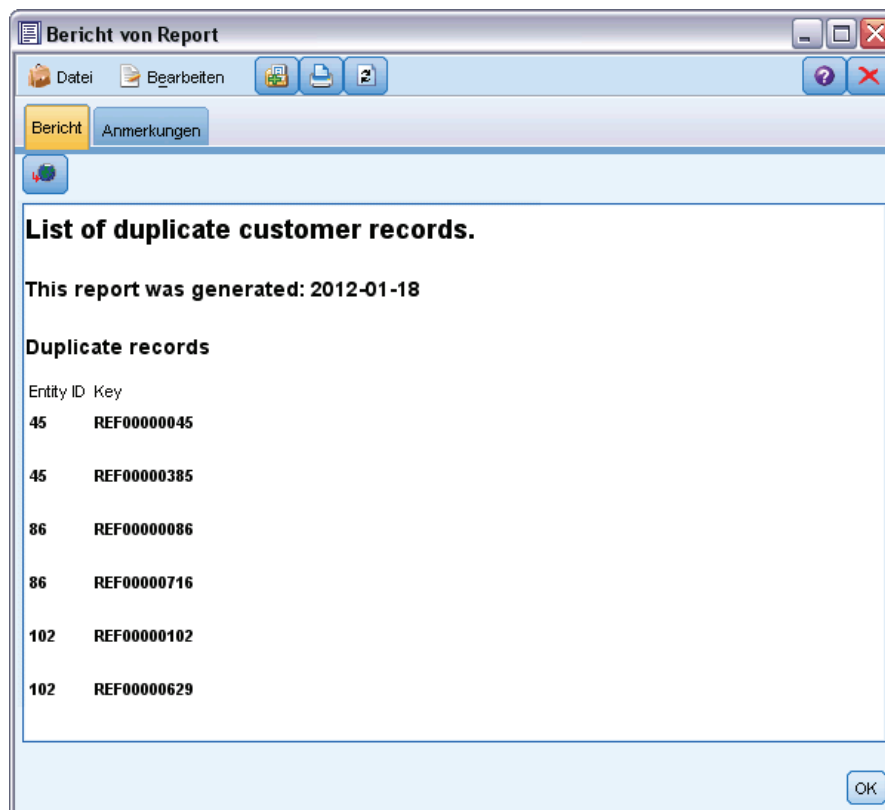
#WHERE IsDuplicate = "Duplicate"

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

</html>
```

Dadurch ergibt sich die folgende Ausgabe:

Abbildung 4-18
Ausgabe aus dem Berichtsknoten



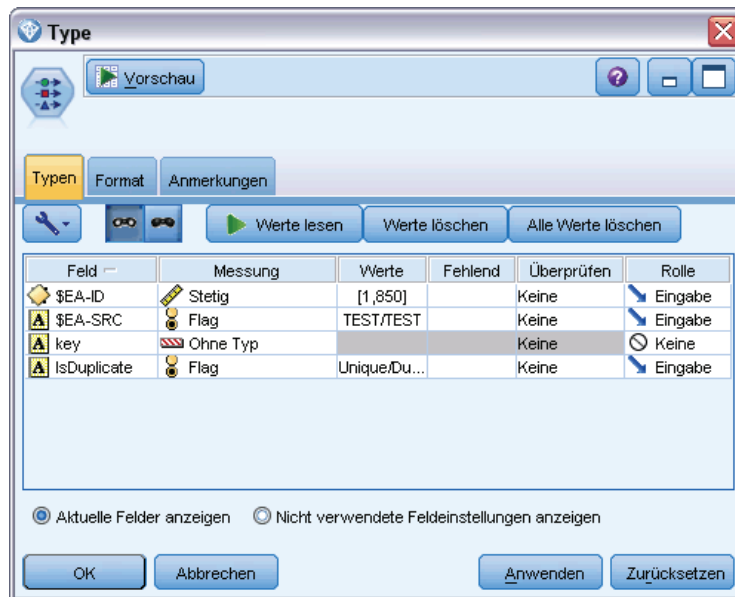
In diesem Fall wird für den Bericht HTML-Format verwendet. Sie könnten jedoch auch XML- oder ASCII-Formatierung verwenden.

Vergleich der Entitätsanalyseausgabe mit dem ursprünglichen Modell

Der letzte Arbeitsschritt in diesem Beispiel besteht darin zu ermitteln, ob sich die Vorhersage durch die Hinzunahme einer Entitätsanalyse gegenüber der ursprünglichen Vorhersage der Bank verändert. Sie erinnern sich vielleicht, dass durch das ursprüngliche Modell 13 Kreditausfälle bei den 150 ausstehenden Anträgen vorhergesagt wurden. Sie verwenden nun einen Zusammenführungsknoten ("Mergen"), um die Ausgabe aus diesem Modell mit Informationen zu doppelten Datensätzen aus der Entitätsanalyse zusammenzuführen, um zu sehen, ob sich die Vorhersage dadurch ändert.

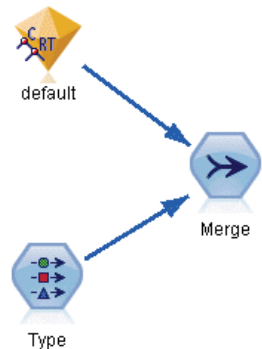
Zunächst müssen Sie sicherstellen, dass die durch die Entitätsanalyse neu hinzugekommenen Felder die richtigen Datentypen (bzw. *Messniveaus* laut SPSS Modeler Terminologie) aufweisen. Fügen Sie einen Typknoten zum Ableitungsknoten IsDuplicate hinzu, öffnen Sie den Typknoten und klicken Sie auf die Schaltfläche Werte lesen.

Abbildung 4-19
Einstellungen für den Typknoten



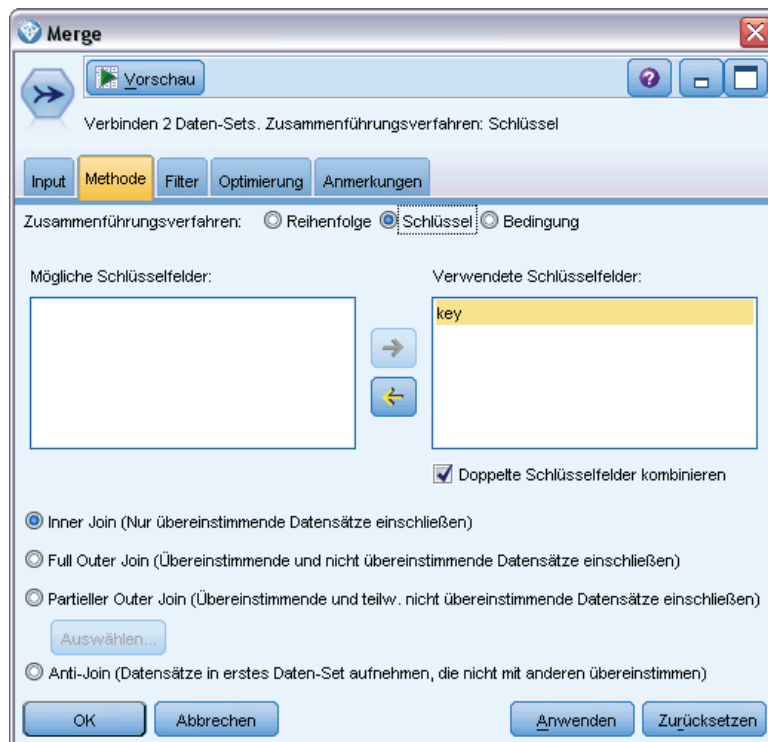
Sie können nun den Zusammenführungsknoten hinzufügen. Verbinden Sie ihn mit dem Typknoten und auch mit dem goldenen Nugget, das das ursprüngliche Modell enthält. Klicken Sie dazu mit der rechten Maustaste auf das goldene Nugget, wählen Sie Verbinden und klicken Sie dann auf den Zusammenführungsknoten, der nun zwei Eingabepfeile aufweisen sollte.

Abbildung 4-20
Eingaben für den Zusammenführungsknoten ("Mergen")



Öffnen Sie den Zusammenführungsknoten, setzen Sie das Zusammenführungsverfahren auf Schlüssel und klicken Sie auf die Schaltfläche mit dem Rechtspfeil, um das Feld key aus dem Bereich Mögliche Schlüsselfelder in den Bereich Verwendete Schlüsselfelder zu verschieben, und klicken Sie dann auf OK.

Abbildung 4-21
Angabe des Schlüsselfelds für die Zusammenführung



Sie sind nun fast so weit, dass Sie den Vergleich durchführen können. Wenn Sie nun jedoch einen Verteilungsknoten hinzufügen und ausführen würden, wären keinerlei Änderungen gegenüber der ursprünglichen Vorhersage zu bemerken. Auch wenn der Stream nun die Ausgaben aus dem ursprünglichen Modell-Nugget mit den neuen Feldern zusammenführt, die von der Entitätsanalyse

erstellt wurden, wurde das Vorhersagefeld selbst (*\$R-default*) im Datenmodell noch nicht mit den neuen Informationen aktualisiert.

Dazu verwenden Sie einen Füllerknoten, der Feldwerte ersetzen kann. Verbinden Sie einen Füllerknoten mit dem Zusammenführungsknoten und öffnen Sie den Füllerknoten.

Klicken Sie auf die obere Schaltfläche rechts neben Felder ausfüllen, blättern Sie ans Ende der Liste, wählen Sie *\$R-default* aus und klicken Sie auf OK. Das ist das Feld, dessen Werte geändert werden sollen, wenn die im restlichen Dialogfeld angegebene Bedingung erfüllt ist.

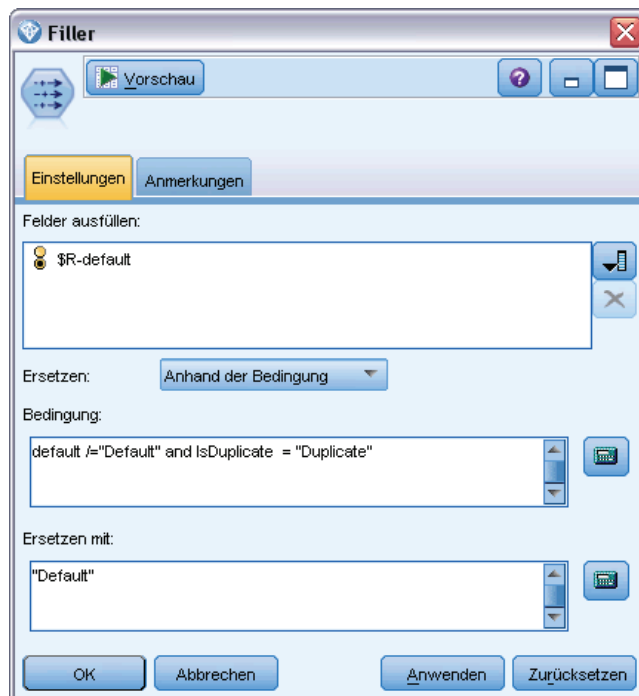
Vergewissern Sie sich zur Angabe der Bedingung, dass Ersetzen anhand der Bedingung gesetzt ist, und geben Sie dann im Feld Bedingung Folgendes ein:

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

Geben Sie Folgendes in das Feld Ersetzen mit ein:

```
"default"
```

Abbildung 4-22
Angabe der Bedingung für das Ersetzen von Feldwerten



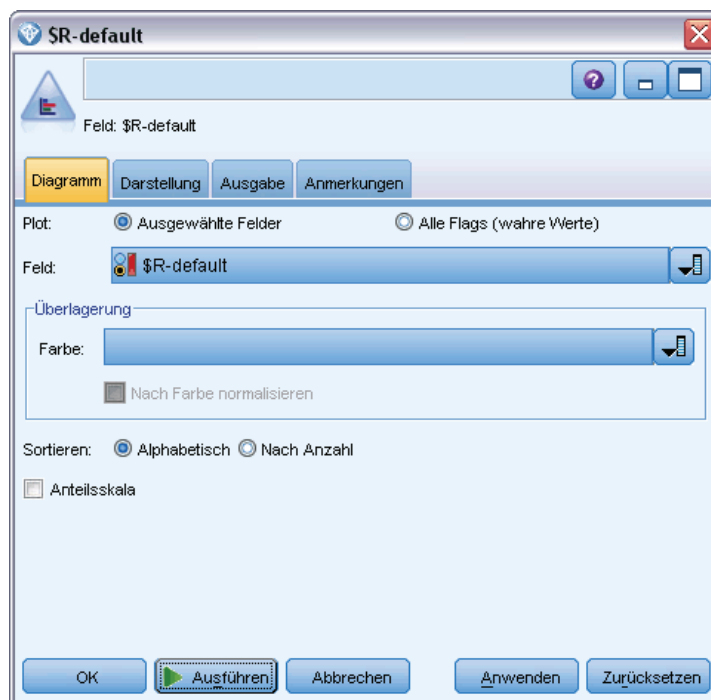
Diese Einstellungen müssen ein wenig erläutert werden. Die Bedingung gibt an, dass für jeden Datensatz, bei dem der Wert des Felds *default* (Zahlungsausfall) aus den ursprünglichen Daten nicht gleich Default ist und bei dem der Datensatz als Duplikat gekennzeichnet wurde, der Wert des Felds *\$R-default* im Modell auf Default“ gesetzt wird.

Das Feld *\$R-default* ist das Feld im Modell, das die Vorhersage enthält, ob ein Kunde den Kredit vermutlich nicht zurückzahlen wird. Auf diese Weise werden Kunden mit doppelten Datensätzen als Kunden mit möglichem Zahlungsausfall zum Modell hinzugefügt.

Klicken Sie auf OK, um den Füllerknoten zu schließen.

Nun können Sie schließlich anzeigen, welche Unterschiede sich durch die Entitätsanalyse ergeben haben. Verbinden Sie in der Diagrammpalette einen Verteilungsknoten mit dem Füllerknoten und öffnen Sie den Verteilungsknoten. Klicken Sie auf die Liste Feld und wählen Sie die Option *\$R-default*.

Abbildung 4-23
Einstellungen für den Verteilungsknoten



Klicken Sie auf Ausführen, um das Diagramm der neuen Vorhersage zu erstellen.

Abbildung 4-24
Ausgabe aus dem Verteilungsknoten nach der Entitätsanalyse

Wert	Anteil	%	Anzahl
Default		10.67	16
Repaid		89.33	134

Es gibt nun 16 risikoreiche Anträge statt bisher 13. Diese zusätzlichen Anträge könnten sehr kostspielig werden, wenn es tatsächlich zu einem Kreditausfall kommen sollte, sodass Sie der Bank den Nutzen einer Ergänzung ihrer Risikobewertung durch Entitätsanalyse grafisch demonstrieren können.

Zusammenfassung

Dieses Beispiel hat gezeigt, wie Sie durch Entitätsanalyse Dopplungen von Datensätzen bei Daten zu Personen bzw. Organisationen ausmerzen und dadurch die Vorhersagequalität verbessern können.

Hinweis: Idealerweise sollten doppelte Datensätze vor allen sonstigen Verarbeitungsschritten entfernt werden. Sie könnten dies weiter ausbauen, indem Sie den Knoten “Automated Data Preparation” (ADP) verwenden, um Ihre Daten zu analysieren und Korrekturen zu identifizieren, problematische oder vermutlich überflüssige Felder auszuschließen, bei Bedarf neue Attribute abzuleiten und die Leistung durch intelligente Screeningverfahren zu verbessern.

Durch eine Kombination aus Entitätsanalyse und automatischer Datenvorbereitung wird sichergestellt, dass Sie mit möglichst sauberen Daten arbeiten.

Hinweise

Diese Informationen wurden für weltweit angebotene Produkte und Dienstleistungen erarbeitet.

IBM bietet die in diesem Dokument behandelten Produkte, Dienstleistungen oder Merkmale möglicherweise nicht in anderen Ländern an. Informationen zu den derzeit in Ihrem Land erhältlichen Produkten und Dienstleistungen erhalten Sie bei Ihrem zuständigen IBM-Mitarbeiter vor Ort. Mit etwaigen Verweisen auf Produkte, Programme oder Dienste von IBM soll nicht behauptet oder impliziert werden, dass nur das betreffende Produkt oder Programm bzw. der betreffende Dienst von IBM verwendet werden kann. Stattdessen können alle funktional gleichwertigen Produkte, Programme oder Dienste verwendet werden, die keine geistigen Eigentumsrechte von IBM verletzen. Es obliegt jedoch der Verantwortung des Benutzers, die Funktionsweise von Produkten, Programmen oder Diensten von Drittanbietern zu bewerten und zu überprüfen.

IBM verfügt möglicherweise über Patente oder hat Patentanträge gestellt, die sich auf in diesem Dokument beschriebene Inhalte beziehen. Durch die Bereitstellung dieses Dokuments werden Ihnen keinerlei Lizenzen an diesen Patenten gewährt. Lizenzanfragen können schriftlich an folgende Adresse gesendet werden:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, U.S.A.

Bei Lizenzanfragen in Bezug auf DBCS-Daten (Double-Byte Character Set) wenden Sie sich an die für geistiges Eigentum zuständige Abteilung von IBM in Ihrem Land. Schriftliche Anfragen können Sie auch an folgende Adresse senden:

Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Japan.

Der folgende Abschnitt findet in Großbritannien und anderen Ländern keine Anwendung, in denen solche Bestimmungen nicht mit der örtlichen Gesetzgebung vereinbar sind: INTERNATIONAL BUSINESS MACHINES STELLT DIESE VERÖFFENTLICHUNG IN DER VERFÜGBAREN FORM OHNE GARANTIEN BEREIT, SEIEN ES AUSDRÜCKLICHE ODER STILLSCHWEIGENDE, EINSCHLIESSLICH JEDOCH NICHT NUR DER GARANTIEN BEZÜGLICH DER NICHT-RECHTSVERLETZUNG, DER GÜTE UND DER EIGNUNG FÜR EINEN BESTIMMTEN ZWECK. Manche Rechtsprechungen lassen den Ausschluss ausdrücklicher oder implizierter Garantien bei bestimmten Transaktionen nicht zu, sodass die oben genannte Ausschlussklausel möglicherweise nicht für Sie relevant ist.

Diese Informationen können technische Ungenauigkeiten oder typografische Fehler aufweisen. An den hierin enthaltenen Informationen werden regelmäßig Änderungen vorgenommen. Diese Änderungen werden in neuen Ausgaben der Veröffentlichung aufgenommen. IBM kann jederzeit und ohne vorherige Ankündigung Optimierungen und/oder Änderungen an den Produkten und/oder Programmen vornehmen, die in dieser Veröffentlichung beschrieben werden.

Jegliche Verweise auf Drittanbieter-Websites in dieser Information werden nur der Vollständigkeit halber bereitgestellt und dienen nicht als Befürwortung dieser. Das Material auf diesen Websites ist kein Bestandteil des Materials zu diesem IBM-Produkt und die Verwendung erfolgt auf eigene Gefahr.

IBM kann die von Ihnen angegebenen Informationen verwenden oder weitergeben, wie dies angemessen erscheint, ohne Ihnen gegenüber eine Verpflichtung einzugehen.

Lizenznehmer dieses Programms, die Informationen dazu benötigen, wie (i) der Austausch von Informationen zwischen unabhängig erstellten Programmen und anderen Programmen und (ii) die gegenseitige Verwendung dieser ausgetauschten Informationen ermöglicht wird, wenden sich an:

IBM Software Group, Attention: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

Derartige Informationen stehen ggf. in Abhängigkeit von den jeweiligen Geschäftsbedingungen sowie in einigen Fällen der Zahlung einer Gebühr zur Verfügung.

Das in diesem Dokument beschriebene lizenzierte Programm und sämtliche dafür verfügbaren lizenzierten Materialien werden von IBM gemäß dem IBM-Kundenvertrag, den Internationalen Nutzungsbedingungen für Programmpakete der IBM oder einer anderen zwischen uns getroffenen Vereinbarung bereitgestellt.

Jegliche hier enthaltene Daten zur Leistung wurden in einer überwachten Umgebung ermittelt. Aus diesem Grund können in anderen Betriebsumgebungen gewonnene Ergebnisse stark davon abweichen. Einige Messungen wurden unter Umständen auf Systemen im Entwicklungsstadium durchgeführt und es kann nicht garantiert werden, dass diese Messungen auf allgemein verfügbaren Systemen zum gleichen Ergebnis führen. Darüber hinaus wurden einige Messungen unter Umständen durch Extrapolation bestimmt. Die tatsächlichen Ergebnisse können hiervon abweichen. Die Benutzer dieses Dokuments sollten die entsprechenden Daten für die jeweils vorliegende Umgebung prüfen.

Informationen zu Produkten von Drittanbietern wurden von den Anbietern des jeweiligen Produkts, aus deren veröffentlichten Ankündigungen oder anderen, öffentlich verfügbaren Quellen bezogen. IBM hat diese Produkte nicht getestet und kann die Genauigkeit bezüglich Leistung, Kompatibilität oder anderen Behauptungen nicht bestätigen, die sich auf Drittanbieter-Produkte beziehen. Fragen bezüglich der Funktionen von Drittanbieter-Produkten sollten an die Anbieter der jeweiligen Produkte gerichtet werden.

Alle Aussagen bezüglich der zukünftigen Ausrichtung von IBM oder der Absichten des Unternehmens können ohne vorherige Ankündigung geändert oder zurückgenommen werden und stellen lediglich Ziele und Vorgaben dar.

Diese Informationen enthalten Beispiele zu Daten und Berichten, die im täglichen Geschäftsbetrieb Verwendung finden. Um diese so vollständig wie möglich zu illustrieren, umfassen die Beispiele Namen von Personen, Unternehmen, Marken und Produkten. Alle diese Namen sind fiktiv und jegliche Ähnlichkeit mit Namen und Adressen realer Unternehmen ist rein zufällig.

Unter Umständen werden Fotografien und farbige Abbildungen nicht angezeigt, wenn Sie diese Informationen nicht in gedruckter Form verwenden.

Marken

IBM, das IBM-Logo, ibm.com und SPSS sind Marken der IBM Corporation und in vielen Ländern weltweit registriert. Eine aktuelle Liste der IBM-Marken finden Sie im Internet unter <http://www.ibm.com/legal/copytrade.shtml>.

Intel, das Intel-Logo, Intel Inside, das Intel Inside-Logo, Intel Centrino, das Intel Centrino-Logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium und Pentium sind Marken oder eingetragene Marken der Intel Corporation oder der Tochtergesellschaften des Unternehmens in den USA und anderen Ländern.

Linux ist eine eingetragene Marke von Linus Torvalds in den USA, anderen Ländern oder beidem.

Microsoft, Windows, Windows NT und das Windows-Logo sind Marken der Microsoft Corporation in den USA, anderen Ländern oder beidem.

UNIX ist eine eingetragene Marke der The Open Group in den USA und anderen Ländern.

Java und alle Java-basierten Marken sowie Logos sind Marken von Sun Microsystems, Inc. in den USA, anderen Ländern oder beidem.

Andere Produkt- und Servicennamen können Marken von IBM oder anderen Unternehmen sein.



Index

- Aufgelöste Identitäten, Analysieren, 33
- Auflösen von Identitäten, Entitätsanalyse, 9

- Datenquelle, Auswahl für Entitätsanalyse, 34
- Datenquellen
 - Anzeigen für die Entitätsanalyse, 18, 40
 - Verbindung mit einer Datenquelle, 5, 15

- Eindeutige Schlüssel
 - Entitäts-Repository, 20
 - Entitätsanalyse, 7
- Entitäts-Repository, 14
 - Einrichten, 14
 - erstellen, 6, 16–17
 - Inhaltslöschung, 45
 - Konfigurieren, 23, 33
 - Konfigurieren von Port-Zuweisungen, 42
 - löschen, 46
 - Optionen, 18
 - Strukturen, 28
 - Verbindung mit IBM SPSS Modeler, 7
 - Verwaltung, 25
 - Verwaltungsaufgaben, 42
- Entitätsabgleich, Festlegen des Schwellenwerts, 32
- Entitätsanalyse
 - definiert, 1
 - Vergleich mit Vorhersageanalyse, 3
 - Verwenden mit IBM SPSS Modeler, 4
- Entitätstypen
 - Entitäts-Repository, 20
 - Entitätsanalyse, 30
- exportieren
 - Daten in ein Entitäts-Repository, 9
- Exportknoten
 - Entity Analytics, 9

- Identitätsauflösung, Entitätsanalyse, 9
- Inhaltslöschung
 - Entitäts-Repository, 45

- Knoten
 - Hinzufügen zu einem Entity Analytics-Stream, 35
- Konfiguration
 - Entitäts-Repository, 23, 33

- löschen
 - Entitäts-Repository, 46

- Marken, 69

- Port-Zuweisungen
 - Konfigurieren für die Entitätsanalyse, 42

- Quell-Tags
 - Entitäts-Repository, 20
- Quellenknoten
 - Entity Analytics, 11

- Rechtliche Hinweise, 67
- Repository
 - Entitätsanalyse, 6–7, 9, 14, 16–20, 22–25, 28, 30, 33, 39, 45–46
 - Verwalten der Entitätsanalyse, 42

- Strukturen
 - Entitäts-Repository, 7, 19–20, 22, 24–25, 28, 39

- Typinformationen, Festlegen für die Entitätsanalyse, 35

- umbenennen
 - Datenfelder für Entitätsanalyse, 34

- Verwendungstypen, Entitätsanalyse, 30

- Zuordnen von Feldern
 - Merkmale des Entitäts-Repository, 7, 19–20, 22, 24, 39