

Spectrum™ Technology Platform

Version 0-SNAPSHOT

Data Quality-Handbuch



Inhalt

1 - Erste Schritte

| | |
|---------------------------------|---|
| Einführung in die Datenqualität | 5 |
|---------------------------------|---|

2 - Parsen

| | |
|---|----|
| Einführung in das Parsing | 8 |
| Definieren von domänenunabhängigen Parsing-Grammatiken in Datenflüssen | 9 |
| Kulturspezifisches Parsen | 10 |
| Analysieren der Ergebnisse vom Parsen | 37 |
| Parsen von Personennamen | 41 |
| Datenflussvorlagen für das Parsen | 43 |

3 - Standardisierung

| | |
|-----------------------------------|----|
| Standardisieren von Begriffen | 59 |
| Standardisieren von Personennamen | 60 |
| Vorlagen für die Standardisierung | 62 |

4 - Vergleichen

| | |
|--|-----|
| Terminologie für Vergleiche | 66 |
| Techniken zur Definition von Vergleichsschlüsseln | 68 |
| Vergleichsregeln | 71 |
| Vergleichen von Datensätzen aus einer einzelnen Quelle | 87 |
| Vergleichen von Datensätzen einer Quelle mit Datensätzen einer anderen Quelle | 93 |
| Vergleichen von Datensätzen zwischen und innerhalb von Quellen | 100 |
| Vergleichen von Datensätzen mit einer Datenbank | 106 |

| | |
|---|-----|
| Vergleichen von Datensätzen mithilfe mehrerer Vergleichsregeln | 109 |
| Erstellen eines allgemeinen Vergleichsdienstes | 113 |
| Verwenden eines Schnellvergleichsschlüssels | 117 |
| Analysieren von Vergleichsergebnissen | 121 |
| Datenflussvorlagen zum Vergleichen | 138 |

5 - Deduplizierung

| | |
|---|-----|
| Herausfiltern von Datensatzdubletten | 147 |
| Erstellen eines „Best of Breed“-Datensatzes | 152 |

6 - Ausnahmedatensätze

| | |
|--|-----|
| Entwerfen eines Datenflusses zur Handhabung von Ausnahmen | 158 |
| Entwerfen eines Datenflusses zur erneuten Überprüfung in Echtzeit | 160 |

7 - Suchtabellen

| | |
|--|-----|
| Einführung in Suchtabellen | 164 |
| Tabellen des Data Normalization-Moduls | 164 |
| Tabellen des Universal Name-Moduls | 170 |
| Anzeigen des Inhalts einer Nachschlagetabelle | 171 |
| Hinzufügen eines Begriffs zu einer Nachschlagetabelle | 172 |
| Entfernen eines Begriffs aus einer Nachschlagetabelle | 173 |
| Ändern der standardisierten Form eines Begriffs | 173 |
| Rückgängigmachen von Tabellenanpassungen | 173 |
| Erstellen einer Nachschlagetabelle | 174 |
| Importieren von Daten | 175 |

8 - Schrittreferenz

| | |
|--------------------------|-----|
| Advanced Matching-Modul | 179 |
| Business Steward-Modul | 254 |
| Data Normalization-Modul | 302 |
| Universal Name-Modul | 322 |

9 - ISO-Ländercodes und Modulunterstützung

| | |
|--|-----|
| ISO-Ländercodes und Modulunterstützung | 360 |
|--|-----|

1 - Erste Schritte

In this section

Einführung in die Datenqualität

5

Einführung in die Datenqualität

Zur Datenqualität gehört es, die Genauigkeit, Aktualität, Vollständigkeit und Konsistenz der von einer Organisation verwendeten Daten sicherzustellen, damit die Daten nutzbar sind. Spectrum™ Technology Platform unterstützt Initiativen für höhere Datenqualität mithilfe der folgenden Funktionen.

Parsen

Parsing ist der Prozess des Analysierens einer Sequenz von Eingabezeichen in einem Feld und des Aufspaltens des Feldes in mehrere Felder. Beispielsweise verfügen Sie möglicherweise über das Feld „Name“, das den Wert „John A. Smith“ enthält. Durch Parsing können Sie das Feld so aufspalten, dass Sie über das Feld „FirstName“ verfügen, das „John“ enthält, über das Feld „MiddleName“, das „A“ enthält“, und über das Feld „LastName“, das „Smith“ enthält.

Standardisierung

Bei der Standardisierung werden Daten desselben Typs einheitlich formatiert. Einige Datentypen, die standardisiert werden können, umfassen Telefonnummern, Datumswerte, Namen, Adressen und Identifikationsnummern. Beispielsweise können Telefonnummern formatiert werden, um nicht numerische Zeichen wie Klammern, Punkte oder Bindestriche zu entfernen.

Sie sollten Ihre Daten standardisieren, bevor Sie Vergleichs- oder Deduplizierungsaktivitäten durchführen, da standardisierte Daten genauer abgeglichen werden als uneinheitlich formatierte Daten.

Vergleichen

Beim Vergleichen werden Datensätze identifiziert, die auf irgendeine Weise zusammenhängen, die für Ihre Zwecke relevant ist. Wenn Sie beispielsweise versuchen, redundante Informationen aus Ihren Kundendaten zu entfernen, können Sie doppelte Datensätze für denselben Kunden identifizieren. Oder: Wenn Sie verhindern möchten, dass doppelte Marketingmaterialien an die gleiche Adresse gehen, können Sie Datensätze von Kunden identifizieren, die im gleichen Haushalt leben.

Deduplizierung

Bei der Deduplizierung werden Datensätze identifiziert, die eine Entität darstellen, aber aus unterschiedlichen Gründen mehrmals in das System eingegeben wurden, manchmal mit geringfügig unterschiedlichen Daten. So kann Ihr System etwa Lieferanteninformationen von verschiedenen Abteilungen Ihrer Organisation enthalten, wobei jede Abteilung eine andere Lieferanten-ID für denselben Lieferanten verwendet. Mit Spectrum™ Technology Platform können Sie diese Datensätze in einem einzelnen Datensatz für jeden Lieferanten konsolidieren.

Überprüfung von Ausnahmedatensätzen

In einigen Fällen können Sie Daten haben, die nicht zuverlässig automatisch verarbeitet und vom Data Steward überprüft werden müssen. Einige Beispiele für Datensätze, die eine manuelle Überprüfung erfordern, sind:

- Fehler bei Adressenüberprüfungen
- Fehler beim Geocoding
- Übereinstimmungen mit niedrigem Zuverlässigkeitsniveau
- Zusammenführungs-/Konsolidierungsentscheidungen

Das Business Steward-Modul besteht aus einer Reihe von Features, mit denen Sie Ausnahmedatensätze identifizieren und auflösen können.

2 - Parsen

In this section

| | |
|--|----|
| Einführung in das Parsing | 8 |
| Definieren von domänenunabhängigen Parsing-Grammatiken in Datenflüssen | 9 |
| Kulturspezifisches Parsen | 10 |
| Analysieren der Ergebnisse vom Parsen | 37 |
| Parsen von Personennamen | 41 |
| Datenflussvorlagen für das Parsen | 43 |

Einführung in das Parsing

Parsing ist der Prozess des Analysierens einer Sequenz von Eingabezeichen in einem Feld und des Aufspaltens des Feldes in mehrere Felder. Beispielsweise verfügen Sie möglicherweise über das Feld „Name“, das den Wert „John A. Smith“ enthält. Durch Parsing können Sie das Feld so aufspalten, dass Sie über das Feld „FirstName“ verfügen, das „John“ enthält, über das Feld „MiddleName“, das „A“ enthält“, und über das Feld „LastName“, das „Smith“ enthält.

Verwenden Sie zum Erstellen eines Datenflusses, der parst, den „Open Parser“-Schritt. Mit Open Parser können Sie Parsing-Regeln schreiben, die als *Grammatiken* bezeichnet werden. Eine Grammatik ist eine Reihe von Ausdrücken, die einer Gruppe von benannten Entitäten, die als *Domänenmuster* bezeichnet werden, eine Sequenz von Zeichen zuordnen. Ein Domänenmuster ist eine Sequenz von einem oder mehreren Token in Ihren Eingabedaten, die Sie als Datenstruktur darstellen möchten, z. B. Name, Adresse oder Kontonummern. Ein Domänenmuster kann aus einer beliebigen Anzahl von Token bestehen, die aus Ihren Eingabedaten geparkt werden können. Ein Domänenmuster wird in der Parsing-Grammatik als <Stamm>-Ausdruck dargestellt. Eingabedaten enthalten oft solche Token in schwierig zu verwendenden oder gemischten Formaten. Beispiel:

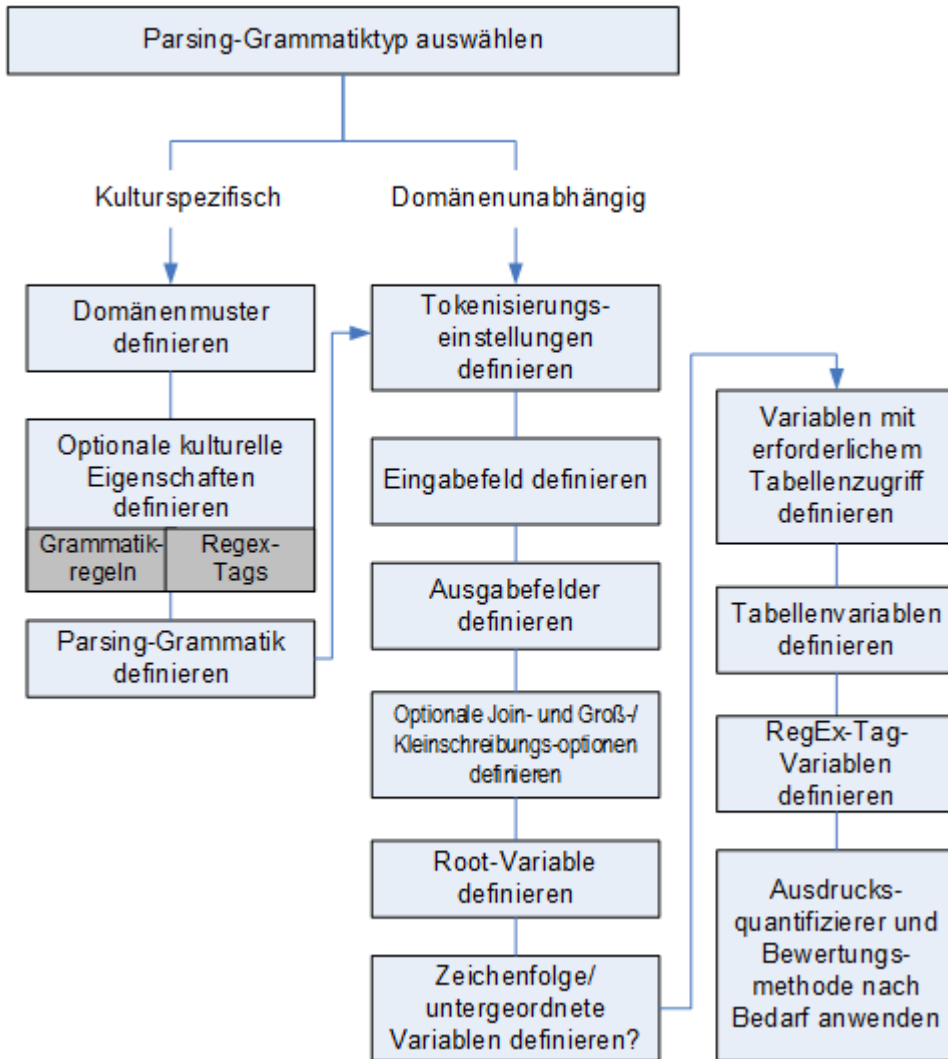
- Ihre Eingabedaten enthalten Namen in einem einzelnen Feld, das Sie in Vorname und Familienname aufteilen möchten.
- Ihre Eingabedaten enthalten Adressen aus mehreren Kulturen und Sie möchten nur die Adressdaten einer bestimmten Kultur extrahieren.
- Ihre Eingabedaten enthalten frei formatierbaren Text, der eingebettete E-Mail-Adressen enthält, und Sie möchten E-Mail-Adressen extrahieren und mit persönlichen Daten verknüpfen und in einer Datenbank speichern.

Es gibt zwei Arten von Grammatiken: die *kulturspezifische* und die *domänenunabhängige* Grammatik. Eine kulturspezifische Parsing-Grammatik ist einer Kultur und/oder Sprache (z. B. Englisch, kanadisches Englisch, Spanisch, mexikanisches Spanisch usw.) und einem bestimmten Datentyp (Telefonnummern, Namen usw.) zugeordnet. Wenn ein „Open Parser“-Schritt für das Durchführen des kulturspezifischen Parsings konfiguriert ist, wird die Parsing-Grammatik der einzelnen Kulturen auf jeden Datensatz angewendet. Die Grammatik mit der besten Parser-Punktzahl (oder die erste mit einer Punktzahl von 100) ist diejenige, deren Ergebnisse zurückgegeben werden. Alternativ können kulturspezifische Parsing-Grammatiken den Wert im Feld „CultureCode“ des Eingabesatzes verwenden und die Daten entsprechend den in der Parsing-Grammatik der Kultur enthaltenen Kultureinstellungen verarbeiten. Kulturspezifische Parsing-Grammatiken können Eigenschaften von einem übergeordneten Element erben. Eine domänenunabhängige Parsing-Grammatik ist weder einer Sprache noch einem bestimmten Datentyp zugeordnet. Domänenunabhängige Parsing-Grammatiken erben keine Eigenschaften von einem übergeordneten Element und ignorieren Informationen aus dem Feld „CultureCode“ in den Eingabedaten.

Open Parser analysiert eine Sequenz von Zeichen in Eingabefeldern und kategorisiert diese durch einen Prozess namens *Tokenisierung* in eine Sequenz von Token. Tokenisierung ist der Prozess der Begrenzung und Klassifizierung von Abschnitten einer Zeichenfolge mit Eingabezeichen in eine

Reihe von Token auf Grundlage von Trennzeichen (auch als Tokenisierungszeichen bezeichnet), wie z. B. Leerzeichen, Bindestrich und andere. Die Token werden dann in von Ihnen angegebenen Ausgabefeldern platziert.

Das folgende Diagramm veranschaulicht den Prozess der Erstellung einer Parsing-Grammatik:



Definieren von domänenunabhängigen Parsing-Grammatiken in Datenflüssen

So definieren Sie domänenunabhängige Parsing-Grammatiken in einem Datenfluss:

1. Fügen Sie im Enterprise Designer einen „Open Parser“-Schritt zu Ihrem Datenfluss hinzu.

2. Doppelklicken Sie auf der Arbeitsfläche auf den „Open Parser“-Schritt.
3. Klicken Sie auf der Registerkarte **Regeln** auf **Domänenunabhängige Grammatik definieren**.
4. Verwenden Sie den Grammatikeditor, um die Grammatikregeln zu erstellen. Sie können Befehle und Variablen in das Textfeld eingeben oder die Befehle auf der Registerkarte **Befehle** verwenden. Weitere Informationen finden Sie unter **Grammatiken** auf Seite 27.
5. Klicken Sie zum Ausschneiden, Kopieren, Einfügen, Suchen und Ersetzen von Textzeichenfolgen in Ihrer Parsing-Grammatik mit der rechten Maustaste in den Grammatikeditor, und wählen Sie den entsprechenden Befehl aus.
6. Klicken Sie zum Überprüfen der von Ihnen erstellten Grammatik auf **Überprüfen**.

Das Überprüfungsfeature listet alle Fehler in Ihrer Grammatiksyntax auf, einschließlich der Zeile und Spalte, in der der Fehler auftritt, eine Beschreibung des Fehlers sowie der Befehlsname oder -wert, der den Fehler verursacht.

7. Klicken Sie auf die Registerkarte **Vorschau**, um die Parsing-Grammatik zu testen.
8. Wenn Sie die Erstellung Ihrer Grammatik abgeschlossen haben, klicken Sie auf **OK**.

Kulturspezifisches Parsen

Definieren einer kulturspezifischen Parsing-Grammatik

Mit einer kulturspezifischen Parsing-Grammatik können Sie verschiedene Parsing-Regeln für verschiedene Sprachen und Kulturen angeben. Dies ermöglicht es Ihnen, in einem einzelnen „Open Parser“-Schritt Daten aus verschiedenen Ländern zu parsen, z. B. Telefonnummern aus den Vereinigten Staaten und Telefonnummern aus Großbritannien. Die einzelnen Eingabedatensätze werden standardmäßig mit der Parsing-Grammatik der jeweiligen Kultur geparkt, und zwar in der Reihenfolge, die im „Open Parser“-Schritt angegeben ist. Wenn Sie möchten, dass die Parsing-Grammatik einer bestimmten Kultur für diesen Datensatz verwendet wird, können Sie auch das Feld „CultureCode“ hinzufügen. Weitere Informationen finden Sie unter **Zuweisen einer Parsing-Kultur zu einem Datensatz** auf Seite 12.

Anmerkung: Wenn Sie eine domänenunabhängige Parsing-Grammatik erstellen möchten, siehe **Definieren von domänenunabhängigen Parsing-Grammatiken in Datenflüssen** auf Seite 9.

1. Rufen Sie im Enterprise Designer die Optionen **Tools > Open Parser Domain Editor** auf.
2. Klicken Sie auf die Registerkarte **Domänen**.
3. Klicken Sie auf **Hinzufügen**.
4. Geben Sie im Feld **Name** einen Domännennamen ein.

5. Geben Sie im Feld **Beschreibung** eine Beschreibung des Domänennamens ein.
6. Klicken Sie auf **OK**, wenn Sie eine neue, leere Domäne erstellen möchten. Gehen Sie wie folgt vor, wenn Sie eine neue Domäne basierend auf einer anderen Domäne erstellen möchten:
 - a) Wählen Sie **Andere Domäne als Vorlage verwenden** aus, wenn Sie eine neue Domäne basierend auf einer anderen Domäne erstellen möchten.
 - b) Wählen Sie eine Domäne aus der Liste aus. Wenn Sie im nächsten Schritt auf **OK** klicken, wird die neue Domäne erstellt. Die neue Domäne enthält die gesamte kulturspezifische Parsing-Grammatik, die in der von Ihnen ausgewählten Domänenvorlage definiert wurde.
 - c) Klicken Sie auf **OK**.
7. Definieren Sie die Parsing-Grammatik für die globale Kultur. Die globale Kultur ist die Standardkultur und wird zum Parsen von Datensätzen verwendet, die eine Kultur aufweisen, für die keine kulturspezifische Parsing-Grammatik definiert worden ist.
 - a) Wählen Sie auf der Registerkarte **Grammatiken** die neue von Ihnen erstellte Domäne aus.
 - b) Wenn Sie aus einer Vorlage eine Domäne erstellt haben, werden möglicherweise bereits Kulturen aufgeführt.
 - Wenn Kulturen aufgeführt werden, wählen Sie **Globale Kultur** aus und klicken Sie anschließend auf **Bearbeiten**.
 - Wenn keine Kulturen aufgeführt werden, klicken Sie auf **Hinzufügen**, wählen Sie **Globale Kultur** aus und klicken Sie anschließend auf **OK**.

- c) Tragen Sie auf der Registerkarte **Grammatik** die Parsing-Grammatik für die globale Kultur ein. Sie können die Registerkarten **Befehle**, **Grammatikregeln** und **RegEx-Tags** verwenden, um vordefinierte Parsing-Grammatikelemente einzufügen. Um ein vordefiniertes Element einzugeben, platzieren Sie den Cursor an der Stelle, an der Sie das Element einfügen möchten, und doppelklicken Sie anschließend auf das Element, das hinzugefügt werden soll.

Auf der Registerkarte **Befehle** werden Parsing-Befehle angezeigt. Für Informationen zu den verfügbaren Befehlen siehe [Grammatiken](#) auf Seite 27.

Auf der Registerkarte **Grammatikregeln** werden die von Ihnen im Dialogfeld „Kultureigenschaften“ erstellten Grammatikregeln angezeigt. Für weitere Informationen zum Erstellen von Grammatikregeln siehe [Definieren der Grammatikregeln einer Kultur](#) auf Seite 32.

Auf der Registerkarte **RegEx-Tags** werden die von Ihnen im Dialogfeld „Kultureigenschaften“ erstellten RegEx-Tags angezeigt. Für weitere Informationen zum Erstellen von RegEx-Tags siehe [Definieren von RegEx-Tags für eine Kultur](#) auf Seite 33.

- d) Klicken Sie auf **Überprüfen**, um die von Ihnen erstellte Syntax zu überprüfen. Das Feature zur Überprüfung der Parsing-Grammatik zeigt alle Fehler in Ihrer Syntax an und umfasst die aufgetretenen Fehler, die Zeile und die Spalte, in denen die Fehler aufgetreten sind, und den Befehl, die Grammatikregel oder den RegEx-Tag, wo die Fehler aufgetreten sind.
- e) Klicken Sie auf die Registerkarte **Vorschau**, um die Ergebnisse Ihrer Grammatik mit Beispieldaten zu testen. Geben Sie unter **Eingabedaten** die Beispieldaten ein, die geparkt werden sollen. Geben Sie einen Datensatz pro Zeile ein. Klicken Sie anschließend auf die

Schaltfläche **Vorschau**. Die geparsten Ausgabefelder werden im Raster **Ergebnisse** angezeigt. Für Informationen zu den Ausgabefeldern siehe **Ausgabe** auf Seite 310. Weitere Informationen zur Ablaufverfolgung finden Sie unter **Ablaufverfolgung von endgültigen Parsing-Ergebnissen** auf Seite 37. Wenn die Ergebnisse nicht Ihren Erwartungen entsprechen, klicken Sie auf die Registerkarte **Grammatiken** und fahren Sie so lange mit der Bearbeitung der Parsing-Grammatik und dem Testen repräsentativer Eingabedaten fort, bis die Parsing-Grammatik die erwarteten Ergebnisse erzeugt.

- f) Klicken Sie auf **OK**, wenn Sie mit dem Definieren der Parsing-Grammatik für die globale Kultur fertig sind.
8. Definieren Sie für jede gewünschte Kultur eine kulturspezifische Grammatik. Um kulturspezifische Grammatiken hinzuzufügen, klicken Sie auf **Hinzufügen** und definieren Sie die Grammatik mit den gleichen Schritten wie für die globale Kultur. Wiederholen Sie diese Schritte nach Bedarf, um so viele Kulturen wie benötigt hinzuzufügen.
9. Wenn Sie mit dem Hinzufügen kulturspezifischer Parsing-Grammatiken fertig sind, klicken Sie auf **OK**.

Die Domäne und die Kulturen, die Sie erstellt haben, können jetzt im „Open Parser“-Schritt zum Durchführen eines Parsings verwendet werden.

Zuweisen einer Parsing-Kultur zu einem Datensatz

Wenn Sie einen „Open Parser“-Schritt für die Verwendung von kulturspezifischen Parsing-Grammatiken konfigurieren, werden die Parsing-Grammatiken für die einzelnen Kulturen in der Reihenfolge auf die einzelnen Eingabedatensätze angewendet, in der die Kulturen im „Open Parser“-Schritt aufgelistet werden. Wenn Sie jedoch eine kulturspezifische Parsing-Grammatik auf einen Datensatz anwenden möchten, können Sie ein Feld mit dem Namen „CultureCode“ hinzufügen. Das Feld muss einen der unterstützten Kulturcodes enthalten, die in der folgenden Tabelle aufgelistet sind.

Kulturcodes

Kulturcodes setzen sich aus einem aus zwei Buchstaben bestehenden Sprachcode in Kleinbuchstaben und einem aus zwei Buchstaben bestehenden Länder- oder Regionscode in Großbuchstaben zusammen. Beispiel: „es-MX“ für Spanisch (Mexiko) und „en-US“ für Englisch (Vereinigte Staaten). In Fällen, in denen kein aus zwei Buchstaben bestehender Sprachcode zur Verfügung steht, wird ein aus drei Buchstaben bestehender Code verwendet, z. B. „uz-Cyrl-UZ“ für Usbekisch (Usbekistan, Kyrillisch). Eine Sprache wird nur durch den zweistelligen Sprachcode in Kleinbuchstaben angegeben. „fr“ gibt beispielsweise die neutrale Kultur für Französisch und „de“ die neutrale Kultur für Deutsch an.

Anmerkung: Es gibt zwei Kulturnamen, die einem anderen Muster folgen. Die Kulturen „zh-Hans“ (Chinesisch, vereinfacht) und „zh-Hant“ (Chinesisch, traditionell) sind neutrale Kulturen. Die Kulturnamen entsprechen dem aktuellen Standard und sollten verwendet

werden, es sei denn, Sie haben einen Grund für die Verwendung der älteren Namen „zh-CHS“ und „zh-CHT“.

Die folgende Tabelle zeigt die unterstützten Kulturcodes.

| Sprache (Kultur/Region) | Kulturcode |
|-------------------------|----------------|
| Globale Kultur | Globale Kultur |
| Afrikaans | af |
| Afrikaans (Südafrika) | af-ZA |
| Albanisch | sq |
| Albanisch (Albanien) | sq-AL |
| Arabisch | ar |
| Arabisch (Algerien) | ar-DZ |
| Arabisch (Bahrain) | ar-BH |
| Arabisch (Ägypten) | ar-EG |
| Arabisch (Irak) | ar-IQ |
| Arabisch (Jordanien) | ar-JO |
| Arabisch (Kuwait) | ar-KW |

| Sprache (Kultur/Region) | Kulturcode |
|-------------------------|------------|
|-------------------------|------------|

| | |
|--------------------|-------|
| Arabisch (Libanon) | ar-LB |
|--------------------|-------|

| | |
|-------------------|-------|
| Arabisch (Libyen) | ar-LY |
|-------------------|-------|

| | |
|--------------------|-------|
| Arabisch (Marokko) | ar-MA |
|--------------------|-------|

| | |
|-----------------|-------|
| Arabisch (Oman) | ar-OM |
|-----------------|-------|

| | |
|------------------|-------|
| Arabisch (Katar) | ar-QA |
|------------------|-------|

| | |
|--------------------------|-------|
| Arabisch (Saudi-Arabien) | ar-SA |
|--------------------------|-------|

| | |
|-------------------|-------|
| Arabisch (Syrien) | ar-SY |
|-------------------|-------|

| | |
|---------------------|-------|
| Arabisch (Tunesien) | ar-TN |
|---------------------|-------|

| | |
|----------------|-------|
| Arabisch (VAE) | ar-AE |
|----------------|-------|

| | |
|------------------|-------|
| Arabisch (Jemen) | ar-YE |
|------------------|-------|

| | |
|-----------|----|
| Armenisch | hy |
|-----------|----|

| | |
|----------------------|-------|
| Armenisch (Armenien) | hy-AM |
|----------------------|-------|

| | |
|-------|----|
| Aseri | az |
|-------|----|

| | |
|------------------------------------|------------|
| Aseri (Aserbaidtschan, Kyrillisch) | az-Cyrl-AZ |
|------------------------------------|------------|

| Sprache (Kultur/Region) | Kulturcode |
|------------------------------------|------------|
| Aseri (Aserbaidtschan, Lateinisch) | az-Latn-AZ |
| Baskisch | eu |
| Baskisch (Baskisch) | eu-ES |
| Weißrussisch | be |
| Weißrussisch (Belarus) | be-BY |
| Bulgarisch | bg |
| Bulgarisch (Bulgarien) | bg-BG |
| katalanisch | ca |
| Katalanisch (Katalanisch) | ca-ES |
| Chinesisch | zh |
| Chinesisch (Hong Kong SAR, VRC) | zh-HK |
| Chinesisch (Macao SVZ) | zh-MO |
| Chinesisch (VRC) | zh-CN |
| Chinesisch (Vereinfacht) | zh-Hans |

| Sprache (Kultur/Region) | Kulturcode |
|-------------------------------------|------------|
| Chinesisch (Singapur) | zh-SG |
| Chinesisch (Taiwan) | zh-TW |
| Chinesisch (Traditionell) | zh-Hant |
| Kroatisch | hr |
| Kroatisch (Kroatien) | hr-HR |
| Tschechisch | cs |
| Tschechisch (Tschechische Republik) | cs-CZ |
| Dänisch | da |
| Dänisch (Dänemark) | da-dk |
| Divehi | dv |
| Divehi (Malediven) | dv-MV |
| Niederländisch | nl |
| Niederländisch (Belgien) | nl-BE |
| Niederländisch (Niederlande) | nl-NL |

| Sprache (Kultur/Region) | Kulturcode |
|-----------------------------------|------------|
| Englisch | de |
| Englisch (Australien) | en-AU |
| Englisch (Belize) | en-BZ |
| Englisch (Kanada) | en-CA |
| Englisch (Karibik) | en-029 |
| Englisch (Irland) | en-IE |
| Englisch (Jamaika) | en-JM |
| Englisch (Neuseeland) | en-NZ |
| Englisch (Philippinen) | en-PH |
| Englisch (Südafrika) | en-ZA |
| Englisch (Trinidad und Tobago) | en-TT |
| Englisch (Vereinigtes Königreich) | en-GB |
| Englisch (Vereinigte Staaten) | en-US |
| Englisch (Simbabwe) | en-ZW |

| Sprache (Kultur/Region) | Kulturcode |
|--------------------------|------------|
| Estnisch | et |
| Estnisch (Estland) | et-EE |
| Färöisch | fo |
| Färöisch (Färöer Inseln) | fo-FO |
| Farsi | fa |
| Farsi (Iran) | fa-IR |
| Finnisch | fi |
| Finnisch (Finnland) | fi-FI |
| Französisch | fr |
| Französisch (Belgien) | fr-BE |
| Französisch (Kanada) | fr-CA |
| Französisch (Frankreich) | fr-FR |
| Französisch (Luxemburg) | fr-LU |
| Französisch (Monaco) | fr-MC |

| Sprache (Kultur/Region) | Kulturcode |
|---------------------------|------------|
| Französisch (Schweiz) | fr-CH |
| Galizisch | gl |
| Galizisch (Spanien) | fi-ES |
| Georgisch | ka |
| Georgisch (Georgien) | ka-GE |
| Deutsch | de |
| Deutsch (Österreich) | de-AT |
| Deutsch (Deutschland) | de-DE |
| Deutsch (Liechtenstein) | de-LI |
| Deutsch (Luxemburg) | de-LU |
| Deutsch (Schweiz) | de-CH |
| Griechisch | el |
| Griechisch (Griechenland) | el-GR |
| Gujarati | gu |

| Sprache (Kultur/Region) | Kulturcode |
|--------------------------|------------|
| Gujarati (Indien) | gu-IN |
| Hebräisch | he |
| Hebräisch (Israel) | he-IL |
| Hindi | hi |
| Hindi (Indien) | hi-IN |
| Ungarisch | hu |
| Ungarisch (Ungarn) | hu-HU |
| Isländisch | is |
| Isländisch (Island) | is-IS |
| Indonesisch | id |
| Indonesisch (Indonesien) | id-ID |
| Italienisch | it |
| Italienisch (Italien) | it-IT |
| Italienisch (Schweiz) | it-CH |

| Sprache (Kultur/Region) | Kulturcode |
|-------------------------|------------|
|-------------------------|------------|

| | |
|-----------|----|
| Japanisch | ja |
|-----------|----|

| | |
|-------------------|-------|
| Japanisch (Japan) | ja-JP |
|-------------------|-------|

| | |
|---------|----|
| Kannada | kn |
|---------|----|

| | |
|-----------------|-------|
| Kanada (Indien) | kn-IN |
|-----------------|-------|

| | |
|------------|----|
| Kasachisch | kk |
|------------|----|

| | |
|-------------------------|-------|
| Kasachisch (Kasachstan) | kk-KZ |
|-------------------------|-------|

| | |
|---------|-----|
| Konkani | kok |
|---------|-----|

| | |
|------------------|--------|
| Konkani (Indien) | kok-IN |
|------------------|--------|

| | |
|------------|----|
| Koreanisch | ko |
|------------|----|

| | |
|--------------------|-------|
| Koreanisch (Korea) | ko-KR |
|--------------------|-------|

| | |
|------------|----|
| Kirgisisch | ky |
|------------|----|

| | |
|--------------------------|-------|
| Kirgisisch (Kirgisistan) | ky-KG |
|--------------------------|-------|

| | |
|----------|----|
| Lettisch | lv |
|----------|----|

| | |
|---------------------|-------|
| Lettisch (Lettland) | lv-LV |
|---------------------|-------|

| Sprache (Kultur/Region) | Kulturcode |
|-------------------------|------------|
|-------------------------|------------|

| | |
|-----------|----|
| Litauisch | lt |
|-----------|----|

| | |
|---------------------|-------|
| Litauisch (Litauen) | lt-LT |
|---------------------|-------|

| | |
|-------------|----|
| Mazedonisch | mk |
|-------------|----|

| | |
|---------------------------------|-------|
| Mazedonisch (Mazedonien, FYROM) | mk-MK |
|---------------------------------|-------|

| | |
|-----------|----|
| Malaiisch | ms |
|-----------|----|

| | |
|-------------------------------|-------|
| Malaiisch (Brunei Darussalam) | ms-BN |
|-------------------------------|-------|

| | |
|----------------------|-------|
| Malaiisch (Malaysia) | ms-my |
|----------------------|-------|

| | |
|---------|----|
| Marathi | mr |
|---------|----|

| | |
|------------------|-------|
| Marathi (Indien) | mr-IN |
|------------------|-------|

| | |
|------------|----|
| Mongolisch | mn |
|------------|----|

| | |
|-----------------------|-------|
| Mongolisch (Mongolei) | mn-MN |
|-----------------------|-------|

| | |
|------------|----|
| Norwegisch | no |
|------------|----|

| | |
|-------------------------------|-------|
| Norwegisch (Bokmål, Norwegen) | nb-NO |
|-------------------------------|-------|

| | |
|--------------------------------|-------|
| Norwegisch (Nynorsk, Norwegen) | nn-NO |
|--------------------------------|-------|

| Sprache (Kultur/Region) | Kulturcode |
|---------------------------|------------|
| Polnisch | pl |
| Polnisch (Polen) | pl-PL |
| Portugiesisch | pt |
| Portugiesisch (Brasilien) | pt-BR |
| Portugiesisch (Portugal) | pt-PT |
| Punjabi | pa |
| Punjabi (Indien) | pa-IN |
| Rumänisch | ro |
| Rumänisch (Rumänien) | ro-RO |
| Russisch | ru |
| Russisch (Russland) | ru-RU |
| Sanskrit | sa |
| Sanskrit (Indien) | sa-IN |
| Serbisch | sr |

| Sprache (Kultur/Region) | Kulturcode |
|------------------------------------|------------|
| Serbisch (Serbien, Kyrillisch) | sr-Cyrl-CS |
| Serbisch (Serbien, Lateinisch) | sr-Latn-CS |
| Slowakisch | sk |
| Slowakisch (Slowakei) | sk-SK |
| Slowenisch | sl |
| Slowenisch (Slowenien) | sl-SI |
| Spanisch | es |
| Spanisch (Argentinien) | es-AR |
| Spanisch (Bolivien) | es-BO |
| Spanisch (Chile) | es-CL |
| Spanisch (Kolumbien) | es-CO |
| Spanisch (Costa Rica) | es-CR |
| Spanisch (Dominikanische Republik) | es-DO |
| Spanisch (Ecuador) | es-EC |

| Sprache (Kultur/Region) | Kulturcode |
|--|--------------|
| Spanisch (El Salvador) | es-SV |
| Spanisch (Guatemala) | es-GT |
| Spanisch (Honduras) | es-HN |
| Spanisch (Mexiko) | es-MX |
| Spanisch (Nicaragua) | es-NI |
| Spanisch (Panama) | es-PA |
| Spanisch (Paraguay) | es-PY |
| Spanisch (Peru) | es-PE |
| Spanisch (Puerto Rico) | es-PR |
| Spanisch (Spanien) | es-ES |
| Spanisch (Spanien, traditionelle Sortierung) | es-ES_tradnl |
| Spanisch (Uruguay) | es-UY |
| Spanisch (Venezuela) | es-VE |
| Suaheli | sw |

| Sprache (Kultur/Region) | Kulturcode |
|-------------------------|------------|
|-------------------------|------------|

| | |
|-----------------|-------|
| Suaheli (Kenia) | sw-KE |
|-----------------|-------|

| | |
|------------|----|
| Schwedisch | sv |
|------------|----|

| | |
|-----------------------|-------|
| Schwedisch (Finnland) | sv-FI |
|-----------------------|-------|

| | |
|-----------------------|-------|
| Schwedisch (Schweden) | sv-SE |
|-----------------------|-------|

| | |
|---------|-----|
| Syrisch | syr |
|---------|-----|

| | |
|------------------|--------|
| Syrisch (Syrien) | syr-SY |
|------------------|--------|

| | |
|-------|----|
| Tamil | ta |
|-------|----|

| | |
|----------------|-------|
| Tamil (Indien) | ta-IN |
|----------------|-------|

| | |
|-------|----|
| Tatar | tt |
|-------|----|

| | |
|------------------|-------|
| Tatar (Russland) | tt-RU |
|------------------|-------|

| | |
|--------|----|
| Telugu | te |
|--------|----|

| | |
|-----------------|-------|
| Telugu (Indien) | te-IN |
|-----------------|-------|

| | |
|------|----|
| Thai | th |
|------|----|

| | |
|-------------------------|-------|
| Thailändisch (Thailand) | th-TH |
|-------------------------|-------|

| Sprache (Kultur/Region) | Kulturcode |
|------------------------------------|------------|
| Türkisch | tr |
| Türkisch (Türkei) | tr-TR |
| Ukrainisch | uk |
| Ukrainisch (Ukraine) | uk-UA |
| Urdu | ur |
| Urdu (Pakistan) | ur-PK |
| Usbekisch | uz |
| Usbekisch (Usbekistan, Kyrillisch) | uz-Cyrl-UZ |
| Usbekisch (Usbekistan, Lateinisch) | Uz-Latn-UZ |
| Vietnamesisch | vi |
| Vietnamesisch (Vietnam) | vi-VN |

Grammatiken

Eine gültige Parsing-Grammatik enthält:

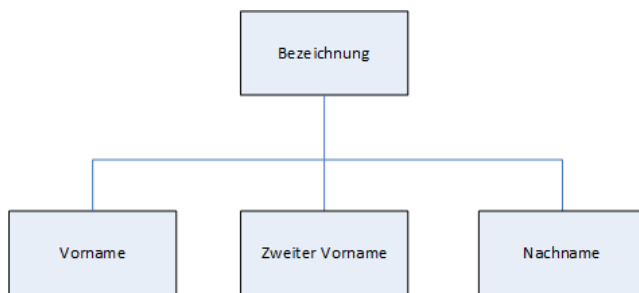
- Eine Stammvariable, die die Sequenz von Token bzw. Domänenmustern als Regelvariablen definiert.

- Regelvariablen, die den gültigen Satz von Zeichen und die Reihenfolge definieren, in der diese Zeichen auftreten können, um als Mitglied eines Domänenmusters betrachtet zu werden. Weitere Informationen finden Sie unter **Befehle des Regelabschnitts** auf Seite 31.
- Das zu parsende Eingabefeld. Eingabefeld bezeichnet das zu parsende Feld in den Quelldatensätzen.
- Die Ausgabefelder für die resultierenden geparsten Daten. Ausgabefelder definieren, wo die einzelnen resultierenden Token, die geparst werden, gespeichert werden sollen.

Eine gültige Parsing-Grammatik enthält zudem weitere optionale Befehle für:

- Zeichen, die für die Tokenisierung der von Ihnen geparsten Eingabedaten verwendet werden. Tokenisierungszeichen sind Zeichen, wie z. B. Leerzeichen und Bindestrich, die den Anfang und das Ende eines Tokens bestimmen. Das Standardtokenisierungszeichen ist ein Leerzeichen. Tokenisierungszeichen stellen die primäre Möglichkeit dar, dass eine Sequenz von Zeichen in eine Reihe von Token aufgeschlüsselt wird. Sie können den Befehl „tokenize“ auf NONE festlegen, um eine Tokenisierung des Feldes zu beenden. Wenn der Befehl „tokenize“ auf „None“ festgelegt ist, müssen die Grammatikregeln alle Leerzeichen innerhalb der zugehörigen Regeldefinition enthalten.
- Optionen für die Beachtung von Groß-/Kleinschreibung für Token in den Eingabedaten.
- Join-Zeichen für die Begrenzung übereinstimmender Token.
- Übereinstimmende Token in Tabellen
- Übereinstimmende zusammengesetzte Token in Tabellen
- Definieren von RegEx-Tags
- Literale Zeichenfolgen in Anführungszeichen
- Ausdrucksquantifizierer (optional). Weitere Informationen zu Ausdrucksquantifizierern finden Sie unter **Befehle des Regelabschnitts** auf Seite 31 und **Ausdrucksquantifizierer: gieriges, widerwilliges und besitzergreifendes Verhalten**.
- Verschiedene andere Indikatoren für Gruppierung, Kommentar und Zuweisung (optional). Weitere Informationen zu gruppierten Ausdrücken finden Sie unter **Grouping-Operator ()**.

Die Regelvariablen in Ihrer Parsing-Grammatik bilden eine überlagerte Baumstruktur der Sequenz von Zeichen oder Token in einem Domänenmuster. Sie können z. B. eine Parsing-Grammatik erstellen, die ein Domänenmuster auf Grundlage von Namenseingabedaten definiert, die die Token `<<FirstName>`, `/><<MiddleName>` und `/><<LastName>` enthalten.



Verwendung der Eingabedaten:

```
Joseph Arnold Cowers
```

Sie können diese Datenzeichenfolge in Form von drei Token in einem Domänenmuster darstellen:

```
<root> = <FirstName><MiddleName><LastName>;
```

Die Regelvariablen für dieses Domänenmuster sind:

```
<FirstName> = <given>;
<MiddleName> = <given>;
<LastName> = @Table("Family Names");
<given> = @Regex("[A-Za-z]+");
```

Basierend auf diesem einfachen Grammatikbeispiel versteht Open Parser Leerzeichen mit Token und interpretiert das Token `Joseph` als Vornamen, da die Zeichen im ersten Token mit der Definition `[A-Za-z]+` übereinstimmen und sich das Token in der definierten Sequenz befindet. Optional kann jedem Ausdruck ein anderer Ausdruck folgen.

Beispiel

```
<variable> = "some leading string" <variable2>;
<variable2> = @Table ("given") @Regex("[0-9]+");
```

Eine Grammatikregel ist eine grammatische Anweisung, wobei eine Variable gleich einem oder mehreren Ausdrücken ist. Jede Grammatikregel folgt der folgenden Form:

```
<rule> = expression [| expression...];
```

Die Grammatikregeln müssen den folgenden Regeln folgen:

- `<root>` Ist ein spezieller Variablenname und die erste Regel, die in der Grammatik ausgeführt wird, da sie das Domänenmuster definiert. Auf `<root>` kann nicht durch eine andere Regel in der Grammatik verwiesen werden.
- Die Variable `<rule>` kann sich nicht direkt oder indirekt auf sich selbst beziehen. Wenn Regel A auf Regel B verweist, die sich auf Regel C bezieht, die auf Regel A verweist, wird ein kreisförmiger Verweis erstellt. Kreisförmige Verweise sind nicht zulässig.
- Die Variable `<rule>` ist gleich einem oder mehrerer Ausdrücke.
- Jeder `expression` wird durch ein OR getrennt, das mit dem Pipe-Zeichen (`|`) angegeben wird.
- Die Ausdrücke werden einzeln überprüft. Der erste übereinstimmende `expression` wird ausgewählt. Es werden keine weiteren Ausdrücke überprüft.
- Der Variablenname kann aus alphabetischen Zeichen, numerischen Zeichen, Unterstrich (`_`) und Bindestrich (`-`) bestehen. Der Name der Variablen kann mit einem beliebigen gültigen Zeichen beginnen. Wenn der angegebene Ausgabefeldname nicht dieser Form entspricht, verwenden Sie das Alias-Feature, um den Variablennamen dem Ausgabefeld zuzuordnen.

Ein Ausdruck kann einer der folgenden Typen sein:

- Eine andere Variable
- Eine Zeichenfolge, die aus einem oder mehreren Zeichen in einfachen oder doppelten Anführungszeichen besteht. Beispiel:

```
"McDonald" 'McDonald' "O'Hara" 'O\'Hara' 'D"har' "D\"har"
```

- Tabelle
- CompoundTable
- RegEx-Befehle

Befehlsmetazeichen

Open Parser unterstützt den Standardsatz an Metazeichen der Java RegEx-Zeichenklasse in den Befehlen „%Tokenize“ und „@RegEx“. Bei einem Metazeichen handelt es sich um ein Zeichen, das eine besondere Bedeutung beim Mustervergleich trägt. Die folgenden Metazeichen werden unterstützt:

`([{ \ ^ - $ |] }) ? * + .`

Es gibt zwei Möglichkeiten, um zu erzwingen, dass ein Metazeichen wie ein normales Zeichen interpretiert wird:

- Dem Metazeichen einen umgekehrten Schrägstrich voranstellen
- Mit \Q (Beginn der Angabe) und \E (Ende der Angabe) einschließen.

„%Tokenize“ agiert gemäß der Regel für Java-Zeichenklassen für reguläre Ausdrücke – nicht für reguläre Java-Ausdrücke im Ganzen.

Im Allgemeinen sind die reservierten Zeichen für einen Zeichensatz die folgenden:

- „[“ und „]“ weisen auf einen anderen Satz hin.
- „-“ ist ein Metazeichen, wenn es sich zwischen zwei anderen Zeichen befindet.
- „^“ ist ein Metazeichen, wenn es das erste Zeichen in einem Satz bildet.
- „&&“ sind Metazeichen, wenn sie sich zwischen zwei anderen Zeichen befinden.
- „\“ bedeutet weiterhin, dass das Zeichen ein Literal ist.

Wenn Ihnen nicht genau bekannt ist, ob ein Zeichen als Metazeichen interpretiert wird, und Sie das Zeichen als Literal interpretieren lassen wollen, führen Sie eine Escape-Sequenz für dieses Zeichen mithilfe eines umgekehrten Schrägstrichs durch.

Befehle des Kopfzeilenbereichs

In diesem Abschnitt werden die Befehle des Kopfzeilenbereichs erläutert. Einige Befehle sind optional. Wenn ein Befehl optional ist, wird der Standardwert oder das Standardverhalten aufgeführt.

- **Tokenize** (optional)
- **Tokenize(None)**
- **InputField** (erforderlich, wenn „Eingabefelder“ nicht verwendet wird)
- **InputFields** (erforderlich, wenn „Eingabefeld“ nicht verwendet wird)
- **OutputFields** (erforderlich)
- **IgnoreCase** (optional)
- **Join** (optional)

Befehle des Regelabschnitts

Die Befehle des Regelabschnitts lauten:

- **RegEx**
- **Tabelle**
- **CompoundTable**
- **Token**
- **Bewertung**
- **RuleID**
- **<root>-Variable**
- **rule|rule**
- **Grouping-Operator ()**
- **Operator für min./max. Vorkommen {min, max}**
- **„Exact Occurrences“-Operator {exact}**
- **Zuweisungsoperator (=)**
- **Operator OR (|)**
- **„End of Rule“-Operator (;)**
- **Kommentierungsoperator (!)**
- **Quantifizierer für null oder ein Vorkommen (*)**
- **Quantifizierer für null oder mehr Vorkommen (*)**
- **Quantifizierer für ein oder mehrere Vorkommen (+)**
- **Ausdrucksquantifizierer: gieriges, widerwilliges und besitzergreifendes Verhalten**

Kulturen

Eine Kultur ist das primäre Konzept für die Organisation kulturspezifischer Parsing-Grammatiken. Sie können Kulturen verwenden, um verschiedene Parsing-Regeln für verschiedene Kulturen und Sprachen zu erstellen. Kultur folgt einer Hierarchie:

- **Globale Kultur:** Die globale Kultur ist kulturunabhängig und sprachagnostisch. Verwenden Sie die globale Kultur, um Parsing-Grammatikregeln zu erstellen, die alle Kulturen und Sprachen umfassen.
- **Sprache:** Eine Sprache ist mit einer Sprache verbunden, aber nicht mit einer bestimmten Kultur/Region. Zum Beispiel Englisch.
 - **Kultur/Region:** Eine Kultur/Region ist einer Sprache und einem Land oder einer Region zugeordnet. Zum Beispiel Englisch in Großbritannien oder Englisch in den Vereinigten Staaten.

In der Kulturhierarchie ist das übergeordnete Element einer Kultur/Region eine Sprache, und das übergeordnete Element einer Sprache ist die globale Kultur.

Kultur/Regionen erben die Eigenschaften der übergeordneten Sprache. Sprachen erben die Eigenschaften der globalen Kultur. Als solche können Sie Parsing-Grammatiken in einer Sprache für die Verwendung in mehreren Ländern definieren, die diese Sprache gemeinsam nutzen. Dann können Sie die Sprachgrammatikregeln mit spezialisierten Parsing-Grammatiken für ein bestimmtes Land oder eine Region außer Kraft setzen, das bzw. die dieselbe Sprache wie die Basissprachkultur hat, aber eine spezifische Adressierung, Namensgebung oder andere länder- oder regionsspezifische Unterschiede aufweist.

Sie können auch eine Kulturvererbung verwenden, um eingehende Datensätze zu parsen, die über einen zugewiesenen Kulturcode verfügen, aber keine definierte Grammatikregel für diesen Kulturcode aufweisen. In diesem Fall sucht Open Parser nach einem Sprachcode mit einer zugewiesenen Grammatikregel. Wenn kein Sprachcode vorhanden ist, sucht Open Parser in der globalen Kultur nach einer zugewiesenen Grammatikregel.

Der Domain Editor verwendet eine Kombination aus einem Sprachcode und einem Kulturcode, um die Sprache bzw. Kultur/Region darzustellen.

Definieren der Grammatikregeln einer Kultur

Sie können die Grammatikregeln einer Kultur verwenden, um einen Teil der Parsing-Grammatik der globalen Kultur durch Zeichenfolgen, Befehle oder Ausdrücke zu ersetzen, die für die Kultur und/oder Sprache typisch sind. Indem Sie eine Grammatikregel definieren, können Sie Teile der Parsing-Grammatik der globalen Kultur auf Basis der Kultur und/oder Sprache des Datensatzes anpassen. Das ist nützlich, wenn Sie keine separate Parsing-Grammatik für jede Kultur erstellen und stattdessen die Grammatik der globalen Kultur verwenden und nur bestimmte Teile der Grammatik der globalen Kultur für jede Kultur anpassen möchten.

In diesem Thema wird beschrieben, wie eine Grammatikregel für eine Kultur erstellt wird.

1. Rufen Sie im Enterprise Designer die Optionen **Tools > Open Parser Domain Editor** auf.
2. Klicken Sie auf die Registerkarte **Kulturen**.

Eine vollständige Liste der unterstützten Kulturen finden Sie unter **Zuweisen einer Parsing-Kultur zu einem Datensatz** auf Seite 12.

3. Wählen Sie die Kultur aus, zu der Sie eine Grammatikregel hinzufügen möchten, und klicken Sie dann auf **Eigenschaften**.
4. Klicken Sie auf die Registerkarte **Grammatikregeln**. Die angezeigten Informationen beinhalten die Grammatikregelnamen, die für die ausgewählte Kultur definiert sind, die zugeordnete Quellenkultur, den definierten Wert der Grammatikregel und die Beschreibung.
5. Klicken Sie auf **Hinzufügen**.
6. Geben Sie in das Feld **Name** einen Namen für die Grammatikregel ein.
7. Geben Sie in das Feld **Beschreibung** eine Beschreibung für die Grammatikregel ein.
8. Geben Sie in das Feld **Wert** die Grammatikregel ein.

Die Grammatikregel kann eine beliebige gültige Variable, Zeichenfolge, ein beliebiger Befehl oder gruppierter Ausdruck sein. Weitere Informationen finden Sie unter **Grammatiken** auf Seite 27.

9. Wählen Sie **Zeilenumbruch aktivieren** aus, um den Wert im Textfeld anzuzeigen, ohne Scrollen zu müssen.
10. Klicken Sie auf **OK**.

Der Grammatikregelwert, den Sie eingegeben haben, wird überprüft. Wenn der Wert grammatikalische Syntaxfehler enthält, werden in einer Nachricht eine Beschreibung der aufgetretenen Fehler, die Zeile und die Spalte sowie der Befehl, die Grammatikregel oder der RegEx-Tag, in denen der Fehler aufgetreten ist, angezeigt.

Beispiel für Grammatikregel

Sie haben eine Grammatik, die westliche Namen parst. Die Struktur des Musters kann für alle Kulturen identisch sein (<FirstName><MiddleName><LastName>), und viele der Regeln können mit demselben Muster oder derselben Tabelle übereinstimmen. Sie haben aber auch kulturspezifische Tabellen für Nachnamen, und es soll die passende Tabelle auf Basis des Kulturcodes des Datensatzes verwendet werden.

Dazu könnten Sie eine Grammatikregel für jede Kultur definieren, die das Element <LastName> in der globalen Kultur durch einen Verweis auf die kulturspezifische Tabelle ersetzt. Beispiel: Wenn Sie eine Tabelle mit niederländischen Nachnamen haben, würden Sie wie folgt eine Grammatikregel für die niederländische Kultur (nl) erstellen:

Name: LastName
Beschreibung: Niederländische Nachnamen
Wert: @Table("Dutch Last Names");

Definieren von RegEx-Tags für eine Kultur

In diesem Thema wird beschrieben, wie beim Definieren einer kulturspezifischen Parsing-Grammatik RegEx-Tags für eine Kultur definiert werden.

1. Rufen Sie im Enterprise Designer die Optionen **Tools > Open Parser Domain Editor** auf.
2. Klicken Sie auf die Registerkarte **Kulturen**. Auf der Registerkarte **Kulturen** wird eine Liste der unterstützten Kulturen angezeigt. Eine vollständige Liste der unterstützten Kulturen finden Sie unter [Zuweisen einer Parsing-Kultur zu einem Datensatz](#) auf Seite 12.
3. Wählen Sie eine Kultur aus der Liste aus und klicken Sie dann auf **Eigenschaften**. Das Dialogfeld **Kultureigenschaften** wird angezeigt.
4. Klicken Sie auf die Registerkarte **Regex-Tags**. Die angezeigten Informationen beinhalten die für die ausgewählte Kultur definierten RegEx-Tagnamen sowie die zugeordnete Quellenkultur, den Wert des RegEx-Tags und die Beschreibung.
5. Klicken Sie auf **Hinzufügen** oder **Ändern**.
6. Geben Sie in das Textfeld **Name** einen Namen für das RegEx-Tag ein.

Wenn Sie einen Namen eingeben, der in der ausgewählten Kultur bereits vorhanden ist, blinkt ein Warnsymbol. Geben Sie einen anderen Namen ein oder schließen Sie das Dialogfeld, löschen Sie das vorhandene RegEx-Tag und klicken Sie danach erneut auf **Hinzufügen**.

7. Geben Sie in das Feld **Beschreibung** eine Beschreibung des RegEx-Tags ein.
8. Geben Sie in das Textfeld **Wert** einen Wert für das RegEx-Tag ein.

Der Wert kann ein beliebiger gültiger regulärer Ausdruck sein, aber keine leere Zeichenfolge.

Domain Editor enthält mehrere vordefinierte RegEx-Tags, die Sie verwenden können, um Kultureigenschaften zu definieren. Sie können diese RegEx-Tags auch verwenden, um Tokenisierungszeichen in Ihrer Parsing-Grammatik zu definieren.

Sie können die vordefinierten RegEx-Tags ändern oder sie kopieren und eigene Varianten erstellen. Sie können auch Überschreibungseigenschaften verwenden, um spezialisierte RegEx-Tags für bestimmte Sprachen zu erstellen.

- Letter: Beliebiger Buchstabe aus beliebiger Sprache. Dieses RegEx-Tag beinhaltet Überschreibungen für mehrere Sprachen aufgrund der unterschiedlichen Schrift, z. B. kyrillische Schrift, asiatisch-sprachige Schriften und thailändische Schrift.
- Lower: Ein Kleinbuchstabe mit einer Großschreibungsvariante.
- Number: Beliebiges numerisches Zeichen in beliebiger Schrift.
- Punctuation: Beliebiges Interpunktionszeichen.
- Upper: Ein Großbuchstabe mit einer Kleinschreibungsvariante.
- Whitespace: Beliebiges Leerzeichen oder unsichtbares Trennzeichen.

9. Klicken Sie auf **OK**.

Importieren und Exportieren von Kulturen

Zusätzlich zum Erstellen von Kulturen können Sie auch Kulturen importieren, die Sie woanders erstellt haben, und Kulturen exportieren, die Sie im Domain Editor erstellt haben.

1. Rufen Sie im Enterprise Designer die Optionen **Tools > Open Parser Domain Editor** auf.
2. Klicken Sie auf die Registerkarte **Kulturen**.
3. Klicken Sie auf **Importieren** oder auf **Exportieren**.
4. Führen Sie eine der folgenden Aktionen aus:
 - Wenn Sie eine Kultur importieren, navigieren Sie zu einer Kultur und wählen Sie diese aus. Klicken Sie auf **Öffnen**. Die importierte Kultur wird im Domain Editor angezeigt.
 - Wenn Sie eine Kultur exportieren, navigieren Sie zu dem Speicherort, an dem die exportierte Kultur gespeichert werden soll, und wählen Sie diesen aus. Klicken Sie auf **Speichern**. Die exportierte Kultur wird gespeichert und der Domain Editor wird wieder angezeigt.

Domänen

Hinzufügen einer Domäne

Eine Domäne stellt einen Datentyp dar, z. B. Daten zu Name, Adresse und Telefonnummer. Sie besteht aus einem Muster, das eine Sequenz von mindestens einem Token in Ihren Eingabedaten darstellt, die Sie häufig parsen müssen und mindestens einer Kultur zuordnen.

In diesem Abschnitt wird beschrieben, wie eine Domäne beim Definieren einer kulturspezifischen Parsing-Grammatik im Domain Editor hinzugefügt wird.

1. Rufen Sie im Enterprise Designer die Optionen **Tools > Open Parser Domain Editor** auf.
2. Klicken Sie auf die Registerkarte **Domänen**.
3. Klicken Sie auf **Hinzufügen**.
4. Geben Sie im Feld **Name** einen Domänennamen ein.
5. Geben Sie im Feld **Beschreibung** eine Beschreibung des Domänennamens ein.
6. Klicken Sie auf **OK**, wenn Sie eine neue, leere Domäne erstellen möchten. Gehen Sie wie folgt vor, wenn Sie eine neue Domäne basierend auf einer anderen Domäne erstellen möchten:
 - a) Wählen Sie **Andere Domäne als Vorlage verwenden** aus, wenn Sie eine neue Domäne basierend auf einer anderen Domäne erstellen möchten.
 - b) Wählen Sie eine Domäne aus der Liste aus. Wenn Sie im nächsten Schritt auf **OK** klicken, wird die neue Domäne erstellt. Die neue Domäne enthält die gesamte kulturspezifische Parsing-Grammatik, die in der von Ihnen ausgewählten Domänenvorlage definiert wurde.
 - c) Klicken Sie auf **OK**.

Ändern einer Domäne

Eine Domäne stellt einen Datentyp dar, z. B. Daten zu Name, Adresse und Telefonnummer. Sie besteht aus einem Muster, das eine Sequenz von mindestens einem Token in Ihren Eingabedaten darstellt, die Sie häufig parsen müssen und mindestens einer Kultur zuordnen.

In diesem Abschnitt wird beschrieben, wie eine Domäne geändert wird.

1. Rufen Sie im Enterprise Designer die Optionen **Tools > Open Parser Domain Editor** auf.
2. Klicken Sie auf die Registerkarte **Domänen**.
3. Wählen Sie eine Domäne aus der Liste aus und klicken Sie anschließend auf **Ändern**. Das Dialogfeld „**Domäne ändern**“ wird angezeigt.
4. Ändern Sie die Informationen in der Beschreibung.
5. Wenn Sie nur die Beschreibung der Domäne ändern möchten, klicken Sie auf **OK**. Wenn Sie Aktualisierungen an der Vorlagendomäne durchgeführt haben und diese Änderungen jetzt zu der Domäne hinzufügen möchten, die Sie gerade ändern, dann fahren Sie mit dem nächsten Schritt fort.

6. Wählen Sie **Andere Domäne als Vorlage verwenden** aus, um an der Vorlagendomäne vorgenommene Änderungen zu übernehmen.
7. Wählen Sie eine Domänenmustervorlage aus der Liste aus. Wenn Sie im nächsten Schritt auf **OK** klicken, wird das Domänenmuster geändert. Das geänderte Domänenmuster enthält die gesamte kulturspezifische Parsing-Grammatik, die in der von Ihnen ausgewählten Domänenmustervorlage ausgewählt wurde. Die gesamte Parsing-Grammatik in der ausgewählten Domäne wird mit der Parsing-Grammatik der Domänenmustervorlage überschrieben.
8. Klicken Sie auf **OK**.

Gehen Sie wie folgt vor, um zu sehen, wie dies funktioniert:

1. Erstellen Sie ein Domänenmuster mit dem Namen **NameParsing** und definieren Sie Parsing-Grammatiken für „Globale Kultur“ („en“ und „en-US“).
2. Erstellen Sie ein Domänenmuster mit dem Namen **NameParsing2** und verwenden Sie **NameParsing** als Domänenmustervorlage. **NameParsing2** wird als genaue Kopie erstellt und enthält Parsing-Grammatiken für „Globale Kultur“ („en“ und „en-US“).
3. Ändern Sie die kulturspezifischen Parsing-Grammatiken für **NameParsing**, indem Sie einige der Grammatikregeln in der Grammatik von „Globale Kultur“ ändern und „en-CA“ als neue Kultur hinzufügen.
4. Wählen Sie auf der Registerkarte „Domänen“ **NameParsing2** aus, klicken Sie auf „Ändern“ und verwenden Sie **NameParsing** erneut als Domänenmustervorlage.

Die Ergebnisse lauten wie folgt:

- Die Parsing-Grammatik von „Globale Kultur“ wird aktualisiert (und Ihre Änderungen werden überschrieben, sofern Änderungen vorgenommen wurden).
- Die Kulturen „en“ und „en-US“ bleiben unverändert (es sei denn, sie wurden in der Zieldomäne geändert; in diesem Fall würden Sie anschließend wieder auf die Namensparsingversion zurückgesetzt).
- Für die Kultur „en-CA“ wird eine kulturspezifische Grammatik hinzugefügt.

Entfernen einer Domäne

Eine Domäne stellt einen Datentyp dar, z. B. Daten zu Name, Adresse und Telefonnummer. Sie besteht aus einem Muster, das eine Sequenz von mindestens einem Token in Ihren Eingabedaten darstellt, die Sie häufig parsen müssen und mindestens einer Kultur zuordnen.

In diesem Abschnitt wird beschrieben, wie eine Domäne entfernt wird.

1. Rufen Sie im Enterprise Designer die Optionen **Tools > Open Parser Domain Editor** auf.
2. Klicken Sie auf die Registerkarte **Domänen**.
3. Wählen Sie eine Domäne aus der Liste aus.
4. Klicken Sie auf **Entfernen**.

Wenn die Domäne mindestens einer kulturspezifischen Parsing-Grammatik zugeordnet ist, wird eine Nachricht angezeigt, in der Sie dazu aufgefordert werden, das Entfernen der Domäne zu bestätigen. Wenn dieser Domäne keine kulturspezifischen Parsing-Grammatiken zugeordnet sind, wird eine Nachricht angezeigt, in der bestätigt wird, dass Sie die ausgewählte Domäne entfernen möchten.

5. Klicken Sie auf **Ja**. Die Domäne und die dieser Domäne zugeordneten kulturspezifischen Parsing-Grammatiken werden entfernt.

Importieren und Exportieren von Domänen

Zusätzlich zum Erstellen von Domänen können Sie auch Domänen importieren, die Sie woanders erstellt haben, und Domänen exportieren, die Sie im Domain Editor erstellt haben.

1. Klicken Sie auf die Registerkarte **Domänen**. Die Registerkarte **Domänen** wird angezeigt.
2. Klicken Sie auf **Importieren** oder auf **Exportieren**.
3. Führen Sie eine der folgenden Aktionen aus:
 - Wenn Sie eine Domäne importieren, navigieren Sie zu einem Domänennamen und wählen Sie diesen aus. Klicken Sie auf **Öffnen**. Die importierte Domäne wird im Domain Editor angezeigt.
 - Wenn Sie eine Domäne exportieren, navigieren Sie zu dem Speicherort, an dem die exportierte Domäne gespeichert werden soll, und wählen Sie diesen aus. Klicken Sie auf **Speichern**. Die exportierte Domäne wird gespeichert und der Domain Editor wird wieder angezeigt.

Analysieren der Ergebnisse vom Parsen

Ablaufverfolgung von endgültigen Parsing-Ergebnissen

Das Feature „Open Parser-Ablaufverfolgungsdetails“ zeigt eine grafische Ansicht davon an, wie das Eingabefeld Token nach Token in die Ausgabefeldwerte geparkt wurde. Die Ablaufverfolgung zeigt übereinstimmende Ergebnisse, nicht übereinstimmende Ergebnisse und Zwischenergebnisse an.

„Endgültige Parsing-Ergebnisse“ zeigt die Parsing-Grammatikstruktur und die resultierende Ausgabe an. Verwenden Sie diese Ansicht, wenn Sie nur die Ergebnisse des Vergleichsprozesses sehen möchten. Dies ist die Standardansicht.

1. Öffnen Sie im Enterprise Designer den Datenfluss, der den „Open Parser“-Schritt mit den Parsing-Ergebnissen enthält, die Sie verfolgen möchten.
2. Doppelklicken Sie auf den „Open Parser“-Schritt auf der Arbeitsfläche.

3. Klicken Sie auf die Registerkarte **Vorschau**.
4. Geben Sie Beispieldaten ein, die Sie parsen möchten, und klicken Sie anschließend auf die Schaltfläche **Vorschau**.
5. Klicken Sie in der Spalte „Ablauf verfolgen“ auf den Link **Hier klicken...**, um das Ablaufverfolgungsdiagramm anzuzeigen.

Die Strukturansicht der Parsing-Grammatik zeigt eine oder mehrere der folgenden Elemente an, je nach ausgewählter Option:

- Die <root>-Variable. Der oberste Knoten in der Struktur ist die <root>-Variable.
- Die in der <root>-Variablen definierten Ausdrücke. Die Knoten der zweiten Ebene sind die Ausdrücke, die in der <root>-Variablen definiert sind. Die <root>-Ausdrücke definieren auch die Namen der Ausgabefelder.
- Die Variablendefinitionen der Knoten der zweiten Ebene. Die Knoten der dritten Ebene und jeder darunterliegenden Ebene sind die Definitionen der einzelnen <root>-Ausdrücke. Ausdrucksdefinitionen können andere Variablen, Aliase oder Regeldefinitionen sein.
- Die Werte und Token, die ausgegeben werden. Der untere Knoten in der Struktur zeigt die Werte, die jedem sequentiellen Token in der Parsing-Grammatik zugewiesen sind.
- Die Parser-Punktzahl für relevante Elemente der Parsing-Grammatik. Parser-Punktzahlen werden vom unteren Teil eines Stammausdrucks nach oben bestimmt. Wenn ein Ausdrucksmuster beispielsweise eine Gewichtung von 80 und eine Vorgängerregel eine Gewichtung von 75 aufweist, ist das Endergebnis für den Vorgängerausdruck das Produkt der untergeordneten Punktzahl und der Vorgängerpunktzahlen, die in diesem Beispiel 60 Prozent betragen würden.
- Das Leerzeichen wird im Textfeld **Eingabedaten** als geschütztes Leerzeichen (Klammer nach oben) angezeigt, sodass Sie Leerzeichen besser erkennen können. Nicht als Token verwendete Begrenzungszeichen werden grau dargestellt.

6. Wählen Sie **Endgültige Parsing-Ergebnisse** im Feld **Informationen** aus.

Anmerkung: Informationen zum Durchlaufen der Parsing-Ereignisse finden Sie unter [Durchlaufen von Parsing-Ereignissen](#) auf Seite 39.

7. Wählen Sie eine der Optionen in der Liste **Detailebene** aus.

- **Ausdrücke ohne Ergebnisse ausblenden** Zeigt die Zweige, die zu einem übereinstimmenden oder nicht übereinstimmenden Ergebnis führen. Jeder Zweig des Stammausdrucks, der nicht zu einer Übereinstimmung führt, wird als Ellipse dargestellt. Wenn Sie einen Zweig anzeigen möchten, der nicht zu einer Übereinstimmung führt, doppelklicken Sie auf die Ellipse.
- **Stammausdrücke ohne Ergebnisse ausblenden** Zeigt alle Zweige der Stammausdrücke an, die übereinstimmende oder nicht übereinstimmende Ergebnisse enthalten. Alle anderen Stammausdrücke werden nicht angezeigt.
- **Alle Stammausdrücke anzeigen.** Zeigt jeden Stammausdruck an. Wenn ein Stammausdruck kein übereinstimmendes Ergebnis hat, wird die Anzeige für diesen Stammausdruck mit dem Ellipsensymbol ausgeblendet.

- **Alle Ausdrücke anzeigen** Zeigt die Stammausdrücke und alle Zweige an. Die Stammausdrücke werden nicht mehr als Ellipse angezeigt. Stattdessen werden die Regeln für jeden Ausdruck in dem Zweig angezeigt.

Wenn Sie eine Ansicht nach Detailebene ausgewählt haben, die Ausdrücke ohne Ergebnisse ausblendet, und Sie einen Stammausdruck auswählen, der derzeit nicht angezeigt wird, ändern die Ablaufverfolgungsdetails die Auswahl der Detailebene in ein Listenelement, das die minimale Anzahl an Stammausdrücken angibt, während der Stammausdruck weiterhin angezeigt wird.

8. Klicken Sie auf **Punktzahlen anzeigen**, um Parser-Punktzahlen für Stammausdrücke, variable Ausdrücke und die resultierenden Übereinstimmungen und nicht übereinstimmenden Ergebnisse anzuzeigen.
9. Wählen Sie im Feld **Zoom** die Größe der Strukturansicht aus.
10. Wählen Sie im Feld **Stammklausel** eine der Optionen aus, um den Zweig des Stammausdrucks anzuzeigen.

Wenn Sie auf einen Ausdruckzweig im Ablaufverfolgungsdiagramm klicken, wird die Liste **Stammklausel** aktualisiert, um die ausgewählte Klausel anzuzeigen. Doppelklicken Sie auf eine Ellipse, um einen ausgeblendeten Ausdruck anzuzeigen.

11. Klicken Sie auf **OK**, wenn Sie fertig sind. Wenn Sie auf **OK** klicken, werden die Einstellungen für „Detailebene“, „Punktzahlen anzeigen“ und „Zoom“ gespeichert.

Durchlaufen von Parsing-Ereignissen

In der Ansicht „Details der Open Parser-Ablaufverfolgung“ können Sie ein Diagramm von „Event-by-Event“-Schritten im Vergleichsprozess anzeigen. Verwenden Sie diese Ansicht, wenn Sie beim Vergleichsprozess eine Fehlerbehebung durchführen und die Auswertung der einzelnen Token, die Parsing-Grammatik-Tokenisierung und die „Token-by-Token“-Vergleichsergebnisse sehen möchten.

1. Öffnen Sie im Enterprise Designer den Datenfluss, der den „Open Parser“-Schritt mit den Parsing-Ergebnissen enthält, die Sie verfolgen möchten.
2. Doppelklicken Sie auf den „Open Parser“-Schritt auf der Arbeitsfläche.
3. Klicken Sie auf die Registerkarte **Vorschau**.
4. Geben Sie Beispieldaten ein, die Sie parsen möchten, und klicken Sie anschließend auf die Schaltfläche **Vorschau**.
5. Klicken Sie in der Spalte „Ablauf verfolgen“ auf den Link **Hier klicken...**, um das Ablaufverfolgungsdiagramm anzuzeigen.

Die Strukturansicht der Parsing-Grammatik zeigt eine oder mehrere der folgenden Elemente an, je nach ausgewählter Option:

- Die <root>-Variable. Der oberste Knoten in der Struktur ist die <root>-Variable.

- Die in der <root>-Variablen definierten Ausdrücke. Die Knoten der zweiten Ebene sind die Ausdrücke, die in der <root>-Variablen definiert sind. Die <root>-Ausdrücke definieren auch die Namen der Ausgabefelder.
 - Die Variablendefinitionen der Knoten der zweiten Ebene. Die Knoten der dritten Ebene und jeder darunterliegenden Ebene sind die Definitionen der einzelnen <root>-Ausdrücke. Ausdrucksdefinitionen können andere Variablen, Aliase oder Regeldefinitionen sein.
 - Die Werte und Token, die ausgegeben werden. Der untere Knoten in der Struktur zeigt die Werte, die jedem sequentiellen Token in der Parsing-Grammatik zugewiesen sind.
 - Die Parser-Punktzahl für relevante Elemente der Parsing-Grammatik. Parser-Punktzahlen werden vom unteren Teil eines Stammausdrucks nach oben bestimmt. Wenn ein Ausdrucksmuster beispielsweise eine Gewichtung von 80 und eine Vorgängerregel eine Gewichtung von 75 aufweist, ist das Endergebnis für den Vorgängerausdruck das Produkt der untergeordneten Punktzahl und der Vorgängerpunktzahlen, die in diesem Beispiel 60 Prozent betragen würden.
 - Das Leerzeichen wird im Textfeld **Eingabedaten** als geschütztes Leerzeichen (Klammer nach oben) angezeigt, sodass Sie Leerzeichen besser erkennen können. Nicht als Token verwendete Begrenzungszeichen werden grau dargestellt.
6. Übereinstimmungen und Nichtübereinstimmungen sind im Ablaufverfolgungsdiagramm farbcodiert:
- Grüne Kästchen geben Übereinstimmungen an, die Bestandteil des endgültigen erfolgreichen Ergebnisses sind.
 - Rote Kästchen geben Nichtübereinstimmungen an.
 - Gelbe Kästchen geben vorläufige Übereinstimmungen an, die schließlich zurückgesetzt werden, wenn die Ereignisse durchlaufen werden. Vorläufige Übereinstimmungen werden nur in Parsing-Ereignisse durchlaufen angezeigt.
 - Graue Kästchen geben vorläufige Übereinstimmungen an, die zurückgesetzt wurden, um dieses Token für einen anderen Ausdruck freizugeben. Vorläufige Übereinstimmungen werden nur in Parsing-Ereignisse durchlaufen angezeigt.
7. Wählen Sie in der Liste **Informationen** den Eintrag **Parsing-Ereignisse durchlaufen** aus.
8. Wählen Sie eine der Optionen in der Liste **Detailebene** aus.
- **Ausdrücke ohne Ergebnisse ausblenden** Zeigt die Zweige, die zu einem übereinstimmenden oder nicht übereinstimmenden Ergebnis führen. Jeder Zweig des Stammausdrucks, der nicht zu einer Übereinstimmung führt, wird als Ellipse dargestellt. Wenn Sie einen Zweig anzeigen möchten, der nicht zu einer Übereinstimmung führt, doppelklicken Sie auf die Ellipse.
 - **Stammausdrücke ohne Ergebnisse ausblenden** Zeigt alle Zweige der Stammausdrücke an, die übereinstimmende oder nicht übereinstimmende Ergebnisse enthalten. Alle anderen Stammausdrücke werden nicht angezeigt.
 - **Alle Stammausdrücke anzeigen**. Zeigt jeden Stammausdruck an. Wenn ein Stammausdruck kein übereinstimmendes Ergebnis hat, wird die Anzeige für diesen Stammausdruck mit dem Ellipsensymbol ausgeblendet.

- **Alle Ausdrücke anzeigen** Zeigt die Stammausdrücke und alle Zweige an. Die Stammausdrücke werden nicht mehr als Ellipse angezeigt. Stattdessen werden die Regeln für jeden Ausdruck in dem Zweig angezeigt.

Wenn Sie eine Ansicht nach Detailebene ausgewählt haben, die Ausdrücke ohne Ergebnisse ausblendet, und Sie einen Stammausdruck auswählen, der derzeit nicht angezeigt wird, ändern die Ablaufverfolgungsdetails die Auswahl der Detailebene in ein Listenelement, das die minimale Anzahl an Stammausdrücken angibt, während der Stammausdruck weiterhin angezeigt wird.

9. Klicken Sie auf **Punktzahlen anzeigen**, um Parser-Punktzahlen für Stammausdrücke, variable Ausdrücke und die resultierenden Übereinstimmungen und nicht übereinstimmenden Ergebnisse anzuzeigen.
10. Wählen Sie im Feld **Zoom** die Größe der Strukturansicht aus.
11. Wählen Sie im Feld **Stammklausel** eine der Optionen aus, um den Zweig des Stammausdrucks anzuzeigen.

Wenn Sie auf einen Ausdruckzweig im Ablaufverfolgungsdiagramm klicken, wird die Liste **Stammklausel** aktualisiert, um die ausgewählte Klausel anzuzeigen. Doppelklicken Sie auf eine Ellipse, um einen ausgeblendeten Ausdruck anzuzeigen.

12. Das Kästchen **Automatisch zu ausgewähltem Knoten springen** ist standardmäßig aktiviert. Wenn dieses Kästchen aktiviert ist und Sie auf die Schaltfläche **Wiedergeben** klicken, werden die Ereignisse von Anfang an ausgeführt und es wird bei dem ersten Ereignis gestoppt, das mit dem ausgewählten Knoten oder einem der zugehörigen untergeordneten Elemente auftritt. Deaktivieren Sie dieses Kästchen, bevor Sie auf die Schaltfläche **Wiedergeben** klicken, um alle Ereignisse ohne Stoppen wiederzugeben.
13. Geben Sie im Feld **Wiedergabeverzögerung (Sekunden)** eine Verzögerung an, um die Geschwindigkeit der Wiedergabegeschwindigkeit zu steuern.
14. Klicken Sie auf die Schaltfläche **Wiedergeben**, um mit der Ausführung der Parsing-Ereignisse zu beginnen.
15. Klicken Sie auf **OK**, wenn Sie fertig sind.

Parsen von Personennamen

Wenn Sie alle Namensdaten in einem Feld haben, möchten Sie den Namen vielleicht parsen, um separate Felder für jeden Namensteil zu erhalten, wie Vorname, Nachname, Titel usw. Diese geparsten Namenselemente können dann in anderen automatisierten Vorgänge verwendet werden, z. B. Namensvergleiche, die Namensstandardisierung oder die Namenskonsolidierung über mehrere Datensätze.

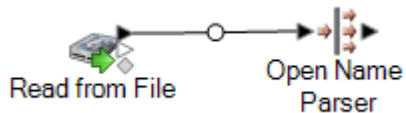
1. Wenn Sie es nicht bereits getan haben, laden Sie die folgenden Tabellen in den Spectrum™ Technology Platform-Server:

- Open Parser Base
- Open Parser Enhanced Names

Verwenden Sie das Database Load Utility des Data Normalization-Moduls, um diese Tabellen zu laden. Anleitungen zum Laden von Tabellen finden Sie im *Installationshandbuch*.

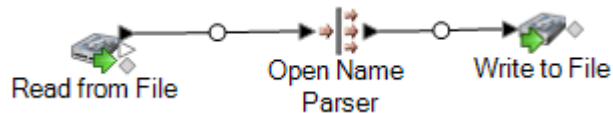
2. Erstellen Sie im Enterprise Designer einen neuen Datenfluss.
3. Ziehen Sie einen Quellschritt auf die Arbeitsfläche.
4. Doppelklicken Sie auf den Quellschritt und konfigurieren Sie ihn. Anweisungen zum Konfigurieren von Quellschritten finden Sie im *Datenfluss-Designer-Handbuch*.
5. Ziehen Sie einen „Open Name Parser“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem Quellschritt.

Wenn Sie beispielsweise einen „Read from File“-Schritt verwenden, sieht Ihr Datenfluss folgendermaßen aus:



6. Ziehen Sie einen Datenladeschritt auf die Arbeitsfläche und verbinden Sie ihn mit „Open Name Parser“.

Wenn Sie beispielsweise eine „Write to File“-Datenladung verwenden, könnte Ihr Datenfluss folgendermaßen aussehen:



7. Doppelklicken Sie auf den Datenladeschritt und konfigurieren Sie ihn. Anweisungen zum Konfigurieren von Quellschritten finden Sie im *Datenfluss-Designer-Handbuch*.

Sie haben einen Datenfluss erstellt, der Personennamen parsen kann, um Namenskomponenten zu erhalten, wobei jeder Namensteil in einem eigenen Feld gespeichert wird.

Datenflussvorlagen für das Parsen

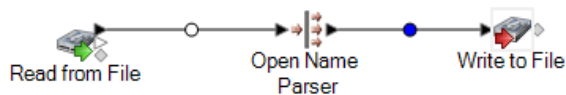
Parsen von englischen Namen

Diese Datenflussvorlage zeigt, wie mit Personennamensdaten (z. B. „John P. Smith“) umgegangen wird, wie sie in Vornamen-, Mittelnamen- und Nachnamenteile geparkt werden und wie Geschlechtsdaten hinzugefügt werden.

Geschäftsszenario

Sie arbeiten für eine Versicherungsgesellschaft, die personalisierte Angebote basierend auf dem Geschlecht der interessierten Kunden herauschicken möchte. Ihre Eingabedaten enthalten Namensdaten als vollständige Namen und Sie möchten die Namensdaten in die Felder „Vorname“, „Zweiter Vorname“ und „Nachname“ parsen. Sie möchten außerdem das Geschlecht der einzelnen Personen in Ihren Eingabedaten bestimmen.

Im folgenden Datenfluss wird eine Lösung für das Geschäftsszenario bereitgestellt:



Diese Datenflussvorlage ist im Enterprise Designer verfügbar. Öffnen Sie **Datei > Neu > Datenfluss > Aus Vorlage**, und wählen Sie **Parse Personal Name** aus.

Dieser Datenfluss erfordert Folgendes:

- Das Universal Name-Modul
- Die Open Parser-Basistabellen
- Die erweiterten Open Parser-Namensabellen

In diesem Datenfluss werden Daten aus einer Datei gelesen und durch den „Open Name Parser“-Schritt verarbeitet. „Open Name Parser“ ist Teil des Universal Naming-Moduls. Für jeden Namen führt der Datenfluss Folgendes aus:

Read from File

Bei diesem Schritt werden der Dateiname, der Speicherort und das Layout der Datei identifiziert, welche die zu parsenden Namen enthält. Die Datei enthält männliche und weibliche Namen.

Open Name Parser

„Open Name Parser“ untersucht Namensfelder und vergleicht sie mit den Namensdaten, die in den Namensdatenbankdateien von Spectrum™ Technology Platform gespeichert sind. Basierend auf

dem Vergleich parst Open Name Parser die Namensdaten in die Felder „Vorname“, „Zweiter Vorname“ und „Nachname“.

Write to File

Die Vorlage enthält einen „Write to File“-Schritt. Neben den Eingabefeldern enthält die Ausgabedatei die Felder „FirstName“, „MiddleName“, „LastName“, „EntityType“, „GenderCode“ und „GenderDeterminationSource“.

Parsen von arabischen Namen

Diese Vorlage zeigt, wie verwestlichte arabische Namen in Komponententeile geparkt werden. Die Parsing-Regel separiert jeden Token im Feld **Name** und kopiert jeden Token in fünf Felder: **Kunya**, **Ism**, **Laqab**, **Nasab**, **Nisba**. Diese Aufgabefelder stellen die fünf Teile eines arabischen Namens dar und werden im Geschäftsszenario beschrieben.

Geschäftsszenario

Sie arbeiten für eine Bank, die das arabische Namenssystem besser verstehen möchten, um den Kundendienst mit arabischsprachigen Kunden zu verbessern. Sie hatten Beschwerden von Kunden, deren Rechnungsinformationen nicht den korrekten Namen enthielten. Bei der Bemühung zur Verbesserung der Kundenprivatsphäre möchte die Marketing-Gruppe, für die Sie arbeiten, arabische Kunden durch Marketing-Kampagnen und Telefon-Support besser ansprechen.

Suchen Sie zum besseren Verständnis des arabischen Namenssystems nach Ressourcen im Internet, die das arabische Namenssystem erklären:

- https://de.wikipedia.org/wiki/Arabischer_Name
- heraldry.sca.org/laurel/names/arabic-naming2.htm

Arabische Namen basieren auf einem Namenssystem, das die folgenden Namensteile enthält: Ism, Kunya, Nasab, Laqab und Nisba.

- Ism ist der Hauptname oder Personennamen einer arabischen Person.
- Kunya, das sich oftmals auf den erstgeborenen Sohn einer Person bezieht, wird als Ersatz für Ism verwendet.
- Nasab ist ein Patronym oder eine Reihe von Patronymen. Es gibt die Herkunft einer Person durch das Wort „ibn“ oder „bin“ an, was Sohn bedeutet. „Bint“ bedeutet Tochter.
- Laqab dient als Beschreibung der Person. „Al-Rashid“ bedeutet z. B. der Rechtschaffende oder der Rechtgeleitete. „Al-Jamil“ bedeutet „wunderschön“.
- Nisba beschreibt den Beruf einer Person, die geografische Heimat oder die Abstammung (Stamm, Familie usw.). Es folgt einer Familie durch mehrere Generationen. Nisba ist gemeinsam mit anderen Komponenten des arabischen Namens vielleicht die naheliegendste Entsprechung des westlichen Nachnamens. „Al-Filistin“ bedeutet z. B. Palästinenser.

Im folgenden Datenfluss wird eine Lösung für das Geschäftsszenario bereitgestellt:



Diese Datenflussvorlage ist im Enterprise Designer verfügbar. Öffnen Sie **Datei > Neu > Datenfluss > Aus Vorlage**, und wählen Sie **ParseArabicNames** aus. Dieser Datenfluss erfordert das Data Normalization-Modul.

In diesem Datenfluss werden Daten aus einer Datei gelesen und über den „Open Parser“-Schritt verarbeitet. Für jede Datenzeile in der Eingabedatei führt dieser Datenfluss Folgendes aus:

Read from File

Bei diesem Schritt werden der Dateiname, der Speicherort und das Layout der Datei identifiziert, welche die zu parsenden Namen enthält. Die Datei enthält männliche und weibliche Namen.

Open Parser

Dieser Schritt definiert, ob eine kulturspezifische Domänengrammatik, die im Domäneneditor erstellt wurde, verwendet wird, oder ob eine domänenunabhängige Grammatik definiert wird. Bei einer von Ihnen im Domain Editor erstellten kulturspezifischen Parsing-Grammatik handelt es sich um eine überprüfte Parsing-Grammatik, die einer Kultur und einer Domäne zugeordnet ist. Bei einer von Ihnen in Open Parser erstellten domänenunabhängigen Parsing-Grammatik handelt es sich um eine überprüfte Parsing-Grammatik, die keiner Kultur und keiner Domäne zugeordnet ist.

In dieser Vorlage ist die Parsing-Grammatik als domänenunabhängige Grammatik definiert.

Der „Open Parser“-Schritt enthält eine Parsing-Grammatik, in der folgende Befehle und Ausdrücke definiert sind:

- %Tokenize ist auf das Leerzeichen eingestellt (\s). Das bedeutet, dass „Open Parser“ das Leerzeichen zur Separierung des Eingabefeldes in Token verwendet. „Abu Mohammed al-Rahim ibn Salamah“ enthält z. B. fünf Token: Abu, Mohammed, al-Rahim, ibn und Salamah.
- %InputField ist so eingestellt, dass Eingabedaten aus dem Feld **Name** geparkt werden.
- %OutputFields ist so eingestellt, das geparkte Daten in fünf Felder kopiert werden: **Kunya, Ism, Laqab, Nasab und Nisba**.
- Der <root>-Ausdruck definiert das Muster für arabische Namen:
- Kein oder ein Vorkommnis von **Kunya**
- Genau ein oder zwei Vorkommnisse von **Ism**
- Kein oder ein Vorkommnis von **Laqab**
- Kein oder ein Vorkommnis von **Nasab**
- Keine oder mehrere Vorkommnisse von **Nisba**

Die Regelvariablen, die die Domäne definieren, müssen die gleichen Namen wie die im erforderlichen Befehl `OutputFields` definierten Ausgabefelder verwenden.

Die Parsing-Grammatik verwendet eine Kombination aus regulären Ausdrücken und Ausdrucksquantifizierer, um ein Muster für arabische Namen aufzubauen. In der Parsing-Grammatik werden folgende Sonderzeichen verwendet:

- Das Zeichen „?“ bedeutet, dass ein regulärer Ausdruck keinmal oder einmal auftreten kann.
- Das Zeichen „*“ bedeutet, dass ein regulärer Ausdruck null oder mehrere Male auftreten kann.
- Das Zeichen „;“ steht für das Ende einer Regel.

Auf der Registerkarte **Befehle** können Sie die Bedeutung der anderen in Parsing-Grammatiken verwendbaren Sonderzeichen herausfinden, indem Sie den Mauszeiger über die Beschreibung bewegen.

Standardmäßig sind die Quantifizierer gierig. Gierig bedeutet, dass der Ausdruck so viele Token wie möglich akzeptiert, aber trotzdem einen erfolgreichen Vergleich zulässt. Sie können dieses Verhalten überschreiben, indem Sie ein „?“ für den Vergleich widerwilligen Verhaltens oder ein „+“ für den Vergleich besitzergreifenden Verhaltens anfügen. Der Vergleich widerwilligen Verhaltens bedeutet, dass der Ausdruck so wenige Token wie möglich akzeptiert, während immer noch ein erfolgreicher Vergleich zulässig ist. Der Vergleich besitzergreifenden Verhaltens bedeutet, dass der Ausdruck so viele Token wie möglich akzeptiert, auch wenn dadurch kein Vergleich möglich ist.

Klicken Sie zum Testen der Parsing-Grammatik auf die Registerkarte **Vorschau**. Geben Sie die unten aufgeführten Namen in das Feld **Name** ein, und klicken Sie auf **Vorschau**.

| Name | Kunya | Ism | Laqab | Nasab | Nisba |
|---|---------------|----------|----------|-----------------------------------|-----------------------|
| Abu Karim Muhammad al-Jamil ibn Nidal ibn Abdulaziz al-Filistini | Abu Karim | Muhammad | al-Jamil | Ibn Nidal ibn Abdulaziz | al-Filistini |
| Layla bint Zuhayr ibn Yazid al-Nahdiyah | | Layla | | bint Zuhayr ibn Yazid | al-Nahdiyah |
| Yazid ibn Abi Hakim | | Yazid | | ibn Abi Hakim | |
| Abu Bishr al-Yaman ibn Abi al-Yaman al-Bandanij | Abu Bishr | al-Yaman | | ibn Abi | al-Yaman al-Bandanij |
| Abu al-Tayyib 'Abd al-Rahim ibn Ahmad al-Harrani | Abu al-Tayyib | 'Abd | al-Rahim | ibn Ahmad | al-Harrani |
| Ahmad ibn Sa'id al-Bahili | | Ahmad | | ibn Sa'id | al-Bahili |
| Abu al-'Abbas Muhammad ibn Ya'qub ibn Yusuf al-Asamm al-Naysaburi | Abu al-'Abbas | Muhammad | | ibn Ya'qub ibn Yusuf | al-Asamm al-Naysaburi |
| Abu al-Qasim Mansur ibn al-Zabiqan ibn Salamah al-Namari | Abu al-Qasim | Mansur | | ibn al-Zabiqan ibn Salamah | al-Namari |
| 'Ubayd ibn Mu'awiyah ibn Zayd ibn Thabit ibn al-Dahhak | | 'Ubayd | | ibn Mu'awiyah ibn Zayd ibn Thabit | ibn al-Dahhak |
| Umm Ja'far Zubaydah | Umm Ja'far | Zubaydah | | | |

Sie können auch andere gültige und ungültige Namen eingeben, um zu sehen, wie die Eingabedaten geparkt werden.

Sie können mithilfe des Features „Ablaufverfolgung“ eine grafische Darstellung eines der endgültigen Parsing-Ergebnisse anzeigen oder die Parsing-Ereignisse durchlaufen. Klicken Sie auf den Link in der Spalte **Ablaufverfolgung**, um die Ablaufverfolgungsdetails für die Datenzeile anzuzeigen.

Write to File

Die Vorlage enthält einen „Write to File“-Schritt. Zusätzlich zum Eingabefeld enthält die Ausgabedatei die Felder **Kunya**, **Ism**, **Laqab**, **Nasab** und **Nisba**.

Parsen chinesischer Namen

Diese Vorlage demonstriert, wie chinesische Namen in Komponententeile geparkt werden. Die Parsing-Regel trennt die einzelnen Token im Feld **Name** und kopiert sie in zwei Felder: **LastName** und **FirstName**.

Geschäftsszenario

Sie arbeiten für ein Finanzdienstleistungsunternehmen, das untersuchen möchte, ob eine Einschließung der chinesischen Zeichen für seine Chinesisch sprechenden Kunden in diverser Korrespondenz umsetzbar ist.

Sie führen im Internet Recherchen durch, um das chinesische Benennungssystem zu verstehen, und finden folgende Ressource, in der erläutert wird, wie sich chinesische Namen zusammensetzen:

en.wikipedia.org/wiki/Chinese_names

Im folgenden Datenfluss wird eine Lösung für das Geschäftsszenario bereitgestellt:



Diese Datenflussvorlage ist im Enterprise Designer verfügbar. Navigieren Sie zu **Datei > Neu > Datenfluss > Aus Vorlage** und wählen Sie **ParseChineseNames** aus. Dieser Datenfluss erfordert das Data Normalization-Modul.

In diesem Datenfluss werden Daten aus einer Datei gelesen und über den „Open Parser“-Schritt verarbeitet. Dieser Datenfluss geht bei jeder Datenzeile in der Eingabedatei wie folgt vor:

Read from File

Bei diesem Schritt werden der Dateiname, der Speicherort und das Layout der Datei identifiziert, welche die zu parsenden Namen enthält. Die Datei enthält männliche und weibliche Namen.

Open Parser

Dieser Schritt definiert, ob eine kulturspezifische Domänengrammatik, die im Domäneneditor erstellt wurde, verwendet wird, oder ob eine domänenunabhängige Grammatik definiert wird. Bei einer von Ihnen im Domain Editor erstellten kulturspezifischen Parsing-Grammatik handelt es sich um eine überprüfte Parsing-Grammatik, die einer Kultur und einer Domäne zugeordnet ist. Bei einer von Ihnen in Open Parser erstellten domänenunabhängigen Parsing-Grammatik handelt es sich um eine überprüfte Parsing-Grammatik, die keiner Kultur und keiner Domäne zugeordnet ist.

In dieser Vorlage ist die Parsing-Grammatik als domänenunabhängige Grammatik definiert.

Der „Open Parser“-Schritt enthält eine Parsing-Grammatik, in der folgende Befehle und Ausdrücke definiert sind:

- `%Tokenize` ist auf „Keine“ festgelegt. Wenn `Tokenize` auf `None` gesetzt ist, muss die Parsing-Grammatikregel alle Leerzeichen und andere Token-Trennzeichen innerhalb der Regeldefinition enthalten.
- `%InputField` ist so eingestellt, dass Eingabedaten aus dem Feld **Name** geparkt werden.
- `%OutputFields` ist zum Kopieren geparkter Daten in zwei Felder festgelegt: **LastName** und **FirstName**.

Der Ausdruck `<root>` definiert das Muster für chinesische Namen:

- Ein Vorkommen von **LastName**
- Ein bis drei Vorkommen von **FirstName**

Die Regelvariablen, die die Domäne definieren, müssen die gleichen Namen wie die im erforderlichen Befehl `OutputFields` definierten Ausgabefelder verwenden.

Die Regelvariable `CJKCharacter` definiert das Zeichenmuster für Chinesisch/Japanisch/Koreanisch (CJK). Das Zeichenmuster ist so definiert, dass nur Zeichen in Form von Buchstaben verwendet werden. Die Regel lautet wie folgt:

```
<CJKCharacter> = @Regex("([\p{InCJKUnifiedIdeographs}&&\p{L}])");
```

- Der reguläre Ausdruck `\p{InX}` wird verwendet, um einen Unicode-Block für eine bestimmte Kultur anzugeben, in dem `X` für die Kultur steht. In dieser Instanz lautet die Kultur „CJKUnifiedIdeographs“.
- In regulären Ausdrücken ist eine Zeichenklasse ein Satz von Zeichen, die Sie abgleichen möchten. Beispiel: `[aeiou]` ist die Zeichenklasse, die nur Vokale enthält. Zeichenklassen können innerhalb anderer Zeichenklassen erscheinen und können sich aus dem Vereinigungsoperator (implizit) und dem Schnittmengenoperator (`&&`) zusammensetzen. Der Vereinigungsoperator bezeichnet eine Klasse mit allen Zeichen, die in mindestens einer der zugehörigen Operandenklassen enthalten sind. Der Schnittmengenoperator bezeichnet eine Klasse mit allen Zeichen, die mit den sich überschneidenden Unicode-Blöcken überlappen.
- Mit dem regulären Ausdruck `\p{L}` wird der Unicode-Block angegeben, der nur Buchstaben umfasst.

Klicken Sie zum Testen der Parsing-Grammatik auf die Registerkarte **Vorschau**. Geben Sie die unten aufgeführten Namen in das Feld **Name** ein, und klicken Sie auf **Vorschau**.

| Name | FirstName | LastName |
|------|-----------|----------|
| 王若琳 | 若琳 | 王 |
| 刘耕宏 | 耕宏 | 刘 |
| 许惠欣 | 惠欣 | 许 |
| 蔡依林 | 依林 | 蔡 |
| 水沫 | 沫 | 水 |
| 蕭亞軒 | 亞軒 | 蕭 |
| 郑元畅 | 元畅 | 郑 |
| 林依晨 | 依晨 | 林 |

Sie können auch andere gültige und ungültige Namen eingeben, um zu sehen, wie die Eingabedaten geparkt werden.

Sie können mithilfe des Features „Ablaufverfolgung“ eine grafische Darstellung eines der endgültigen Parsing-Ergebnisse anzeigen oder die Parsing-Ereignisse durchlaufen. Klicken Sie auf den Link in der Spalte **Ablaufverfolgung**, um die Ablaufverfolgungsdetails für die Datenzeile anzuzeigen.

Write to File

Die Vorlage enthält einen „Write to File“-Schritt. Neben dem Eingabefeld enthält die Ausgabedatei die Felder **LastName** und **FirstName**. Wählen Sie in der Liste der Vergleichsergebnisse ein Vergleichsergebnis aus, und klicken Sie dann auf **Entfernen**.

Parsen von spanischen und deutschen Namen

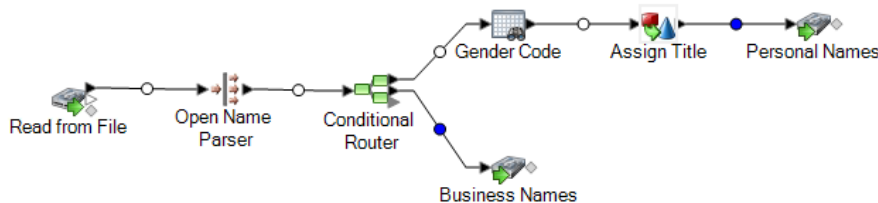
Diese Vorlage demonstriert, wie Namen gemischter Kulturen, z. B. spanische und deutsche Namen, in Komponententeile geparkt werden. Die Parsing-Regel trennt die einzelnen Token im Feld **Name** und kopiert sie in die Felder, die in der Parsing-Grammatik für Personen- und Unternehmensnamen definiert sind. Wählen Sie für weitere Informationen zu dieser Parsing-Grammatik **Tools > Open Parser Domain Editor** aus. Wählen Sie anschließend die Domäne **Personen- und Unternehmensnamen** und entweder die Kultur **Deutsch (de)** oder **Spanisch (es)** aus.

In dieser Vorlage werden auch über in „Tabellenverwaltung“ enthaltene Tabellendaten Geschlechtscodes auf Personennamen angewendet. Wählen Sie für weitere Informationen zu „Tabellenverwaltung“ **Tools > Tabellenverwaltung** aus.

Geschäftsszenario

Sie arbeiten für ein Pharmazieunternehmen in Brüssel, das seine Tätigkeiten in Deutschland und Spanien konsolidiert hat. Ihr Unternehmen möchte eine Datenbank für gemischte Kulturen mit Namensdaten implementieren. Sie haben den Auftrag, die Namensvarianten zwischen den zwei Kulturen zu analysieren.

Im folgenden Datenfluss wird eine Lösung für das Geschäftsszenario bereitgestellt:



Diese Datenflussvorlage ist im Enterprise Designer verfügbar. Navigieren Sie zu **Datei > Neu > Datenfluss > Aus Vorlage** und wählen Sie **ParseSpanisch&GermanNames** aus. Dieser Datenfluss erfordert das Data Normalization-Modul.

In diesem Datenfluss werden Daten aus einer Datei gelesen und über den „Open Parser“-Schritt verarbeitet. Dieser Datenfluss geht bei jeder Datenzeile in der Eingabedatei wie folgt vor:

Read from File

Bei diesem Schritt werden der Dateiname, der Speicherort und das Layout der Datei identifiziert, welche die zu parsenden Namen enthält. Die Datei enthält männliche und weibliche Namen und umfasst „CultureCode“-Informationen für die einzelnen Namen. Die „CultureCode“-Informationen bezeichnen die Eingabenamen als deutsch (de) oder spanisch (es).

Open Name Parser

„Open Name Parser“ untersucht Namensfelder und vergleicht sie mit den Namensdaten, die in den Namensdatenbankdateien von Spectrum™ Technology Platform gespeichert sind. Basierend auf

dem Vergleich parst Open Name Parser die Namensdaten in die Felder „Vorname“, „Zweiter Vorname“ und „Nachname“.

Conditional Router

In diesem Schritt wird die Eingabe weitergeleitet, damit Personennamen an den Schritt „Geschlechtscodes“ und Unternehmensnamen an den Schritt „Unternehmensnamen“ weitergeleitet werden.

Geschlechtscode

Doppelklicken Sie auf diesen Schritt auf der Arbeitsfläche, und klicken Sie anschließend auf **Ändern**, um Optionen für „Table Lookup“-Regeln anzuzeigen.

Die Option **Kategorisieren** verwendet den Quellwert als Schlüssel und kopiert den entsprechenden Wert aus dem Tabelleneintrag in das aus der Zielliste ausgewählte Feld. In dieser Vorlage wird **Vollständiges Feld** ausgewählt, und **Quelle** wird auf die Verwendung des Feldes **FirstName** festgelegt. Beim Table Lookup wird das gesamte Feld als eine Zeichenfolge behandelt und der Datensatz gekennzeichnet, wenn die Zeichenfolge als Ganzes kategorisiert werden kann.

Als **Ziel** wird das Feld **GenderCode** festgelegt. Zudem werden die in der Tabelle **Geschlechtscodes** enthaltenen Suchbegriffe für die Kategorisierung männlicher und weiblicher Namen verwendet. Wenn ein Begriff in den Eingabedaten nicht gefunden wurde, weist Table Lookup ihm den Wert **U** zu. Dieser Wert steht für „Unbekannt“. Wählen Sie für ein besseres Verständnis der Funktionsweise **Tools > Tabellenverwaltung** aus und wählen Sie dann die Tabelle „Geschlechtscodes“ aus.

Write to File

Die Vorlage enthält zwei „Write to File“-Schritte: einen für Personennamen und einen für Unternehmensnamen. Neben dem Eingabefeld enthält auch die Ausgabedatei für Personennamen die Felder **Name**, **TitleOfRespect**, **FirstName**, **MiddleName**, **LastName**, **PaternalLastName**, **MaternalLastName**, **MaturitySuffix**, **GenderCode**, **CultureUsed** und **ParserScore**.

Die Ausgabedatei für Unternehmensnamen enthält die Felder **Name**, **FirmName**, **FirmSuffix**, **CultureUsed** und **ParserScore**.

Parsen von E-Mail-Adressen

Diese Vorlage demonstriert, wie E-Mail-Adressen in Komponententeile geparkt werden. Die Parsing-Regel trennt die einzelnen Token im Feld **Email** und kopiert sie in drei Felder: **Local-Part**, **DomainName** und **DomainExtension**. **Local-Part** stellt den Teil des Domännennamens der E-Mail-Adresse dar, **DomainName** den Domännennamen der E-Mail-Adresse und **DomainExtension** die Domänenerweiterung der E-Mail-Adresse. Beispiel: Bei `pb.com` steht „pb“ für den Domännennamen und „com“ für die Domänenerweiterung.

Das Internet ist eine großartige Quelle für öffentlich zugängliche Informationen, die Sie bei Ihren offenen Parsing-Aufgaben unterstützen können. In diesem Beispiel wurden von verschiedenen

Internetressourcen Informationen zur E-Mail-Formatierung abgerufen und anschließend zur Erstellung einer Tabelle mit Domänenwerten in „Tabellenverwaltung“ importiert. Die Aufgabe zur Domänenerweiterung, die Sie in dieser Vorlagenaktivität durchführen, demonstriert die Nützlichkeit dieser Methode.

Zudem demonstriert diese Vorlage, wie Tabellendaten, die Sie in „Tabellenverwaltung“ geladen haben, effektiv im Rahmen Ihrer Parsing-Aufgaben für Tabellensuchvorgänge genutzt werden können.

Geschäftsszenario

Sie arbeiten für ein Versicherungsunternehmen, das seine erste E-Mail-Marketingkampagne plant. Ihre Datenbank enthält die E-Mail-Adressen Ihrer Kunden, und Sie wurden darum gebeten, einen Weg zu finden, um sicherzustellen, dass diese E-Mail-Adressen ein gültiges SMTP-Format aufweisen.

Bevor Sie diesen Datenfluss erstellen, müssen Sie eine Tabelle mit gültigen Domännennamenerweiterungen in „Tabellenverwaltung“ laden, damit Sie im Rahmen des Überprüfungsprozesses nach Domännennamenerweiterungen suchen können.

Im folgenden Datenfluss wird eine Lösung für das Geschäftsszenario bereitgestellt:



Diese Datenflussvorlage ist im Enterprise Designer verfügbar. Navigieren Sie zu **Datei > Neu > Datenfluss > Aus Vorlage** und wählen Sie **ParseEmail** aus. Dieser Datenfluss erfordert das Data Normalization-Modul.

In diesem Datenfluss werden Daten aus einer Datei gelesen und über den „Open Parser“-Schritt verarbeitet. Für jede Datenzeile in der Eingabedatei führt dieser Datenfluss Folgendes aus:

Erstellen einer Tabelle mit Domänenerweiterungen

Die erste Aufgabe besteht darin, unter „Tabellenverwaltung“ eine „Open Parser“-Tabelle zu erstellen, mit der Sie überprüfen können, ob die Domänenerweiterungen in Ihrer E-Mail-Adresse gültig sind.

1. Wählen Sie aus dem Menü **Tools** den Eintrag **Tabellenverwaltung** aus.
2. Wählen Sie aus der Liste **Typ** den Eintrag **Open Parser** aus.
3. Klicken Sie auf **Neu**.
4. Geben Sie im Dialogfeld **Benutzerdefinierte Tabelle hinzufügen** `EmailDomains` in das Feld **Tabellenname** ein. Stellen Sie sicher, dass in der Liste **Kopieren von** der Eintrag **Keine** ausgewählt ist und klicken Sie anschließend auf **OK**.
5. Klicken Sie, wenn **EmailDomains** in der Liste **Name** angezeigt wird, auf **Importieren**.
6. Klicken Sie im Dialogfeld **Importieren** auf **Durchsuchen** und suchen Sie die Quelldatei für die Tabelle. Die Standardspeicherort ist: `<drive>:\Program Files\Pitney Bowes\Spectrum\server\modules\coretemplates\data\Email_Domains.txt`. In „Tabellenverwaltung“ wird eine Vorschau der Begriffe angezeigt, die in der Importdatei enthalten sind.

7. Klicken Sie auf **OK**. „Tabellenverwaltung“ importiert die Quelldateien und zeigt eine Liste der internen Domänenenerweiterungen an.
8. Klicken Sie auf **Schließen**. Die Tabelle `EmailDomains` wird erstellt. Erstellen Sie jetzt mithilfe der Vorlage „ParseEmail“ den Datenfluss.

Read from File

In diesem Schritt wird der Dateiname, der Speicherort und das Layout der Datei identifiziert, die die zu parsenden Adressen enthält.

Open Parser

In der Parsing-Grammatik des „Open Parser“-Schrittes werden folgende Befehle und Ausdrücke definiert:

- `%Tokenize` ist auf „Keine“ festgelegt. Wenn `Tokenize` auf `None` gesetzt ist, muss die Parsing-Grammatikregel alle Leerzeichen und andere Token-Trennzeichen innerhalb der Regeldefinition enthalten.
- `%InputField` ist zum Parsen von Eingabedaten aus dem Feld **Email_Address** festgelegt.
- `%OutputFields` ist zum Kopieren geparster Daten in drei Felder festgelegt: **Local-Part**, **DomainName** und **DomainExtension**.
- Der Stammausdruck definiert das Muster der Token, die geparst werden:

```
<root> = <Local-Part>"@"<DomainName>". "<DomainExtension>;
```

Die Regelvariablen, die die Domäne definieren, müssen die gleichen Namen wie die im erforderlichen Befehl `OutputFields` definierten Ausgabefelder verwenden.

- In der restlichen Parsing-Grammatik werden die einzelnen Regelvariablen als Ausdrücke definiert.

```
<Local-Part> = (<alphanum> ".")* <alphanum> | (<alphanum> "_")*
<alphanum> ;
<DomainName> = (<alphanum> ".")? <alphanum>;
<DomainExtension> = @Table("EmailDomains")* ". "? @Table("EmailDomains");
<alphanum>=@RegEx("[A-Za-z0-9]+");
```

Die Variable `<Local-Part>` ist als Textzeichenfolge definiert, die die Variable `<alphanum>`, den Punkt und eine weitere `<alphanum>`-Variable enthält.

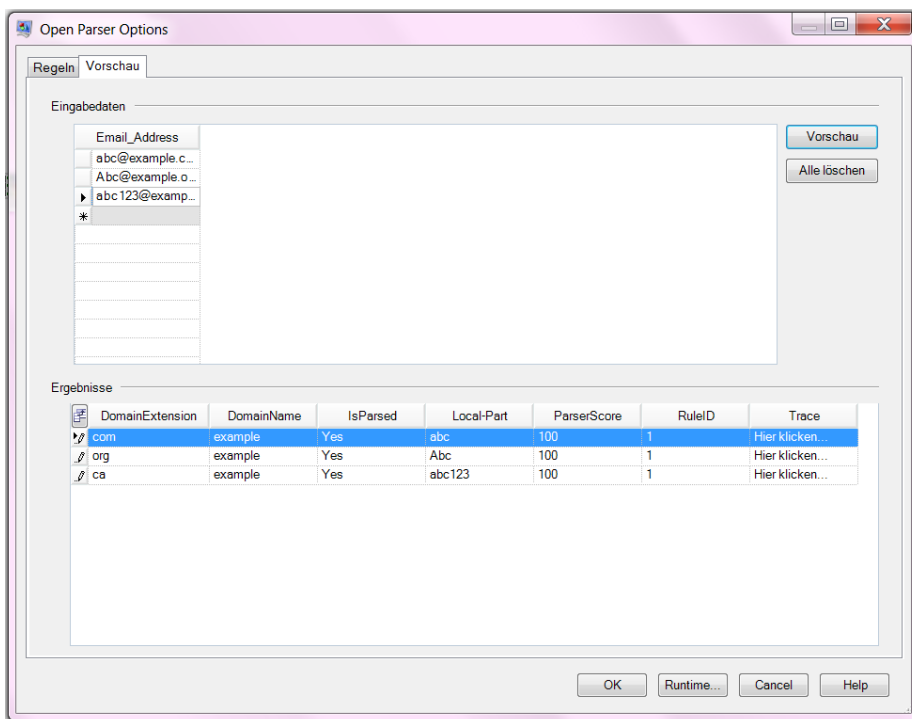
Bei der Definition der Variable `<alphanum>` handelt es sich um einen regulären Ausdruck, also um eine beliebige Zeichenfolge mit Zeichen von A bis Z, a bis z und 0 – 9. Die Variable `<alphanum>` wird in dieser gesamten Parsing-Grammatik verwendet und wird einmal in der letzten Zeile der Parsing-Grammatik definiert.

Zur Erstellung eines Musters für E-Mail-Adressen verwendet die Parsing-Grammatik eine Kombination aus regulären Ausdrücken und Literalzeichen. Bei allen Zeichen, die in dieser Parsing-Grammatik in doppelten Anführungszeichen (") stehen, handelt es sich um Literalzeichen, Namen von Tabellen für Suchvorgänge oder reguläre Ausdrücke. In der Parsing-Grammatik werden folgende Sonderzeichen verwendet:

- Das Zeichen „+“ bedeutet, dass ein regulärer Ausdruck mindestens einmal auftreten kann.
- Das Zeichen „?“ bedeutet, dass ein regulärer Ausdruck keinmal oder einmal auftreten kann.
- Das Zeichen „|“ bedeutet, dass die Variable eine OR-Bedingung aufweist.
- Das Zeichen „;“ steht für das Ende einer Regel.

Auf der Registerkarte **Befehle** können Sie die Bedeutung der anderen in Parsing-Grammatiken verwendbaren Sonderzeichen herausfinden, indem Sie den Mauszeiger über die Beschreibung bewegen.

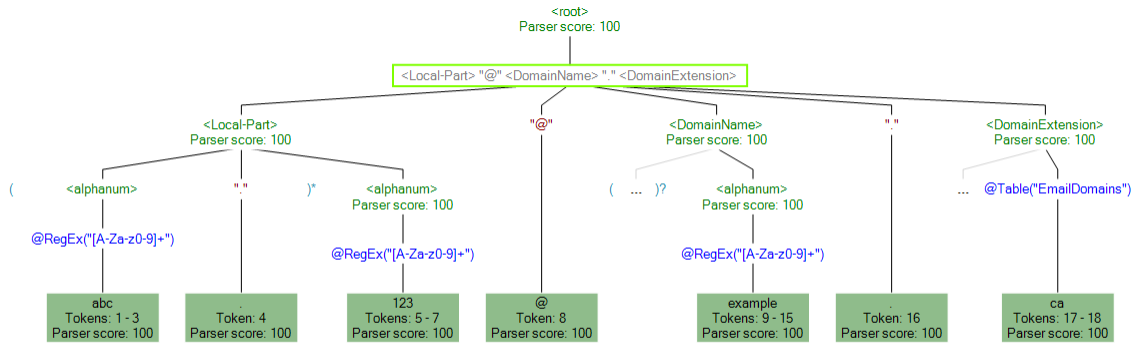
Klicken Sie zum Testen der Parsing-Grammatik auf die Registerkarte **Vorschau**. Geben Sie die unten im Feld **E-Mail-Adresse** angezeigten E-Mail-Adressen ein und klicken Sie anschließend auf **Vorschau**.



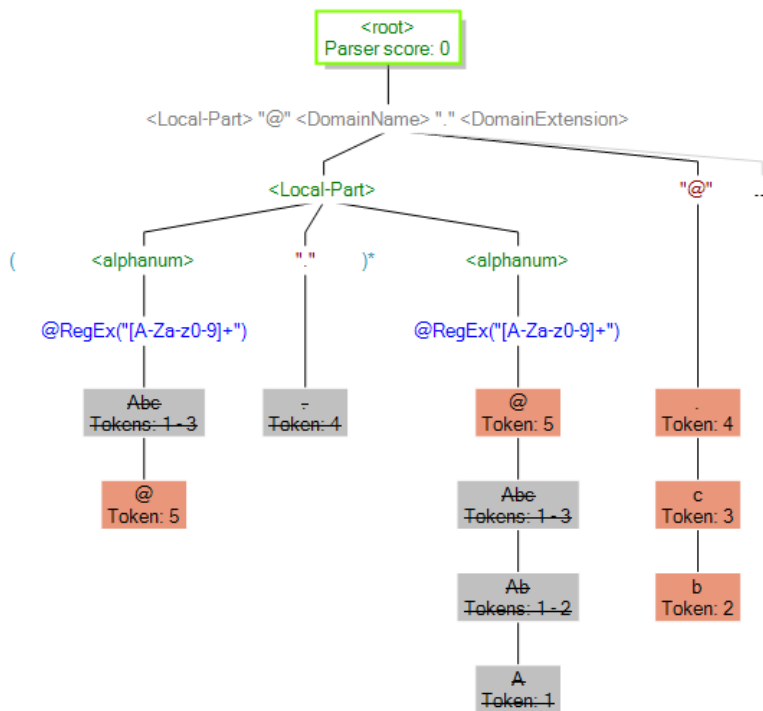
Sie können auch andere E-Mail-Adressen eingeben, um zu sehen, wie die Eingabedaten geparkt werden.

Zudem können Sie mithilfe des Features „Ablaufverfolgung“ eine grafische Darstellung eines der endgültigen Parsing-Ergebnisse anzeigen oder die Parsing-Ereignisse durchlaufen. Klicken Sie auf den Link in der Spalte **Ablaufverfolgung**, um die Ablaufverfolgungsdetails für die Datenzeile anzuzeigen.

Unter „Ablaufverfolgungsdetails“ wird ein Vergleichsergebnis angezeigt. Vergleichen Sie die Token miteinander, die für die einzelnen Ausdrücke in der Parsing-Grammatik abgeglichen wurden.



Sie können auch über „Ablaufverfolgung“ nicht übereinstimmende Ergebnisse anzeigen. In der folgenden Grafik wird ein nicht übereinstimmendes Ergebnis dargestellt. Vergleichen Sie die Token miteinander, die für die einzelnen Ausdrücke in der Parsing-Grammatik abgeglichen wurden. Der Grund für die fehlende Übereinstimmung dieser Eingabedaten (Abc.example.com) ist, dass nicht alle abzugleichenden Token darin enthalten waren: Das Zeichen „@“ zum Trennen des Tokens „Local-Part“ von den Token vom Typ „Domain“ ist nicht vorhanden.



Write to File

Die Vorlage enthält einen „Write to File“-Schritt. Neben dem Eingabefeld die Ausgabedatei die Felder **Local-Part**, **DomainName**, **DomainExtension**, **IsParsed** und **ParserScore**.

Parsen von US-amerikanischen Telefonnummern

Diese Vorlage demonstriert, wie US-amerikanische Telefonnummern in Komponententeile geparkt werden. Die Parsing-Regel trennt die einzelnen Token im Feld **PhoneNumber** und kopiert sie in vier Felder: **CountryCode**, **AreaCode**, **Exchange** und **Number**.

Geschäftsszenario

Sie arbeiten für einen Mobilfunkanbieter und wurden einem Projekt zugewiesen, in dem Sie die Daten eingehender Telefonnummern für einen wachsenden Bereich Ihres Unternehmens analysieren sollen.

Im folgenden Datenfluss wird eine Lösung für das Geschäftsszenario bereitgestellt:



Diese Datenflussvorlage ist im Enterprise Designer verfügbar. Navigieren Sie zu **Datei > Neu > Datenfluss > Aus Vorlage** und wählen Sie **ParseUSPhoneNumbers** aus. Dieser Datenfluss erfordert das Data Normalization-Modul.

In diesem Datenfluss werden Daten aus einer Datei gelesen und über den „Open Parser“-Schritt verarbeitet. Dieser Datenfluss geht bei jeder Datenzeile in der Eingabedatei wie folgt vor:

Read from File

In diesem Schritt wird der Dateiname, der Speicherort und das Layout der Datei identifiziert, die die zu parsenden Telefonnummern enthält.

Open Parser

Dieser Schritt definiert, ob eine kulturspezifische Domänengrammatik, die im Domäneneditor erstellt wurde, verwendet wird, oder ob eine domänenunabhängige Grammatik definiert wird. Bei einer von Ihnen im Domain Editor erstellten kulturspezifischen Parsing-Grammatik handelt es sich um eine überprüfte Parsing-Grammatik, die einer Kultur und einer Domäne zugeordnet ist. Bei einer von Ihnen in Open Parser erstellten domänenunabhängigen Parsing-Grammatik handelt es sich um eine überprüfte Parsing-Grammatik, die keiner Kultur und keiner Domäne zugeordnet ist.

In dieser Vorlage ist die Parsing-Grammatik als domänenunabhängige Grammatik definiert.

Der „Open Parser“-Schritt enthält eine Parsing-Grammatik, in der folgende Befehle und Ausdrücke definiert sind:

- `%Tokenize` ist auf „Keine“ festgelegt. Wenn `Tokenize` auf `None` gesetzt ist, muss die Parsing-Grammatikregel alle Leerzeichen und andere Token-Trennzeichen innerhalb der Regeldefinition enthalten.

- %InputField ist zum Parsen von Eingabedaten aus dem Feld **PhoneNumber** festgelegt.
- %OutputFields ist zum Teilen geparster Daten in vier Felder festgelegt: **CountryCode**, **AreaCode**, **Exchange** und **Number**.
- Der Ausdruck <root> definiert Muster von geparsten Token und umfasst OR-Anweisungen (|), wie z. B., dass sich eine gültige Telefonnummer wie folgt zusammensetzt:
- **CountryCode**, **AreaCode**, **Exchange** und **Number** OR
- **AreaCode**, **Exchange** und **Number** OR
- **Exchange** und **Number**

Zur Erstellung eines Musters für Telefonnummern verwendet die Parsing-Grammatik eine Kombination aus regulären Ausdrücken und Literalzeichen. Bei allen Zeichen, die in dieser Parsing-Grammatik in doppelten Anführungszeichen (") stehen, handelt es sich um Literalzeichen oder reguläre Ausdrücke.

Das in diesem <root>-Befehl verwendete Pluszeichen (+) wird als Literalzeichen definiert, da es in Anführungszeichen steht. Für die Angabe eines Literalzeichens können Sie einfache oder doppelte Anführungszeichen verwenden. Wird das Pluszeichen ohne Anführungsstriche verwendet, bedeutet dies, dass der Ausdruck, hinter dem es steht, einmal oder mehrere Male auftreten kann.

Die Domänenregel für Telefonnummern sind so definiert, dass sie folgenden Zeichenmustern entsprechen:

- Null oder einem Vorkommen des Zeichens „+“.
- Der Regel **CountryCode**, die eine einzelne Ziffer zwischen 0 und 9 darstellt.
- Null oder einem Vorkommen einer geöffneten Klammer, eines Bindestrichs oder eines Leerzeichens (). Wenn zwei dieser Zeichen nacheinander auftreten, führt dies zu einer Nichtübereinstimmung, d. h. zu einer ungültigen Telefonnummer.
- Der Regel **AreaCode**, die eine Sequenz von genau drei Ziffern zwischen 0 und 9 darstellt.
- Null oder einem Vorkommen einer geöffneten Klammer, eines Bindestrichs oder eines Leerzeichens (). Wenn zwei dieser Zeichen nacheinander auftreten, führt dies zu einer Nichtübereinstimmung, d. h. zu einer ungültigen Telefonnummer.
- Der Regel **Exchange**, die eine Sequenz von genau drei Ziffern zwischen 0 und 9 darstellt.
- Null oder einem Vorkommen einer geöffneten Klammer, eines Bindestrichs oder eines Leerzeichens (). Wenn zwei dieser Zeichen nacheinander auftreten, führt dies zu einer Nichtübereinstimmung, d. h. zu einer ungültigen Telefonnummer.
- Der Regel **Number**, die eine Sequenz von genau vier Ziffern zwischen 0 und 9 darstellt.

Die Regelvariablen, die die Domäne definieren, müssen die gleichen Namen wie die im erforderlichen Befehl `OutputFields` definierten Ausgabefelder verwenden.

Reguläre Ausdrücke und Ausdrucksquantifizierer

Zur Erstellung eines Musters für US-amerikanische Telefonnummern verwendet die Parsing-Grammatik eine Kombination aus regulären Ausdrücken und Ausdrucksquantifizierern. In der Parsing-Grammatik werden folgende Sonderzeichen verwendet:

- Das Zeichen „?“ bedeutet, dass ein regulärer Ausdruck keinmal oder einmal auftreten kann.
- Das Zeichen (|) gibt eine OR-Bedingung an.

- Das Zeichen „;“ steht für das Ende einer Regel.

Auf der Registerkarte **Befehle** können Sie die Bedeutung der anderen in Parsing-Grammatiken verwendbaren Sonderzeichen herausfinden, indem Sie den Mauszeiger über die Beschreibung bewegen.

Verwenden der Registerkarte „Vorschau“

Klicken Sie zum Testen der Parsing-Grammatik auf die Registerkarte Vorschau. Geben Sie die unten im Feld **PhoneNumber** angezeigten Telefonnummern ein und klicken Sie anschließend auf **Vorschau**.

| PhoneNumber | CountryCode | AreaCode | Exchange | Number |
|-----------------|-------------|----------|----------|--------|
| 1(410)286-7334 | 1 | 410 | 286 | 7334 |
| 14042867534 | 1 | 404 | 286 | 7534 |
| (410)286-7256 | | 410 | 286 | 7256 |
| 301-868-9999 | | 301 | 868 | 9999 |
| 1-222-458-7799 | 1 | 222 | 458 | 7799 |
| +1(410)286-7334 | 1 | 410 | 286 | 7334 |
| 901 888 9990 | | 901 | 888 | 9990 |
| 1 410 888 2345 | 1 | 410 | 888 | 2345 |
| 234-4567 | | | 234 | 4567 |
| 234 6789 | | | 234 | 6789 |

Sie können auch andere gültige und ungültige Telefonnummern eingeben, um zu sehen, wie die Eingabedaten geparkt werden.

Zudem können Sie mithilfe des Features „Ablaufverfolgung“ eine grafische Darstellung eines der endgültigen Parsing-Ergebnisse anzeigen oder die Parsing-Ereignisse durchlaufen. Klicken Sie auf den Link in der Spalte **Ablaufverfolgung**, um die Ablaufverfolgungsdetails für die Datenzeile anzuzeigen.

Write to File

Die Vorlage enthält einen „Write to File“-Schritt. Neben dem Eingabefeld enthält die Ausgabedatei die Felder **CountryCode**, **AreaCode**, **Exchange** und **Number**.

3 - Standardisierung

In this section

| | |
|-----------------------------------|----|
| Standardisieren von Begriffen | 59 |
| Standardisieren von Personennamen | 60 |
| Vorlagen für die Standardisierung | 62 |

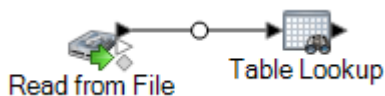
Standardisieren von Begriffen

Die inkonsistente Verwendung von Begriffen kann ein Datenqualitätsproblem sein, das Schwierigkeiten beim Parsen, bei Suchvorgängen und mehr verursacht. Sie können einen Datenfluss erstellen, der Begriffe in Ihren Daten findet, die inkonsistent verwendet werden, und sie standardisieren. Wenn Ihre Daten etwa die Begriffe „Incorporated“, „Inc.“ und „Inc“ enthalten, können Sie einen Datenfluss erstellen, um eine Form als Standardform festzulegen (z. B. „Inc.“).

Anmerkung: Vor Ausführen dieser Prozedur muss Ihr Administrator die Datenbank des Data Normalization-Moduls installieren, die standardisierte Begriffe enthält, die auf Ihre Daten angewendet werden sollen. Eine Anleitung zum Installieren von Datenbanken finden Sie im *Installationshandbuch*.

1. Erstellen Sie im Enterprise Designer einen neuen Datenfluss.
2. Ziehen Sie einen Quellschritt auf die Arbeitsfläche.
3. Doppelklicken Sie auf den Quellschritt und konfigurieren Sie ihn. Anweisungen zum Konfigurieren von Quellschritten finden Sie im *Datenfluss-Designer-Handbuch*.
4. Ziehen Sie einen „Table Lookup“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem Quellschritt.

Bei Verwendung eines „Read from File“-Quellschrittes würde Ihr Datenfluss so aussehen:

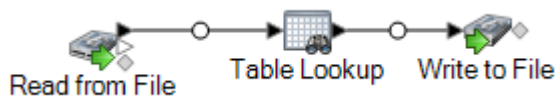


5. Doppelklicken Sie auf den „Table Lookup“-Schritt auf der Arbeitsfläche.
6. Um die Optionen für „Table Lookup“ anzugeben, erstellen Sie eine Regel. Sie können mehrere Regeln erstellen und dann die Reihenfolge angeben, in der die Regeln angewendet werden sollen. Klicken Sie auf **Hinzufügen**, um eine Regel zu erstellen.
7. Lassen Sie im Feld **Aktion** die Standardoption **Standardisieren** ausgewählt.
8. Lassen Sie im Feld **Für** die Option **Vollständiges Feld** ausgewählt, wenn das ganze Feld der Begriff ist, den Sie standardisieren möchten. Oder wählen Sie **Einzelne Begriffe im Feld**, um einzelne Wörter im Feld zu standardisieren.
9. Wählen Sie im Feld **Quelle** das Feld aus, das Sie standardisieren möchten.
10. Wählen Sie im Feld **Ziel** das Feld aus, das den standardisierten Begriff enthalten soll. Wenn Sie dasselbe Feld wie das Quellenfeld angeben, wird der Wert des Quellenfeldes durch den standardisierten Begriff ersetzt.
11. Wählen Sie im Feld **Tabelle** die Tabelle aus, die die standardisierten Begriffe enthält.

Anmerkung: Wenn Sie die erforderliche Tabelle nicht sehen, setzen sich mit Ihrem Systemadministrator in Verbindung. Die Datenbank des Data Normalization-Moduls muss geladen sein.

12. Wählen Sie im Feld **Wenn Tabellenwert nicht gefunden wird, Zielwert festlegen auf** die Option **Quellwert** aus.
13. Klicken Sie auf **OK**.
14. Definieren Sie zusätzliche Regeln, wenn Werte in mehr Feldern standardisiert werden sollen. Klicken Sie auf **OK**, wenn Sie alle Regeln definiert haben.
15. Ziehen Sie einen Datenladeschritt auf die Arbeitsfläche und verbinden Sie ihn mit „Table Lookup“.

Bei Verwendung von „Write to File“ würde Ihr Datenfluss so aussehen:



16. Doppelklicken Sie auf den Datenladeschritt und konfigurieren Sie ihn.
Informationen zum Konfigurieren von Datenladeschritten finden Sie im *Datenfluss-Designer-Handbuch*.

Sie haben jetzt einen Datenfluss, der Begriffe standardisiert.

Standardisieren von Personennamen

Diese Schritte zeigen, wie Sie einen Datenfluss erstellen, der anhand von Personennamensdaten (z. B. „John P. Smith“) gemeinsame Rufnamen desselben Namens identifiziert und eine Standardversion des Namens erstellt, mit der sich dann redundante Datensätze konsolidieren lassen.

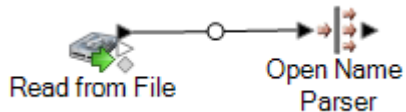
Anmerkung: Vergewissern Sie sich vorab, dass Ihre Eingabedaten ein Feld mit der Bezeichnung „Name“ umfassen, das den vollständigen Namen der Person enthält.

1. Wenn Sie es nicht bereits getan haben, laden Sie die folgenden Tabellen in den Spectrum™ Technology Platform-Server:
 - Open Parser Base
 - Open Parser Enhanced Names

Verwenden Sie das Database Load Utility des Data Normalization-Moduls, um diese Tabellen zu laden. Anleitungen zum Laden von Tabellen finden Sie im *Installationshandbuch*.
2. Erstellen Sie im Enterprise Designer einen neuen Datenfluss.

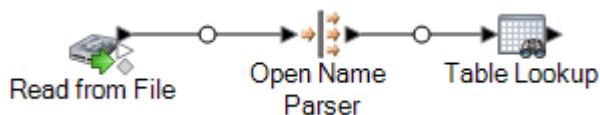
3. Ziehen Sie einen Quellschritt auf die Arbeitsfläche.
4. Doppelklicken Sie auf den Quellschritt und konfigurieren Sie ihn. Anweisungen zum Konfigurieren von Quellschritten finden Sie im *Datenfluss-Designer-Handbuch*.
5. Ziehen Sie einen „Open Name Parser“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem Quellschritt.

Wenn Sie beispielsweise einen „Read from File“-Schritt verwenden, sieht Ihr Datenfluss folgendermaßen aus:



6. Ziehen Sie einen „Table Lookup“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem „Open Name Parser“-Schritt.

Ihr Datenfluss sollte nun folgendermaßen aussehen:

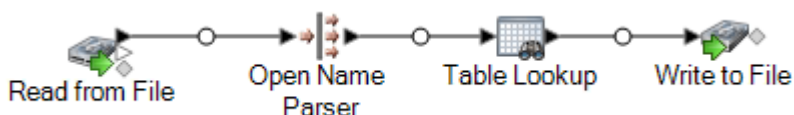


7. Doppelklicken Sie auf den „Table Lookup“-Schritt auf der Arbeitsfläche.
8. Wählen Sie im Feld **Quelle** den Eintrag **FirstName** aus.
9. Wählen Sie im Feld **Ziel** den Eintrag **FirstName** aus.

Durch Angeben desselben Feldes sowohl als Quelle als auch als Ziel wird das Feld mit der standardisierten Version des Namens aktualisiert.

10. Wählen Sie im Feld **Tabelle** den Eintrag **NickNames.xml** aus.
11. Klicken Sie auf **OK**.
12. Klicken Sie erneut auf **OK**, um das Fenster **Table Lookup-Optionen** zu schließen.
13. Ziehen Sie einen Datenladeschritt auf die Arbeitsfläche und verbinden Sie ihn mit dem „Table Lookup“-Schritt.

Würden Sie beispielsweise eine „Write to File“-Datenladung verwenden, würde Ihr Datenfluss jetzt so aussehen:



14. Doppelklicken Sie auf den Datenladeschritt und konfigurieren Sie ihn. Anweisungen zum Konfigurieren von Quellschritten finden Sie im *Datenfluss-Designer-Handbuch*.

Sie haben jetzt einen Datenfluss, der anhand von Personennamen den Vornamen standardisiert und Rufnamen durch die Standardform des Namens ersetzt.

Vorlagen für die Standardisierung

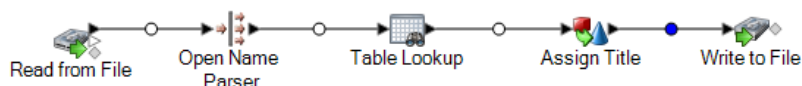
Formalisierung von Personennamen

Diese Datenflussvorlage zeigt, wie mit Personennamensdaten (z. B. „John P. Smith“) umgegangen wird, wie häufige Spitznamen desselben Namens identifiziert werden und wie eine Standardversion des Namens erstellt wird, mit der sich dann redundante Datensätze konsolidieren lassen. Außerdem zeigt sie, wie Sie Titel-/Anredatedaten basierend auf den Geschlechtsdaten hinzufügen können.

Geschäftsszenario

Sie arbeiten für eine Wohltätigkeitsorganisation, die Einladungen für einen Gala-Abend verschicken möchte. Ihre Eingabedaten enthalten Namensdaten als vollständige Namen und Sie möchten die Namensdaten in die Felder „Vorname“, „Zweiter Vorname“ und „Nachname“ parsen sowie das Feld „Anrede“ hinzufügen, um Ihre Einladungen formaler zu gestalten. Sie möchten außerdem Spitznamen in Ihren Namensdaten ersetzen, um eine höflichere Variante des Namens zu nutzen.

Im folgenden Datenfluss wird eine Lösung für das Geschäftsszenario bereitgestellt:



Diese Datenflussvorlage ist im Enterprise Designer verfügbar. Öffnen Sie **Datei > Neu > Datenfluss > Aus Vorlage**, und wählen Sie **StandardizePersonalNames** aus. Dieser Datenfluss erfordert das Data Normalization-Modul und das Universal Name-Modul.

Dieser Datenfluss geht bei jeder Datenzeile in der Eingabedatei wie folgt vor:

Read from File

Bei diesem Schritt werden der Dateiname, der Speicherort und das Layout der Datei identifiziert, welche die zu parsenden Namen enthält. Die Datei enthält männliche und weibliche Namen.

Name Parser

In dieser Vorlage ist der „Name Parser“-Schritt als „Parse Personal Name“ benannt. Der „Parse Personal Name“-Schritt untersucht Namensfelder und vergleicht sie mit den Namensdaten, die in den Namensdatenbankdateien von Spectrum™ Technology Platform gespeichert sind. Basierend auf dem Vergleich parst er die Namensdaten in den Feldern „Vorname“, „Zweiter Vorname“ und

„Nachname“ und weist jedem Namen einen Entitätstyp sowie ein Geschlecht zu. Er verwendet neben den Namensdaten außerdem die Mustererkennung.

In dieser Vorlage wird der „Parse Personal Name“-Schritt wie folgt konfiguriert.

- „Personennamen parsen“ wird ausgewählt und „Unternehmensnamen parsen“ wird gelöscht. Bei der Auswahl dieser Optionen werden Vornamen nach Geschlecht, Reihenfolge und Interpunktion ausgewertet und es wird keine Bewertung der Unternehmensnamen durchgeführt.
- „Quelle für Geschlechtsbestimmung“ ist auf „Standard“ eingestellt. In den meisten Fällen ist „Standard“ die beste Einstellung für die Geschlechtsbestimmung, da mit ihr ein breites Spektrum von Namen abgedeckt wird. Wenn Sie jedoch Namen aus einer bestimmten Kultur verarbeiten, wählen Sie diese Kultur aus. Die Auswahl einer bestimmten Kultur hilft bei der Sicherstellung, dass den Namen das richtige Geschlecht zugewiesen wird. Beispiel: Wenn Sie „Standard“ aktiviert lassen, wird der Name Jean als weiblicher Name identifiziert. Wenn Sie jedoch Französisch auswählen, wird der Name als männlicher Name identifiziert.
- „Reihenfolge“ ist auf „Natürlich“ eingestellt. Die Namensfelder sind nach Titel, Vorname, zweiter Vorname, Nachname und Suffix sortiert.
- „Punkte beibehalten“ ist deaktiviert. Das heißt, dass keinerlei Interpunktion in den Namensdaten beibehalten wird.

Transformer

In dieser Vorlage wird der Transformer-Schritt mit „Assign Titles“ benannt: Der „Assign Titles“-Schritt verwendet ein benutzerdefiniertes Skript, um jede Zeile in der Datenstromausgabe mithilfe des „Parse Personal Name“-Schrittes zu durchsuchen und einen **TitleOfRespect**-Wert basierend auf dem **GenderCode**-Wert zuzuweisen.

Das benutzerdefinierte Skript lautet wie folgt:

```
if (row.get('TitleOfRespect') == '')
{
  if (row.get('GenderCode') == 'M')
    row.set('TitleOfRespect', 'Mr')
  if (row.get('GenderCode') == 'F')
    row.set('TitleOfRespect', 'Ms')
```

Jedes Mal, wenn der „Assign Titles“-Schritt den Buchstaben **M** im Feld **GenderCode** erkennt, wird der Wert für **TitleOfRespect** auf **Mr** (Herr) eingestellt. Jedes Mal, wenn der „Assign Titles“-Schritt den Buchstaben **F** im Feld **GenderCode** erkennt, wird der Wert für **TitleOfRespect** auf **Ms** (Frau) eingestellt.

Standardisierung

In dieser Vorlage wird der Standardization-Schritt mit „Standardize Nicknames“ benannt: Der „Standardize Nicknames“-Schritt sucht zunächst nach Vornamen in der „Nicknames.xml“-Datenbank und ersetzt alle Spitznamen durch eine regulärere Form des Namens. Der Name Tommy wird z. B. durch Thomas ersetzt.

Write to File

Die Vorlage enthält einen „Write to File“-Schritt. Neben den Eingabefeldern enthält die Ausgabedatei die Felder „TitleOfRespect“, „FirstName“, „MiddleName“, „LastName“, „EntityType“, „GenderCode“ und „GenderDeterminationSource“.

4 - Vergleichen

In this section

| | |
|---|-----|
| Terminologie für Vergleiche | 66 |
| Techniken zur Definition von Vergleichsschlüsseln | 68 |
| Vergleichsregeln | 71 |
| Vergleichen von Datensätzen aus einer einzelnen Quelle | 87 |
| Vergleichen von Datensätzen einer Quelle mit Datensätzen einer anderen Quelle | 93 |
| Vergleichen von Datensätzen zwischen und innerhalb von Quellen | 100 |
| Vergleichen von Datensätzen mit einer Datenbank | 106 |
| Vergleichen von Datensätzen mithilfe mehrerer Vergleichsregeln | 109 |
| Erstellen eines allgemeinen Vergleichsdienstes | 113 |
| Verwenden eines Schnellvergleichsschlüssels | 117 |
| Analysieren von Vergleichsergebnissen | 121 |
| Datenflussvorlagen zum Vergleichen | 138 |

Terminologie für Vergleiche

| | |
|--|--|
| Durchschnittliche Punktzahl | Der durchschnittliche Übereinstimmungswert aller Dubletten. Möglich sind Werte zwischen 0 und 100, wobei 0 eine niedrige Übereinstimmung und 100 eine genaue Übereinstimmung angibt. |
| Basis | Das ausgewählte Vergleichsergebnis, das mit einem anderen Vergleichsergebnis verglichen wird. |
| Kandidatengruppe | Kopfdubletten- und Kandidatendatensätze, die zusammen nach einer ID gruppiert sind, die durch Candidate Finder zugewiesen wurde. Die Kopfdublette (der erste Datensatz in der Gruppe) ist ein Datensatz, der aus einer Input-Quelle gelesen wird, während Kandidaten in der Regel Datensätze sind, die mithilfe einer SQL-Abfrage in einer Datenbank ermittelt werden. |
| Kandidatendatensätze | Alle Datensätze in einer Übereinstimmungsgruppe oder Kandidatengruppe, bei denen es sich nicht um Kopfdubletten handelt. |
| Drop | Eine Abnahme der Dubletten. |
| Übereinstimmender Datensatz – Details | Ein einzelner Datensatz, der mit einem Datensatz übereinstimmt, der bei einem Vergleichsschritt verarbeitet wurde. Jeder Datensatz stellt Informationen darüber bereit, ob es sich bei dem Datensatz um eine Kopfdublette, einen eindeutigen Datensatz oder eine Dublette handelt. Außerdem werden Informationen über seine Übereinstimmungsgruppe oder Kandidatengruppe und Ausgabesammlung angezeigt. Kandidatendatensätze bieten Informationen zu den Gründen, warum der Eingabedatensatz mit seiner Kopfdublette übereinstimmte oder nicht damit übereinstimmte. |
| Sammlungsduplikate | Ein Sammlungsduplikat besteht aus einer Kopfdublette und ihren Datensatzdubletten, welche wiederum nach einer Sammlungsnummer gruppiert sind. Eindeutige Datensätze gehören immer zu Sammlungsnummer 0. |
| Datensatzduplikate | Ein Datensatz, der mit einem anderen Datensatz innerhalb einer Übereinstimmungsgruppe übereinstimmt. Dies kann eine Kopfdublette oder ein Kandidat sein. |
| Ausdrückliche Übereinstimmungen | Ein Schnellvergleich erfolgt, wenn eine Kopfdublette und ein Kandidat eine genaue Übereinstimmung hinsichtlich der Inhalte eines festgelegten Feldes aufweisen. In der Regel ist dies ein vom Match Key Generator bereitgestellter ExpressMatchKey. Wenn ein Schnellvergleich durchgeführt wird, wird die Verarbeitung nicht fortgesetzt, um zu ermitteln, ob die Kopfdublette und der Kandidat Dubletten sind. |
| Eingabedatensätze | Die Reihenfolge der Datensätze im Vergleichsschritt, bevor die Vergleichssortierung durchgeführt wird. |
| Interflow Match | Ein Vergleichsschritt, der Übereinstimmungen zwischen ähnlichen Datensätzen zwischen zwei Eingabedatensatzstreams ermittelt. Der |

| | |
|---------------------------------------|---|
| | erste Datensatzstream bildet eine Quelle für Kopfdubletten-Datensätze und der zweite Stream ist eine Quelle für Kandidatendatensätze. |
| Intraflow Match | Ein Vergleichsschritt, der Übereinstimmungen zwischen ähnlichen Datensätzen innerhalb eines einzelnen Eingabestreams ermittelt. |
| Lift | Ein Anstieg in Dubletten. |
| Übereinstimmungsgruppen | (Gruppieren nach) Entweder nach einem Vergleichsschlüssel oder nach einem gleitenden Fenster gruppierte Datensätze. |
| Vergleichsergebnisse | (oder Ressourcengruppe) Logische Gruppierung von Dateien, die durch einen Schritt erzeugt wurden. Diese Daten werden für jede Ausführung eines Schrittes auf der Festplatte gespeichert. Durch nachfolgende Ausführungen werden die Ergebnisse einer vorherigen Ausführung nicht überschrieben oder geändert. In MAT dienen die Gruppen dazu, Informationen zu den zusammenfassenden und detaillierten Ergebnissen sowie Informationen zu den Einstellungen zu liefern. |
| Liste der Vergleichsergebnisse | Liste der Vergleichsergebnisse eines einzelnen Typs, die MAT in der aktuellen Analysesitzung analysieren kann. |
| Typ der Vergleichsergebnisse | Gibt die Inhalte der Vergleichsergebnisse an. Bei MAT dient der Typ der Vergleichsergebnisse zum Ermitteln, wie die Daten zu verwenden sind. |
| Vergleichsschritt | Ein Schritt auf der Arbeitsfläche, der Vergleichsvorgänge ausführt. Die Vergleichsschritte sind „Interflow Match“, „Intraflow Match“ und „Transactional Match“. |
| Übersehene Übereinstimmung | Ein Datensatz, der zuvor eine Kopfdublette oder eine Dublette war und nun als eindeutig eingeordnet wird. |
| Neue Übereinstimmung | Ein Datensatz, der zuvor als eindeutig eingeordnet wurde und nun eine Kopfdublette oder eine Dublette ist. |
| Gleitendes Fenster | Die Vergleichsmethode „Gleitendes Fenster“ füllt der Reihe nach eine vorbestimmte Puffergröße, Fenster genannt, mit der entsprechenden Menge an Datenzeilen. Bei jedem Hinzufügen einer Zeile in das Fenster wird sie mit jedem Element verglichen, das sich bereits im Fenster befindet. |
| Kopfdubletten-Datensätze | Ein Treiberdatensatz, der mit Kandidaten innerhalb einer Übereinstimmungsgruppe oder Kandidatengruppe verglichen wird. |
| Transactional Match | Ein Vergleichsschritt, der Kopfdubletten-Datensätze mit Kandidatendatensätzen abgleicht, die von Candidate Finder oder von einer externen Anwendung zurückgegeben wurden. |
| Eindeutige Datensätze | Ein Kopfdubletten- oder Kandidatendatensatz, der mit keinen anderen Datensätzen in einer Übereinstimmungsgruppe übereinstimmt. Wenn er den einzigen Datensatz in einer Übereinstimmungsgruppe bildet, wird der Kopfdubletten-Datensatz automatisch als eindeutig eingeordnet. |

Techniken zur Definition von Vergleichsschlüsseln

Ein effektiver und effizienter Vergleich erfordert das richtige Gleichgewicht zwischen Genauigkeit und Leistung. Der akkurateste Ansatz zum Vergleichen wäre die Analyse jedes Datensatzes im Vergleich zu allen anderen Datensätzen. Dies ist unpraktisch, da die Anzahl der Datensätze, die verarbeitet werden müssten, zu einer unannehmbar schlechten Leistung führen würde. Ein besserer Ansatz besteht darin, die Anzahl der im Vergleichsprozess beteiligten Datensätze auf diejenigen zu beschränken, die am wahrscheinlichsten übereinstimmen. Sie können dies mithilfe von Vergleichsschlüsseln tun. Ein Vergleichsschlüssel ist ein Wert, der mithilfe eines von Ihnen definierten Algorithmus für jeden Datensatz erstellt wird. Der Algorithmus nimmt Werte aus dem Datensatz und verwendet sie, um einen Vergleichsschlüsselwert zu erzeugen, der als neues Feld im Datensatz gespeichert wird.

Beispiel: Wenn der eingehende Datensatz wie folgt lautet:

Vorname – Fred
 Nachname – Mertz
 Postleitzahl – 21114-1687
 Geschlechtscode – M

Und Sie eine Vergleichsschlüsselregel definieren, die einen Vergleichsschlüssel generiert, indem Daten aus dem Datensatz wie folgt kombiniert werden:

| Eingabefeld | Startposition | Länge |
|-----------------|---------------|-------|
| Postleitzahl | 1 | 5 |
| Postleitzahl | 7 | 4 |
| Nachname | 1 | 5 |
| Vorname | 1 | 5 |
| Geschlechtscode | 1 | 1 |

Dann ergäbe sich der Schlüssel:

211141687MertzFredM

Alle Datensätze mit demselben Vergleichsschlüssel werden in einer Übereinstimmungsgruppe platziert. Der Vergleichsprozess vergleicht anschließend die Datensätze in der Gruppe miteinander, um Übereinstimmungen zu finden.

Wenn Sie einen Vergleichsschlüssel erstellen möchten, verwenden Sie einen „Match Key Generator“-Schritt, wenn Sie Datensätze mithilfe von Interflow Match oder Intraflow Match vergleichen. Wenn Sie mithilfe von Transactional Match vergleichen, verwenden Sie den „Candidate Finder“-Schritt, um Übereinstimmungsgruppen zu erstellen.

Anmerkung: Die folgenden Richtlinien können auf „Match Key Generator“-Schlüssel und „Candidate Finder“-Abfragen angewendet werden. Für „Candidate Finder“ gelten diese Richtlinien für die Definition der SELECT-Anweisung.

Größe und Leistung einer Übereinstimmungsgruppe

Der Vergleichsschlüssel bestimmt die Größe der Übereinstimmungsgruppe und damit die Leistung Ihres Datenflusses. Wenn die Größe der Übereinstimmungsgruppe verdoppelt wird, verdoppelt sich die Ausführungszeit. Wenn Sie beispielsweise einen Vergleichsschlüssel definieren, der eine Gruppe von 20 potenziell übereinstimmenden Datensätzen erzeugt, dauert die Verarbeitung doppelt so lange, als wenn Sie den Vergleichsschlüssel so ändern, dass die Übereinstimmungsgruppe nur 10 potenziell übereinstimmende Datensätze enthält. Der Nachteil einer „Verschärfung“ der Vergleichsschlüsselregel für eine kleinere Übereinstimmungsgruppe ist, dass Sie den Ausschluss von übereinstimmenden Datensätzen riskieren. Das „Lockern“ der Vergleichsschlüsselregeln verringert die Wahrscheinlichkeit, dass ein übereinstimmender Datensatz aus der Gruppe ausgeschlossen wird. Dadurch erhöht sich jedoch die Gruppengröße. Um das richtige Gleichgewicht für Ihre Daten zu finden, ist es wichtig, dass Sie mit einer Vielzahl von Vergleichsschlüsselregeln testen. Verwenden Sie dabei repräsentative Daten, die Sie in der Produktion verarbeiten möchten.

Dichte

Beim Entwurf eines Vergleichsschlüssels ist es wichtig, die Dichte der Daten zu berücksichtigen. Dichte bezieht sich auf den Grad, in dem die Daten auf Übereinstimmungsgruppen verteilt werden können. Da die Leistung durch die Anzahl der vom System durchgeführten Vergleiche bestimmt wird, führen die Vergleichsschlüssel, die eine kleine Anzahl von großen Übereinstimmungsgruppen erzeugen, zu einer schlechteren Leistung als die Vergleichsschlüssel, die eine große Anzahl von kleinen Übereinstimmungsgruppen erzeugen.

Um dieses Konzept zu veranschaulichen, betrachten Sie eine Situation, in der Sie einen Satz aus einer Million zu vergleichender Namens- und Adressdatensätze haben. Sie können einen Vergleichsschlüssel als die ersten drei Byte der Postleitzahl und den ersten Buchstaben des Nachnamens definieren. Wenn die Datensätze alle aus den USA stammen, würde der Vergleichsschlüssel eine gute Anzahl von Übereinstimmungsgruppen erzeugen und wahrscheinlich eine annehmbare Leistung aufweisen. Wenn jedoch alle Datensätze aus New York stammen, würden die Postleitzahlen alle mit „100“ beginnen und Sie würden letztlich nur 26 Übereinstimmungsgruppen erhalten. Dies würde große Übereinstimmungsgruppen erzeugen, die durchschnittlich etwa 38.000 Datensätze enthalten.

Sie können die maximale Anzahl der für jede Übereinstimmungsgruppe durchgeführten Vergleiche berechnen, indem Sie die folgende Formel anwenden:

$$N * (N-1) / 2$$

Dabei steht N für die Anzahl der Datensätze in der Übereinstimmungsgruppe.

Wenn Sie also über 26 Übereinstimmungsgruppen mit jeweils 38.000 Datensätzen verfügen, würde die maximale Anzahl der durchgeführten Vergleiche etwa 18,7 Milliarden betragen. So wird diese Zahl berechnet:

Bestimmen Sie zunächst die maximale Anzahl der Vergleiche pro Übereinstimmungsgruppe:

$$38,000 * (38,000-1) / 2 = 721,981,000$$

Multiplizieren Sie anschließend diesen Betrag mit der Anzahl der Übereinstimmungsgruppen:

$$721,981,000 * 26 = 18,771,506,000$$

Gäbe es stattdessen 100 eindeutige Werte für die ersten 3 Byte der Postleitzahl, so hätten Sie 2.600 Übereinstimmungsgruppen mit durchschnittlich 380 Datensätzen. In diesem Fall würde die maximale Anzahl der Vergleiche 187 Millionen betragen. Das ist 100 Mal weniger. Wenn also die Datensätze nur aus New York stammen, könnten Sie die ersten vier oder sogar fünf Bytes der Postleitzahl für den Vergleichsschlüssel verwenden, um mehr Übereinstimmungsgruppen zu erzeugen und die Anzahl der Vergleiche zu reduzieren. Dabei könnten Sie einige Übereinstimmungen überspringen, aber der Vorteil wäre eine enorm kürzere Ausführungszeit.

In Wirklichkeit führt ein Vergleichsschlüssel wie in diesem Beispiel aufgrund von Abweichungen in den Daten nicht zu Übereinstimmungsgruppen gleicher Größe. Es gibt z. B. viel mehr Personen, deren Nachname mit „S“ beginnt als mit „X“. Aus diesem Grund sollten Sie Ihre Bemühungen auf die Größenverringerung der größten Übereinstimmungsgruppen konzentrieren. Eine Übereinstimmungsgruppe mit 100.000 Datensätzen ist 10 Mal größer als eine Übereinstimmungsgruppe mit 10.000 Datensätzen. Sie erfordert jedoch 100 Mal mehr Vergleiche und wird 100 Mal so lange dauern. Nehmen wir beispielsweise an, dass Sie eine Postleitzahl mit fünf Byte und das Feld „AddressLine1“ mit sechs Byte für Ihren Vergleichsschlüssel verwenden. Oberflächlich sieht das nach einem relativ guten Vergleichsschlüssel aus. Das Problem sind jedoch Postfachadressen. Während die meisten Übereinstimmungsgruppen eine akzeptable Größe haben können, gibt es einige sehr große Übereinstimmungsgruppen mit Schlüsseln wie 10002PO BOX, die eine sehr große Anzahl von Datensätzen enthalten. Um die großen Übereinstimmungsgruppen aufzubrechen, können Sie den Vergleichsschlüssel ändern, um die ersten Ziffern der Postfachnummer einzubeziehen.

Abstimmen des Vergleichsschlüssels mit der Vergleichsregel

Um genaueste Ergebnisse zu erzielen, sollten Sie den Vergleichsschlüssel so gestalten, dass er gut mit der Vergleichsregel zusammenarbeitet, mit der Sie ihn verwenden werden. Dazu müssen Sie berücksichtigen, wie die Vergleichsregel definiert ist.

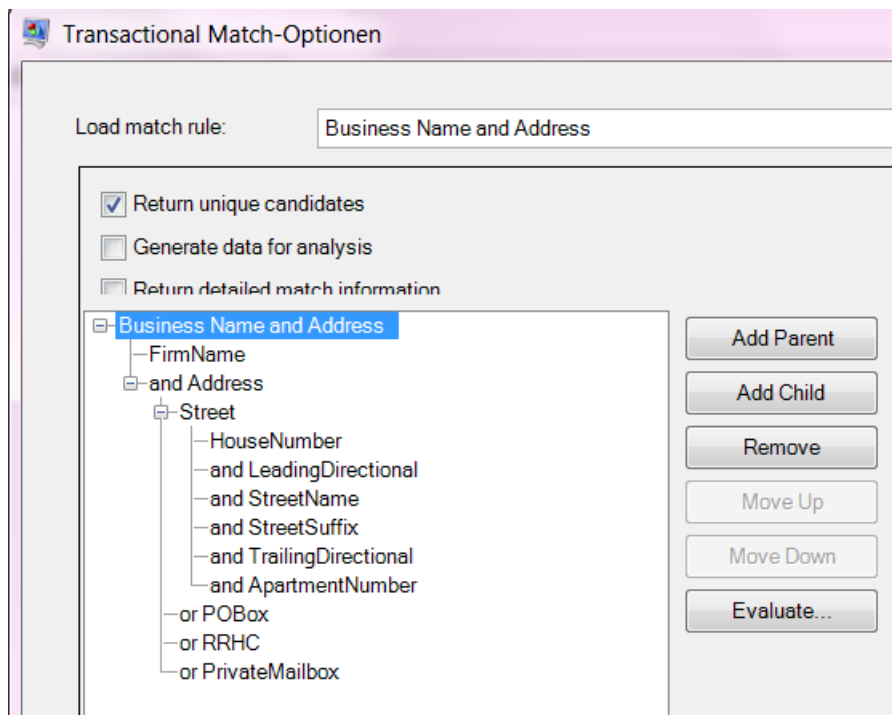
- Der Vergleichsschlüssel sollte alle Felder enthalten, die die Vergleichsregel für eine genaue Übereinstimmung erfordert.
- Der Vergleichsschlüssel sollte die gleiche Art von Algorithmus wie die Vergleichsregel verwenden. Wenn Sie beispielsweise einen Vergleichsschlüssel für die Verwendung einer Vergleichsregel mit einem phonetischen Algorithmus erstellen, sollte die Vergleichsschlüssel ebenfalls einen phonetischen Algorithmus verwenden.
- Der Vergleichsschlüssel sollte mithilfe der Daten aus allen Feldern erstellt werden, die in der Vergleichsregel verwendet werden.

- Überlegen Sie, inwieweit der Vergleichsschlüssel betroffen ist, wenn Daten aus einem oder mehreren Feldern fehlen, die für den Vergleichsschlüssel verwendet werden. Gehen wir davon aus, dass Sie eine mittlere Initiale als Teil des Vergleichsschlüssels verwenden, und Sie haben einen Datensatz für Max A. Meier und einen anderen für Max Meier. Sie haben die Vergleichsregel so konfiguriert, dass sie leere Werte im Feld für die mittlere Initiale ignoriert. Dadurch würden diese beiden Datensätze entsprechend Ihrer Vergleichsregel übereinstimmen. Da der Vergleichsschlüssel jedoch die mittlere Initiale verwendet, würden die beiden Datensätze in verschiedenen Übereinstimmungsgruppen landen und nicht miteinander verglichen werden. Dies würde den Zweck Ihrer Vergleichsregel aufheben.

Vergleichsregeln

Jeder der Vergleichsschritte („Interflow Match“, „Intraflow Match“ und „Transactional Match“) erfordert, dass Sie eine Vergleichsregel konfigurieren. Vergleichsregeln definieren die Kriterien, anhand denen bestimmt wird, ob ein Datensatz mit einem anderen übereinstimmt. Sie geben an, welche Felder wie verglichen werden sollen. Darüber hinaus definieren sie eine Hierarchie von Vergleichen, um komplexe Regeln für Vergleiche zu ermöglichen.

Die Erstellung eines hierarchischen Satzes von Vergleichen ermöglicht Ihnen, verschachtelte boolesche Vergleichsregeln zu bilden. Betrachten Sie zum Beispiel die folgende Vergleichsregel:



In diesem Beispiel versucht die Vergleichsregel, Datensätze auf Basis eines Unternehmensnamens und einer Adresse zu vergleichen. Das erste Element der Vergleichsregel ist das Feld „FirmName“.

Dieses Element bedeutet, dass der Wert im Feld „FirmName“ übereinstimmen muss, damit Datensätze übereinstimmen. Das zweite Element wertet die Adresse aus. Beachten Sie, dass ihm der logische Operator AND vorangestellt ist, was bedeutet, dass sowohl „FirmName“ als auch „Address“ übereinstimmen müssen, damit Datensätze übereinstimmen. Der „Address“-Teil der Vergleichsregel besteht aus untergeordneten Regeln, die vier Typen von Adressen auswerten: „Street“, „POBox“, „RRHC“ (Rural Route/Highway Contract) und „PrivateMailbox“. Das untergeordnete Element „Street“ analysiert die Datenflussfelder „HouseNumber“, „LeadingDirectional“, „StreetName“, „StreetSuffix“, „TrailingDirectional“ und „ApartmentNumber“. Wenn diese alle übereinstimmen, werden die übergeordnete Regel „Street“ und deren übergeordnete Regel „Address“ alle als „wahr“ (true) ausgewertet. Wenn die Regel „Street“ nicht als wahr ausgewertet wird, wird das Feld „POBox“, danach „RRHC“ und danach „PrivateMailbox“ ausgewertet. Wenn eines dieser drei übereinstimmt, dann stimmt auch das übergeordnete Element „Address“ überein.

Erstellen einer Vergleichsregel

Vergleichsregeln werden in „Interflow Match“, „Intraflow Match“ und „Transactional Match“ verwendet, um die Kriterien zu definieren, die bestimmen, ob ein Datensatz mit einem anderen übereinstimmt. Vergleichsregeln geben an, welche Felder wie verglichen werden sollen. Darüber hinaus definieren sie eine Hierarchie von Vergleichen, um komplexe Regeln für Vergleiche zu ermöglichen.

Sie können Vergleichsregeln in „Interflow Match“, „Intraflow Match“ und „Transactional Match“ erstellen. Sie können Vergleichsregeln auch im Tool „Vergleichsregelverwaltung“ des Enterprise Designer erstellen. Beim Erstellen einer Regel im Tool „Vergleichsregelverwaltung“ wird die Regel zur Verwendung in einem beliebigen Datenfluss sowie auch für andere Benutzer verfügbar. Beim Erstellen einer Vergleichsregel in einem der vier Vergleichsschritte wird die Regel nur für den betreffenden Schritt verfügbar, außer Sie speichern sie durch Klicken auf die Schaltfläche **Speichern**, wodurch sie auch für andere Schritte und Benutzer verfügbar wird.

1. Öffnen Sie den Enterprise Designer.
2. Führen Sie eine der folgenden Aktionen aus:
 - Wenn Sie eine Vergleichsregel in „Interflow Match“, „Intraflow Match“ oder „Transactional Match“ definieren möchten, doppelklicken Sie auf den Vergleichsschritt, für den Sie eine Vergleichsregel definieren möchten. Wählen Sie im Feld **Vergleichsregel laden** eine vordefinierte Vergleichsregel als Ausgangspunkt. Wenn Sie mit einer leeren Vergleichsregel beginnen möchten, klicken Sie auf **Neu**.
 - Wenn Sie eine Vergleichsregel im Tool „Vergleichsregelverwaltung“ definieren möchten, wählen Sie **Tools > Vergleichsregelverwaltung** aus. Wenn Sie eine vorhandene Regel als Ausgangspunkt für Ihre Regel verwenden möchten, aktivieren Sie das Kästchen **Kopieren von** und wählen Sie die gewünschte Regel aus.
3. Geben Sie die Datenflussfelder an, die Sie in der Vergleichsregel verwenden möchten, und ebenso die Vergleichsregelhierarchie.
 - a) Klicken Sie auf **Übergeordnetes Element hinzufügen**.

- b) Geben Sie einen Namen für das übergeordnete Element ein. Der Name muss eindeutig sein, und es darf kein Feld sein. Das erste übergeordnete Element in der Hierarchie wird als Vergleichsregelname im Feld **Vergleichsregel laden** verwendet. Alle benutzerdefinierten Vergleichsregeln, die Sie erstellen, und alle vordefinierten Regeln, die Sie ändern, werden mit „Benutzerdefiniert“ als Präfix vor dem Namen gespeichert.
- c) Klicken Sie auf **Untergeordnetes Element hinzufügen**. Ein Dropdown-Menü wird in der Regelhierarchie angezeigt. Wählen Sie ein Feld aus, das dem übergeordneten Element hinzugefügt werden soll.

Anmerkung: Alle untergeordneten Elemente unter einem übergeordneten Element müssen denselben logischen Operator verwenden. Wenn Sie andere logische Operatoren zwischen Feldern verwenden möchten, müssen Sie zuerst temporäre übergeordnete Elemente erstellen.

- d) Wiederholen Sie diese Schritte, um Ihre Vergleichshierarchie abzuschließen.
4. Übergeordnete Optionen definieren Übergeordnete Optionen werden rechts von der Regelhierarchie angezeigt, wenn ein übergeordneter Knoten ausgewählt wird.
- a) Klicken Sie auf **Übereinstimmung, wenn nicht wahr**, um den logischen Operator für das übergeordnete Element von AND in AND NOT zu ändern. Wenn Sie diese Option auswählen, stimmen Datensätze nur überein, wenn sie der in diesem übergeordneten Element definierten Logik nicht entsprechen.

Anmerkung: Die Aktivierung der Option **Übereinstimmung, wenn nicht wahr** hat den Effekt, dass die Optionen von **Vergleichsmethode** negiert werden. Weitere Informationen finden Sie unter **Negative Vergleichsbedingungen** auf Seite 83.

- b) Geben Sie im Feld **Vergleichsmethode** an, wie ermittelt werden soll, ob ein übergeordnetes Element eine Übereinstimmung oder keine Übereinstimmung ist. Zur Auswahl stehen:

Alle Übereinstimmungen Ein übergeordnetes Element wird als eine Übereinstimmung gewertet, wenn alle untergeordneten Elemente als übereinstimmend erkannt werden. Über diese Methode wird eine AND-Verbindungsstraße zwischen den untergeordneten Elementen erstellt.

Beliebige Übereinstimmung Ein übergeordnetes Element wird als eine Übereinstimmung gewertet, wenn mindestens ein untergeordnetes Element als übereinstimmend erkannt wird. Über diese Methode wird eine OR-Verbindungsstraße zwischen den untergeordneten Elementen erstellt.

Basierend auf Schwellenwert Ein übergeordnetes Element wird als Übereinstimmung gewertet, wenn die Punktzahl des übergeordneten Elements größer oder gleich dem Schwellenwert des übergeordneten Elements ist. Wenn Sie diese Option auswählen, wird der Schieberegler **Schwellenwert** angezeigt. Verwenden Sie diesen Schieberegler, um einen Schwellenwert anzugeben. Die Bewertungsmethode bestimmt,

welche logische Verbindung verwendet wird. Schwellenwerte übergeordneter Elemente können nicht größer sein als die der untergeordneten Elemente.

Anmerkung: Der hier festgelegte Schwellenwert kann zur Laufzeit im Dialogfeld „Datenflussoptionen“ überschrieben werden. Rufen Sie **Bearbeiten > Datenflussoptionen** auf und klicken Sie auf **Hinzufügen**. Erweitern Sie den Schritt, klicken Sie auf **Schwellenwert für übergeordnetes Element** und geben Sie den Schwellenwert in das Feld **Standardwert** ein.

- c) Geben Sie im Feld **Fehlende Daten** an, wie leere Daten in einem Feld bewertet werden sollen. Zur Auswahl stehen:

| | |
|--------------------------------|---|
| Leerzeichen ignorieren | Ignoriert das Feld, wenn es leere Daten enthält. |
| Als 0 zählen | Bewertet das Feld mit 0, wenn es leere Daten enthält. |
| Als 100 zählen | Bewertet das Feld mit 100, wenn es leere Daten enthält. |
| Leerzeichen vergleichen | Bewertet die Kopfdoubletten- und Kandidatenfelder mit 100, wenn beide leere Daten enthalten; bewertet andernfalls die Kopfdoubletten- und Kandidatenfelder mit 0. |

- d) Wählen Sie im Feld **Bewertungsmethode** die Methode aus, die verwendet wird, um die Punktzahl für die Übereinstimmung zu ermitteln. Zur Auswahl stehen:

| | |
|-------------------------------|--|
| Gewichteter Mittelwert | Verwendet die Gewichtung jedes untergeordneten Elements, um die durchschnittliche Punktzahl für die Übereinstimmung zu bestimmen. |
| Durchschnitt | Verwendet die durchschnittliche Punktzahl von jedem untergeordneten Element, um die Punktzahl eines übergeordneten Elements zu bestimmen. |
| Maximum | Verwendet die höchste Punktzahl untergeordneter Elemente, um die Punktzahl eines übergeordneten Elements zu bestimmen. |
| Minimum | Verwendet die niedrigste Punktzahl untergeordneter Elemente, um die Punktzahl eines übergeordneten Elements zu bestimmen. |
| Vektoraddition | Verwendet die Vektoraddition der Punktzahlen aller untergeordneten Elemente, um die Punktzahl des übergeordneten Elements zu bestimmen. Die Formel zur Berechnung ist: $\sqrt{a^2 + b^2 + c^2} / \sqrt{n}$ Datei gilt: a, b und c sind die Punktzahlen von drei untergeordneten Elementen, und n ist die Anzahl der untergeordneten Elemente. |

Die folgende Tabelle zeigt die logische Beziehung zwischen Vergleichsmethoden und Bewertungsmethoden und wie sich durch jede Kombination die bei der Vergleichsverarbeitung verwendete Logik ändert.

Tabelle 1: Matrix Vergleichsmethode zu Bewertungsmethode

| Bewertungsmethode | Vergleichsmethode | | | Anmerkungen |
|------------------------|---------------------------|------------------------|-----------------------------|---|
| | Beliebige Übereinstimmung | Alle Übereinstimmungen | Basierend auf Schwellenwert | |
| Gewichteter Mittelwert | nicht verfügbar | AND | AND | Nur verfügbar, wenn „Alle Übereinstimmungen“ oder „Basierend auf Schwellenwert“ als Vergleichsmethode ausgewählt wird. |
| Durchschnitt | nicht verfügbar | AND | AND | |
| Vektoraddition | nicht verfügbar | AND | AND | |
| Maximum | OR | nicht verfügbar | OR | Nur verfügbar, wenn „Beliebige Übereinstimmung“ oder „Basierend auf Schwellenwert“ als Vergleichsmethode ausgewählt wird. |
| Minimum | OR | nicht verfügbar | OR | |

5. Definieren Sie untergeordnete Optionen. Untergeordnete Optionen werden rechts von der Regelhierarchie angezeigt, wenn ein untergeordnetes Element ausgewählt wird.
 - a) Aktivieren Sie die Option **Kandidatenfeld**, um den ausgewählten untergeordneten Datensatz einem Feld in der Eingabedatei zuzuordnen.
 - b) Aktivieren Sie die Option **Kreuzprobe mit** und wählen Sie ein oder mehr Elemente aus der Dropdown-Liste aus, um unterschiedliche Felder zwischen zwei Datensätzen miteinander zu vergleichen. Wenn Sie das Tool „Vergleichsregelverwaltung“ verwenden, um eine Vergleichsregel zu erstellen oder zu bearbeiten, wird keine Dropdown-Liste angezeigt. Stattdessen müssen Sie dann jeden Feldnamen, getrennt durch Kommas, eingeben.
 - c) Klicken Sie auf **Übereinstimmung, wenn nicht wahr**, um den logischen Operator von AND in NOT zu ändern. Wenn Sie diese Option auswählen, wird die Vergleichsregel nur als wahr ausgewertet, wenn die Datensätze nicht mit der in diesem untergeordneten Element definierten Logik übereinstimmen.

Wenn Sie beispielsweise Personen identifizieren möchten, die mehreren Konten zugeordnet sind, könnten Sie eine Vergleichsregel erstellen, die einen Vergleich anhand des Namens durchführt, nicht aber anhand der Kontonummer. Sie würden die Option **Übereinstimmung, wenn nicht wahr** für das untergeordnete Element verwenden, das mit der Kontonummer übereinstimmt.

- d) Geben Sie im Feld **Fehlende Daten** an, wie leere Daten in einem Feld bewertet werden sollen. Zur Auswahl stehen:

| | |
|--------------------------------|---|
| Leerzeichen ignorieren | Ignoriert das Feld, wenn es leere Daten enthält. |
| Als 0 zählen | Bewertet das Feld mit 0, wenn es leere Daten enthält. |
| Als 100 zählen | Bewertet das Feld mit 100, wenn es leere Daten enthält. |
| Leerzeichen vergleichen | Bewertet die Kopfdubletten- und Kandidatenfelder mit 100, wenn beide leere Daten enthalten; bewertet andernfalls die Kopfdubletten- und Kandidatenfelder mit 0. |

- e) Geben Sie im Feld **Schwellenwert** den Schwellenwert an, der auf Einzelfeldebene erfüllt sein muss, damit dieses Feld als Übereinstimmung erkannt wird.

- f) Wählen Sie im Feld **Bewertungsmethode** die Methode aus, die verwendet wird, um die Punktzahl für die Übereinstimmung zu ermitteln. Zur Auswahl stehen:

| | |
|-------------------------------|--|
| Gewichteter Mittelwert | Verwendet die Gewichtung jedes Algorithmus, um die durchschnittliche Punktzahl für die Übereinstimmung zu bestimmen. |
| Durchschnitt | Verwendet die durchschnittliche Punktzahl jedes Algorithmus, um die Punktzahl für die Übereinstimmung zu bestimmen. |
| Maximum | Verwendet die höchste Algorithmuspunktzahl, um die Punktzahl für die Übereinstimmung zu bestimmen. |
| Minimum | Verwendet die niedrigste Algorithmuspunktzahl, um die Punktzahl für die Übereinstimmung zu bestimmen. |
| Vektoraddition | Verwendet die Vektoraddition der Punktzahl jedes Algorithmus, um die Punktzahl für die Übereinstimmung zu bestimmen. Diese Bewertungsmethode ist nützlich, wenn Sie möchten, dass eine höhere Punktzahl für die Übereinstimmung in einem oder mehr Algorithmen in der endgültigen Punktzahl für die Übereinstimmung proportional dargestellt wird. Die zur Berechnung der endgültigen Punktzahl verwendete Formel ist: |

$\sqrt{a^2 + b^2 + c^2} / \sqrt{n}$. Dabei gilt: a, b und c sind die Punktzahlen von drei verschiedenen Algorithmen, und n ist die Anzahl der verwendeten Algorithmen.

- g) Wählen Sie einen oder mehr Algorithmen, die verwendet werden sollen, um zu ermitteln, ob die Werte im Feld übereinstimmen. Die folgenden Optionen stehen zur Auswahl:

| | |
|----------------|---|
| Akronym | Bestimmt, ob ein Unternehmensname mit dem zugehörigen Akronym übereinstimmt, indem nach Akronymdaten gesucht wird; andernfalls wird anhand des ersten Zeichens jedes Wortes ein Akronym erstellt. Beispiel: Internal Revenue Service und das zugehörige Akronym IRS würden als Übereinstimmung gewertet und eine Punktzahl für die Übereinstimmung von 100 zurückgeben. |
|----------------|---|

| | |
|--------------------------------|--|
| Zeichenhäufigkeit | Bestimmt die Häufigkeit, mit der jedes Zeichen in einer Zeichenfolge vorkommt, und vergleicht die Gesamthäufigkeit zwischen zwei Zeichenfolgen. |
| Daitch-Mokotoff Soundex | Phonetischer Algorithmus, der eine höhere Genauigkeit beim Vergleich germanischer, slawischer und jiddischer Nachnamen mit ähnlicher Aussprache, aber unterschiedlicher Schreibweise ermöglicht. Codierte Namen sind sechs Stellen lang, und es können mehrere mögliche Codierungen für einen einzelnen Namen zurückgegeben werden. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex bei der Verarbeitung von germanischen oder slawischen Nachnamen zu reagieren. |
| Date | <p>Vergleicht Datumsfelder unabhängig vom Datumsformat in den Eingabedatensätzen. Klicken Sie in der Spalte „Optionen“ auf „Bearbeiten“, um Folgendes anzugeben:</p> <ul style="list-style-type: none"> • Monat erforderlich: Verhindert, dass ein Datum, das nur aus einem Jahr besteht, übereinstimmt. • Tag erforderlich: Verhindert, dass ein Datum, das nur aus einem Monat und Jahr besteht, übereinstimmt. • Mit transponiertem MM/TT vergleichen: Wenn Monat und Tag im numerischen Format bereitgestellt werden, werden Kopfdublettenmonat mit Kandidatentag und Kopfdublettentag mit Kandidatenmonat verglichen sowie der Standardvergleich von Kopfdublettenmonat mit Kandidatenmonat und Kopfdublettentag mit Kandidatentag durchgeführt. • Format TT/MM/JJJJ vor MM/TT/JJJJ bevorzugen: Trägt zum Parsen von Datumswerten bei, bei denen sowohl Monat als auch Tag im numerischen Format bereitgestellt werden und die anhand des Kontexts nicht identifiziert werden können. Beispiel: Bei den Zahlen 5 und 13 ordnet der Parser automatisch 5 dem Monat und 13 dem Tag zu, weil es nur 12 Monate im Jahr gibt. Aber bei den Zahlen 5 und 12 (bzw. zwei beliebigen Zahlen kleiner oder gleich 12) nimmt der Parser an, dass die erste Zahl der Monat ist. Durch Aktivieren dieser Option wird sichergestellt, dass der Parser die erste Zahl als Tag und nicht als Monat liest. • Bereichsoptionen – Insgesamt: Ermöglicht Ihnen, die maximale Anzahl von Tagen zwischen übereinstimmenden Daten festzulegen. Beispiel: Wenn Sie einen Gesamtbereich von 35 Tagen eingeben und Ihr Kandidatendatum am 31. Dezember 2000 ist, wäre das Kopfdublettendatum 5. Februar 2001 eine Übereinstimmung, das Kopfdublettendatum 6. Februar hingegen nicht. Wenn Sie einen Gesamtbereich von 1 Tag eingeben und Ihr |

Kandidatendatum Januar 2000 ist, wäre das Kopfdublettendatum 1999 eine Übereinstimmung (Vergleich mit 31. Dezember 1999), das Kopfdublettendatum 2001 hingegen nicht.

- **Bereichsoptionen – Jahr:** Ermöglicht Ihnen, die Anzahl von Jahren zwischen übereinstimmenden Daten, unabhängig von Monat und Tag, festzulegen. Beispiel: Wenn Sie einen Jahresbereich von 3 eingeben und Ihr Kandidatendatum 31. Januar 2000 ist, wäre das Kopfdublettendatum 31. Januar 2003 eine Übereinstimmung, das Kopfdublettendatum Februar 2003 hingegen nicht. Ähnlich gilt: Wenn Ihr Kandidatendatum 2000 ist, wäre das Kopfdublettendatum März 2003 eine Übereinstimmung, weil die Monate keinen Konflikt verursachen und es innerhalb des Dreijahresbereichs liegt.
- **Bereichsoptionen – Monat:** Ermöglicht Ihnen, die Anzahl von Monaten zwischen übereinstimmenden Daten, unabhängig von Jahr und Tag, festzulegen. Beispiel: Wenn Sie einen Monatsbereich von 4 eingeben und Ihr Kandidatendatum 1. Januar 2000 ist, wäre das Kopfdublettendatum Mai 2000 eine Übereinstimmung, weil die Tage keinen Konflikt verursachen und es innerhalb des Viermonatsbereichs liegt, das Kopfdublettendatum 2. Mai 2000 hingegen nicht, weil die Tage einen Konflikt verursachen.
- **Bereichsoptionen – Tag:** Ermöglicht Ihnen, die Anzahl von Tagen zwischen übereinstimmenden Daten, unabhängig von Jahr und Monat, festzulegen. Beispiel: Wenn Sie einen Tagesbereich von 5 eingeben und Ihr Kandidatendatum 1. Januar 2000 ist, wäre das Kopfdublettendatum Januar 2000 eine Übereinstimmung, weil die Tage keinen Konflikt verursachen, das Kopfdublettendatum 27. Dezember 1999 hingegen nicht, weil die Monate einen Konflikt verursachen.

Double Metaphone

Bestimmt die Ähnlichkeit zwischen zwei Zeichenfolgen auf Grundlage einer phonetischen Darstellung ihrer Zeichen. „Double Metaphone“ ist eine verbesserte Version des Algorithmus „Metaphone“ und versucht, viele der in unterschiedlichen Sprachen vorkommenden Unregelmäßigkeiten zu berücksichtigen.

Entfernung bearbeiten

Bestimmt die Ähnlichkeit zwischen zwei Zeichenfolgen anhand der Anzahl der Löschungen, Einfügungen oder Ersetzungen, die zum Transformieren einer Zeichenfolge in eine andere erforderlich sind.

| | |
|---------------------------------|---|
| Euklidischer Abstand | Bietet ein Maß der Ähnlichkeit zwischen zwei Zeichenfolgen unter Verwendung des Vektorraums kombinierter Begriffe als Dimensionen. Er bestimmt auch den größten gemeinsamen Teiler von zwei Ganzzahlen. Er akzeptiert ein Paar positiver Ganzzahlen und bildet ein neues Paar, das aus der kleineren Zahl und der Differenz zwischen der größeren und der kleineren Zahl besteht. Der Prozess wird wiederholt, bis die Zahlen gleich sind. Diese Zahl ist dann der größte allgemeine Teiler des ursprünglichen Paares. Beispiel: 21 ist der größte gemeinsame Teiler von 252 und 105: ($252 = 12 \times 21$; $105 = 5 \times 21$); da $252 - 105 = (12 - 5) \times 21 = 147$, ist der ggT von 147 und 105 ebenfalls 21. |
| Genauere Übereinstimmung | Bestimmt, ob zwei Zeichenfolgen identisch sind. |
| Initialen | Dient dazu, die Ähnlichkeit geparster Personennamen zu vergleichen. |
| Jaro–Winkler-Abstand | Bestimmt die Ähnlichkeit zwischen zwei Zeichenfolgen anhand der Anzahl der Zeichenersetzungen, die zum Transformieren einer Zeichenfolge in eine andere erforderlich sind. Diese Option wurde für kurze Zeichenfolgen wie Personennamen entwickelt. |
| Abstand auf Tastatur | Bestimmt die Ähnlichkeit zwischen zwei Zeichenfolgen basierend auf der Anzahl der Löschungen, Einfügungen oder Ersetzungen, die erforderlich sind, um eine Zeichenfolge in eine andere zu transformieren. Dabei erfolgt eine Gewichtung entsprechend der Position der Tasten auf der Tastatur. Klicken Sie in der Spalte „Optionen“ auf „Bearbeiten“, um den Typ der verwendeten Tastatur anzugeben: QWERTY (Vereinigte Staaten). QWERTZ (Österreich und Deutschland) oder AZERTY (Frankreich). |
| Köln | Indiziert Namen nach Klang, so wie sie auf Deutsch ausgesprochen werden. Ermöglicht die Codierung von Namen mit derselben Aussprache in derselben Darstellung, sodass sie trotz kleiner Unterschiede bei der Schreibweise verglichen werden können. Das Ergebnis ist immer eine Folge von Zahlen; Sonderzeichen und Leerzeichen werden ignoriert. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren. |
| Kullback-Liebler-Abstand | Bestimmt die Ähnlichkeit zwischen zwei Zeichenfolgen basierend auf den Unterschieden in der Verteilung der Wörter in den beiden Zeichenfolgen. |

| | |
|-----------------------------|---|
| Metaphone | Bestimmt die Ähnlichkeit zwischen zwei englischen Zeichenfolgen auf Grundlage einer phonetischen Darstellung ihrer Zeichen. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren. |
| Metaphone (Spanisch) | Bestimmt die Ähnlichkeit zwischen zwei Zeichenfolgen auf Grundlage einer phonetischen Darstellung ihrer Zeichen. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren. |
| Metaphone 3 | Übertrifft die Algorithmen „Metaphone“ und „Double Metaphone“ mit genaueren Einstellungen für Konsonanten und inneren Vokalen, was Ihnen ermöglicht, Wörter oder Namen auf phonetischer Basis mehr oder weniger eng mit Suchbegriffen abgeglichen zu erzeugen. Metaphone 3 erhöht die Genauigkeit der phonetischen Codierung auf 98 %. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren. |
| Namensvariante | Bestimmt, ob zwei Namen Varianten voneinander sind. Der Algorithmus gibt eine Punktzahl für die Übereinstimmung von 100 zurück, wenn zwei Namen Varianten voneinander sind, und eine Punktzahl für die Übereinstimmung von 0, wenn zwei Namen keine Varianten voneinander sind. So ist beispielsweise JOHN eine Variante von JAKE und liefert eine Punktzahl für die Übereinstimmung von 100. JOHN ist keine Variante von HENRY und liefert eine Punktzahl für die Übereinstimmung von 0. Klicken Sie in der Spalte „Optionen“ auf Bearbeiten , um Optionen für Namensvarianten auszuwählen. Weitere Informationen finden Sie unter Name Variant Finder auf Seite 345. |
| N-Gramm-Abstand | Berechnet die Wahrscheinlichkeit des nächsten Begriffs in Text oder Sprache basierend auf n vorherigen Begriffen. Diese können Phoneme, Silben, Buchstaben, Wörter oder Basispaare umfassen und aus einer beliebigen Buchstabenkombination bestehen. Dieser Algorithmus beinhaltet eine Option, um die Größe des N-Gramms einzugeben; der Standardwert ist 2. |
| N-Gramm-Ähnlichkeit | Bestimmt die Ähnlichkeit zwischen zwei Zeichenfolgen basierend auf der Länge der längsten gemeinsamen Teilfolge von Phonemen, Silben, Buchstaben, Wörtern oder Basispaaren. Der Algorithmus beinhaltet die folgenden Optionen: <ul style="list-style-type: none"> • N-Gramm-Größe: Geben Sie die Größe des N-Gramms ein. Der Standardwert ist 2. |

- **Sonderzeichen entfernen:** Aktivieren Sie das Kästchen, um die Interpunktion durch Leerzeichen zu ersetzen.
- **Leerzeichen entfernen:** Aktivieren Sie das Kästchen, um Wörter zusammenzuführen.

Numerische Zeichenfolge

Vergleicht Adresszeilen durch Trennen der numerischen Attribute einer Adresszeile von den Buchstaben. Beispiel: In der Zeichenfolgenadresse 1234 Main Street Apt 567 werden die numerischen Attribute der Zeichenfolge (1234567) geparkt und anders als der restliche Zeichenfolgenwert (Main Street Apt) behandelt. Der Algorithmus vergleicht zuerst die numerischen Daten in der Zeichenfolge mit dem numerischen Algorithmus. Falls die Übereinstimmung für numerische Daten 100 ist, werden die alphabetischen Daten anhand von „Entfernung bearbeiten“ und „Zeichenhäufigkeit“ verglichen. Die endgültige Punktzahl für die Übereinstimmung wird wie folgt berechnet:

$$\frac{(\text{numericScore} + (\text{EditDistanceScore} + \text{CharacterFrequencyScore}) / 2) / 2}$$

So ist beispielsweise die Punktzahl für die Übereinstimmung dieser zwei Adressen 95,5, die wie folgt berechnet wird:

123 Main St Apt 567
123 Maon St Apt 567

Numerische Punktzahl = 100
Entfernung bearbeiten = 91
Zeichenhäufigkeit = 91

$$\begin{aligned} 91 + 91 &= 182 \\ 182/2 &= 91 \\ 100 + 91 &= 191 \\ 191/2 &= 95,5 \end{aligned}$$

NYSIIS

Phonetischer Codealgorithmus, der eine ungefähr übereinstimmende Aussprache mit der exakten Schreibweise vergleicht und Wörter indiziert, deren Aussprache ähnlich ist. Bestandteil des New York State Identification and Intelligence System. Angenommen, Sie suchen die Daten einer Person in einer Personendatenbank. Sie meinen, dass der Name der Person wie „John Smith“ klingt, er wird aber „Jon Smath“ geschrieben. Würden Sie eine Suche durchführen, in der nach einer genauen Übereinstimmung mit „John Smith“ gesucht wird, würden keine Ergebnisse zurückgegeben. Wenn Sie jedoch die Datenbank mit dem Algorithmus NYSIIS indizieren und eine erneute Suche mit dem Algorithmus NYSIIS

durchführen, wird die richtige Übereinstimmung zurückgegeben, weil vom Algorithmus sowohl „John Smith“ als auch „Jon Smath“ als „JANSNATH“ indiziert werden. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren; sie verarbeitet manche N-Gramme mit mehreren Zeichen und behält die relativen Vokalpositionen bei, was bei Soundex nicht der Fall ist.

Anmerkung: Dieser Algorithmus verarbeitet keine nicht alphanumerischen Zeichen; Datensätze, die solche enthalten, schlagen bei der Verarbeitung fehl.

| | |
|-----------------------|--|
| Phonix | Vorverarbeitet Namenszeichenfolgen, indem mehr als 100 Transformationsregeln auf einzelne Zeichen oder auf Zeichenfolgen angewendet werden. 19 dieser Regeln werden nur angewendet, wenn das bzw. die Zeichen am Anfang der Zeichenfolge stehen, während 12 der Regeln nur angewendet werden, wenn sie in der Mitte der Zeichenfolge stehen. 28 der Regeln werden nur angewendet, wenn sie am Ende der Zeichenfolge stehen. Die transformierte Namenszeichenfolge wird als Code codiert, der aus einem Anfangsbuchstaben gefolgt von drei Stellen besteht (Nullen und doppelt vorhandene Zahlen werden entfernt). Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren; sie ist komplexer und deshalb langsamer als Soundex. |
| Soundex | Bestimmt die Ähnlichkeit zwischen zwei Zeichenfolgen auf Grundlage einer phonetischen Darstellung ihrer Zeichen. |
| SubString | Bestimmt, ob eine Zeichenfolge in einer anderen Zeichenfolge enthalten ist. |
| Silbenabgleich | Kombiniert phonetische Informationen mit auf „Entfernung bearbeiten“ basierenden Berechnungen. Wandelt die zu vergleichenden Zeichenfolgen in Sequenzen von Silben um und berechnet die Anzahl der Bearbeitungen, die erforderlich sind, um eine Silbensequenz in eine andere umzuwandeln. |

Die folgende Tabelle beschreibt die logische Beziehung zwischen der Anzahl von Algorithmen, die Sie basierend auf der ausgewählten übergeordneten Bewertungsmethode verwenden können.

Tabelle 2: Matrix Algorithmus zu Bewertungsmethode

| Bewertungsmethode | Algorithmen | |
|------------------------|-----------------|---------|
| | Einzel | Mehrere |
| Gewichteter Mittelwert | nicht verfügbar | Ja |
| Durchschnitt | nicht verfügbar | Ja |
| Maximum | Ja | Ja |
| Minimum | nicht verfügbar | Ja |
| Vektoraddition | nicht verfügbar | Ja |

6. Wenn Sie eine Regel in „Interflow Match“, „Intraflow Match“ oder „Transactional Match“ definieren und die Regel mit anderen Schritten und/oder Benutzern gemeinsam nutzen möchten, klicken Sie oben im Fenster auf die Schaltfläche **Speichern**.

Negative Vergleichsbedingungen

Vergleichsbedingungen sind Anweisungen, die angeben, welche Felder übereinstimmen müssen, damit zwei Datensätze als Übereinstimmung betrachtet werden. In manchen Fällen möchten Sie aber vielleicht eine Bedingung definieren, die besagt, dass zwei Felder *nicht* übereinstimmen dürfen, damit sie als Übereinstimmung betrachtet werden. Diese Methode, als *Negation* bezeichnet, kehrt die Logik einer Bedingung innerhalb einer Vergleichsregel um.

Angenommen, Sie haben Kundensupport-Datensätze für ein Callcenter und möchten Kunden identifizieren, die das Callcenter kontaktiert haben, dies jedoch für mehrere Konten. Anders gesagt, sollen Personen identifiziert werden, die mehreren Konten zugeordnet sind. Um Kunden zu identifizieren, die mehrere Konten haben, könnten Sie Datensätze vergleichen, bei denen der Name übereinstimmt, die Kontonummer hingegen *nicht*. In diesem Fall würden Sie eine Negation in einer Vergleichsbedingung für die Kontonummer verwenden.

Um eine Negation zu verwenden, aktivieren Sie das Kästchen **Übereinstimmung, wenn nicht wahr** beim Definieren der Vergleichsregel. Diese Option ist sowohl für übergeordnete Elemente (Gruppen von Bedingungen) als auch für untergeordnete Elemente (einzelne Bedingungen) in der Vergleichsregel verfügbar. Die Auswirkung dieser Option ist geringfügig unterschiedlich, wenn sie für ein übergeordnetes statt für ein untergeordnetes Element verwendet wird. Wenn sie für ein übergeordnetes Element verwendet wird, kehrt die Option **Übereinstimmung, wenn nicht wahr** die Vergleichsmethodeoption praktisch wie folgt um:

- Die Vergleichsmethode **Alle Übereinstimmungen** wird praktisch zu „beliebige falsch“. Die Vergleichsregel kann nur Datensätze vergleichen, wenn mindestens eines der untergeordneten Elemente unter dem übergeordneten Element als falsch ausgewertet wird, sodass auch das

übergeordnete Element als falsch ausgewertet wird. Da die Option **Übereinstimmung, wenn nicht wahr** aktiviert ist, resultiert diese Auswertung als falsch in einer Übereinstimmung.

- Die Vergleichsmethode **Beliebige Übereinstimmung** wird praktisch zu „keine Übereinstimmung“. Die Vergleichsregel kann nur Datensätze zuordnen, bei denen keine der untergeordneten Elemente als wahr ausgewertet werden. Andernfalls ist nämlich das übergeordnete Element wahr. Mit aktivierter Option **Übereinstimmung, wenn nicht wahr** ergibt diese Auswertung als wahr keine Übereinstimmung. Nur wenn keine der untergeordneten Elemente wahr sind, wodurch auch das übergeordnete Element als nicht wahr ausgewertet wird, kann die Regel eine Übereinstimmung finden.
- Die Vergleichsmethode **Basierend auf Schwellenwert** arbeitet praktisch anders, indem statt Datensätzen, die gleich oder größer als ein angegebener Schwellenwert sind, Datensätze zugeordnet werden, die kleiner als der Schwellenwert sind. Da liegt daran, dass Datensätze mit einem Schwellenwert, der kleiner ist als der angegebene, als falsch ausgewertet werden. Da **Übereinstimmung, wenn nicht wahr** aktiviert ist, ergibt das eine Übereinstimmung.

Die Option **Übereinstimmung wenn nicht wahr** ist einfacher zu verstehen, wenn sie auf untergeordnete Elemente in einer Vergleichsregel angewendet wird. Sie gibt einfach an, dass zwei Datensätze als Übereinstimmung gewertet werden, wenn der Algorithmus *keine* Übereinstimmung ergibt.

Testen einer Vergleichsregel

Nachdem Sie eine Vergleichsregel definiert haben, sollten Sie sie testen, um ihre Ergebnisse anzuzeigen. Hierfür können Sie die Vergleichsregelauswertung verwenden, um die Auswirkungen einer Vergleichsregel an einem kleinen Satz an Beispieldaten zu prüfen.

1. Öffnen Sie den Datenfluss im Enterprise Designer.
2. Klicken Sie doppelt auf den Schritt mit der zu testenden Vergleichsregel.
Vergleichsregeln werden in Interflow Match, Intraflow Match und Transactional Match verwendet.
3. Wählen Sie in der Vergleichsregelhierarchie den zu testenden Knoten aus, und klicken Sie auf **Auswerten**.
4. Geben Sie die Testdaten auf der Registerkarte **Importieren** ein (eine Kopfdublette und bis zu 10 Kandidaten). Testdaten können über zwei Methoden eingegeben werden.
 - Um die Testdaten manuell einzugeben, müssen Sie unter **Kopfdublette** einen Kopfdubletten-Datensatz und unter **Kandidat** bis zu zehn Kandidatendatensätze eingeben. Nachdem Sie die Datensätze eingegeben haben, können Sie auf **Exportieren** klicken, um die Datensätze in einer Datei zu speichern, die Sie später importieren können, statt die Daten erneut manuell eingeben zu müssen.
 - Um Testdaten aus einer Datei zu importieren, müssen Sie auf **Importieren...** klicken und die Datei mit den Beispieldatensätzen auswählen. Dateien können durch Kommata, senkrechte Striche oder Tabstopps getrennt werden und sollten über einen Header-Datensatz verfügen,

der mit den unter **Kandidaten** angezeigten Feldnamen übereinstimmt. Ein Beispiel für einen Header-Datensatz für die Eingabe „Household“ würde folgendermaßen aussehen:

```
Name, AddressLine1, City, StateProvince
```

5. Prüfen Sie die Regel anhand einer der folgenden Methoden:

- Klicken Sie auf **Aktuelle Regel**. Hierdurch wird die Regel ausgeführt, die auf der Registerkarte **Vergleichsregel** definiert wurde. Ergebnisse werden für ein Kopfdubletten-/Kandidatenpaar gleichzeitig angezeigt. Klicken Sie auf die Pfeilschaltflächen, um die Ergebnisse durchzublättern. Punktzahlen für Felder und Algorithmen werden in einem Baumformat ähnlich der Vergleichsregelsteuerung angezeigt. Optional können die Ergebnisse in eine XML-Datei exportiert werden.

Anmerkung: Wenn Sie Änderungen an der Vergleichsregel vornehmen und die Regeln auf die Vergleichsregel des Schrittes anwenden möchten, klicken Sie auf **Speichern**.

- Klicken Sie auf **Alle Algorithmen**. Dadurch wird die Vergleichsregel ignoriert. Stattdessen werden alle Algorithmen auf jedem Feld für Kopfdubletten-/Kandidatenpaare ausgeführt. Die Ergebnisse werden für ein Kopfdubletten-/Kandidatenpaar gleichzeitig angezeigt und können mithilfe der Pfeilschaltflächen durchgeblättert werden.

Um die Ergebnisse automatisch zu aktualisieren, während Sie die Änderungen an der Vergleichsregel und/oder der Eingabe vornehmen, müssen Sie das Kästchen **Automatisch aktualisieren** aktivieren. Wenn dieses Feature zusammen mit der Option **Alle Algorithmen** verwendet wird, werden die Ergebnisse lediglich durch Änderungen an der Eingabe aktualisiert.

Die unter Punktzahlen angezeigten Ergebnisse weisen die folgende Farbcodierung auf:

- Grün: Die Regel hat eine Übereinstimmung ergeben.
- Rot: Die Regel hat keine Übereinstimmung ergeben.
- Grau: Die Regel wurde ignoriert.
- Blau: Die Ergebnisse für einzelne Algorithmen innerhalb der Regel.

Klicken Sie auf **Exportieren**, um die Ergebnisse der Auswertung im XML-Format zu exportieren.

Gemeinsames Nutzen von Vergleichsregeln

Sie können Vergleichsregeln erstellen, die von Modulen, Schritten, Datenflüssen und Benutzern gemeinsam genutzt werden können. Indem Sie eine Vergleichsregel gemeinsam nutzen, können Sie die Entwicklung von Datenflüssen vereinfachen, indem Sie eine Vergleichsregel einmal definieren und sie dann nach Bedarf referenzieren. Dies trägt auch dazu bei, sicherzustellen, dass die Vergleichsregeln, die die gleichen Funktionen ausführen sollen, datenflussübergreifend konsistent sind.

Die in der Datenbank für Vergleichsregeln gespeicherten Vergleichsregeln können mit der Administrationsumgebung als MR- oder JSON-Dateien exportiert werden. Dies ist besonders nützlich, wenn Sie versuchen, diese Regeln mit anderen Spectrum-Installationen gemeinsam zu nutzen oder sie im Spectrum Big Data Quality SDK zu verwenden.

Anmerkung: Standardmäßig werden die Vergleichsregeln als MR-Datei exportiert.

- Um eine in „Interflow Match“, „Intraflow Match“ oder „Transactional Match“ erstellte Vergleichsregel gemeinsam zu nutzen, klicken Sie oben im Optionsfenster des Schrittes auf die Schaltfläche **Speichern**.
- Wenn Sie die Regel im Tool „Vergleichsregelverwaltung“ erstellen, ist die Regel automatisch zur Nutzung in Datenflüssen durch alle Benutzer verfügbar. Wählen Sie zum Anzeigen des Tools „Vergleichsregelverwaltung“ im Enterprise Designer **Tools > Vergleichsregelverwaltung** aus.

Anzeigen von gemeinsam genutzten Vergleichsregeln

Im Enterprise Designer können Sie alle gemeinsam genutzten, für Sie auf Ihrem Spectrum™ Technology Platform-System verfügbaren Vergleichsregeln durchsuchen. Diese Vergleichsregeln können durch die „Interflow Match“- , „Intraflow Match“- und „Transactional Match“-Schritte in einem Datenfluss verwendet werden, um Vergleiche durchzuführen.

Führen Sie die nachfolgenden Schritte aus, um die Vergleichsregeln in der Datenbank für Vergleichsregeln zu durchsuchen.

1. Öffnen Sie den Enterprise Designer.
2. Wählen Sie **Tools > Vergleichsregelverwaltung** aus.
3. Wählen Sie die Regel aus, die angezeigt werden soll, und klicken Sie auf **Anzeigen**.

Erstellen einer benutzerdefinierten Vergleichsregel als JSON-Objekt

Vergleichsregeln können konfiguriert und zur Laufzeit übergeben werden, wenn sie als Datenflussoptionen verfügbar gemacht werden. Das ermöglicht Ihnen, Vergleichsregeln für mehrere Systeme gemeinsam zu nutzen und vorhandene Vergleichsregeln durch Vergleichsregel-Zeichenfolgen im JSON-Format zu überschreiben. Sie können auch Schrittoptionen festlegen, wenn Sie den Auftrag über einen Prozessfluss oder über das Befehlszeilentool „Job Executor“ aufrufen.

Sie finden Schemas für „MatchRule“ und „MatchInfo“ im folgenden Ordner:

```
<Spectrum Location>\server\modules\jsonSchemas\matcher.
```

1. Speichern Sie den Datenfluss, der die Vergleichsregel enthält, und machen Sie ihn verfügbar.
2. Öffnen Sie den Datenfluss, der die Vergleichsregel nutzt.
3. Öffnen Sie **Bearbeiten > Datenflussoptionen**.

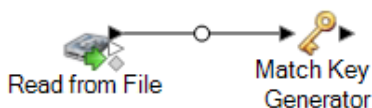
4. Klicken Sie in der Tabelle **Datenflussoptionen zu Schritten zuordnen** auf den Vergleichsschritt, der die Vergleichsregel nutzt, und aktivieren Sie das Kästchen **Benutzerdefinierte Vergleichsregel**.
5. Optional: Ändern Sie den Namen der Vergleichsregel im Feld **Optionsbeschriftung** von „Benutzerdefinierte Vergleichsregel“ in einen Namen Ihrer Wahl.
6. Klicken Sie zweimal auf **OK**.

Vergleichen von Datensätzen aus einer einzelnen Quelle

Diese Schritte beschreiben, wie Sie einen „Intraflow Match“-Schritt verwenden können, um Datensatzgruppen innerhalb einer einzelnen Datenquelle (wie eine Datei oder Datenbanktabelle) identifizieren können, die bezüglich Ihrer angegebenen Vergleichskriterien übereinstimmen. Der Datenfluss gruppiert Datensätze zu Sammlungen und schreibt die Sammlungen in eine Ausgabedatei.

1. Erstellen Sie im Enterprise Designer einen neuen Datenfluss.
2. Ziehen Sie einen Quellschritt auf die Arbeitsfläche.
3. Doppelklicken Sie auf den Quellschritt und konfigurieren Sie ihn. Anweisungen zum Konfigurieren von Quellschritten finden Sie im *Datenfluss-Designer-Handbuch*.
4. Ziehen Sie einen „Match Key Generator“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem Quellschritt.

Wenn Sie beispielsweise einen „Read from File“-Quellschritt verwenden, sieht Ihr Datenfluss jetzt folgendermaßen aus:



Match Key Generator erstellt einen nicht eindeutigen Schlüssel für jeden Datensatz, der dann von Vergleichsschritten verwendet werden kann, um Gruppen potenzieller Datensatzdubletten zu identifizieren. Vergleichsschlüssel erleichtern den Vergleichsprozess, da sie Ihnen erlauben, Datensätze nach Vergleichsschlüssel zu gruppieren und dann nur Datensätze innerhalb dieser Gruppen zu vergleichen.

5. Doppelklicken Sie auf den „Match Key Generator“-Schritt.
6. Klicken Sie auf **Hinzufügen**.
7. Definieren Sie die Regel, die zur Generierung eines Vergleichsschlüssels für jeden Datensatz verwendet werden soll.

Tabelle 3: Match Key Generator-Optionen

| Name der Option | Beschreibung/gültige Werte |
|-----------------|----------------------------|
|-----------------|----------------------------|

| | |
|-------------|--|
| Algorithmus | |
|-------------|--|

Name der Option

Beschreibung/gültige Werte

Gibt den Algorithmus an, der verwendet werden soll, um den Vergleichsschlüssel zu generieren. Zur Auswahl stehen:

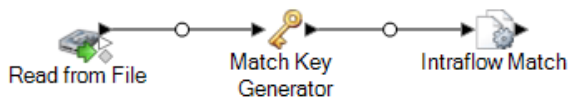
| | |
|-----------------------------|--|
| Consonant | Gibt die angegebenen Felder mit entfernten Konsonanten zurück. |
| Double Metaphone | Gibt einen auf einer phonetischen Darstellung ihrer Zeichen basierenden Code zurück. „Double Metaphone“ ist eine verbesserte Version des Algorithmus „Metaphone“ und versucht, viele der in unterschiedlichen Sprachen vorkommenden Unregelmäßigkeiten zu berücksichtigen. |
| Köln | Indiziert Namen nach Klang, so wie sie auf Deutsch ausgesprochen werden. Ermöglicht die Codierung von Namen mit derselben Aussprache in derselben Darstellung, sodass sie trotz kleiner Unterschiede bei der Schreibweise verglichen werden können. Das Ergebnis ist immer eine Folge von Zahlen; Sonderzeichen und Leerzeichen werden ignoriert. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren. |
| MD5 | Ein Message Digest-Algorithmus, der einen 128-Bit-Hashwert erzeugt. Dieser Algorithmus wird im Allgemeinen verwendet, um die Datenintegrität zu überprüfen. |
| Metaphone | Gibt einen per Metaphone codierten Schlüssel ausgewählter Felder zurück. Metaphone ist ein Algorithmus, um Wörter anhand ihrer englischen Aussprache zu codieren. |
| Metaphone (Spanisch) | Gibt einen per Metaphone codierten Schlüssel ausgewählter Felder für die spanische Sprache zurück. Dieser Metaphone-Algorithmus codiert Wörter anhand ihrer spanischen Aussprache. |
| Metaphone 3 | Übertrifft die Algorithmen „Metaphone“ und „Double Metaphone“ mit genaueren Einstellungen für Konsonanten und inneren Vokalen, was Ihnen ermöglicht, Wörter oder Namen auf phonetischer Basis mehr oder weniger eng mit Suchbegriffen abgeglichen zu erzeugen. Metaphone 3 erhöht die Genauigkeit der phonetischen Codierung auf 98 %. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren. |
| NYSIIS | Phonetischer Codealgorithmus, der eine ungefähr übereinstimmende Aussprache mit der exakten Schreibweise vergleicht und Wörter indiziert, deren Aussprache ähnlich ist. Bestandteil des New York State Identification and Intelligence System. Angenommen, Sie suchen die Daten einer Person in einer Personendatenbank. Sie meinen, dass der Name der Person wie „John Smith“ klingt, er wird aber „Jon Smyth“ geschrieben. Würden Sie eine Suche durchführen, in der nach einer genauen Übereinstimmung mit „John Smith“ gesucht wird, würden keine Ergebnisse zurückgegeben. Wenn Sie jedoch die Datenbank mit dem Algorithmus NYSIIS |

| Name der Option | Beschreibung/gültige Werte |
|-------------------------|--|
| | indizieren und eine erneute Suche mit dem Algorithmus NYSIIS durchführen, wird die richtige Übereinstimmung zurückgegeben, weil vom Algorithmus sowohl „John Smith“ als auch „Jon Smyth“ als „JAN SNATH“ indiziert werden. |
| Phonix | Vorverarbeitet Namenszeichenfolgen, indem mehr als 100 Transformationsregeln auf einzelne Zeichen oder auf Zeichenfolgen angewendet werden. 19 dieser Regeln werden nur angewendet, wenn das bzw. die Zeichen am Anfang der Zeichenfolge stehen, während 12 der Regeln nur angewendet werden, wenn sie in der Mitte der Zeichenfolge stehen. 28 der Regeln werden nur angewendet, wenn sie am Ende der Zeichenfolge stehen. Die transformierte Namenszeichenfolge wird als Code codiert, der aus einem Anfangsbuchstaben gefolgt von drei Stellen besteht (Nullen und doppelt vorhandene Zahlen werden entfernt). Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren; sie ist komplexer und deshalb langsamer als Soundex. |
| Soundex | Gibt einen Soundex-Code ausgewählter Felder zurück. Soundex erzeugt einen auf der englischen Aussprache eines Wortes basierenden Code mit fester Länge. |
| Teilzeichenfolge | Gibt einen angegebenen Teil des ausgewählten Feldes zurück. |
| Feldname | Gibt das Feld an, auf das Sie den ausgewählten Algorithmus anwenden möchten, um den Vergleichsschlüssel zu generieren. Wenn Sie beispielsweise ein Feld mit Namen „LastName“ auswählen und den Soundex-Algorithmus wählen, würde der Soundex-Algorithmus auf die Daten im Feld „LastName“ angewendet, um einen Vergleichsschlüssel zu erzeugen. |
| Startposition | Gibt die Startposition innerhalb des angegebenen Feldes an. Nicht alle Algorithmen erlauben Ihnen, eine Startposition anzugeben. |
| Länge | Gibt die Länge der Zeichen an, die ab der Startposition eingeschlossen werden sollen. Nicht alle Algorithmen erlauben Ihnen, eine Länge anzugeben. |
| Sonderzeichen entfernen | Entfernt alle nicht numerischen und nicht alphabetischen Zeichen wie Bindestriche, Leerzeichen und andere Sonderzeichen aus einem Eingabefeld. |

| Name der Option | Beschreibung/gültige Werte |
|-------------------|---|
| Eingabe sortieren | Sortiert alle Zeichen in einem Eingabefeld oder alle Begriffe in einem Eingabefeld in alphabetischer Reihenfolge. <ul style="list-style-type: none"> Zeichen Sortiert die Zeichenwerte aus einem Eingabefeld vor dem Erstellen einer eindeutigen ID. Begriffe Sortiert jeden Begriffswert aus einem Eingabefeld vor dem Erstellen einer eindeutigen ID. |

8. Klicken Sie auf **OK**, wenn Sie mit dem Definieren der Regel fertig sind.
9. Wenn Sie weitere Vergleichsregeln hinzufügen möchten, klicken Sie auf **Hinzufügen** und fügen Sie sie hinzu. Klicken Sie andernfalls auf **OK**, wenn Sie fertig sind.
10. Ziehen Sie einen „Intraflow Match“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem „Match Key Generator“-Schritt.

Wenn Sie beispielsweise einen „Read from File“-Quellschritt verwenden, sieht Ihr Datenfluss jetzt folgendermaßen aus:



11. Doppelklicken Sie auf den „Intraflow Match“-Schritt.
12. Wählen Sie im Feld **Vergleichsregel laden** eine der vordefinierten Vergleichsregeln aus, die Sie entweder unverändert verwenden oder nach Bedarf ändern können. Wenn Sie eine neue Vergleichsregel erstellen möchten, ohne eine der vordefinierten Vergleichsregeln als Startpunkt zu verwenden, klicken Sie auf **Neu**. Im Datenfluss darf lediglich eine benutzerdefinierte Regel enthalten sein.

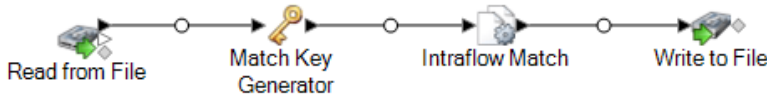
Anmerkung: Das Feature „Datenflussoptionen“ im Enterprise Designer ermöglicht es, die Vergleichsregel für die Konfiguration zur Laufzeit verfügbar zu machen.

13. Wählen Sie im Feld **Gruppieren nach** den **Vergleichsschlüssel** aus.

Datensätze mit demselben Vergleichsschlüssel werden dann zusammen gruppiert. Die Vergleichsregel wird auf Datensätze innerhalb einer Gruppe angewendet, um herauszufinden, ob Dubletten vorhanden sind. Der Vergleichsschlüssel für jeden Datensatz wird vom „Generate Match Key“-Schritt generiert, den Sie weiter oben konfiguriert haben.
14. Weitere Informationen zum Ändern der anderen Optionen finden Sie unter [Erstellen einer Vergleichsregel](#) auf Seite 72.
15. Klicken Sie auf **OK**, um Ihre „Intraflow Match“-Konfiguration zu speichern und zur Datenfluss-Arbeitsfläche zurückzukehren.

16. Ziehen Sie einen Datenladeschritt auf die Arbeitsfläche und verbinden Sie ihn mit dem „Generate Match Key“-Schritt.

Bei Verwendung eines „Write to File“-Datenladeschrittes würde Ihr Datenfluss so aussehen:



17. Doppelklicken Sie auf den Datenladeschritt und konfigurieren Sie ihn.

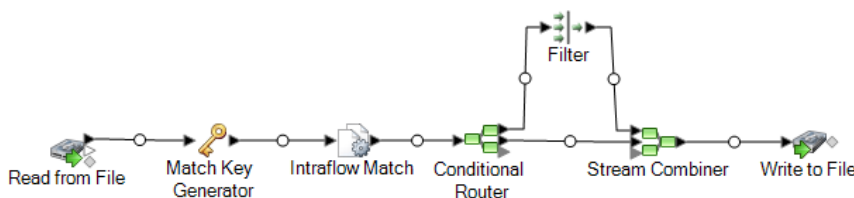
Informationen zum Konfigurieren von Datenladeschritten finden Sie im *Datenfluss-Designer-Handbuch*.

Sie verfügen jetzt über einen Datenfluss, der Datensätze aus einer einzelnen Quelle vergleicht.

Beispiel zum Vergleichen von Datensätzen aus einer einzelnen Quelle

Als Data Steward in einer Kreditkartengesellschaft möchten Sie Ihre Kundendatenbank analysieren und herausfinden, welche Adressen mehrmals vorhanden sind und welche Namen dort verzeichnet sind, damit Sie die Anzahl der doppelten Kreditkartenangebote, die an denselben Haushalt gesendet werden, minimieren können.

In diesem Beispiel wird gezeigt, wie Sie Mitglieder desselben Haushalts identifizieren, indem Sie Informationen innerhalb einer einzelnen Eingabedatei vergleichen und eine Ausgabedatei erstellen, die einen Datensatz pro Haushalt enthält.



Der „Read from File“-Schritt liest Daten ein, die sowohl eindeutige Datensätze für Haushalte als auch Datensätze enthält, die sich möglicherweise auf denselben Haushalt beziehen. Die Eingabedatei enthält Namen und Adressen.

Der „Match Key Generator“ erstellt einen Vergleichsschlüssel, der bei ähnlichen Datensätzen einen nicht eindeutigen Schlüssel darstellt, um mögliche Dubletten zu identifizieren.

Der „Intraflow Match“-Schritt vergleicht Datensätze, die denselben Vergleichsschlüssel aufweisen, und markiert jeden Datensatz entweder als eindeutigen Datensatz oder als einen von mehreren Datensätzen für denselben Haushalt.

Der „Conditional Router“ sendet Datensätze, die Sammlungen von Datensätzen für jeden Haushalt sind, zum Filter-Schritt, der für jeden Haushalt alle Datensätze bis auf einen herausfiltert und diesen weiter zum „Stream Combiner“-Schritt leitet. Der

„Conditional Router“-Schritt sendet außerdem eindeutige Datensätze direkt an den Stream Combiner.

Am Ende erstellt der „Write to File“-Schritt eine Ausgabedatei, die einen Datensatz für jeden Haushalt enthält.

Vergleichen von Datensätzen einer Quelle mit Datensätzen einer anderen Quelle

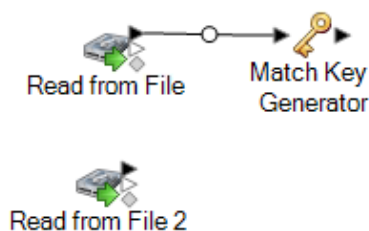
In dieser Prozedur wird beschrieben, wie ein „Interflow Match“-Schritt verwendet wird, um Datensätze in einer Quelle zu identifizieren, die Datensätzen in einer anderen Quelle entsprechen. Die erste Quelle enthält verdächtige Datensätze und die zweite Quelle enthält Kandidatendatensätze. Der Datenfluss vergleicht nur Datensätze einer Quelle mit Datensätzen einer anderen Quelle. Es versucht nicht, Datensätze aus derselben Quelle miteinander zu vergleichen. Der Datenfluss gruppiert Datensätze in Sammlungen mit übereinstimmenden Datensätzen und schreibt diese Sammlungen in eine Ausgabedatei.

1. Erstellen Sie im Enterprise Designer einen neuen Datenfluss.
2. Ziehen Sie zwei Quellschritte auf die Arbeitsfläche. Konfigurieren Sie eine von ihnen so, dass auf die Quelle der verdächtigen Datensätze verwiesen wird, und konfigurieren Sie die andere so, dass auf die Quelle der Kandidatendatensätze verwiesen wird.

Anweisungen zum Konfigurieren von Quellschritten finden Sie im *Datenfluss-Designer-Handbuch*.

3. Ziehen Sie einen „Match Key Generator“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit einem der Quellschritte.

Wenn Sie beispielsweise einen „Read from File“-Quellschritt verwenden, sieht Ihr Datenfluss jetzt folgendermaßen aus:



Match Key Generator erstellt einen nicht eindeutigen Schlüssel für jeden Datensatz, der dann von Vergleichsschritten verwendet werden kann, um Gruppen potenzieller Datensatzdubletten zu identifizieren. Vergleichsschlüssel erleichtern den Vergleichsprozess, da sie Ihnen erlauben, Datensätze nach Vergleichsschlüssel zu gruppieren und dann nur Datensätze innerhalb dieser Gruppen zu vergleichen.

Anmerkung: Sie fügen später einen zweiten „Match Key Generator“-Schritt hinzu. Fürs Erste benötigen Sie nur einen Schritt auf der Arbeitsfläche.

4. Doppelklicken Sie auf den „Match Key Generator“-Schritt.
5. Klicken Sie auf **Hinzufügen**.
6. Definieren Sie die Regel, die zur Generierung eines Vergleichsschlüssels für jeden Datensatz verwendet werden soll.

Tabelle 4: Match Key Generator-Optionen

| Name der Option | Beschreibung/gültige Werte |
|-----------------|----------------------------|
|-----------------|----------------------------|

| | |
|-------------|--|
| Algorithmus | |
|-------------|--|

Name der Option

Beschreibung/gültige Werte

Gibt den Algorithmus an, der verwendet werden soll, um den Vergleichsschlüssel zu generieren. Zur Auswahl stehen:

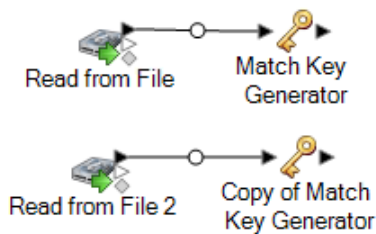
| | |
|-----------------------------|--|
| Consonant | Gibt die angegebenen Felder mit entfernten Konsonanten zurück. |
| Double Metaphone | Gibt einen auf einer phonetischen Darstellung ihrer Zeichen basierenden Code zurück. „Double Metaphone“ ist eine verbesserte Version des Algorithmus „Metaphone“ und versucht, viele der in unterschiedlichen Sprachen vorkommenden Unregelmäßigkeiten zu berücksichtigen. |
| Köln | Indiziert Namen nach Klang, so wie sie auf Deutsch ausgesprochen werden. Ermöglicht die Codierung von Namen mit derselben Aussprache in derselben Darstellung, sodass sie trotz kleiner Unterschiede bei der Schreibweise verglichen werden können. Das Ergebnis ist immer eine Folge von Zahlen; Sonderzeichen und Leerzeichen werden ignoriert. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren. |
| MD5 | Ein Message Digest-Algorithmus, der einen 128-Bit-Hashwert erzeugt. Dieser Algorithmus wird im Allgemeinen verwendet, um die Datenintegrität zu überprüfen. |
| Metaphone | Gibt einen per Metaphone codierten Schlüssel ausgewählter Felder zurück. Metaphone ist ein Algorithmus, um Wörter anhand ihrer englischen Aussprache zu codieren. |
| Metaphone (Spanisch) | Gibt einen per Metaphone codierten Schlüssel ausgewählter Felder für die spanische Sprache zurück. Dieser Metaphone-Algorithmus codiert Wörter anhand ihrer spanischen Aussprache. |
| Metaphone 3 | Übertrifft die Algorithmen „Metaphone“ und „Double Metaphone“ mit genaueren Einstellungen für Konsonanten und inneren Vokalen, was Ihnen ermöglicht, Wörter oder Namen auf phonetischer Basis mehr oder weniger eng mit Suchbegriffen abgeglichen zu erzeugen. Metaphone 3 erhöht die Genauigkeit der phonetischen Codierung auf 98 %. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren. |
| NYSIIS | Phonetischer Codealgorithmus, der eine ungefähr übereinstimmende Aussprache mit der exakten Schreibweise vergleicht und Wörter indiziert, deren Aussprache ähnlich ist. Bestandteil des New York State Identification and Intelligence System. Angenommen, Sie suchen die Daten einer Person in einer Personendatenbank. Sie meinen, dass der Name der Person wie „John Smith“ klingt, er wird aber „Jon Smyth“ geschrieben. Würden Sie eine Suche durchführen, in der nach einer genauen Übereinstimmung mit „John Smith“ gesucht wird, würden keine Ergebnisse zurückgegeben. Wenn Sie jedoch die Datenbank mit dem Algorithmus NYSIIS |

| Name der Option | Beschreibung/gültige Werte |
|-------------------------|--|
| | indizieren und eine erneute Suche mit dem Algorithmus NYSIIS durchführen, wird die richtige Übereinstimmung zurückgegeben, weil vom Algorithmus sowohl „John Smith“ als auch „Jon Smyth“ als „JAN SNATH“ indiziert werden. |
| Phonix | Vorverarbeitet Namenszeichenfolgen, indem mehr als 100 Transformationsregeln auf einzelne Zeichen oder auf Zeichenfolgen angewendet werden. 19 dieser Regeln werden nur angewendet, wenn das bzw. die Zeichen am Anfang der Zeichenfolge stehen, während 12 der Regeln nur angewendet werden, wenn sie in der Mitte der Zeichenfolge stehen. 28 der Regeln werden nur angewendet, wenn sie am Ende der Zeichenfolge stehen. Die transformierte Namenszeichenfolge wird als Code codiert, der aus einem Anfangsbuchstaben gefolgt von drei Stellen besteht (Nullen und doppelt vorhandene Zahlen werden entfernt). Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren; sie ist komplexer und deshalb langsamer als Soundex. |
| Soundex | Gibt einen Soundex-Code ausgewählter Felder zurück. Soundex erzeugt einen auf der englischen Aussprache eines Wortes basierenden Code mit fester Länge. |
| Teilzeichenfolge | Gibt einen angegebenen Teil des ausgewählten Feldes zurück. |
| Feldname | Gibt das Feld an, auf das Sie den ausgewählten Algorithmus anwenden möchten, um den Vergleichsschlüssel zu generieren. Wenn Sie beispielsweise ein Feld mit Namen „LastName“ auswählen und den Soundex-Algorithmus wählen, würde der Soundex-Algorithmus auf die Daten im Feld „LastName“ angewendet, um einen Vergleichsschlüssel zu erzeugen. |
| Startposition | Gibt die Startposition innerhalb des angegebenen Feldes an. Nicht alle Algorithmen erlauben Ihnen, eine Startposition anzugeben. |
| Länge | Gibt die Länge der Zeichen an, die ab der Startposition eingeschlossen werden sollen. Nicht alle Algorithmen erlauben Ihnen, eine Länge anzugeben. |
| Sonderzeichen entfernen | Entfernt alle nicht numerischen und nicht alphabetischen Zeichen wie Bindestriche, Leerzeichen und andere Sonderzeichen aus einem Eingabefeld. |

| Name der Option | Beschreibung/gültige Werte |
|-------------------|--|
| Eingabe sortieren | Sortiert alle Zeichen in einem Eingabefeld oder alle Begriffe in einem Eingabefeld in alphabetischer Reihenfolge. Zeichen Sortiert die Zeichenwerte aus einem Eingabefeld vor dem Erstellen einer eindeutigen ID. Begriffe Sortiert jeden Begriffswert aus einem Eingabefeld vor dem Erstellen einer eindeutigen ID. |

- Klicken Sie auf **OK**, wenn Sie mit dem Definieren der Regel fertig sind.
- Klicken Sie mit der rechten Maustaste auf den „Match Key Generator“-Schritt auf der Arbeitsfläche, und wählen Sie **Schritt kopieren** aus.
- Klicken Sie mit der rechten Maustaste in einen leeren Bereich der Arbeitsfläche, und wählen Sie **Einfügen** aus.
- Verbinden Sie die Kopie von „Match Key Generator“ mit dem anderen Quellschritt.

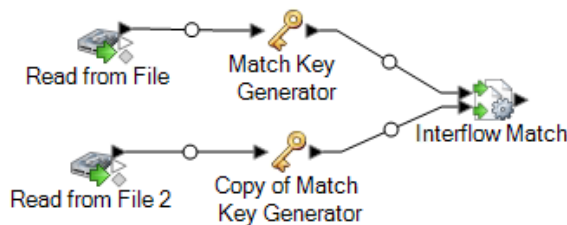
Bei Verwendung eines „Read from File“-Eingabeschrittes etwa würde Ihr Datenfluss jetzt folgendermaßen aussehen:



Der Datenfluss enthält nun zwei „Match Key Generator“-Schritte, die für jede Quelle mit exakt den gleichen Regeln Vergleichsschlüssel erzeugen. Die identische Konfiguration der „Match Key Generator“-Schritte ist wesentlich dafür, dass dieser Datenfluss ordnungsgemäß funktioniert.

- Ziehen Sie einen „Interflow Match“-Schritt auf die Arbeitsfläche und verbinden Sie die einzelnen „Match Key Generator“-Schritte mit ihm.

Bei Verwendung eines „Read from File“-Eingabeschrittes etwa würde Ihr Datenfluss jetzt folgendermaßen aussehen:



12. Doppelklicken Sie auf den „Interflow Match“-Schritt.
13. Wählen Sie im Feld **Vergleichsregel laden** eine der vordefinierten Vergleichsregeln aus, die Sie entweder unverändert verwenden oder nach Bedarf ändern können. Wenn Sie eine neue Vergleichsregel erstellen möchten, ohne eine der vordefinierten Vergleichsregeln als Startpunkt zu verwenden, klicken Sie auf **Neu**. Im Datenfluss darf lediglich eine benutzerdefinierte Regel enthalten sein.

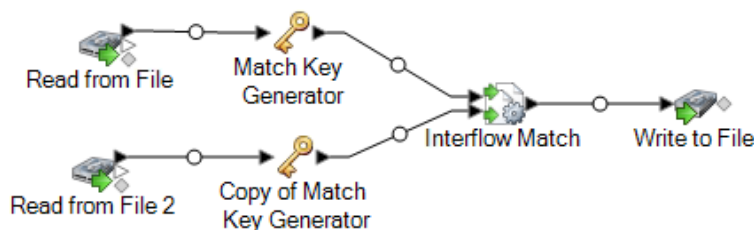
Anmerkung: Das Feature „Datenflussoptionen“ im Enterprise Designer ermöglicht es, die Vergleichsregel für die Konfiguration zur Laufzeit verfügbar zu machen.

14. Wählen Sie im Feld **Gruppieren nach** den Eintrag **MatchKey** aus.

Datensätze mit demselben Vergleichsschlüssel werden dann zusammen gruppiert. Die Vergleichsregel wird auf Datensätze innerhalb einer Gruppe angewendet, um herauszufinden, ob Dubletten vorhanden sind. Der Vergleichsschlüssel für jeden Datensatz wird von den „Generate Match Key“-Schritten generiert, die Sie weiter oben konfiguriert haben.

15. Weitere Informationen zum Ändern der anderen Optionen finden Sie unter [Erstellen einer Vergleichsregel](#) auf Seite 72.
16. Ziehen Sie einen Datenladeschritt auf die Arbeitsfläche und verbinden Sie ihn mit dem „Interflow Match“-Schritt.

Bei Verwendung eines „Write to File“-Datenladeschrittes würde Ihr Datenfluss so aussehen:



17. Doppelklicken Sie auf den Datenladeschritt und konfigurieren Sie ihn.

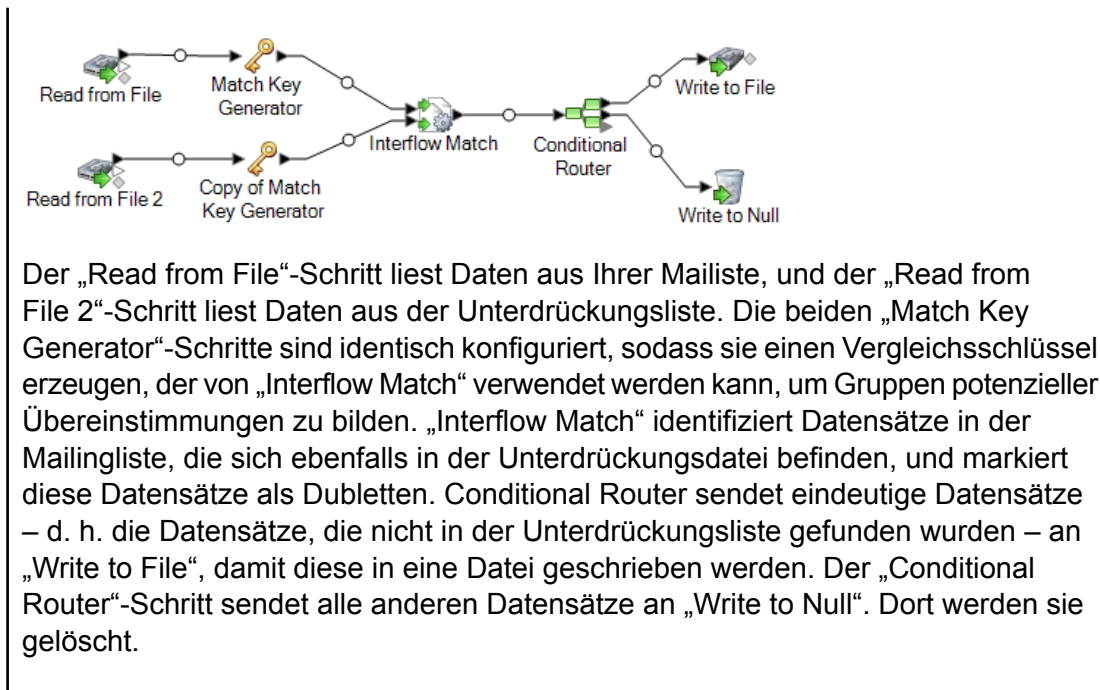
Informationen zum Konfigurieren von Datenladeschritten finden Sie im *Datenfluss-Designer-Handbuch*.

Sie haben jetzt einen Datenfluss, der Datensätze aus zwei Datenquellen miteinander vergleicht.

Beispiel für den Vergleich von Datensätzen aus mehreren Quellen

Als Werbesendungsunternehmen möchten Sie Personen identifizieren, die sich in einer „Do-Not-Mail“-Liste befinden, um keine Werbesendung an sie zu senden. Sie verfügen über eine Liste von Empfängern in einer Datei und eine Liste von Personen, die keine Werbesendung in einer anderen Datei (eine Unterdrückungsdatei) erhalten möchten.

Im folgenden Datenfluss wird eine Lösung für dieses Geschäftsszenario bereitgestellt:

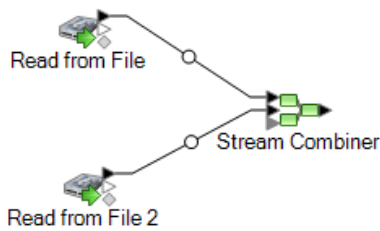


Vergleichen von Datensätzen zwischen und innerhalb von Quellen

Diese Schritte beschreiben, wie Sie einen „Intraflow Match“-Schritt verwenden können, um Datensätze in einer Datei zu identifizieren, die mit Datensätzen in einer anderen oder in derselben Datei übereinstimmen. Wenn Sie beispielsweise zwei Dateien haben (Datei A und Datei B), können Sie feststellen, ob in Datei A Datensätze vorhanden sind, die mit Datensätzen in Datei B übereinstimmen. Darüber hinaus können Sie herausfinden, ob in Datei A Datensätze vorhanden sind, die mit anderen Datensätzen in Datei A übereinstimmen. Dies erreichen Sie durch die Verwendung eines „Stream Combiner“- und einer „Intraflow Match“-Schrittes.

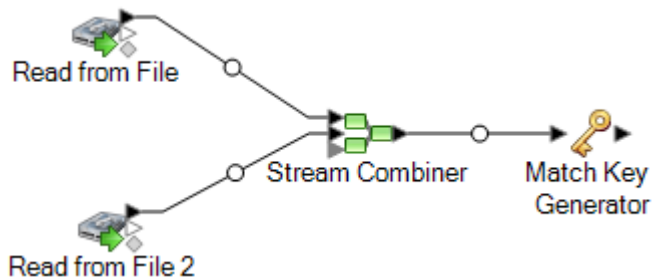
1. Erstellen Sie im Enterprise Designer einen neuen Datenfluss.
2. Ziehen Sie einen Quellschritt auf die Arbeitsfläche.
3. Doppelklicken Sie auf den Quellschritt und konfigurieren Sie ihn. Anweisungen zum Konfigurieren von Quellschritten finden Sie im *Datenfluss-Designer-Handbuch*.
4. Ziehen Sie einen zweiten Quellschritt auf die Arbeitsfläche und konfigurieren Sie ihn, die zweite Datenquelle in den Datenfluss einzulesen.
5. Ziehen Sie einen „Stream Combiner“-Schritt auf die Arbeitsfläche und verbinden Sie die beiden Quellschritte mit diesem.

Wenn Ihr Datenfluss beispielsweise aus zwei „Read from File“-Schritten bestehen würde, sähe der Datenfluss nach dem Hinzufügen des „Stream Combiner“-Schrittes folgendermaßen aus:



- Ziehen Sie einen „Match Key Generator“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem „Stream Combiner“-Schritt.

Ihr Datenfluss könnte jetzt beispielsweise aussehen wie folgt:



Match Key Generator erstellt einen nicht eindeutigen Schlüssel für jeden Datensatz, der dann von Vergleichsschritten verwendet werden kann, um Gruppen potenzieller Datensatzdubletten zu identifizieren. Vergleichsschlüssel erleichtern den Vergleichsprozess, da sie Ihnen erlauben, Datensätze nach Vergleichsschlüssel zu gruppieren und dann nur Datensätze innerhalb dieser Gruppen zu vergleichen.

- Doppelklicken Sie auf den „Match Key Generator“-Schritt.
- Klicken Sie auf **Hinzufügen**.
- Definieren Sie die Regel, die zur Generierung eines Vergleichsschlüssels für jeden Datensatz verwendet werden soll.

Tabelle 5: Match Key Generator-Optionen

| Name der Option | Beschreibung/gültige Werte |
|-----------------|----------------------------|
|-----------------|----------------------------|

| | |
|-------------|--|
| Algorithmus | |
|-------------|--|

Name der Option

Beschreibung/gültige Werte

Gibt den Algorithmus an, der verwendet werden soll, um den Vergleichsschlüssel zu generieren. Zur Auswahl stehen:

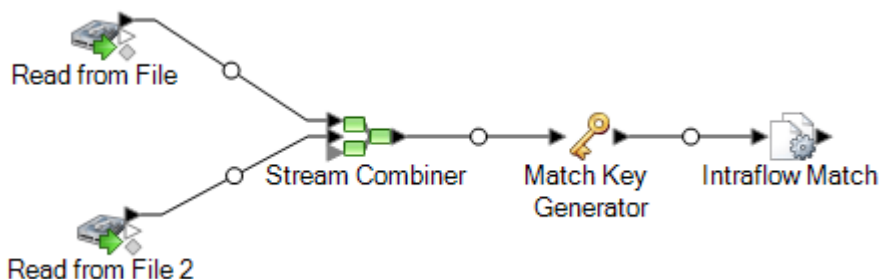
| | |
|-----------------------------|--|
| Consonant | Gibt die angegebenen Felder mit entfernten Konsonanten zurück. |
| Double Metaphone | Gibt einen auf einer phonetischen Darstellung ihrer Zeichen basierenden Code zurück. „Double Metaphone“ ist eine verbesserte Version des Algorithmus „Metaphone“ und versucht, viele der in unterschiedlichen Sprachen vorkommenden Unregelmäßigkeiten zu berücksichtigen. |
| Köln | Indiziert Namen nach Klang, so wie sie auf Deutsch ausgesprochen werden. Ermöglicht die Codierung von Namen mit derselben Aussprache in derselben Darstellung, sodass sie trotz kleiner Unterschiede bei der Schreibweise verglichen werden können. Das Ergebnis ist immer eine Folge von Zahlen; Sonderzeichen und Leerzeichen werden ignoriert. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren. |
| MD5 | Ein Message Digest-Algorithmus, der einen 128-Bit-Hashwert erzeugt. Dieser Algorithmus wird im Allgemeinen verwendet, um die Datenintegrität zu überprüfen. |
| Metaphone | Gibt einen per Metaphone codierten Schlüssel ausgewählter Felder zurück. Metaphone ist ein Algorithmus, um Wörter anhand ihrer englischen Aussprache zu codieren. |
| Metaphone (Spanisch) | Gibt einen per Metaphone codierten Schlüssel ausgewählter Felder für die spanische Sprache zurück. Dieser Metaphone-Algorithmus codiert Wörter anhand ihrer spanischen Aussprache. |
| Metaphone 3 | Übertrifft die Algorithmen „Metaphone“ und „Double Metaphone“ mit genaueren Einstellungen für Konsonanten und inneren Vokalen, was Ihnen ermöglicht, Wörter oder Namen auf phonetischer Basis mehr oder weniger eng mit Suchbegriffen abgeglichen zu erzeugen. Metaphone 3 erhöht die Genauigkeit der phonetischen Codierung auf 98 %. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren. |
| NYSIIS | Phonetischer Codealgorithmus, der eine ungefähr übereinstimmende Aussprache mit der exakten Schreibweise vergleicht und Wörter indiziert, deren Aussprache ähnlich ist. Bestandteil des New York State Identification and Intelligence System. Angenommen, Sie suchen die Daten einer Person in einer Personendatenbank. Sie meinen, dass der Name der Person wie „John Smith“ klingt, er wird aber „Jon Smyth“ geschrieben. Würden Sie eine Suche durchführen, in der nach einer genauen Übereinstimmung mit „John Smith“ gesucht wird, würden keine Ergebnisse zurückgegeben. Wenn Sie jedoch die Datenbank mit dem Algorithmus NYSIIS |

| Name der Option | Beschreibung/gültige Werte |
|-------------------------|---|
| Phonix | <p>indizieren und eine erneute Suche mit dem Algorithmus NYSIIS durchführen, wird die richtige Übereinstimmung zurückgegeben, weil vom Algorithmus sowohl „John Smith“ als auch „Jon Smyth“ als „JAN SNATH“ indiziert werden.</p> <p>Vorverarbeitet Namenszeichenfolgen, indem mehr als 100 Transformationsregeln auf einzelne Zeichen oder auf Zeichenfolgen angewendet werden. 19 dieser Regeln werden nur angewendet, wenn das bzw. die Zeichen am Anfang der Zeichenfolge stehen, während 12 der Regeln nur angewendet werden, wenn sie in der Mitte der Zeichenfolge stehen. 28 der Regeln werden nur angewendet, wenn sie am Ende der Zeichenfolge stehen. Die transformierte Namenszeichenfolge wird als Code codiert, der aus einem Anfangsbuchstaben gefolgt von drei Stellen besteht (Nullen und doppelt vorhandene Zahlen werden entfernt). Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren; sie ist komplexer und deshalb langsamer als Soundex.</p> |
| Soundex | <p>Gibt einen Soundex-Code ausgewählter Felder zurück. Soundex erzeugt einen auf der englischen Aussprache eines Wortes basierenden Code mit fester Länge.</p> |
| Teilzeichenfolge | <p>Gibt einen angegebenen Teil des ausgewählten Feldes zurück.</p> |
| Feldname | <p>Gibt das Feld an, auf das Sie den ausgewählten Algorithmus anwenden möchten, um den Vergleichsschlüssel zu generieren. Wenn Sie beispielsweise ein Feld mit Namen „LastName“ auswählen und den Soundex-Algorithmus wählen, würde der Soundex-Algorithmus auf die Daten im Feld „LastName“ angewendet, um einen Vergleichsschlüssel zu erzeugen.</p> |
| Startposition | <p>Gibt die Startposition innerhalb des angegebenen Feldes an. Nicht alle Algorithmen erlauben Ihnen, eine Startposition anzugeben.</p> |
| Länge | <p>Gibt die Länge der Zeichen an, die ab der Startposition eingeschlossen werden sollen. Nicht alle Algorithmen erlauben Ihnen, eine Länge anzugeben.</p> |
| Sonderzeichen entfernen | <p>Entfernt alle nicht numerischen und nicht alphabetischen Zeichen wie Bindestriche, Leerzeichen und andere Sonderzeichen aus einem Eingabefeld.</p> |

| Name der Option | Beschreibung/gültige Werte |
|-------------------|---|
| Eingabe sortieren | Sortiert alle Zeichen in einem Eingabefeld oder alle Begriffe in einem Eingabefeld in alphabetischer Reihenfolge. <ul style="list-style-type: none"> Zeichen Sortiert die Zeichenwerte aus einem Eingabefeld vor dem Erstellen einer eindeutigen ID. Begriffe Sortiert jeden Begriffswert aus einem Eingabefeld vor dem Erstellen einer eindeutigen ID. |

10. Klicken Sie auf **OK**, wenn Sie mit dem Definieren der Regel fertig sind.
11. Wenn Sie weitere Vergleichsregeln hinzufügen möchten, klicken Sie auf **Hinzufügen** und fügen Sie sie hinzu. Klicken Sie andernfalls auf **OK**, wenn Sie fertig sind.
12. Ziehen Sie einen „Intraflow Match“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem „Match Key Generator“-Schritt.

Ihr Datenfluss könnte jetzt beispielsweise aussehen wie folgt:



13. Doppelklicken Sie auf den „Intraflow Match“-Schritt.
14. Wählen Sie im Feld **Vergleichsregel laden** eine der vordefinierten Vergleichsregeln aus, die Sie entweder unverändert verwenden oder nach Bedarf ändern können. Wenn Sie eine neue Vergleichsregel erstellen möchten, ohne eine der vordefinierten Vergleichsregeln als Startpunkt zu verwenden, klicken Sie auf **Neu**. Im Datenfluss darf lediglich eine benutzerdefinierte Regel enthalten sein.

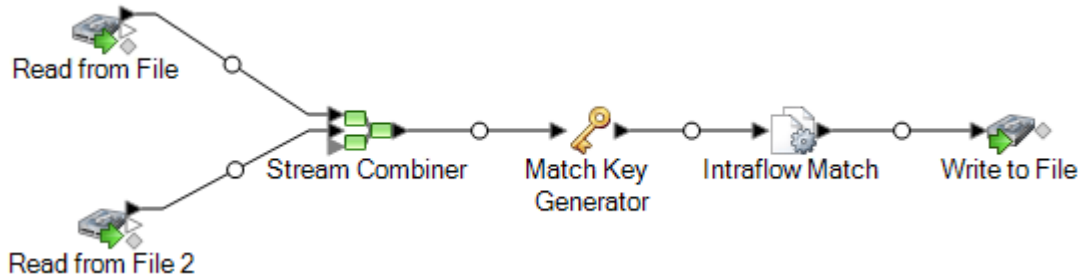
Anmerkung: Das Feature „Datenflussoptionen“ im Enterprise Designer ermöglicht es, die Vergleichsregel für die Konfiguration zur Laufzeit verfügbar zu machen.

15. Wählen Sie im Feld **Gruppieren nach** den **Vergleichsschlüssel** aus.

Datensätze mit demselben Vergleichsschlüssel werden dann zusammen gruppiert. Die Vergleichsregel wird auf Datensätze innerhalb einer Gruppe angewendet, um herauszufinden, ob Dubletten vorhanden sind. Der Vergleichsschlüssel für jeden Datensatz wird vom „Generate Match Key“-Schritt generiert, den Sie weiter oben konfiguriert haben.

16. Weitere Informationen zum Ändern der anderen Optionen finden Sie unter [Erstellen einer Vergleichsregel](#) auf Seite 72.
17. Klicken Sie auf **OK**, um Ihre „Intraflow Match“-Konfiguration zu speichern und zur Datenfluss-Arbeitsfläche zurückzukehren.
18. Ziehen Sie einen Datenladeschritt auf die Arbeitsfläche und verbinden Sie ihn mit dem „Generate Match Key“-Schritt.

Bei Verwendung eines „Write to File“-Datenladeschrittes würde Ihr Datenfluss so aussehen:



19. Doppelklicken Sie auf den Datenladeschritt und konfigurieren Sie ihn.

Informationen zum Konfigurieren von Datenladeschritten finden Sie im *Datenfluss-Designer-Handbuch*.

Vergleichen von Datensätzen mit einer Datenbank

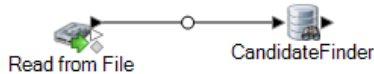
Diese Schritte beschreiben, wie Datensätze verglichen werden, in denen die Kopfdubletten-Datensätze aus einer Quelle wie einer Datei oder Datenbank stammen und sich die Kandidatendatensätze in einer Datenbank mit weiteren Datensätzen ohne Bezug dazu befinden. Für jeden Eingabedatensatz fragt der Datenfluss die Datenbank auf Kandidaten für diesen Datensatz hin ab und verwendet dann einen „Transactional Match“-Schritt, um Datensätze zu vergleichen. Schließlich schreibt der Datenfluss die Sammlungen übereinstimmender Datensätze in eine Ausgabedatei.

Anmerkung: Transactional Match gleicht lediglich Kopfdubletten-Datensätze mit Kandidaten ab. Es vergleicht keine Kopfdubletten-Datensätze mit anderen Kopfdubletten-Datensätzen, wie es bei Intraflow Match der Fall ist.

1. Erstellen Sie im Enterprise Designer einen neuen Datenfluss.
2. Ziehen Sie einen Quellschritt auf die Arbeitsfläche.
3. Doppelklicken Sie auf den Quellschritt und konfigurieren Sie ihn. Anweisungen zum Konfigurieren von Quellschritten finden Sie im *Datenfluss-Designer-Handbuch*.

- Ziehen Sie einen „Candidate Finder“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem Quellschritt.

Bei Verwendung des „Read from File“-Quellschrittes würde Ihr Datenfluss so aussehen:



„Candidate Finder“ erhält die Kandidatendatensätze, die die Gruppe potenzieller Übereinstimmungen bilden, die „Transactional Match“ später im Datenfluss auswertet.

- Doppelklicken Sie auf den „Candidate Finder“-Schritt auf der Arbeitsfläche.
- Wählen Sie im Feld **Verbindung** die Datenbank aus, die abgefragt werden soll, um Kandidatendatensätze zu finden. Wenn die gewünschte Datenbank nicht aufgelistet ist, öffnen Sie Management Console und definieren Sie dort zuerst die Datenbankverbindung.
- Geben Sie in das Feld „SQL“ eine SQL-SELECT-Anweisung ein, die basierend auf dem Wert in einem der Datenflussfelder Kandidatendatensätze findet. Verwenden Sie zum Referenzieren von Datenflussfeldern das Format `${FieldName}`. Dabei ist `FieldName` der Name des Feldes, das Sie referenzieren möchten.

Wenn Sie etwa Datensätze in der Datenbank finden möchten, bei denen der Wert in der Spalte „LastName“ identisch mit dem Feld „Customer_LastName“ der Datensätze im Datenfluss ist, würden Sie eine solche SQL-Anweisung schreiben:

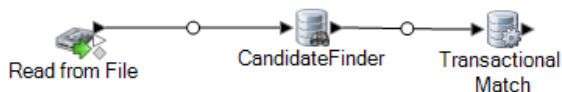
```
SELECT FirstName, LastName, Address, City, State, PostalCode
FROM Customer_Table
WHERE LastName = ${Customer_LastName};
```

- Wählen Sie auf der Registerkarte **Feldzuordnung** aus, welche Felder im Datenfluss die Daten aus jeder Datenbankspalte enthalten sollen.

In der Spalte **Ausgewählte Felder** werden die Datenbankspalten und in **Schrittfelder** die Felder im Datenfluss aufgelistet.

- Klicken Sie auf **OK**.
- Ziehen Sie einen „Transactional Match“-Schritt auf die Arbeitsfläche und verbinden Sie den „Candidate Finder“-Schritt damit.

Bei Verwendung eines „Read from File“-Eingabeschrittes etwa würde Ihr Datenfluss jetzt so aussehen:



Transactional Match vergleicht Kopfdubletten-Datensätze mit Kandidatendatensätzen, die vom „Candidate Finder“-Schritt zurückgegeben wurden. Transactional Match verwendet Vergleichsregeln, um den Kopfdubletten-Datensatz mit allen Kandidatendatensätzen mit

derselben (durch Candidate Finder zugewiesenen) Kandidatengruppennummer zu vergleichen, um Dubletten zu finden.

11. Doppelklicken Sie auf den „Transactional Match“-Schritt auf der Arbeitsfläche.
12. Wählen Sie im Feld **Vergleichsregel laden** eine der vordefinierten Vergleichsregeln aus, die Sie entweder unverändert verwenden oder nach Bedarf ändern können. Wenn Sie eine neue Vergleichsregel erstellen möchten, ohne eine der vordefinierten Vergleichsregeln als Startpunkt zu verwenden, klicken Sie auf **Neu**. Im Datenfluss darf lediglich eine benutzerdefinierte Regel enthalten sein.

Anmerkung: Das Feature „Datenflussoptionen“ im Enterprise Designer ermöglicht es, die Vergleichsregel für die Konfiguration zur Laufzeit verfügbar zu machen.

13. Weitere Informationen zum Ändern der anderen Optionen finden Sie unter [Erstellen einer Vergleichsregel](#) auf Seite 72.
14. Wenn Sie mit der Konfiguration des „Transactional Match“-Schrittes fertig sind, klicken Sie auf **OK**.
15. Ziehen Sie einen Datenladeschritt auf die Arbeitsfläche und verbinden Sie ihn mit dem „Transactional Match“-Schritt.

Bei Verwendung eines „Write to File“-Datenladeschrittes würde Ihr Datenfluss so aussehen:



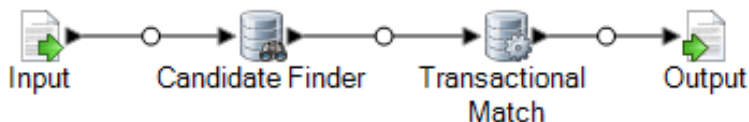
16. Doppelklicken Sie auf den Datenladeschritt und konfigurieren Sie ihn.
Informationen zum Konfigurieren von Datenladeschritten finden Sie im *Datenfluss-Designer-Handbuch*.

Sie haben jetzt einen Datenfluss, der Datensätze aus zwei Datenquellen vergleicht.

Beispiel für den Vergleich von Datensätzen mit einer Datenbank

Als Verkaufsleiter eines Online-Handelsunternehmens möchten Sie herausfinden, ob ein potenzieller Online-Kunde ein Bestandskunde oder ein Neukunde ist.

Der folgende Datenflusssdienst bietet eine Lösung für das Geschäftsszenario:



Dieser Datenfluss ist ein Dienst, der die an ihn gesendeten Interessentendaten über einen API- oder Webservice-Aufruf auswertet. Er wertet die Daten anhand der Kundendaten in einer Kundendatenbank aus, um zu ermitteln, ob ein Interessent schon Kunde ist.

Der Input-Schritt ist so konfiguriert, dass der Datenfluss die folgenden Eingabefelder akzeptiert: AddressLine1, City, Name, PostalCode und StateProvince. „AddressLine1“ und „Name“ sind die Felder, die für die Datenflussverarbeitung in dieser Vorlage entscheidend sind.

Im „Candidate Finder“-Schritt werden die Kandidatendatensätze abgerufen, die die Gruppe potenzieller Übereinstimmungen bilden, die im „Transactional Match“-Schritt ausgewertet wird.

Der „Transactional Match“-Schritt vergleicht Kopfdoubletten-Datensätze mit Kandidatendatensätzen, die vom „Candidate Finder“-Schritt zurückgegeben werden. Transactional Match verwendet Vergleichsregeln, um den Kopfdoubletten-Datensatz mit allen Kandidatendatensätzen mit derselben (durch Candidate Finder zugewiesenen) Kandidatengruppennummer zu vergleichen, um Doubletten zu finden. In diesem Beispiel vergleicht Transactional Match „LastName“ und „AddressLine1“.

Der Output-Schritt gibt die Ergebnisse des Datenflusses über eine API- oder Webservice-Antwort zurück.

Vergleichen von Datensätzen mithilfe mehrerer Vergleichsregeln

■ [Beispieldatenfluss herunterladen](#)

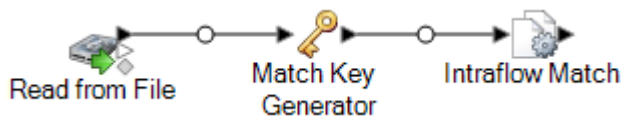
Wenn Sie über zu vergleichende Datensätze verfügen und Sie mehr als einen Vergleichsvorgang verwenden möchten, können Sie einen Datenfluss erstellen, der mehr als einen Vergleichsschlüssel verwendet und dann die Ergebnisse kombiniert, um mehrere separate Kriterien effektiv zu vergleichen. Beispiel: Sie möchten einen Datenfluss erstellen, der Datensätze anhand folgender Kriterien vergleicht:

Name und Adresse stimmen überein.
OR
Geburtsdatum und Ausweisnummer stimmen überein.

Um einen Vergleich anhand dieser Logik durchzuführen, erstellen Sie einen Datenfluss, der in einem Schritt Name und Adresse und in einem anderen Schritt Geburtsdatum und Ausweisnummer vergleicht und dann die entsprechenden Datensätze in einer einzelnen Sammlung kombiniert.

In diesem Thema wird die allgemeine Vorgehensweise beschrieben, wie ein Datenfluss eingerichtet wird, der zwei Vergleichsschritte verwendet. Diese Prozedur verwendet zur Veranschaulichung „Intraflow Match“-Schritte. Sie können diese Technik jedoch auch mit „Interflow Match“-Schritten anwenden.

1. Erstellen Sie im Enterprise Designer einen neuen Datenfluss.
2. Ziehen Sie einen Quellschritt auf die Arbeitsfläche.
3. Doppelklicken Sie auf den Quellschritt und konfigurieren Sie ihn. Anweisungen zum Konfigurieren von Quellschritten finden Sie im *Datenfluss-Designer-Handbuch*.
4. Definieren Sie den ersten Vergleichsdurchlauf. Die Ergebnisse dieses ersten Vergleichsdurchlaufs bestehen aus Datensätzen, die bezüglich Ihres ersten Satzes an Vergleichskriterien übereinstimmen, in diesem Fall im Namen und in der Adresse.
 - a) Ziehen Sie einen „Match Key Generator“- und einen „Intraflow Match“-Schritt auf die Arbeitsfläche und verbinden Sie sie gemäß dem folgenden Beispiel:



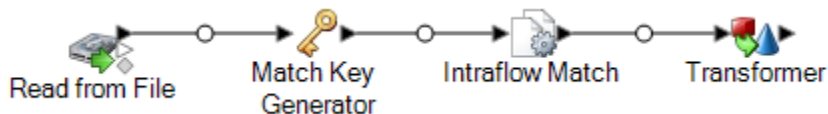
- a) Definieren Sie im „Match Key Generator“-Schritt den Vergleichsschlüssel, der für den ersten Vergleichsdurchlauf verwendet werden soll.

Wenn Sie beispielsweise im ersten Vergleichsdurchlauf Name und Adresse vergleichen möchten, können Sie einen Vergleichsschlüssel erstellen, der auf den Feldern mit dem Nachnamen und der Postleitzahl basiert.

- b) Definieren Sie im „Intraflow Match“-Schritt die Vergleichsregeln, die im ersten Vergleichsdurchlauf angewendet werden sollen.

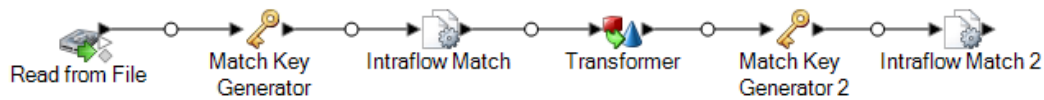
Sie können diese Vergleichsschritte zum Beispiel konfigurieren, Name und Adresse zu vergleichen.

5. Speichern Sie die Sammlungsnummern aus dem ersten Vergleichsdurchlauf in einem anderen Feld. Dies ist erforderlich, da das Feld „CollectionNumber“ im Verlauf des zweiten Vergleichsdurchlaufs überschrieben wird. Das Feld „CollectionNumber“ muss umbenannt werden, um die Ergebnisse des ersten Vergleichsdurchlaufs zu bewahren.
 - a) Ziehen Sie einen Transformer-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem „Intraflow Match“-Schritt, sodass sich folgender Datenfluss ergibt:



- b) Konfigurieren Sie den Transformer-Schritt, das Feld „CollectionNumber“ in „CollectionNumberPass1“ umzubenennen.

6. Definieren Sie den zweiten Vergleichsdurchlauf. Die Ergebnisse dieses zweiten Vergleichsdurchlaufs bestehen aus Datensätzen, die bezüglich Ihres zweiten Satzes an Vergleichskriterien übereinstimmen, in diesem Fall im Geburtsdatum und in der Ausweisnummer.
 - a) Ziehen Sie einen „Match Key Generator“- und einen „Intraflow Match“-Schritt auf die Arbeitsfläche und verbinden Sie sie gemäß dem folgenden Beispiel:



- b) Definieren Sie im zweiten „Match Key Generator“-Schritt den Vergleichsschlüssel, der für den zweiten Vergleichsdurchlauf verwendet werden soll.

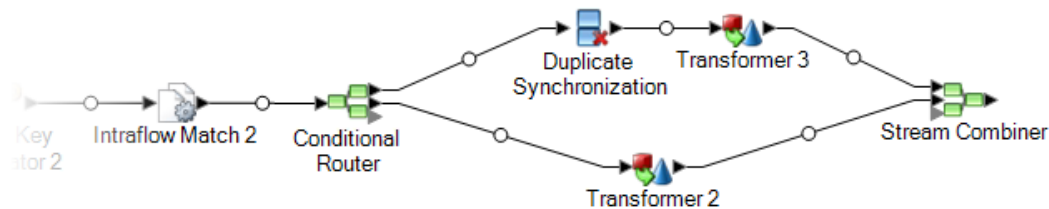
Wenn Sie beispielsweise im zweiten Vergleichsdurchlauf Geburtsdatum und Ausweisnummer vergleichen möchten, können Sie einen Vergleichsschlüssel erstellen, der auf den Feldern mit dem Geburtstag und der Ausweisnummer basiert.

- c) Definieren Sie im zweiten „Intraflow Match“-Schritt die Vergleichsregel, die im zweiten Vergleichsdurchlauf angewendet werden soll.

Sie können diesen Vergleichsschritt zum Beispiel konfigurieren, Geburtsdatum und Ausweisnummer zu vergleichen.

7. Prüfen Sie, ob im zweiten Vergleichsdurchlauf gefundene Datensatzduplikate auch im ersten Vergleichsdurchlauf gefunden wurden.

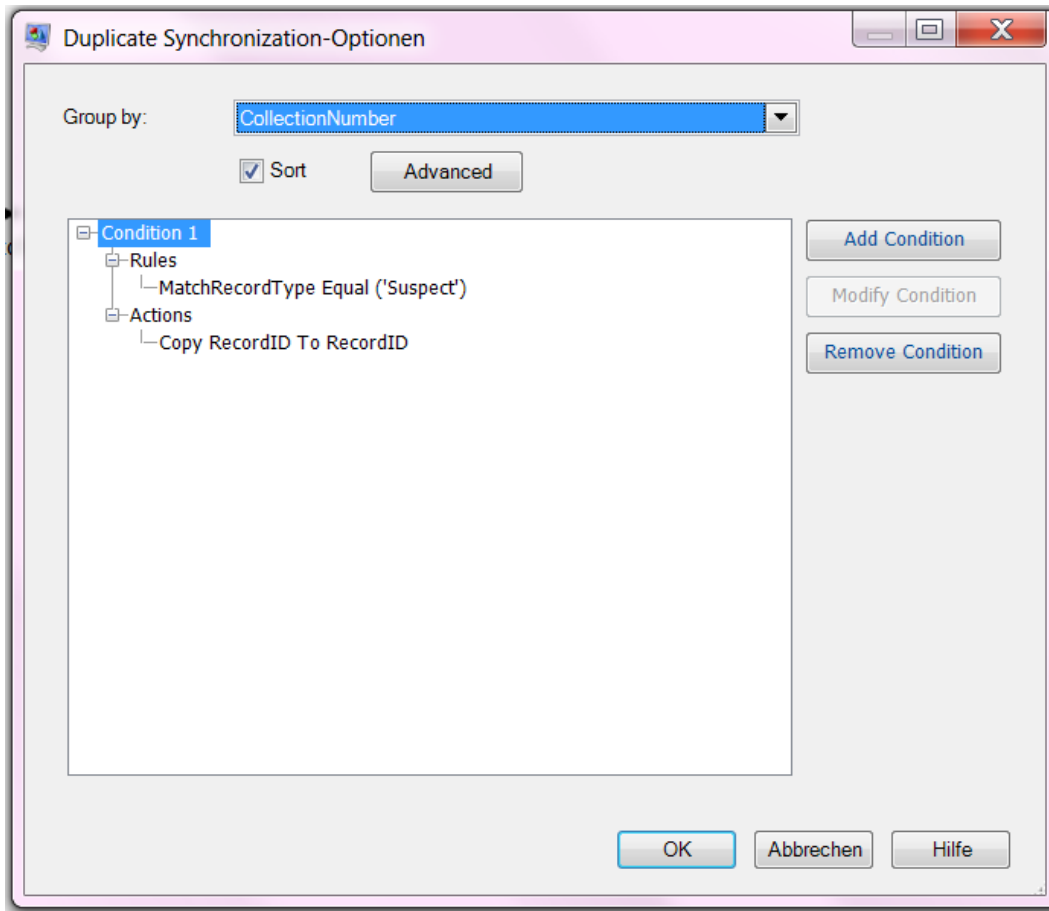
- a) Führen Sie den Datenfluss nach dem zweiten „Intraflow Match“-Schritt wie unten dargestellt fort:



- b) Konfigurieren Sie den „Conditional Router“-Schritt so, dass Datensätze, bei denen das Feld „CollectionNumber“ ungleich „0“ ist, zum „Duplicate Synchronization“-Schritt weitergeleitet werden.

So werden die Dubletten aus dem ersten Vergleichsschritt dem „Duplicate Synchronization“-Schritt geleitet.

- c) Konfigurieren Sie den „Duplicate Synchronization“-Schritt so, dass Datensätze nach dem Feld „CollectionNumber“ (nach der Sammlungsnummer aus dem zweiten Vergleichsdurchlauf) gruppiert werden. Prüfen Sie dann bei jeder Sammlung, ob Datensätze der Sammlung auch im ersten Vergleichsdurchlauf als Dubletten identifiziert wurden. Wenn welche gefunden wurden, kopieren Sie die Sammlungsnummer aus dem ersten Durchlauf in ein neues Feld namens „CollectionNumberConsolidated“. Konfigurieren Sie dazu Duplicate Synchronization wie folgt:



- d) Erstellen Sie im Transformer-Schritt nach dem „Duplicate Synchronization“-Schritt eine benutzerdefinierte Transformation unter Verwendung des folgenden Skripts:

```
if (data['CollectionNumberConsolidated'] == null) {
  data['CollectionNumberConsolidated'] = data['CollectionNumber']
}
```

- e) Konfigurieren Sie im Transformer, der direkt auf Conditional Router folgt (Transformer 2 im Beispieldatenfluss), eine Transformation, die „CollectionNumberPass1“ nach „CollectionNumberConsolidated“ kopiert.

Dabei wird bei den eindeutigen Datensätzen aus dem zweiten Vergleichsdurchlauf „CollectionNumberPass1“ nach „CollectionNumberConsolidated“ kopiert.

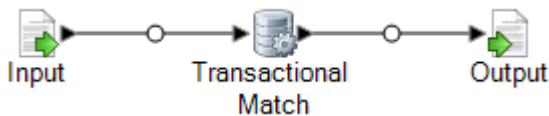
8. Nach dem Stream Combiner verfügen Sie über Sammlungen von Datensätzen, die in einer der Vergleichsdurchläufe Übereinstimmungen aufwiesen. Das Feld „CollectionNumberConsolidated“ zeigt die übereinstimmenden Datensätze an. Nach dem „Stream Combiner“-Schritt können Sie eine Datenladung oder eine beliebige zusätzliche Verarbeitung anfügen.

Erstellen eines allgemeinen Vergleichsdienstes

Bei einem allgemeinen Vergleichsdienst handelt es sich um einen Dienst, bei dem Ihre Vergleichsregeln zum Ausführen eines Vergleichs verwendet und Eingabefelder akzeptiert werden können. Der Dienst verwendet den Namen einer Vergleichsregel als Eingabeoption. Dadurch können Sie die Vergleichsregel angeben, die Sie im API-Aufruf oder in der Webservice-Anforderung verwenden möchten. Es gibt kein vordefiniertes Eingabeschema für den Dienst. Daher können Sie für die Typen von Datensätzen, die Sie miteinander vergleichen möchten, alle geeigneten Felder einschließen. Durch Erstellen eines allgemeinen Vergleichsdienstes können Sie vermeiden, dass es für jede Vergleichsregel einen separaten Dienst gibt. Dadurch können Sie neue Vergleichsregeln hinzufügen, ohne einen Dienst hinzufügen zu müssen.

In dieser Prozedur wird beschrieben, wie ein allgemeiner Vergleichsdienst erstellt wird. Sie beinhaltet zudem ein Beispiel für eine Webservice-Anforderung an den allgemeinen Vergleichsdienst.

1. Erstellen Sie einen neuen Dienstdatenfluss im Enterprise Designer.
2. Ziehen Sie einen Input-Schritt, einen „Transactional Match“-Schritt und einen Output-Schritt auf die Arbeitsfläche und verbinden Sie diese, sodass Sie über einen Datenfluss verfügen, der wie folgt aussieht:



3. Doppelklicken Sie auf den „Transactional Match“-Schritt.
4. Wählen Sie im Feld **Vergleichsregel laden** eine Vergleichsregel aus. Sie können beispielsweise die Standardvergleichsregel **Household** auswählen.

Auch wenn Sie die Vergleichsregel in der Dienstanforderung angeben, müssen Sie den „Transactional Match“-Schritt mit einer Standardvergleichsregel konfigurieren, damit der Datenfluss gültig ist. Wenn Sie keine Vergleichsregel auswählen, schlägt die Überprüfung des Datenflusses fehl, der folglich nicht verfügbar gemacht werden kann.

5. Klicken Sie auf **OK**.
6. Doppelklicken Sie auf den Output-Schritt.
7. Wählen Sie aus, dass die Felder `MatchRecordType` und `MatchScore` verfügbar gemacht werden sollen.
8. Klicken Sie auf **OK**.

Anmerkung: Im Input-Schritt müssen keine Felder verfügbar gemacht werden, da Eingabefelder in der Dienstanforderung als benutzerdefinierte Felder angegeben werden.

9. Klicken Sie auf **Bearbeiten > Datenflussoptionen**.

10. Klicken Sie auf **Hinzufügen**.
11. Erweitern Sie **Transactional Match** und aktivieren Sie das Kästchen neben **Vergleichsregel**.
Dadurch wird die Vergleichsregeloption als Laufzeitoption verfügbar gemacht, wodurch die Vergleichsregel in der Dienstanforderung angegeben werden kann.
12. Klicken Sie auf **OK** und klicken Sie anschließend erneut auf **OK**, um das Fenster „**Datenflussoptionen**“ zu schließen.
13. Speichern Sie den Datenfluss und machen Sie ihn verfügbar.

Sie verfügen jetzt über einen allgemeinen Vergleichsdienst, mit dem Sie mit beliebigen Vergleichsregeln, die im Vergleichsregelverwaltungstool im Enterprise Designer definiert wurden, einen Vergleich ausführen können. Geben Sie beim Aufrufen des Dienstes bei der Option `MatchRule` die Vergleichsregel an und geben Sie die Eingabefelder als benutzerdefinierte Felder an.

Beispiel: Aufrufen des allgemeinen Vergleichsdienstes

Sie haben im Vergleichsregelverwaltungstool eine Vergleichsregel mit dem Namen `AddressAndBirthday` erstellt. Diese Vergleichsregel führt anhand der Felder „Adresse“ und „Geburtstag“ einen Vergleich durch. Sie möchten unter Verwendung des allgemeinen Vergleichsdienstes mit dieser Regel über eine SOAP-Webservice-Anforderung einen Vergleich durchführen.

Hierfür würden Sie über eine SOAP-Anforderung verfügen, in der `AddressAndBirthday` im Element `MatchRule` und die Felder des Datensatzes im Element `user_fields` angegeben sind.

```
<soapenv:Envelope
xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/"
xmlns:univ="http://www.pb.com/spectrum/services/UniversalMatchingService">

  <soapenv:Header/>
  <soapenv:Body>
    <univ:UniversalMatchingServiceRequest>
      <univ:options>

<univ:MatchRule>AddressAndBirthday</univ:MatchRule>
      </univ:options>
      <univ:Input>
        <univ:Row>
          <univ:user_fields>
            <univ:user_field>
              <univ:name>Name</univ:name>
              <univ:value>Bob Smith</univ:value>
            </univ:user_field>
            <univ:user_field>
              <univ:name>Address</univ:name>
              <univ:value>4200 Parliament
Pl</univ:value>
            </univ:user_field>
```

```

        <univ:user_field>
          <univ:name>Birthday</univ:name>
          <univ:value>1973-6-15</univ:value>
        </univ:user_field>
      </univ:user_fields>
    </univ:Row>
    <univ:Row>
      <univ:user_fields>
        <univ:user_field>
          <univ:name>Name</univ:name>
          <univ:value>Robert M. Smith</univ:value>

        </univ:user_field>
        <univ:user_field>
          <univ:name>Address</univ:name>
          <univ:value>4200 Parliament
Pl</univ:value>
        </univ:user_field>
        <univ:user_field>
          <univ:name>Birthday</univ:name>
          <univ:value>1973-6-15</univ:value>
        </univ:user_field>
      </univ:user_fields>
    </univ:Row>
    <univ:Row>
      <univ:user_fields>
        <univ:user_field>
          <univ:name>Name</univ:name>
          <univ:value>Bob Smith</univ:value>
        </univ:user_field>
        <univ:user_field>
          <univ:name>Address</univ:name>
          <univ:value>424 Washington
Blvd</univ:value>
        </univ:user_field>
        <univ:user_field>
          <univ:name>Birthday</univ:name>
          <univ:value>1959-2-19</univ:value>
        </univ:user_field>
      </univ:user_fields>
    </univ:Row>
  </univ:Input>
</univ:UniversalMatchingServiceRequest>
</soapenv:Body>
</soapenv:Envelope>

```

Aus dieser Anforderung ergäbe sich die folgende Antwort:

```

<soap:Envelope
xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/">
  <soap:Body>

```

```

    <ns3:UniversalMatchingServiceResponse
xmlns:ns2="http://spectrum.pb.com/"

xmlns:ns3="http://www.pb.com/spectrum/services/UniversalMatchingService">

    <ns3:Output>
      <ns3:Row>
        <ns3:MatchScore/>

<ns3:MatchRecordType>Suspect</ns3:MatchRecordType>
      <ns3:user_fields>
        <ns3:user_field>
          <ns3:name>Name</ns3:name>
          <ns3:value>Bob Smith</ns3:value>
        </ns3:user_field>
        <ns3:user_field>
          <ns3:name>Birthday</ns3:name>
          <ns3:value>1973-6-15</ns3:value>
        </ns3:user_field>
        <ns3:user_field>
          <ns3:name>Address</ns3:name>
          <ns3:value>4200 Parliament Pl</ns3:value>

          </ns3:user_field>
        </ns3:user_fields>
      </ns3:Row>
      <ns3:Row>
        <ns3:MatchScore>100</ns3:MatchScore>

<ns3:MatchRecordType>Duplicate</ns3:MatchRecordType>
      <ns3:user_fields>
        <ns3:user_field>
          <ns3:name>Name</ns3:name>
          <ns3:value>Robert M. Smith</ns3:value>
        </ns3:user_field>
        <ns3:user_field>
          <ns3:name>Birthday</ns3:name>
          <ns3:value>1973-6-15</ns3:value>
        </ns3:user_field>
        <ns3:user_field>
          <ns3:name>Address</ns3:name>
          <ns3:value>4200 Parliament Pl</ns3:value>

          </ns3:user_field>
        </ns3:user_fields>
      </ns3:Row>
    </ns3:Output>
  </ns3:UniversalMatchingServiceResponse>
</soap:Body>
</soap:Envelope>

```

Verwenden eines Schnellvergleichsschlüssels

Der schnelle Schlüsselvergleich kann ein nützliches Tool sein, um die Anzahl der ausgeführten Vergleichsvorgänge zu reduzieren und dadurch die Ausführungsgeschwindigkeit bei Datenflüssen zu verbessern, in denen ein „Interflow Match“- oder „Intraflow Match“-Schritt verwendet wird. Wenn der Schnellvergleichsschlüssel in zwei Datensätzen genau übereinstimmt, wird der Kandidat als 100-prozentige Übereinstimmung gewertet und keine weiteren Übereinstimmungsversuche unternommen. Wenn der Schnellvergleichsschlüssel in zwei Datensätzen nicht übereinstimmt, werden diese anhand der regelbasierten Methode verglichen. Ein loser Schnellvergleichsschlüssel kann allerdings zu vielen falschen positiven Übereinstimmungen führen.

1. Öffnen Sie Ihren Datenfluss im Enterprise Designer.
2. Doppelklicken Sie auf den „Match Key Generator“-Schritt.
3. Aktivieren Sie das Kästchen **Schnellvergleichsschlüssel generieren**.
4. Klicken Sie auf **Hinzufügen**.
5. Füllen Sie die folgenden Felder aus:

Tabelle 6: Match Key Generator-Optionen

| Name der Option | Beschreibung/gültige Werte |
|-----------------|----------------------------|
|-----------------|----------------------------|

| | |
|-------------|--|
| Algorithmus | |
|-------------|--|

Name der Option

Beschreibung/gültige Werte

Gibt den Algorithmus an, der verwendet werden soll, um den Vergleichsschlüssel zu generieren. Zur Auswahl stehen:

| | |
|-----------------------------|--|
| Consonant | Gibt die angegebenen Felder mit entfernten Konsonanten zurück. |
| Double Metaphone | Gibt einen auf einer phonetischen Darstellung ihrer Zeichen basierenden Code zurück. „Double Metaphone“ ist eine verbesserte Version des Algorithmus „Metaphone“ und versucht, viele der in unterschiedlichen Sprachen vorkommenden Unregelmäßigkeiten zu berücksichtigen. |
| Köln | Indiziert Namen nach Klang, so wie sie auf Deutsch ausgesprochen werden. Ermöglicht die Codierung von Namen mit derselben Aussprache in derselben Darstellung, sodass sie trotz kleiner Unterschiede bei der Schreibweise verglichen werden können. Das Ergebnis ist immer eine Folge von Zahlen; Sonderzeichen und Leerzeichen werden ignoriert. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren. |
| MD5 | Ein Message Digest-Algorithmus, der einen 128-Bit-Hashwert erzeugt. Dieser Algorithmus wird im Allgemeinen verwendet, um die Datenintegrität zu überprüfen. |
| Metaphone | Gibt einen per Metaphone codierten Schlüssel ausgewählter Felder zurück. Metaphone ist ein Algorithmus, um Wörter anhand ihrer englischen Aussprache zu codieren. |
| Metaphone (Spanisch) | Gibt einen per Metaphone codierten Schlüssel ausgewählter Felder für die spanische Sprache zurück. Dieser Metaphone-Algorithmus codiert Wörter anhand ihrer spanischen Aussprache. |
| Metaphone 3 | Übertrifft die Algorithmen „Metaphone“ und „Double Metaphone“ mit genaueren Einstellungen für Konsonanten und inneren Vokalen, was Ihnen ermöglicht, Wörter oder Namen auf phonetischer Basis mehr oder weniger eng mit Suchbegriffen abgeglichen zu erzeugen. Metaphone 3 erhöht die Genauigkeit der phonetischen Codierung auf 98 %. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren. |
| NYSIIS | Phonetischer Codealgorithmus, der eine ungefähr übereinstimmende Aussprache mit der exakten Schreibweise vergleicht und Wörter indiziert, deren Aussprache ähnlich ist. Bestandteil des New York State Identification and Intelligence System. Angenommen, Sie suchen die Daten einer Person in einer Personendatenbank. Sie meinen, dass der Name der Person wie „John Smith“ klingt, er wird aber „Jon Smyth“ geschrieben. Würden Sie eine Suche durchführen, in der nach einer genauen Übereinstimmung mit „John Smith“ gesucht wird, würden keine Ergebnisse zurückgegeben. Wenn Sie jedoch die Datenbank mit dem Algorithmus NYSIIS |

| Name der Option | Beschreibung/gültige Werte |
|-------------------------|--|
| | indizieren und eine erneute Suche mit dem Algorithmus NYSIIS durchführen, wird die richtige Übereinstimmung zurückgegeben, weil vom Algorithmus sowohl „John Smith“ als auch „Jon Smyth“ als „JAN SNATH“ indiziert werden. |
| Phonix | Vorverarbeitet Namenszeichenfolgen, indem mehr als 100 Transformationsregeln auf einzelne Zeichen oder auf Zeichenfolgen angewendet werden. 19 dieser Regeln werden nur angewendet, wenn das bzw. die Zeichen am Anfang der Zeichenfolge stehen, während 12 der Regeln nur angewendet werden, wenn sie in der Mitte der Zeichenfolge stehen. 28 der Regeln werden nur angewendet, wenn sie am Ende der Zeichenfolge stehen. Die transformierte Namenszeichenfolge wird als Code codiert, der aus einem Anfangsbuchstaben gefolgt von drei Stellen besteht (Nullen und doppelt vorhandene Zahlen werden entfernt). Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren; sie ist komplexer und deshalb langsamer als Soundex. |
| Soundex | Gibt einen Soundex-Code ausgewählter Felder zurück. Soundex erzeugt einen auf der englischen Aussprache eines Wortes basierenden Code mit fester Länge. |
| Teilzeichenfolge | Gibt einen angegebenen Teil des ausgewählten Feldes zurück. |
| Feldname | Gibt das Feld an, auf das Sie den ausgewählten Algorithmus anwenden möchten, um den Vergleichsschlüssel zu generieren. Wenn Sie beispielsweise ein Feld mit Namen „LastName“ auswählen und den Soundex-Algorithmus wählen, würde der Soundex-Algorithmus auf die Daten im Feld „LastName“ angewendet, um einen Vergleichsschlüssel zu erzeugen. |
| Startposition | Gibt die Startposition innerhalb des angegebenen Feldes an. Nicht alle Algorithmen erlauben Ihnen, eine Startposition anzugeben. |
| Länge | Gibt die Länge der Zeichen an, die ab der Startposition eingeschlossen werden sollen. Nicht alle Algorithmen erlauben Ihnen, eine Länge anzugeben. |
| Sonderzeichen entfernen | Entfernt alle nicht numerischen und nicht alphabetischen Zeichen wie Bindestriche, Leerzeichen und andere Sonderzeichen aus einem Eingabefeld. |

| Name der Option | Beschreibung/gültige Werte |
|-------------------|---|
| Eingabe sortieren | Sortiert alle Zeichen in einem Eingabefeld oder alle Begriffe in einem Eingabefeld in alphabetischer Reihenfolge. |
| | Zeichen Sortiert die Zeichenwerte aus einem Eingabefeld vor dem Erstellen einer eindeutigen ID. |
| | Begriffe Sortiert jeden Begriffswert aus einem Eingabefeld vor dem Erstellen einer eindeutigen ID. |

6. Klicken Sie auf **OK**.
7. Wenn Sie ein zusätzliches Feld und/oder einen zusätzlichen Algorithmus zum Generieren des Schnellvergleichsschlüssels angeben möchten, klicken Sie auf **Hinzufügen**. Klicken Sie anderenfalls auf **OK**.
8. Doppelklicken Sie auf den „Interflow Match“- oder „Intraflow Match“-Schritt auf der Arbeitsfläche.
9. Wählen Sie die Option **Schnellvergleich aktiviert** aus und dann das Feld **ExpressMatchKey**. Dieses Feld enthält den von „Match Key Generator“ erzeugten Schnellvergleichsschlüssel.
10. Klicken Sie auf **OK**.
11. Speichern Sie Ihren Datenfluss und führen Sie ihn aus.

Um zu ermitteln, ob ein Kandidat mithilfe eines Schnellvergleichsschlüssels verglichen wurde, sehen Sie nach, ob das Feld **ExpressKeyIdentified** „Y“ (ja) oder „N“ (nein) enthält. Beachten Sie, dass Datensatzkopfdubletten immer den Wert „N“ in **ExpressKeyIdentified** enthalten.

Analysieren von Vergleichsergebnissen

Das Vergleichsanalyse-Tool im Enterprise Designer zeigt die Ergebnisse von einem oder mehreren Vergleichsschritten desselben Typs an. Das Tool liefert zusammenfassende Vergleichsergebnisse für einen Datenfluss und ermöglicht es Ihnen außerdem, Vergleichsergebnisse von Datensatz zu Datensatz anzuzeigen. Sie können diese Informationen zur Problembehandlung oder Feinabstimmung Ihrer Vergleichsregeln verwenden, um die gewünschten Ergebnisse zu erhalten.

Über das Vergleichsanalyse-Tool werden die folgenden Features zur Verfügung gestellt:

- Zusammenfassung der Vergleichsergebnisse: Zeigt Datensatzsummenzähler für ein einzelnes Vergleichsergebnis oder Vergleiche zwischen zwei Vergleichsergebnissen an.
- Erhöhen/Verringern-Diagramme: Verwendet Balkendiagramme, um einen Anstieg oder eine Abnahme der Übereinstimmungen anzuzeigen.

- **Vergleichsregeln:** Zeigt die Vergleichsregeln, die für ein einzelnes Vergleichsergebnis verwendet werden, oder die Änderungen an, die beim Vergleich zweier Vergleichsergebnisse an den Vergleichsregeln vorgenommen wurden.
- **Details der Vergleichsergebnisse:** Zeigt die Verarbeitungsdetails des Datensatzes für ein einzelnes Vergleichsergebnis oder den Vergleich zwischen zwei Vergleichsergebnissen an.

Anzeigen der Zusammenfassung der Vergleichsergebnisse

Das Vergleichsanalyse-Tool kann zusammenfassende Informationen zu den Vergleichsprozessen in einem Datenfluss anzeigen, z. B. die Anzahl an Datensatzdubletten, die durchschnittliche Punktzahl usw. Sie können die Ergebnisse eines einzelnen Auftrags anzeigen oder die Ergebnisse mehrerer Aufträge vergleichen.

1. Öffnen Sie im Enterprise Designer den zu analysierenden Datenfluss.
2. Klicken Sie für jeden „Interflow Match“-, „Intraflow Match“- oder „Transactional Match“-Schritt, dessen Vergleich Sie analysieren möchten, doppelt auf den Schritt, und aktivieren Sie das Kästchen **Daten für Analyse generieren**.

Wichtig: Die Leistung wird beeinträchtigt, wenn die Option **Daten für Analyse generieren** aktiviert wird. Sie sollten diese Option deaktivieren, wenn Sie mit der Nutzung des Vergleichsanalyse-Tools fertig sind.

3. Wählen Sie **Ausführen > Aktuellen Fluss ausführen** aus.

Anmerkung: Für optimale Ergebnisse sollten Sie Daten verwenden, die höchstens 100.000 Datensätze erzeugen. Die Leistung des Vergleichsanalyse-Tools verlangsamt sich, je mehr Vergleichsergebnisse es gibt.

4. Wählen Sie **Tools > Vergleichsanalyse** aus, wenn die Ausführung des Datenflusses abgeschlossen ist.

Im Dialogfeld **Vergleichsergebnisse durchsuchen** wird eine Liste von Datenflüssen mit Vergleichsergebnissen angezeigt, die im Vergleichsanalyse-Tool angezeigt werden können. Wenn der zu analysierende Auftrag nicht aufgeführt wird, müssen Sie den Datenfluss öffnen und sicherstellen, dass für den Vergleichsschritt das Kästchen **Daten für Analyse generieren** aktiviert ist.

Tipp: Wenn viele Datenflüsse vorhanden sind und Sie die Datenflüsse filtern möchten, müssen Sie eine Filteroption aus der Dropdown-Liste **Nur Aufträge anzeigen, bei denen** auswählen.

5. Klicken Sie auf das Symbol „+“ neben dem anzuzeigenden Datenfluss, um ihn zu erweitern.
6. Unter dem Datenfluss ist ein Eintrag für jeden Vergleichsschritt im Datenfluss vorhanden. Wählen Sie den Schritt aus, dessen Ergebnisse Sie anzeigen möchten, und klicken Sie auf **Hinzufügen**.

Das Vergleichsanalyse-Tool wird im unteren Bereich des Fensters des Enterprise Designer angezeigt.

7. Für einen Seite-an-Seite-Vergleich der Vergleichsergebnisse mit den Ergebnissen aus einem anderen Vergleich:
- Klicken Sie auf **Add** (Hinzufügen).
 - Wählen Sie den Vergleich aus, dessen Ergebnisse Sie vergleichen möchten.
 - Klicken Sie auf **Add** (Hinzufügen).
 - Wählen Sie in der Datenflussliste den Vergleich aus, den Sie gerade hinzugefügt haben, und klicken Sie auf **Vergleichen**.

Auf der Registerkarte **Zusammenfassung** wird die Vergleichsstatistik für den Auftrag angezeigt. Je nach Vergleichsschritttyp, der im Datenfluss verwendet wird, werden unterschiedliche Informationen angezeigt.

Für Intraflow Match werden die folgenden zusammenfassenden Informationen angezeigt:

| | |
|--|---|
| Eingabedatensätze | Die Anzahl der insgesamt während des Vergleichsschrittes verarbeiteten Datensätze. |
| Eindeutige Datensätze | Ein Kopfdubletten- oder Kandidatendatensatz, der mit keinen anderen Datensätzen in einer Übereinstimmungsgruppe übereinstimmt. Wenn er den einzigen Datensatz in einer Übereinstimmungsgruppe bildet, wird der Kopfdubletten-Datensatz automatisch als eindeutig eingeordnet. |
| Übereinstimmungsgruppen | (Gruppieren nach) Entweder nach einem Vergleichsschlüssel oder nach einem gleitenden Fenster gruppierte Datensätze. |
| Sammlungsduplikate | Ein Sammlungsduplikat besteht aus einer Kopfdublette und ihren Datensatzdubletten, welche wiederum nach einer Sammlungsnummer gruppiert sind. Eindeutige Datensätze gehören immer zu Sammlungsnummer 0. |
| Ausdrückliche Übereinstimmungen | Ein Schnellvergleich erfolgt, wenn eine Kopfdublette und ein Kandidat eine genaue Übereinstimmung hinsichtlich der Inhalte eines festgelegten Feldes aufweisen. In der Regel ist dies ein vom Match Key Generator bereitgestellter ExpressMatchKey. Wenn ein Schnellvergleich durchgeführt wird, wird die Verarbeitung nicht fortgesetzt, um zu ermitteln, ob die Kopfdublette und der Kandidat Dubletten sind. |
| Durchschnittliche Punktzahl | Der durchschnittliche Übereinstimmungswert aller Dubletten. Möglich sind Werte zwischen 0 und 100, wobei 0 eine niedrige Übereinstimmung und 100 eine genaue Übereinstimmung angibt. |

Für Interflow Match werden die folgenden zusammenfassenden Informationen angezeigt:

| | |
|--|--|
| Sammlungsduplikate | Ein Sammlungsduplikat besteht aus einer Kopfdublette und ihren Datensatzdubletten, welche wiederum nach einer Sammlungsnummer gruppiert sind. Eindeutige Datensätze gehören immer zu Sammlungsnummer 0. |
| Ausdrückliche Übereinstimmungen | Ein Schnellvergleich erfolgt, wenn eine Kopfdublette und ein Kandidat eine genaue Übereinstimmung hinsichtlich der Inhalte eines festgelegten Feldes aufweisen. In der Regel ist dies ein vom Match Key Generator bereitgestellter ExpressMatchKey. Wenn ein |

Schnellvergleich durchgeführt wird, wird die Verarbeitung nicht fortgesetzt, um zu ermitteln, ob die Kopfdublette und der Kandidat Dubletten sind.

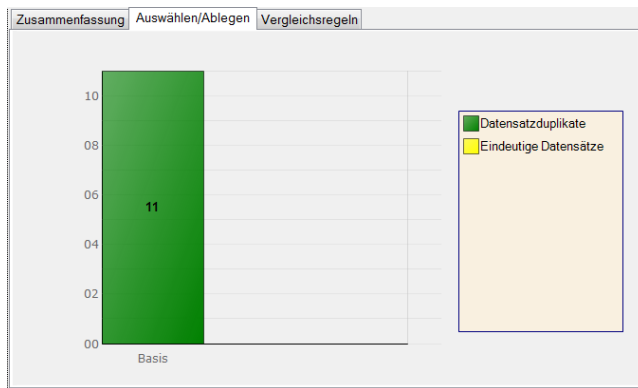
| | |
|--------------------------------------|--|
| Durchschnittliche Punktzahl | Der durchschnittliche Übereinstimmungswert aller Dubletten. Möglich sind Werte zwischen 0 und 100, wobei 0 eine niedrige Übereinstimmung und 100 eine genaue Übereinstimmung angibt. |
| Eingabe-Kopfdubletten | Die Anzahl der Datensätze im Eingabestream, die der Vergleichsvorgang mit anderen Datensätzen abgleichen wollte. |
| Kopfdubletten mit Dubletten | Die Anzahl der Eingabe-Kopfdubletten, die mit mindestens einem der Kandidatendatensätze übereingestimmt haben. |
| Eindeutige Kopfdubletten | Die Anzahl der Eingabe-Kopfdubletten, die mit keinem der Kandidatendatensätze übereingestimmt haben. |
| Kopfdubletten mit Kandidaten | Die Anzahl der Eingabe-Kopfdubletten, in deren Übereinstimmungsgruppe sich mindestens ein Kandidatendatensatz befand und die dementsprechend mindestens einen Vergleichsversuch aufweisen. |
| Kopfdubletten ohne Kandidaten | Die Anzahl der Eingabe-Kopfdubletten, in deren Übereinstimmungsgruppe sich keine Kandidatendatensätze befanden und die dementsprechend keine Vergleichsversuche aufweisen. |

Für Transactional Match werden die folgenden zusammenfassenden Informationen angezeigt:

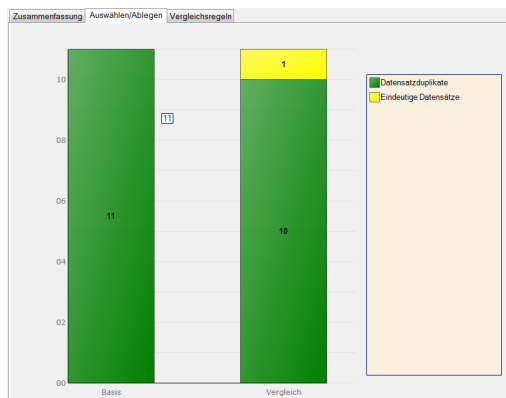
| | |
|--------------------------------------|--|
| Durchschnittliche Punktzahl | Der durchschnittliche Übereinstimmungswert aller Dubletten. Möglich sind Werte zwischen 0 und 100, wobei 0 eine niedrige Übereinstimmung und 100 eine genaue Übereinstimmung angibt. |
| Eingabe-Kopfdubletten | Die Anzahl der Datensätze im Eingabestream, die der Vergleichsvorgang mit anderen Datensätzen abgleichen wollte. |
| Kopfdubletten mit Dubletten | Die Anzahl der Eingabe-Kopfdubletten, die mit mindestens einem der Kandidatendatensätze übereingestimmt haben. |
| Eindeutige Kopfdubletten | Die Anzahl der Eingabe-Kopfdubletten, die mit keinem der Kandidatendatensätze übereingestimmt haben. |
| Kopfdubletten mit Kandidaten | Die Anzahl der Eingabe-Kopfdubletten, in deren Übereinstimmungsgruppe sich mindestens ein Kandidatendatensatz befand und die dementsprechend mindestens einen Vergleichsversuch aufweisen. |
| Kopfdubletten ohne Kandidaten | Die Anzahl der Eingabe-Kopfdubletten, in deren Übereinstimmungsgruppe sich keine Kandidatendatensätze befanden und die dementsprechend keine Vergleichsversuche aufweisen. |

Die Registerkarte **Erhöhen/Verringern** des Vergleichsanalyse-Tools zeigt die Anzahl duplizierter und eindeutiger Datensätze in einem Balkendiagramm für die ausgewählten Basis- und optional auch Vergleichsergebnisse an. *Erhöhen* ist der Anstieg in der Anzahl an Datensatzdubletten. *Verringern* ist die Abnahme in der Anzahl an Datensatzdubletten. Eindeutige Datensätze werden in Gelb und Datensatzdubletten in Grün dargestellt.

Wenn lediglich ein Basisauftrag ausgewählt wird, werden im Diagramm die Ergebnisse für diesen einen Auftrag angezeigt:



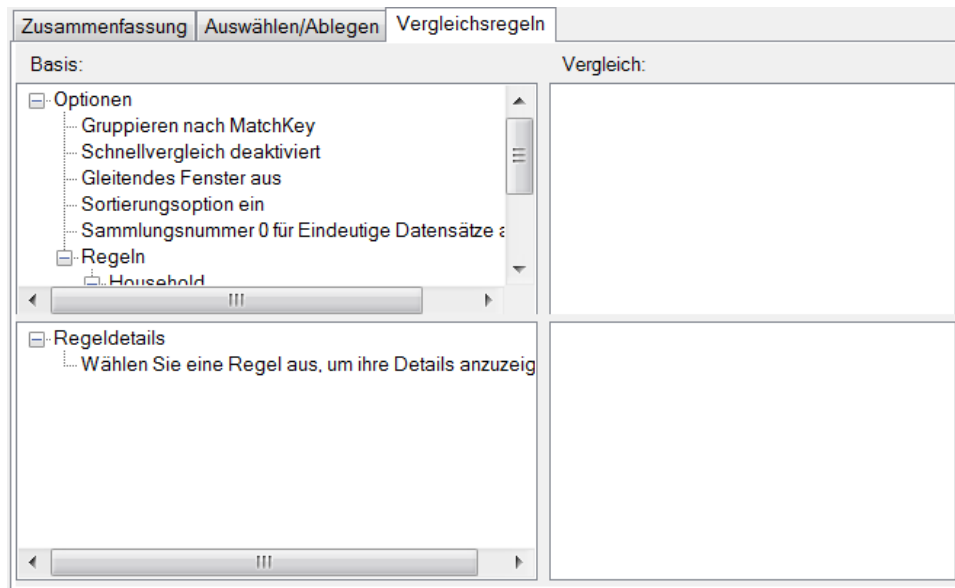
Wenn ein Basisauftrag und ein Vergleichsauftrag ausgewählt werden, wird ein Diagramm für die Basis- und Vergleichsaufträge nebeneinander angezeigt:



Die Registerkarte **Vergleichsregeln** des Vergleichsanalyse-Tools zeigt die Vergleichsregeln, die für ein einzelnes Vergleichsergebnis verwendet werden, oder die Änderungen an, die beim Vergleich zweier Vergleichsergebnisse an den Vergleichsregeln vorgenommen wurden.

Vergleichsregeln werden in einer hierarchischen Struktur dargestellt, die der Anzeige im Schritt ähnelt, in der sie erstellt wurden. Die Regelhierarchie enthält zwei Knoten: Optionen und Regeln. Im Knoten „Optionen“ werden die Schritteinstellungen für das ausgewählte Vergleichsergebnis angezeigt. Im Knoten „Regeln“ werden die Vergleichsregeln für das ausgewählte Vergleichsergebnis angezeigt.

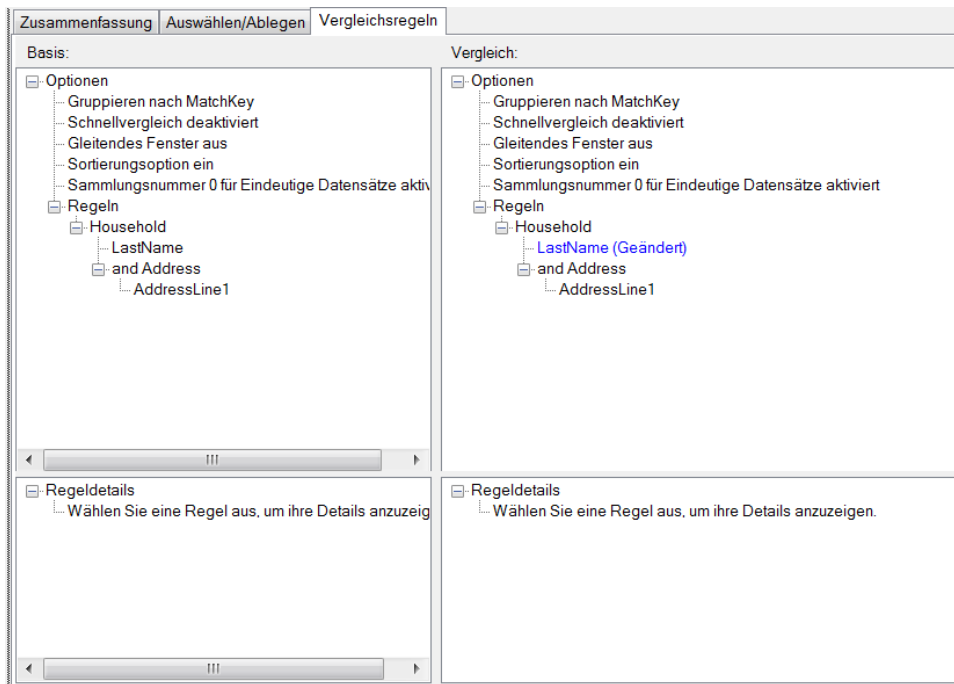
Wählen Sie einen Knoten in der Hierarchie, um die Regeldetails anzuzeigen.



Wenn Sie Vergleichsregeln zwischen mehreren Aufträgen vergleichen, werden die Unterschiede zwischen den Vergleichsergebnissen der Basis und des Vergleichs anhand des folgenden Farbcodes dargestellt:

- Blau** Gibt an, dass die Vergleichsregel beim Vergleich der Vergleichsergebnisse geändert wurde.
- Grün** Gibt an, dass die Vergleichsregel beim Vergleich der Vergleichsergebnisse hinzugefügt wurde.
- Rot** Gibt an, dass die Vergleichsregel beim Vergleich der Vergleichsergebnisse ausgelassen wurde.

Beispiel:



Anzeigen der Vergleichsergebnisse auf Datensatzebene

In den detaillierten Ergebnissen wird eine Sammlung von Details über Vergleichsdatensätze für Vergleichsergebnisgruppen angezeigt.

So zeigen Sie detaillierte Ergebnisse an:

1. Geben Sie im Vergleichsanalyse-Tool einen Basisauftrag und optional einen Vergleichsauftrag an.
2. Klicken Sie auf **Details**.

Die Basis-Vergleichsergebnisse werden basierend auf der ausgewählten Ansicht in der Dropdown-Liste „Anzeigen“ angezeigt. In der folgenden Tabelle sind die für jeden Vergleichsschritttyp angezeigten Spalten aufgelistet.

Tabelle 7: Angezeigte detaillierte Ergebnisdaten

| Detailbezogene Ergebnisse | Intraflow | Interflow | Transactional |
|------------------------------|-----------|-----------|---------------|
| Anzahl der Eingabedatensätze | X | X | X |
| Übereinstimmungsgruppe | X | X | |

| Detailbezogene Ergebnisse | Intraflow | Interflow | Transactional |
|--|-----------|-----------|---------------|
| Schnellschlüssel | X | X | |
| Treiberdatensatz des Schnellschlüssels | X | X | |
| Sammlungsnummer | X | X | X |
| Vergleichsdatensatztyp | X | X | X |
| Von den Regeln verwendete Felder | X | X | X |
| Gesamtpunktzahl der Regel (oberste Ebene) | X | | |
| Kandidatengruppe | | X | X |
| MatchScore: Wählen Sie ein Vergleichsergebnis in der Liste der Vergleichsergebnisse aus, und klicken Sie dann auf Entfernen . | | X | X |

Informationen zum Übereinstimmungsratendiagramm finden Sie unter [Übereinstimmungsratendiagramm](#) auf Seite 130.

- Wählen Sie im Feld **Analysieren** eine der folgenden Optionen aus:
 - Basis** Zeigt die Vergleichsergebnisse aus der Basisausführung an.
 - Vergleich** Zeigt die Vergleichsergebnisse aus der Vergleichsausführung an.
- Wählen Sie einen der folgenden Werte aus der Liste **Anzeigen** aus, und klicken Sie anschließend auf **Aktualisieren**. Wenn Sie Basisergebnisse analysieren, stehen Ihnen die folgenden Optionen zur Verfügung:
 - Kopfdubletten mit Kandidaten: (Alle Übereinstimmungen) Zeigt Kopfdubletten-Datensätze und alle Kandidatendatensätze an, die einen Abgleich mit jeder Kopfdublette versucht haben.
 - Kopfdubletten mit Dubletten: (Alle Übereinstimmungen) Zeigt alle Kopfdubletten-Datensätze und Kandidatendatensätze an, die mit jeder Kopfdublette übereingestimmt haben.
 - Kopfdubletten mit ausdrücklichen Übereinstimmungen: (Interflow Match und Intraflow Match, wenn der Schnellvergleichsschlüssel aktiviert ist) Zeigt Kopfdubletten-Datensätze und

Kandidatendatensätze an, die basierend auf dem Schnellvergleichsschlüssel miteinander übereinstimmen.

- Sammlungsduplikate: (Intraflow und Interflow) Zeigt alle Sammlungsduplikate nach Sammlungsnummer an.
- Übereinstimmungsgruppen: (Intraflow und Interflow) Zeigt Datensätze nach Übereinstimmungsgruppen an.
- Kandidatengruppen: (Transactional Match) Zeigt Datensätze nach Kandidatengruppen an.
- Eindeutige Kopfdoubletten: (Interflow und Transactional Match) Zeigt alle Kopfdoubletten-Datensätze an, die mit keinen Kandidatendatensätzen übereingestimmt haben.
- Eindeutige Datensätze: (Intraflow) Zeigt alle nicht übereinstimmenden Datensätze an.
- Kopfdoubletten ohne Kandidaten: (Interflow und Transactional Match) Zeigt alle Kopfdoubletten an, die keine Kandidaten für einen Vergleich enthalten haben.
- Alle Datensätze: Zeigt alle durch den Vergleichsschritt verarbeiteten Datensätze an.

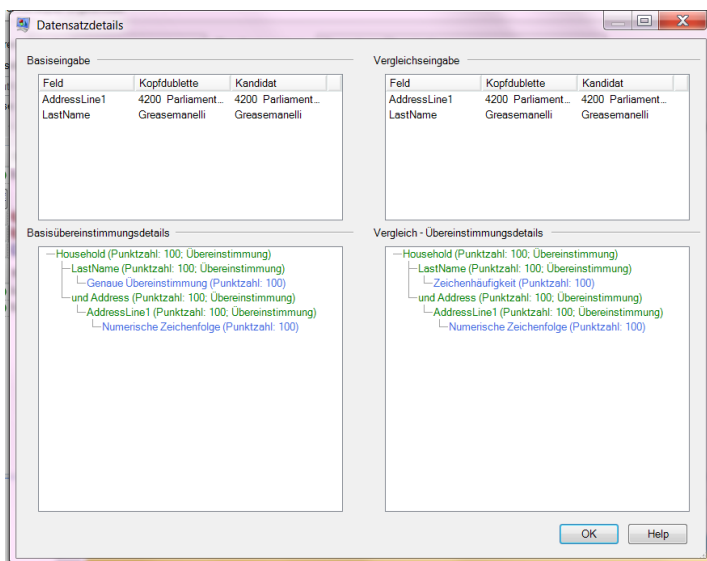
Wenn Sie Vergleichsergebnisse analysieren, stehen Ihnen die folgenden Anzeigoptionen zur Verfügung:

- Neue Übereinstimmungen: (Intraflow) Zeigt alle neuen Übereinstimmungen und ihre zugehörigen Kopfdoubletten an. In dieser Ansicht werden die Ergebnisse von „Kopfdoubletten mit neuen Doubletten“ und „Neue Kopfdoubletten“ in einer Ansicht zusammengeführt.
 - Neue verglichene Kopfdoubletten: (Interflow und Transactional Match) Zeigt Kopfdoubletten an, die keine Doubletten in der Basis aufgewiesen haben, aber mindestens eine Doublette im Vergleich aufweisen.
 - Neue eindeutige Kopfdoubletten: (Interflow und Transactional Match) Zeigt Kopfdoubletten an, die Doubletten in der Basis aufgewiesen haben, aber keine im Vergleich.
 - Übersehene Übereinstimmungen: (Intraflow) Zeigt alle übersehenen Übereinstimmungen an. In dieser Ansicht werden die Ergebnisse von „Kopfdoubletten mit übersehenen Doubletten“ und „Übersehene Kopfdoubletten“ in einer Ansicht zusammengeführt.
 - Kopfdoubletten mit neuen Doubletten: (Alle Übereinstimmungen) Zeigt Datensätze an, die neue Doubletten für Datensätze sind, die Kopfdoubletten in der Basis waren und Kopfdoubletten im Vergleich geblieben sind.
 - Kopfdoubletten mit übersehenen Doubletten: (Alle Übereinstimmungen) Zeigt Datensätze an, die übersehene Doubletten für Datensätze sind, die Kopfdoubletten in der Basis waren und Kopfdoubletten im Vergleich geblieben sind.
 - Neue Kopfdoubletten: (Intraflow) Zeigt Datensätze an, die Kopfdoubletten im Vergleichsergebnis sind, aber keine Kopfdoubletten in der Basis waren.
 - Übersehene Kopfdoubletten: (Intraflow) Zeigt Datensätze an, die keine Kopfdoubletten im Vergleichsergebnis sind, aber Kopfdoubletten in der Basis waren.
5. Erweitern Sie einen Kopfdoubletten-Datensatz, um die zugehörigen Kandidaten anzuzeigen.
 6. Wählen Sie einen Kandidatendatensatz aus, und klicken Sie auf **Details**.

Anmerkung: Diese Option ist nicht verfügbar, wenn „Gleitendes Fenster“ in „Intraflow Match“-Schritten aktiviert ist.

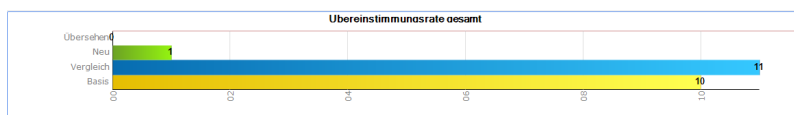
Im Fenster **Datensatzdetails** werden Daten auf Feldebene sowie der MatchScore für jede Vergleichsregel angezeigt. Wenn Sie einen Auftragslauf sowohl für die Basis als auch für den Vergleich festgelegt haben, können Sie die Ergebnisse des Datensatzes für Basis- und Vergleichsläufe anzeigen.

- **Basiseingabe:** Zeigt die beim Vergleich verwendeten Daten auf Feldebene sowohl von der Kopfdublette als auch dem Kandidaten an.
- **Basisübereinstimmungsdetails:** Zeigt Informationen zur Punktzahl für jeden Knoten in den Vergleichsregeln an.
- **Vergleichseingabe:** Zeigt die beim Vergleich verwendeten Daten auf Feldebene sowohl von der Kopfdublette als auch dem Kandidaten an.
- **Vergleich – Übereinstimmungsdetails:** Zeigt Informationen zur Punktzahl für jeden Knoten in den Vergleichsregeln an. Grüner Text stellt eine Übereinstimmung für einen Knoten in den Regeln dar. Roter Text stellt eine fehlende Übereinstimmung für einen Knoten in den Regeln dar.



Übereinstimmungsratendiagramm

Übereinstimmungsratendiagramme stellen Übereinstimmungsinformationen grafisch in Detailansichten dar.



Für Intraflow-Übereinstimmungen wird ein Diagramm angezeigt, das die gesamten Übereinstimmungen darstellt:


- **Basisübereinstimmungen:** Gesamtzahl der Übereinstimmungen im Basisergebnis.
- **Vergleichsübereinstimmungen:** Gesamtzahl der Übereinstimmungen im Basisergebnis.

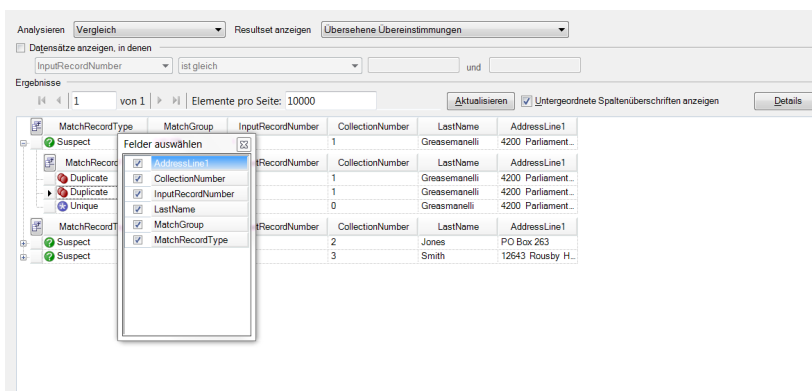
- Neue Übereinstimmungen: Die Anzahl aller Datensätze, die im Basisergebnis eindeutig waren, aber eine Kopfdublette oder Dublette im Vergleichsergebnis sind.
- Übersehene Übereinstimmungen: Die Anzahl aller Datensätze, die Kopfdubletten oder Dubletten im Basisergebnis waren, aber im Vergleichsergebnis eindeutig sind.

Für Interflow- und Transactional-Übereinstimmungen werden zwei Diagramme angezeigt:

- Übereinstimmungsrate gesamt
- Basisübereinstimmungen: Gesamtzahl der Übereinstimmungen im Basisergebnis.
- Vergleichsübereinstimmungen: Gesamtzahl der Übereinstimmungen im Basisergebnis.
- Neue Übereinstimmungen: Die Anzahl aller Datensätze, die im Basisergebnis eindeutig waren, aber eine Kopfdublette oder Dublette im Vergleichsergebnis sind.
- Übersehene Übereinstimmungen: Die Anzahl aller Datensätze, die Kopfdubletten oder Dubletten im Basisergebnis waren, aber im Vergleichsergebnis eindeutig sind.
- Übereinstimmungsrate für Kopfdublette
- Basisübereinstimmungen: Die Anzahl aller Kopfdubletten, die in der Basis nicht eindeutig waren.
- Vergleichsübereinstimmungen: Die Anzahl aller Kopfdubletten, die im Vergleich nicht eindeutig waren.
- Neue Übereinstimmungen: Die Anzahl aller Kopfdubletten, die in der Basis eindeutig waren, aber Übereinstimmungen im Vergleichsergebnis sind.
- Übersehene Übereinstimmungen: Die Anzahl aller Kopfdubletten, die Übereinstimmungen in der Basis waren, aber im Vergleichsergebnis eindeutig sind.

Verwenden der Feldauswahl

Klicken Sie auf das Symbol „Feldauswahl“ , um die ausgewählten Spalten in den Vergleichsanalyse-Ergebnissen anzuzeigen. Die Feldauswahl zeigt die Ergebnisse auf übergeordneter und untergeordneter Ebene an. Sie können Anzeigespalten für übergeordnete und untergeordnete Elemente unabhängig anzeigen.



| MatchRecordType | MatchGroup | InputRecordNumber | CollectionNumber | LastName | AddressLine1 |
|-----------------|------------|-------------------|------------------|------------------|--------------------|
| Suspect | | | 1 | Greasemanelli | 4200 Parliament... |
| MatchRecord | | | RecordNumber | CollectionNumber | LastName |
| Duplicate | | | 1 | Greasemanelli | 4200 Parliament... |
| Duplicate | | | 1 | Greasemanelli | 4200 Parliament... |
| Unique | | | 0 | Greasemanelli | 4200 Parliament... |
| MatchRecord | | | RecordNumber | CollectionNumber | LastName |
| Suspect | | | 2 | Jones | PO Box 263 |
| Suspect | | | 3 | Smith | 12643 Rousby H... |

Filtern von Datensätzen

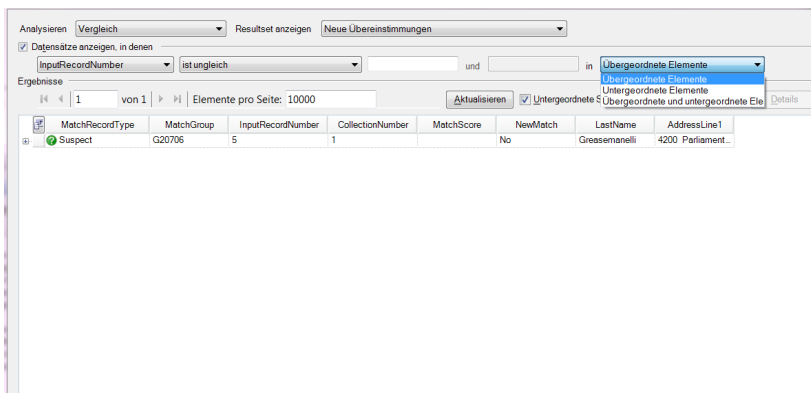
Mithilfe des Kästchens **Datensätze anzeigen, in denen** können Sie die angezeigten detaillierten Vergleichsdatsätze filtern. Sie können Datensätze basierend auf mehreren Operatoren filtern, um vom Benutzer bereitgestellte Werte mit Daten in einem Feld eines jeden detaillierten Vergleichsdatsatzes zu filtern.

Die folgenden Operatoren stehen zur Auswahl:

- Felder des Typs Zeichenfolge (MatchGroup, MatchRecordType, beliebige Übereinstimmungsdaten)
- enthält
- ist zwischen
- ist gleich
- ist ungleich
- beginnt mit
- Numerische Felder (CollectionNumber, InputRecordNumber, MatchScore)
- ist zwischen
- ist gleich
- ist ungleich
- ist größer als
- ist größer oder gleich
- ist kleiner als
- ist kleiner oder gleich

So filtern Sie Datensätze:

1. Wählen Sie ein Vergleichsergebnis aus der Basis oder dem Vergleich aus der Ansicht „Vergleichsanalyse-Ergebnisse“ aus, und klicken Sie auf **Aktualisieren**.
2. Aktivieren Sie das Kästchen **Datensätze anzeigen, in denen**.



3. Wählen Sie ein Feld aus dem Feld „Feldliste“ aus.
4. Wählen Sie einen Operator aus.
5. Geben Sie einen Wert für den ausgewählten Operatortypen ein. Wenn Sie **ist zwischen** auswählen, müssen Sie einen Wertebereich eingeben.

6. Beim Filtern von Kopfdublettenansichten können Sie für Folgendes Filter anwenden:
 - Übergeordnete Elemente: Nur übergeordnete Elemente (Kopfdubletten) filtern, alle untergeordneten Elemente zurückgeben.
 - Untergeordnete Elemente: Alle untergeordneten Elemente herausfiltern, die nicht innerhalb des Filterbereichs liegen. Übergeordnete Knoten (Kopfdubletten) werden zurückgegeben.
 - Übergeordnete und untergeordnete Elemente: Übergeordnete Elemente (Kopfdubletten) filtern. Wenn übergeordnete Elemente zurückgegeben werden, deren untergeordnete Elemente filtern.
7. Klicken Sie auf **Aktualisieren**. Datensätze, die innerhalb des Bereichs der Optionen und Werte liegen, werden angezeigt. Wenn keine Datensätze innerhalb des Bereichs der ausgewählten Optionen und Werte liegen, zeigt eine Nachricht an, dass keine Datensätze zurückgegeben wurden.

Analysieren von Änderungen an der Vergleichsregel

Mithilfe des Vergleichsanalyse-Tools im Enterprise Designer können Sie detailliert die Auswirkungen anzeigen, die eine Änderung in einer Vergleichsregel in den Vergleichsergebnissen des Datenflusses bewirkt. Führen Sie dafür den Datenfluss aus, nehmen Sie die Änderungen vor, führen Sie den Datenfluss erneut aus, und zeigen Sie anschließend die Ergebnisse im Vergleichsanalyse-Tool an. In dieser Prozedur wird beschrieben, wie dies durchgeführt werden kann.

Wichtig: Beim Abgleich der Vergleichsergebnisse müssen die Eingabedaten, die für die Basis- und Vergleichsläufe verwendet wurden, identisch sein. Wenn unterschiedliche Eingabedaten verwendet werden, können irreführende Ergebnisse die Folge sein. Beachten Sie Folgendes, um einen genauen Vergleich sicherzustellen:

- Dieselben Eingabedateien oder Tabellen verwenden
 - Die Daten vor dem Vergleichsschritt auf dieselbe Art und Weise sortieren
 - Dieselben Candidate Finder-Abfragen bei der Nutzung von Transactional Match verwenden
1. Öffnen Sie im Enterprise Designer den zu analysierenden Datenfluss.
 2. Klicken Sie für jeden „Interflow Match“-, „Intraflow Match“- oder „Transactional Match“-Schritt, dessen Vergleich Sie analysieren möchten, doppelt auf den Schritt, und aktivieren Sie das Kästchen **Daten für Analyse generieren**.

Wichtig: Die Leistung wird beeinträchtigt, wenn die Option **Daten für Analyse generieren** aktiviert wird. Sie sollten diese Option deaktivieren, wenn Sie mit der Nutzung des Vergleichsanalyse-Tools fertig sind.

3. Wählen Sie **Ausführen > Aktuellen Fluss ausführen** aus.

Anmerkung: Für optimale Ergebnisse sollten Sie Daten verwenden, die höchstens 100.000 Datensätze erzeugen. Die Leistung des Vergleichsanalyse-Tools verlangsamt sich, je mehr Vergleichsergebnisse es gibt.

4. Nehmen Sie im Vergleichsschritt oder den Vergleichsschritten des Datenflusses die gewünschten Änderungen an der Vergleichsregel vor, und führen Sie den Datenfluss anschließend erneut aus.

Wenn Sie beispielsweise die Auswirkungen einer Anhebung des Schwellenwertes testen möchten, müssen Sie den Schwellenwert ändern und anschließend den Datenfluss erneut ausführen.

5. Wählen Sie **Tools > Vergleichsanalyse** aus, wenn die Ausführung des Datenflusses abgeschlossen ist.

Im Dialogfeld **Vergleichsergebnisse durchsuchen** wird eine Liste von Datenflüssen mit Vergleichsergebnissen angezeigt, die im Vergleichsanalyse-Tool angezeigt werden können. Wenn der zu analysierende Auftrag nicht aufgeführt wird, müssen Sie den Datenfluss öffnen und sicherstellen, dass für den Vergleichsschritt das Kästchen **Daten für Analyse generieren** aktiviert ist.

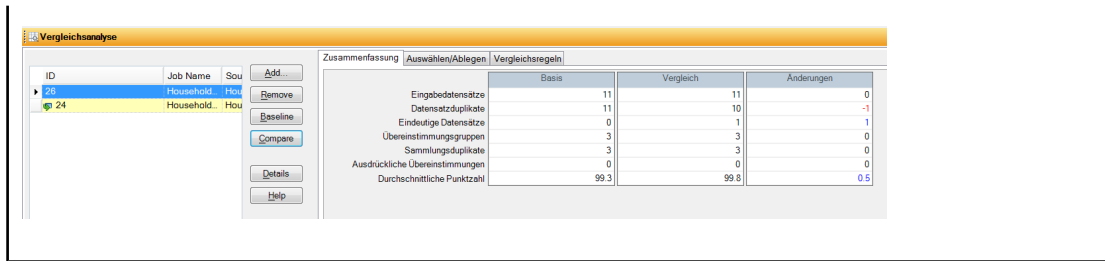
Tipp: Wenn viele Datenflüsse vorhanden sind und Sie die Datenflüsse filtern möchten, müssen Sie eine Filteroption aus der Dropdown-Liste **Nur Aufträge anzeigen, bei denen** auswählen.

6. Auf der linken Seite des Vergleichsanalyse-Bereichs befindet sich eine Liste der Vergleichsschritte (einer pro Ausführung). Wählen Sie die Vergleichsschritte in der Ausführung aus, die Sie als Basis für den Vergleich verwenden möchten, und klicken Sie anschließend auf **Basis**. Wählen Sie anschließend die Ausführung aus, mit der Sie die Basis vergleichen möchten, und klicken Sie auf **Vergleichen**.

Sie können nun zusammenfassende Vergleichsergebnisse vergleichen, z. B. die Summe der Datensatzdubletten. Sie können ebenso detaillierte Informationen auf Datensatzebene vergleichen, die anzeigen, wie jeder Datensatz gegen die Vergleichsregeln ausgewertet wurde.

Beispiel eines Vergleichs der Vergleichsergebnisse

Nehmen wir beispielsweise an, dass Sie einen Auftrag mit der Bezeichnung „HouseholdRelationshipsAnalysis“ ausführen. Sie möchten die Auswirkungen einer Änderung am Schritt „Household Match 2“ testen. Zunächst führen Sie den Auftrag mit den ursprünglichen Einstellungen aus, anschließend ändern Sie die Vergleichsregeln im Schritt „Household Match 2“ und führen den Auftrag erneut aus. Im Vergleichsanalyse-Tool ist die Ausführung mit der Auftrags-ID 10 die Ausführung mit den ursprünglichen Einstellungen. Dementsprechend legen Sie sie als Basis fest. Die Ausführung mit der Auftrags-ID 13 ist die Ausführung mit der geänderten Vergleichsregel. Wenn Sie auf **Vergleichen** klicken, können Sie sehen, dass die geänderte Vergleichsregel (Auftrags-ID 13) eine zusätzliche Datensatzdublette und einen eindeutigen Datensatz weniger als die ursprüngliche Vergleichsregel erzeugt hat.



Hinzufügen von Vergleichsergebnissen

Wenn Sie einen Auftrag ausführen, während das Tool „Vergleichsanalyse“ geöffnet ist, und die Liste der Vergleichsergebnisse leer ist, werden die Vergleichsergebnisse automatisch der Liste hinzugefügt. Nachdem ein Vergleichsergebnis hinzugefügt wurde, fügt das Tool „Vergleichsanalyse“ nur Vergleichsergebnisse desselben Vergleichstyps („Interflow Match“, „Intraflow Match“ oder „Transactional Match“) hinzu.

Wenn Sie Vergleichsergebnisse eines anderen als des derzeit im Tool „Vergleichsanalyse“ ausgewählten Typs analysieren möchten, führen Sie diese Schritte aus.

1. Wählen Sie alle Vergleichsergebnisse in der Liste der Vergleichsergebnisse aus, und klicken Sie dann auf **Entfernen**.
2. Öffnen Sie einen Auftrag von Server-Explorer, der einen anderen Vergleichsschritt verwendet, oder klicken Sie auf die Registerkarte über der Arbeitsfläche, wenn der Auftrag bereits geöffnet ist.
3. Führen Sie den Auftrag aus.

Wenn die Ausführung des Auftrags beendet ist, werden die Vergleichsergebnisse der letzten Auftragsinstanz der Liste der Vergleichsergebnisse hinzugefügt.

Entfernen von Vergleichsergebnissen

Um ein Vergleichsergebnis aus der Liste der Vergleichsergebnisse zu entfernen, wählen Sie es in der Liste aus und klicken Sie auf **Entfernen**.

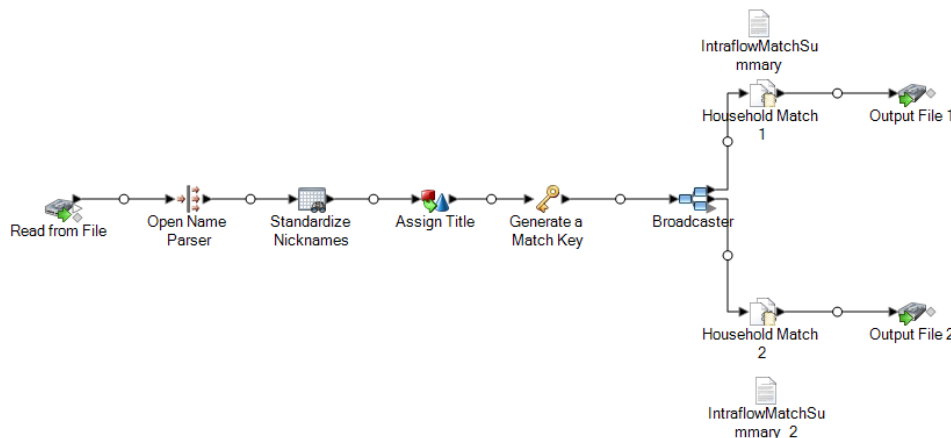
Das System aktualisiert die Liste der Vergleichsergebnisse und die Registerkarte „Zusammenfassung“ wie folgt:

- Wenn das entfernte Vergleichsergebnis weder das Vergleichsergebnis „Basis“ noch „Vergleich“ war, wird das Vergleichsergebnis entfernt und erfolgen keine Änderungen an der Registerkarte „Zusammenfassung“.
- Wenn das entfernte Vergleichsergebnis als „Basis“ festgelegt wurde, legt das System das nächstälteste Vergleichsergebnis als neue Basis fest und aktualisiert die Registerkarte „Zusammenfassung“, sodass nur die neuen Basisdaten angezeigt werden.

- Wenn das entfernte Vergleichsergebnis als „Vergleich“ festgelegt wurde, aktualisiert das System die Registerkarte „Zusammenfassung“, sodass nur die bestehenden Basisdaten angezeigt werden.
- Wenn das entfernte Vergleichsergebnis eines von zwei in der Liste der Vergleichsergebnisse angezeigt ist, wird das verbleibende Vergleichsergebnis als neue Basis festgelegt und aktualisiert das System die Registerkarte „Zusammenfassung“, sodass nur die neuen Basisdaten angezeigt werden.

Beispiel: Verwenden von Match Analysis

Dieses Beispiel zeigt, wie das „Match Analysis“-Tool verwendet wird, um die An- und Abstiegswahrscheinlichkeiten von zwei unterschiedlichen Übereinstimmungen zu vergleichen. Bevor die Daten an das Vergleichstool gesendet werden, werden sie mithilfe eines Broadcasters in zwei Streams aufgeteilt. Jeder Stream wird anschließend durch einen „Intraflow Match“-Schritt geschickt. Jeder Datenstream enthält identische Kopien der verarbeiteten Daten. Jeder „Intraflow Match“-Schritt verwendet einen anderen Vergleichsalgorithmus und generiert „Match Analysis“-Daten, die Sie zum Vergleich des An-/Abstiegs von verschiedenen Übereinstimmungen nutzen können.

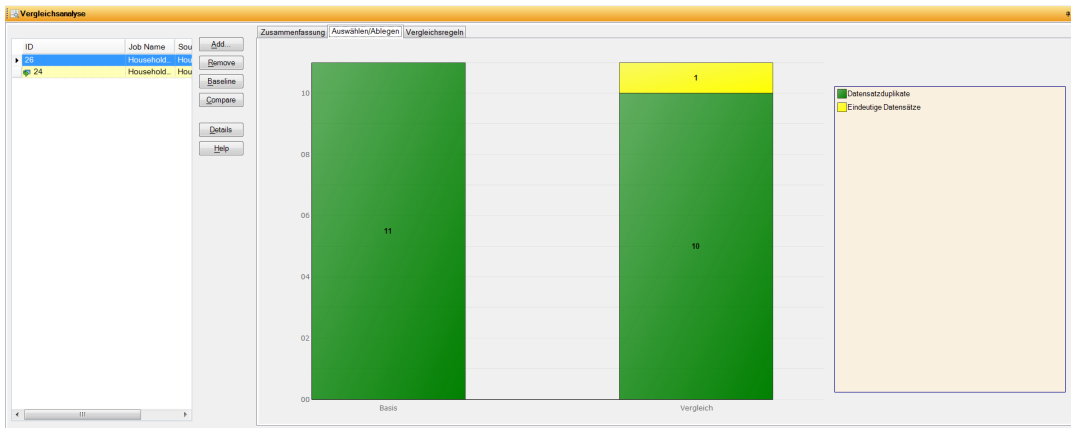


Dieser Beispieldatenfluss steht im Enterprise Designer zur Verfügung. Öffnen Sie **Datei > Neu > Datenfluss > Aus Vorlage**, und wählen Sie **HouseholdRelationshipsAnalysis** aus. Dieser Datenfluss erfordert die folgenden Module: Advanced Matching-Modul, Data Normalization-Modul und Universal Name-Modul. Er erfordert außerdem das Laden der „Table Lookup“-Kerndatenbank und der „Open Parser“-Basistabellen.

Schauen Sie sich zur korrekten Anwendung das Beispiel an:

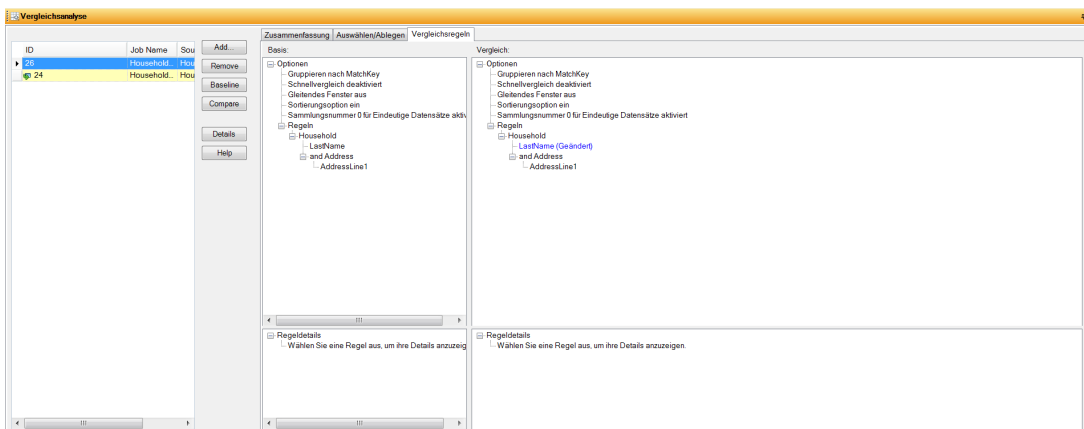
1. Führen Sie den Datenfluss aus.
2. Wählen Sie **Tools > Match Analysis** aus.
3. Erweitern Sie im Fenster **Vergleichsergebnisse durchsuchen** das Element **HouseholdRelationshipAnalysis**, und wählen Sie **Household Match 1** und **Household Match 2** aus der Liste „Quelle“ aus. Klicken Sie anschließend auf **Hinzufügen**.
4. Wählen Sie aus der Liste der Vergleichsergebnisse den Eintrag **Household Match 1** aus, und klicken Sie auf **Vergleichen**. Die Zusammenfassungsergebnisse werden angezeigt.

5. Klicken Sie auf die Registerkarte **An-/Abstieg**. Das Diagramm „An-/Abstieg“ wird angezeigt.



Dieses Diagramm zeigt die Unterschiede zwischen Dubletten- und eindeutigen Datensätzen, die für die unterschiedlichen verwendeten Vergleichsregeln erstellt wurden.

6. Klicken Sie auf die Registerkarte **Vergleichsregeln**. Der Vergleich der Vergleichsregeln wird angezeigt.



Auf dieser Registerkarte sehen Sie, dass der Algorithmus geändert wurde. Die Zeichenhäufigkeit wurde ausgelassen und „Genauere Übereinstimmung“ wurde hinzugefügt.

7. Klicken Sie auf **Details**.
8. Wählen Sie **Dublettensammlungen** aus der Anzeigeliste aus, und klicken Sie auf **Aktualisieren**.
9. Erweitern Sie jedes **CollectionNumber**, um die Kopfdubletten- und Dublettendatensätze für jede Dublettensammlung anzuzeigen.

| MatchRecordType | MatchGroup | InputRecordNumber | CollectionNumber | LastName | AddressLine1 |
|-----------------|------------|-------------------|------------------|--------------|--------------------|
| Suspect | G20706 | 5 | 1 | Gressmanelli | 4200 Parliament... |
| Duplicate | G20706 | 6 | 1 | Gressmanelli | 4200 Parliament... |
| Duplicate | G20706 | 11 | 1 | Gressmanelli | 4200 Parliament... |
| Unique | G20706 | 10 | 0 | Gressmanelli | 4200 Parliament... |
| Suspect | J20612 | 7 | 2 | Jones | PO Box 263 |
| Duplicate | J20612 | 8 | 2 | Jones | PO Box 263 |
| Duplicate | J20612 | 9 | 2 | Jones | PO Box 263 |
| Suspect | S20657 | 1 | 3 | Smith | 12643 Rousby H... |
| Duplicate | S20657 | 2 | 3 | Smith | 12643 Rousby Ha... |
| Duplicate | S20657 | 3 | 3 | Smith | 12643 Rousby H... |
| Duplicate | S20657 | 4 | 3 | Smith | 12643 Rousby H... |

10. Vergleichen Sie die Sammlungen in der Ansicht „Detail“ mit der erstellten Ausgabedatei.

Datenflussvorlagen zum Vergleichen

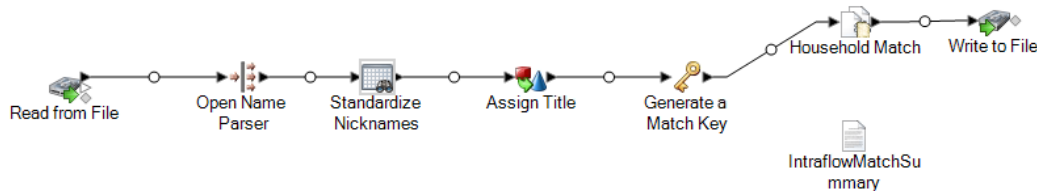
Identifizieren von Mitgliedern eines Haushalts

Diese Datenflussvorlage zeigt, wie Sie Mitglieder desselben Haushalts identifizieren, indem Sie Informationen innerhalb einer einzelnen Eingabedatei vergleichen und eine Ausgabedatei für die Haushaltssammlungen erstellen.

Geschäftsszenario

Als Data Steward in einer Kreditkartengesellschaft möchten Sie Ihre Kundendatenbank analysieren und herausfinden, welche Adressen mehrmals vorhanden sind und welche Namen dort verzeichnet sind, damit Sie die Anzahl der doppelten Zustellungen und Kreditkartenangebote, die an dieselbe Adresse gesendet werden, minimieren können.

Im folgenden Datenfluss wird eine Lösung für das Geschäftsszenario bereitgestellt:



Diese Datenflussvorlage ist im Enterprise Designer verfügbar. Öffnen Sie **Datei > Neu > Datenfluss > Aus Vorlage**, und wählen Sie **HouseholdRelationships** aus. Dieser Datenfluss erfordert die folgenden Module: Advanced Matching-Modul, Data Normalization-Modul und Universal Name-Modul.

Für jeden Datensatz in der Eingabedatei führt dieser Datenfluss Folgendes aus:

Read from File

Bei diesem Schritt werden der Dateiname, der Speicherort und das Layout der Datei identifiziert, welche die zu parsenden Namen enthält. Die Datei enthält männliche und weibliche Namen.

Open Name Parser

„Open Name Parser“ untersucht Namensfelder und vergleicht sie mit den Namensdaten, die in den Namensdatenbankdateien von Spectrum™ Technology Platform gespeichert sind. Basierend auf dem Vergleich parst er die Namensdaten in den Feldern „Vorname“, „Zweiter Vorname“ und „Nachname“ und weist jedem Namen einen Entitätstyp sowie ein Geschlecht zu. Er verwendet neben den Namensdaten außerdem die Mustererkennung.

Standardize Nicknames

In dieser Vorlage wird der „Table Lookup“-Schritt mit „Standardize Nicknames“ benannt: Der „Standardize Nicknames“-Schritt sucht zunächst nach Vornamen in der „Nicknames.xml“-Datenbank und ersetzt alle Spitznamen durch eine regulärere Form des Spitznamens. Der Name Tommy wird z. B. durch Thomas ersetzt.

Transformer

In dieser Vorlage wird der Transformer-Schritt mit „Assign Titles“ benannt: Der „Assign Titles“-Schritt verwendet ein benutzerdefiniertes Skript, um jede Zeile in der Datenstromausgabe mithilfe des „Parse Personal Name“-Schrittes zu durchsuchen und einen **TitleOfRespect**-Wert basierend auf dem **GenderCode**-Wert zuzuweisen.

Das benutzerdefinierte Skript lautet wie folgt:

```
if (row.get('TitleOfRespect') == '')
{
  if (row.get('GenderCode') == 'M')
    row.set('TitleOfRespect', 'Mr')
  if (row.get('GenderCode') == 'F')
    row.set('TitleOfRespect', 'Ms')
```

Jedes Mal, wenn der „Assign Titles“-Schritt den Buchstaben **M** im Feld **GenderCode** erkennt, wird der Wert für **TitleOfRespect** auf **Mr** (Herr) eingestellt. Jedes Mal, wenn der „Assign Titles“-Schritt den Buchstaben **F** im Feld **GenderCode** erkennt, wird der Wert für **TitleOfRespect** auf **Ms** (Frau) eingestellt.

Match Key Generator

Der Match Key Generator verarbeitet benutzerdefinierte Regeln, die aus Algorithmen und Eingabequellenfeldern bestehen, um das Vergleichsschlüsselfeld zu generieren. Ein Vergleichsschlüssel ist ein nicht eindeutiger Schlüssel, der von ähnlichen Datensätzen geteilt wird, die Datensätze als potenzielle Dubletten identifizieren. Der Vergleichsschlüssel wird zur Vereinfachung des Vergleichsprozesses verwendet, indem nur die Datensätze miteinander verglichen

werden, die den gleichen Vergleichsschlüssel enthalten. Ein Vergleichsschlüssel besteht aus Eingabefeldern. Jedes angegebene Eingabefeld verfügt über einen ausgewählten Algorithmus, der dafür ausgeführt wird. Das Ergebnis jedes Feldes wird dann verkettet, um ein einzelnes Vergleichsschlüsselfeld zu erstellen.

In dieser Vorlage sind zwei Vergleichsschlüsselfelder definiert: SubString (LastName (1:3)) und SubString (PostalCode (1:5)).

Beispiel: Wenn die eingehende Adresse wie folgt lauten würde:

FirstName – Fred

LastName – Mertz

PostalCode – 21114-1687

Und die Regeln Folgendes festlegen würden:

| Eingabefeld | Startposition | Länge |
|-------------|---------------|-------|
| LastName | 1 | 3 |
| PostalCode | 1 | 5 |

Dann würde der Schlüssel basierend auf den oben angeführten Regeln und Eingabedaten wie folgt lauten:

Mer21114

Household Match

In dieser Datenflussvorlage wird der „Intraflow Match“-Schritt als „Household Match“ benannt. Dieser Schritt ermittelt Übereinstimmungen zwischen ähnlichen Datensätzen innerhalb eines einzelnen Eingabestreams. Vergleichene Datensätze können ebenfalls mithilfe von Nicht-Name-/Nicht-Adresse-Informationen begrenzt werden. Die Vergleichsengine ermöglicht die Erstellung von hierarchischen Regeln basierend auf beliebigen Feldern, die in anderen Schritten definiert oder erstellt wurden.

Ein zu vergleichender Datensatz-Stream sowie Einstellungen, die angeben, welche Felder verglichen werden sollten, wie Punktzahlen bewertet werden sollten und aus was im Allgemeinen eine erfolgreiche Übereinstimmung besteht.

Mit dieser Vorlage erstellen Sie eine benutzerdefinierte Vergleichsregel, die „LastName“ und „AddressLine1“ vergleicht. Aktivieren Sie das Kontrollkästchen **Daten für Analyse generieren**, um Daten für den Interflow-Zusammenfassungsbericht zu generieren.

Hier sind einige Richtlinien zur Erstellung Ihrer Vergleichshierarchie:

- Ein übergeordneter Knoten muss einen eindeutigen Namen haben. Es darf kein Feld sein.
- Das untergeordnete Feld muss ein Spectrum™ Technology Platform-Datentypfeld sein. Das heißt, dass es durch eine oder mehrere Komponenten verfügbar ist.
- Alle untergeordneten Elemente unter einem übergeordneten Element müssen dieselben logischen Operatoren verwenden. Zur Kombination von Connectors müssen Sie zunächst übergeordnete Zwischenknoten erstellen.
- Schwellenwerte übergeordneter Elemente können nicht größer sein als die der untergeordneten Elemente.
- Übergeordnete Knoten müssen keinen Schwellenwert aufweisen.

Write to File

Die Vorlage enthält einen „Write to File“-Schritt, der eine Textdatei erstellt, die die Adressen als eine Sammlung von Haushalten zeigt.

Intraflow-Zusammenfassungsbericht

Die Vorlage enthält den „Intraflow Match“-Zusammenfassungsbericht. Erweitern Sie nach Ausführung des Auftrags die Option **Berichte** im Fenster Ausführungsdetails, und klicken Sie auf **IntraflowMatchSummary**.

Der „Intraflow Match“-Zusammenfassungsbericht enthält die Statistik für die verarbeiteten Datensätze und zeigt ein Balkendiagramm, das die Datensatzanzahl und die Gesamtpunktzahl der Übereinstimmungen darstellt.

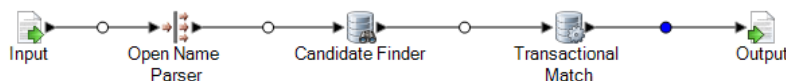
Bestimmen, ob ein Interessent ein Kunde ist

Diese Datenflussvorlage zeigt, wie Interessentendaten einer Eingabedatei für Kundendaten in einer Kundendatenbank bewertet werden, um zu bestimmen, ob ein Interessent ein Kunde ist. Dies ist ein Dienstdatenfluss, d. h. dass der Datenfluss über die API oder Webservices zugänglich ist.

Geschäftsszenario

Als Verkaufsleiter eines Onlinehandel-Unternehmens möchten Sie herausfinden, ob ein potenzieller Onlinekunde ein Bestandskunde oder ein Neukunde ist.

Der folgende Datenflussdienst bietet eine Lösung für das Geschäftsszenario:



Diese Datenflussvorlage ist im Enterprise Designer verfügbar. Öffnen Sie **Datei > Neu > Datenfluss > Aus Vorlage**, und wählen Sie **ProspectMatching** aus. Dieser Datenfluss erfordert das Advanced Matching-Modul und das Universal Name-Modul.

Für jeden Datensatz in der Eingabedatei führt dieser Datenfluss Folgendes aus:

Eingabe

Die ausgewählten Eingabefelder für diese Vorlage sind „AddressLine1“, „City“, „Name“, „PostalCode“ und „StateProvince“. „AddressLine1“ und „Name“ sind die Felder, die für die Datenflussverarbeitung in dieser Vorlage entscheidend sind.

Name Parser

In dieser Vorlage ist der „Name Parser“-Schritt als „Parse Personal Name“ benannt. Der „Parse Personal Name“-Schritt untersucht Namensfelder und vergleicht sie mit den Namensdaten, die in den Namensdatenbankdateien von Spectrum™ Technology Platform gespeichert sind. Basierend auf dem Vergleich parst er die Namensdaten in den Feldern „Vorname“, „Zweiter Vorname“ und „Nachname“ und weist jedem Namen einen Entitätstyp sowie ein Geschlecht zu. Er verwendet neben den Namensdaten außerdem die Mustererkennung.

In dieser Vorlage wird der „Parse Personal Name“-Schritt wie folgt konfiguriert.

- „Personennamen parsen“ wird ausgewählt und „Unternehmensnamen parsen“ wird gelöscht. Bei der Auswahl dieser Optionen werden Vornamen nach Geschlecht, Reihenfolge und Interpunktion ausgewertet und es wird keine Bewertung der Unternehmensnamen durchgeführt.
- „Quelle für Geschlechtsbestimmung“ ist auf „Standard“ eingestellt. In den meisten Fällen ist „Standard“ die beste Einstellung für die Geschlechtsbestimmung, da mit ihr ein breites Spektrum von Namen abgedeckt wird. Wenn Sie jedoch Namen aus einer bestimmten Kultur verarbeiten, wählen Sie diese Kultur aus. Die Auswahl einer bestimmten Kultur hilft bei der Sicherstellung, dass den Namen das richtige Geschlecht zugewiesen wird. Beispiel: Wenn Sie „Standard“ aktiviert lassen, wird der Name Jean als weiblicher Name identifiziert. Wenn Sie jedoch Französisch auswählen, wird der Name als männlicher Name identifiziert.
- „Reihenfolge“ ist auf „Natürlich“ eingestellt. Die Namensfelder sind nach Titel, Vorname, zweiter Vorname, Nachname und Suffix sortiert.
- „Punkte beibehalten“ ist deaktiviert. Das heißt, dass keinerlei Interpunktion in den Namensdaten beibehalten wird.

Candidate Finder

Der „Candidate Finder“-Schritt wird zur Kombination mit dem „Transactional Match“-Schritt verwendet.

Im „Candidate Finder“-Schritt werden die Kandidatendatensätze abgerufen, die die Gruppe potenzieller Übereinstimmungen bilden, die im „Transactional Match“-Schritt ausgewertet wird. Zusätzlich muss je nach Format Ihrer Daten „Candidate Finder“ gegebenenfalls auch den Namen oder die Adresse des Kopfdubletten-Datensatzes, der Kandidatendatensätze oder beides parsen.

Wählen Sie als Teil der Konfiguration von „Candidate Finder“ die Datenbankverbindung aus, über die die angegebene Abfrage ausgeführt werden soll. Sie können eine beliebige, auf der Management Console konfigurierte Verbindung auswählen. Um eine nicht aufgeführte Datenbank zu verbinden, müssen Sie in der Management Console eine Verbindung zu dieser Datenbank konfigurieren, und anschließend „Candidate Finder“ schließen und erneut öffnen, um die Liste der Verbindungen zu aktualisieren.

Zur Definition einer SQL-Abfrage können Sie eine beliebige gültige SQL SELECT-Anweisung in das Textfeld im Dialogfeld „Candidate Finder“-Optionen“ eingeben. Angenommen, Sie haben eine Tabelle mit Namen „Customer_Table“ in Ihrer Datenbank, die die folgenden Spalten aufweist:

Customer_Table

Cust_Name

Cust_Address

Cust_City

Cust_State

Cust_Zip

Anmerkung: Sie können jede gültige SQL-Auswahl eingeben, `Select *` ist jedoch in dieser Steuerung ungültig.

Um alle Zeilen aus der Datenbank abzurufen, könnten Sie eine Abfrage ähnlich der folgenden erstellen:

```
select Cust_Name, Cust_Address, Cust_City, Cust_State, Cust_Zip from
Customer_Table;
```

Es ist jedoch unwahrscheinlich, dass Sie Ihre Transaktion mit allen Zeilen der Datenbank vergleichen möchten. Um nur relevante Kandidatendatensätze zurückzugeben, müssten Sie eine WHERE-Klausel mithilfe einer Variablenersetzung hinzufügen. Eine Variablenersetzung ist eine besondere Notation, mit der die Candidate Selection-Engine die Variable mit den tatsächlichen Daten aus Ihrem Kopfdubletten-Datensatz ersetzt.

Um die Variablenersetzung zu verwenden, schließen Sie den Feldnamen in geschweiften Klammern ein, denen Sie ein Dollarzeichen voranstellen: **#{Feldname}**. Beispiel: Die folgende Abfrage gibt nur Datensätze mit einem Wert in „Cust_Zip“ zurück, der mit dem Wert in „PostalCode“ im Kopfdubletten-Datensatz übereinstimmt.

```
select Cust_Name, Cust_Address, Cust_City, Cust_State, Cust_Zip
from Customer_Table
where Cust_Zip = #{PostalCode};
```

Als nächstes müssen Sie die Datenbankspalten den Schrittfeldern zuordnen, wenn die Spaltennamen in Ihrer Datenbank nicht genau mit den „Component Field“-Namen übereinstimmen. Wenn sie nicht übereinstimmen, werden sie automatisch den entsprechenden Schrittfeldern zugeordnet. Sie müssen „Ausgewählte Felder“ (Spalten aus der Datenbank) verwenden, um sie den Schrittfeldern (im Datenfluss definierte Feldnamen) zuzuordnen.

Berücksichtigen Sie erneut „Customer_Table“ aus dem obigen Beispiel:

Customer_Table

Cust_Name

Cust_Address

Cust_City

Cust_State

Cust_Zip

Wenn Sie diese Datensätze aus der Datenbank abrufen, müssen Sie die Spaltennamen den Feldnamen zuordnen, die durch den „Transactional Match“-Schritt und andere Schritte in Ihrem Datenfluss verwendet werden. So könnte beispielsweise „Cust_Address“ „AddressLine1“ und „Cust_Zip“ „PostalCode“ zugeordnet werden.

1. Wählen Sie in der Ansicht „Candidate Finder“-Optionen“ die Dropdown-Liste unter **Ausgewählte Felder** aus. Wählen Sie danach die Datenbankspalte „Cust_Zip“ aus.
2. Wählen Sie die Dropdown-Liste unter **Schrittfelder** aus. Wählen Sie danach das Feld aus, das Sie zuordnen möchten.

Wenn Sie beispielsweise „Cust_Zip“ „Postal Code“ zuordnen möchten, wählen Sie zuerst „Cust_Zip“ unter „Ausgewählte Felder“ und dann „PostalCode“ in der entsprechenden Zeile „Schrittfeld“ aus.

Zusätzlich zum Zuordnen von Feldern wie oben beschrieben können Sie die spezielle Notation in Ihrer SQL-Abfrage nutzen, um die Zuordnung durchzuführen. Geben Sie dazu den Namen des Schrittfeldes in Klammern nach dem Spaltennamen in die Abfrage ein. Dabei werden die ausgewählten Felder automatisch den entsprechenden Schrittfeldern zugeordnet.

Hier folgt ein Beispiel unter Anwendung der Abfrage aus dem vorherigen Beispiel:

```
select Cust_Name {Name}, Cust_Address {AddressLine1},
       Cust_City {City}, Cust_State {StateProvince},
       Cust_Zip {PostalCode}
from Customer
where Cust_Zip = ${PostalCode};
```

Transactional Match

Der „Transactional Match“-Schritt wird zur Kombination mit dem „Candidate Finder“-Schritt verwendet.

Mit dem „Transactional Match“-Schritt können Sie Kopfdoubletten-Datensätze mit Kandidatendatensätzen vergleichen, die vom „Candidate Finder“-Schritt zurückgegeben werden.

„Transactional Match“ verwendet Vergleichsregeln, um den Kopfdoubletten-Datensatz mit allen Kandidatendatensätzen mit derselben (durch Candidate Finder zugewiesenen) Kandidatengruppennummer zu vergleichen, um Doubletten zu finden. Wenn es sich bei dem Kandidatendatensatz um eine Doublette handelt, wird ihm eine Sammlungsnummer zugewiesen, der übereinstimmende Datensatztyp wird als eine Doublette gekennzeichnet und der Datensatz wird anschließend ausgeschrieben. Allen Kandidaten in der Gruppe ohne Übereinstimmung wird die Sammlungsnummer 0 zugewiesen und sie werden als eindeutig markiert und ebenfalls ausgeschrieben.

Mit dieser Vorlage erstellen Sie eine benutzerdefinierte Vergleichsregel, die „LastName“ und „AddressLine1“ vergleicht.

Hier sind einige Richtlinien zur Erstellung Ihrer Vergleichshierarchie:

- Ein übergeordneter Knoten muss einen eindeutigen Namen haben. Es darf kein Feld sein.
- Das untergeordnete Feld muss ein Spectrum™ Technology Platform-Datentypfeld sein. Das heißt, das es durch eine oder mehrere Schritte verfügbar ist.
- Alle untergeordneten Elemente unter einem übergeordneten Element müssen dieselben logischen Operatoren verwenden. Zur Kombination von Connectors müssen Sie zunächst übergeordnete Zwischenknoten erstellen.
- Schwellenwerte übergeordneter Elemente können nicht größer sein als die der untergeordneten Elemente.
- Übergeordnete Knoten müssen keinen Schwellenwert aufweisen.

Ausgabe

Als Dienst schickt diese Vorlage alle verfügbaren Felder zur Ausgabe. Sie können die Ausgabe basierend auf Ihren Bedürfnissen eingrenzen.

5 - Deduplizierung

In this section

| | |
|---|-----|
| Herausfiltern von Datensatzdubletten | 147 |
| Erstellen eines „Best of Breed“-Datensatzes | 152 |

Herausfiltern von Datensatzdubletten

Sie entfernen Datensatzdubletten am einfachsten, wenn Sie nach einem Vergleichsschritt einen Filter-Schritt zu Ihrem Datenfluss hinzufügen. Der Filter-Schritt entfernt Datensätze aus Sammlungen mit Datensatzdubletten basierend auf den von Ihnen festgelegten Einstellungen.

1. Erstellen Sie im Enterprise Designer einen Datenfluss, der doppelte Datensätze per Vergleich identifiziert.

Der Vergleich ist der erste Schritt der Deduplizierung, weil Sie Datensätze identifizieren müssen, die ähnlich sind, etwa Datensätze mit derselben Kontonummer oder demselben Namen. Die folgenden Themen bieten eine Anleitung zum Erstellen eines Datenflusses, mit dem Datensätze verglichen werden.

[Vergleichen von Datensätzen aus einer einzelnen Quelle](#) auf Seite 87

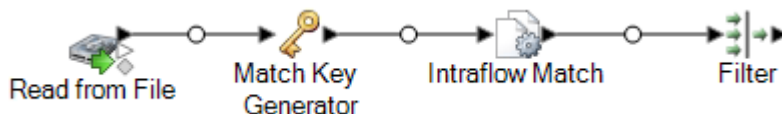
[Vergleichen von Datensätzen einer Quelle mit Datensätzen einer anderen Quelle](#) auf Seite 93

[Vergleichen von Datensätzen mit einer Datenbank](#) auf Seite 106

Anmerkung: Sie müssen den Datenfluss nur bis zu dem Punkt erstellen, an dem er Daten liest und mithilfe eines „Interflow Match“- , „Intraflow Match“- oder „Transactional Match“-Schrittes einen Vergleich durchführt. Sobald Sie einen Datenfluss bis zu diesem Punkt erstellt haben, fahren Sie mit den folgenden Schritten fort.

2. Wenn Sie einen Datenfluss definiert haben, der Daten liest und Datensätze vergleicht, ziehen Sie einen Filter-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem Schritt, der den Vergleich durchführt (Interflow Match, Intraflow Match oder Transactional Match).

Wenn Ihr Datenfluss beispielsweise Daten aus einer Datei liest und einen Vergleich mithilfe von Intraflow Match durchführt, würde Ihr Datenfluss nach dem Hinzufügen eines Filter-Schrittes wie folgt aussehen:



3. Doppelklicken Sie in der Arbeitsfläche auf den Filter-Schritt.
4. Wählen Sie im **Gruppierungsfeld** den Eintrag **CollectionNumber** aus.
5. Lassen Sie die Option **Anzahl zurückgegebener doppelter Datensätze beschränken** ausgewählt und den Wert auf 1 gesetzt. Dies sind die Standardeinstellungen.
6. Entscheiden Sie, ob Sie den ersten Datensatz in jeder Sammlung beibehalten möchten oder ob Sie eine Regel definieren möchten, um festzulegen, welcher Datensatz aus jeder Sammlung beibehalten werden soll. Wenn Sie den ersten Datensatz in jeder Sammlung behalten möchten,

überspringen Sie diesen Schritt. Wenn Sie eine Regel definieren möchten, wählen Sie im Regelbaum das Element **Regeln** aus, und gehen Sie folgendermaßen vor:

a) Klicken Sie auf **Regel hinzufügen**.

Datensätze in jeder Gruppe werden dahingehend überprüft, ob sie die Regeln erfüllen, die Sie hier definieren. Wenn ein Datensatz die Regel erfüllt, ist er der „überlebende“ Datensatz und die anderen Datensätze in der Gruppe werden verworfen.

b) Definieren Sie eine Regel, um den beizubehaltenden Datensatz aus jeder Gruppe zu identifizieren.

Verwenden Sie die folgenden Optionen, um eine Regel zu definieren:

| Option | Beschreibung |
|----------|--|
| Feldname | Gibt den Namen des Datenflussfeldes an, dessen Wert Sie auswerten möchten, um zu bestimmen, ob der Datensatz gefiltert werden soll. |
| Feldtyp | Gibt den Datentyp im Feld an. Zur Auswahl stehen: <ul style="list-style-type: none"> Nicht numerisch Wählen Sie diese Option aus, wenn das Feld nicht numerische Daten enthält (z. B. string-Daten). Numerisch Wählen Sie diese Option aus, wenn das Feld numerische Daten enthält (z. B. double, float usw.). |

| Option | Beschreibung |
|--------|--------------|
|--------|--------------|

| | |
|----------|--|
| Operator | |
|----------|--|

| Option | Beschreibung |
|----------------------------|---|
| | Gibt den Vergleichstyp an, den Sie zur Auswertung des Feldes verwenden möchten. Zur Auswahl stehen: |
| Enthält | Bestimmt, ob das Feld den angegebenen Wert enthält. Zum Beispiel enthält „Segelboot“ den Wert „boot“. |
| Gleich | Bestimmt, ob das Feld exakt den angegebenen Wert enthält. |
| Größer als | Bestimmt, ob der Feldwert größer als der angegebene Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Größer oder gleich | Bestimmt, ob der Feldwert größer als oder gleich dem angegebenen Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Höchste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den höchsten Wert im Feld aufweist. Wenn die Felder in der Gruppe beispielsweise Werte von 10, 20, 30 und 100 enthalten, würde der Datensatz mit dem Feldwert 100 ausgewählt werden. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Ist leer | Bestimmt, ob das Feld keinen Wert enthält. |
| Ist nicht leer | Bestimmt, ob das Feld einen Wert enthält. |
| Kleiner als | Bestimmt, ob der Feldwert kleiner als der angegebene Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Kleiner oder gleich | Bestimmt, ob der Feldwert kleiner als oder gleich dem angegebenen Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Längste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den längsten Wert (in Bytes) im Feld aufweist. Wenn die Gruppe beispielsweise die Werte „Mike“ und „Michael“ enthält, wird der Datensatz mit dem Wert „Michael“ ausgewählt. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Niedrigste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den niedrigsten Wert im Feld aufweist. Wenn die Felder in der Gruppe beispielsweise Werte von 10, 20, 30 und 100 enthalten, würde der Datensatz mit dem Feldwert 10 ausgewählt werden. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Häufigste | Bestimmt, ob der Feldwert den Wert enthält, der in diesem Feld unter den Datensätzen in der Gruppe am häufigsten auftritt. Wenn zwei oder mehr Werte am häufigsten auftreten, werden keine Aktionen durchgeführt. |

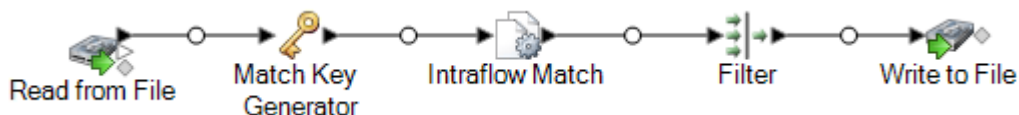
| Option | Beschreibung |
|---------|---|
| | <p>Ungleich Bestimmt, ob der Feldwert nicht mit dem angegebenen Wert übereinstimmt.</p> |
| Werttyp | <p>Gibt den Werttyp an, den Sie mit dem Wert des Feldes vergleichen möchten. Zur Auswahl stehen:</p> <p>Anmerkung: Diese Option ist nicht verfügbar, wenn Sie den Operator „Höchste“, „Niedrigste“ oder „Längste“ auswählen.</p> <p>Feld Wählen Sie diese Option, wenn Sie den Wert eines anderen Datenflussfeldes mit dem Feld vergleichen möchten.</p> <p>Zeichenfolge Wählen Sie diese Option, wenn Sie das Feld mit einem spezifischen Wert vergleichen möchten.</p> |
| Wert | <p>Gibt den Wert an, der mit dem Wert des Feldes verglichen werden soll. Wenn Sie im Feld Feldtyp den Eintrag „Feld“ ausgewählt haben, müssen Sie ein Datenflussfeld auswählen. Wenn Sie im Feld Werttyp den Eintrag „Zeichenfolge“ ausgewählt haben, müssen Sie den bei dem Vergleich zu verwendenden Wert eingeben.</p> <p>Anmerkung: Diese Option ist nicht verfügbar, wenn Sie den Operator „Höchste“, „Niedrigste“ oder „Längste“ auswählen.</p> |

c) Klicken Sie auf **OK**.

Sie haben „Filter“ nun mit einer Regel konfiguriert. Sie können bei Bedarf zusätzliche Regeln hinzufügen.

- Klicken Sie auf **OK**, um das Fenster **Filteroptionen** zu schließen.
- Ziehen Sie einen Datenladeschritt auf die Arbeitsfläche und verbinden Sie ihn mit dem Filter-Schritt.

Bei Verwendung eines „Write to File“-Datenladeschrittes würde Ihr Datenfluss so aussehen:



- Doppelklicken Sie auf den Datenladeschritt und konfigurieren Sie ihn.

Informationen zum Konfigurieren von Datenladeschritten finden Sie im *Datenfluss-Designer-Handbuch*.

Sie verfügen nun über einen Datenfluss, der die übereinstimmenden Datensätze identifiziert und alle Datensätze der Dublettengruppe mit Ausnahme von einem Datensatz entfernt. Dies führt zu einer Ausgabedatei, die Daten ohne Dubletten enthält.

Erstellen eines „Best of Breed“-Datensatzes

Um doppelte Datensätze aus Ihren Daten zu entfernen, können Sie Daten aus Gruppen doppelter Datensätze in einem einzelnen „Best of Breed“-Datensatz zusammenführen. Dieser Ansatz ist nützlich, wenn die einzelnen Datensatzdubletten Daten des gleichen Typs enthalten (z. B. Telefonnummern oder Namen) und Sie die besten Daten aus den einzelnen Datensätzen in dem erhalten gebliebenen Datensatz beibehalten möchten.

Diese Schritte beschreiben, wie ein Datenfluss erstellt wird, der doppelte Datensätze in einem Best of Breed-Datensatz zusammenführt.

1. Erstellen Sie im Enterprise Designer einen Datenfluss, der doppelte Datensätze per Vergleich identifiziert.

Der Vergleich ist der erste Schritt der Deduplizierung, weil Sie Datensätze identifizieren müssen, die ähnlich sind, etwa Datensätze mit derselben Kontonummer oder demselben Namen. Die folgenden Themen bieten eine Anleitung zum Erstellen eines Datenflusses, mit dem Datensätze verglichen werden.

[Vergleichen von Datensätzen aus einer einzelnen Quelle](#) auf Seite 87

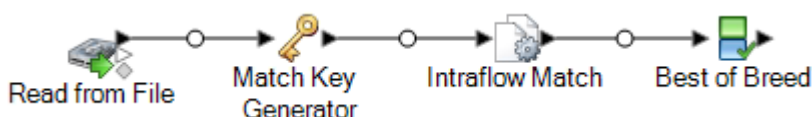
[Vergleichen von Datensätzen einer Quelle mit Datensätzen einer anderen Quelle](#) auf Seite 93

[Vergleichen von Datensätzen mit einer Datenbank](#) auf Seite 106

Anmerkung: Sie müssen den Datenfluss nur bis zu dem Punkt erstellen, an dem er Daten liest und mithilfe eines „Interflow Match“- , „Intraflow Match“- oder „Transactional Match“-Schrittes einen Vergleich durchführt. Sobald Sie einen Datenfluss bis zu diesem Punkt erstellt haben, fahren Sie mit den folgenden Schritten fort.

2. Sobald Sie einen Datenfluss definiert haben, der Daten liest und Datensätze vergleicht, ziehen Sie einen „Best of Breed“-Schritt auf die Arbeitsfläche. Verbinden Sie ihn mit dem Schritt, der den Vergleich durchführt (Interflow Match, Intraflow Match oder Transactional Match).

Wenn Ihr Datenfluss beispielsweise Daten aus einer Datei liest und den Vergleich mit Intraflow Match durchführt, sieht Ihr Datenfluss nach Hinzufügen eines „Best of Breed“-Schrittes so aus:



3. Doppelklicken Sie auf den „Best of Breed“-Schritt auf der Arbeitsfläche.

4. Wählen Sie im Feld **Gruppieren nach CollectionNumber** aus.
5. Wählen Sie unter **Best-of-Breed-Einstellungen Regeln** im Bedingungsbaum aus.
6. Klicken Sie auf **Regel hinzufügen**.

Datensätze in jeder Gruppe werden ausgewertet, um festzustellen, ob sie die hier definierten Regeln erfüllen. Wenn ein Datensatz einer Regel entspricht, können seine Daten in den Best of Breed-Datensatz kopiert werden. Dies hängt davon ab, wie Sie die der Regel zugeordneten Aktionen konfigurieren. Sie definieren diese Aktionen später.

7. Definieren Sie eine Regel, die ein doppelter Datensatz erfüllen muss, damit seine Daten in den Best of Breed-Datensatz kopiert werden.

Definieren Sie die Regel anhand der folgenden Optionen:

| Option | Beschreibung |
|----------|--|
| Feldname | Gibt den Namen des Datenflussfeldes mit dem auszuwertenden Wert an, um zu bestimmen, ob die Bedingung erfüllt ist und die dazugehörigen Aktionen ausgeführt werden sollen. |
| Feldtyp | Gibt den Datentyp im Feld an. Zur Auswahl stehen: <ul style="list-style-type: none"> Nicht numerisch Wählen Sie diese Option aus, wenn das Feld nicht numerische Daten enthält (z. B. string-Daten). Numerisch Wählen Sie diese Option aus, wenn das Feld numerische Daten enthält (z. B. double, float usw.). |

| Option | Beschreibung |
|----------------------------|---|
| Operator | Gibt den Vergleichstyp an, den Sie zur Auswertung des Feldes verwenden möchten. Zur Auswahl stehen: |
| Enthält | Bestimmt, ob das Feld den angegebenen Wert enthält. Zum Beispiel enthält „Segelboot“ den Wert „boot“. |
| Gleich | Bestimmt, ob das Feld exakt den angegebenen Wert enthält. |
| Größer als | Bestimmt, ob der Feldwert größer als der angegebene Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Größer oder gleich | Bestimmt, ob der Feldwert größer als oder gleich dem angegebenen Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Höchste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den höchsten Wert im Feld aufweist. Wenn die Felder in der Gruppe beispielsweise Werte von 10, 20, 30 und 100 enthalten, würde der Datensatz mit dem Feldwert 100 ausgewählt werden. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Ist leer | Bestimmt, ob das Feld keinen Wert enthält. |
| Ist nicht leer | Bestimmt, ob das Feld einen Wert enthält. |
| Kleiner als | Bestimmt, ob der Feldwert kleiner als der angegebene Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Kleiner oder gleich | Bestimmt, ob der Feldwert kleiner als oder gleich dem angegebenen Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Längste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den längsten Wert (in Bytes) im Feld aufweist. Wenn die Gruppe beispielsweise die Werte „Mike“ und „Michael“ enthält, wird der Datensatz mit dem Wert „Michael“ ausgewählt. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Niedrigste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den niedrigsten Wert im Feld aufweist. Wenn die Felder in der Gruppe beispielsweise Werte von 10, 20, 30 und 100 enthalten, würde der Datensatz mit dem Feldwert 10 ausgewählt werden. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Häufigste | Bestimmt, ob der Feldwert den Wert enthält, der in diesem Feld unter den Datensätzen in der Gruppe am häufigsten auftritt. Wenn zwei oder mehr Werte am häufigsten auftreten, werden keine Aktionen durchgeführt. |
| Ungleich | Bestimmt, ob der Feldwert nicht mit dem angegebenen Wert übereinstimmt. |

| Option | Beschreibung |
|---------|---|
| Werttyp | <p>Gibt den Werttyp an, den Sie mit dem Wert des Feldes vergleichen möchten. Zur Auswahl stehen:</p> <p>Anmerkung: Diese Option ist nicht verfügbar, wenn Sie den Operator „Höchste“, „Niedrigste“ oder „Längste“ auswählen.</p> <p>Feld Wählen Sie diese Option, wenn Sie den Wert eines anderen Datenflussfeldes mit dem Feld vergleichen möchten.</p> <p>Zeichenfolge Wählen Sie diese Option, wenn Sie das Feld mit einem spezifischen Wert vergleichen möchten.</p> |

| | |
|------|--|
| Wert | <p>Gibt den Wert an, der mit dem Wert des Feldes verglichen werden soll. Wenn Sie im Feld Feldtyp den Eintrag „Feld“ ausgewählt haben, müssen Sie ein Datenflussfeld auswählen. Wenn Sie im Feld Werttyp den Eintrag „Zeichenfolge“ ausgewählt haben, müssen Sie den bei dem Vergleich zu verwendenden Wert eingeben.</p> <p>Anmerkung: Diese Option ist nicht verfügbar, wenn Sie den Operator „Höchste“, „Niedrigste“ oder „Längste“ auswählen.</p> |
|------|--|

8. Klicken Sie auf **OK**.
9. Klicken Sie im Baum auf den Knoten **Aktionen**.
10. Klicken Sie auf **Aktion hinzufügen**.
11. Geben Sie die Daten an, die in den Best of Breed-Datensatz kopiert werden sollen, wenn der Datensatz den in der Regel definierten Kriterien entspricht.

| Option | Beschreibung |
|------------|---|
| Quelltyp | <p>Gibt den Datentyp an, der in den „Best of Breed“-Datensatz kopiert werden soll. Die folgenden Optionen stehen zur Auswahl:</p> <p>Feld Wählen Sie diese Option aus, wenn Sie einen Wert aus einem Feld in den „Best of Breed“-Datensatz kopieren möchten.</p> <p>Zeichenfolge Wählen Sie diese Option aus, wenn Sie einen konstanten Wert in den „Best of Breed“-Datensatz kopieren möchten.</p> |
| Quelldaten | <p>Gibt die Daten an, die in den „Best of Breed“-Datensatz kopiert werden sollen. Wenn der Quelltyp „Feld“ ist, müssen Sie das Feld auswählen, dessen Wert in das Zielfeld kopiert werden soll. Wenn der Quelltyp „Zeichenfolge“ ist, müssen Sie einen konstanten Wert angeben, der in das Zielfeld kopiert werden soll.</p> |

| Option | Beschreibung |
|-----------------------|--|
| Ziel | Gibt das Feld im „Best of Breed“-Datensatz an, in das die im Feld Quelldaten angegebenen Daten kopiert werden sollen. |
| Quelldaten kumulieren | <p>Wenn es sich bei den Daten im Feld Quelldaten um numerische Daten handelt, können Sie diese Option aktivieren, um die Quelldaten für alle Datensatzdubletten zu kombinieren und den Gesamtwert in den „Best of Breed“-Datensatz einzufügen.</p> <p>Beispiel: Es sind drei Datensatzdubletten in der Gruppe vorhanden und weisen die folgenden Werte im Feld „Einzahlungen“ auf:</p> <p>100,00 20,00 5,00</p> <p>Alle drei Werte werden addiert und die Summe (125,00) wird in das Feld „Einzahlungen“ des „Best of Breed“-Datensatzes eingefügt.</p> |

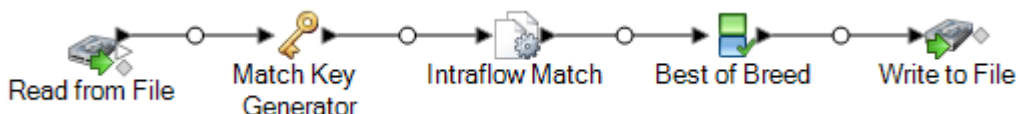
12. Klicken Sie auf **OK**.

Sie haben jetzt „Best of Breed“ mit einer Regel und einer Aktion konfiguriert. Sie können bei Bedarf zusätzliche Regeln und Aktionen hinzufügen.

13. Klicken Sie auf **OK**, um das Fenster **Best of Breed** zu schließen.

14. Ziehen Sie einen Datenladeschritt auf die Arbeitsfläche und verbinden Sie ihn mit dem „Best of Breed“-Schritt.

Bei Verwendung eines „Write to File“-Datenladeschrittes würde Ihr Datenfluss so aussehen:



15. Doppelklicken Sie auf den Datenladeschritt und konfigurieren Sie ihn.

Informationen zum Konfigurieren von Datenladeschritten finden Sie im *Datenfluss-Designer-Handbuch*.

Sie haben jetzt einen Datenfluss, der übereinstimmende Datensätze identifiziert und Datensätze innerhalb einer Sammlung in einem einzelnen Best of Breed-Datensatz zusammenführt.

6 -

Ausnahmedatensätze

In this section

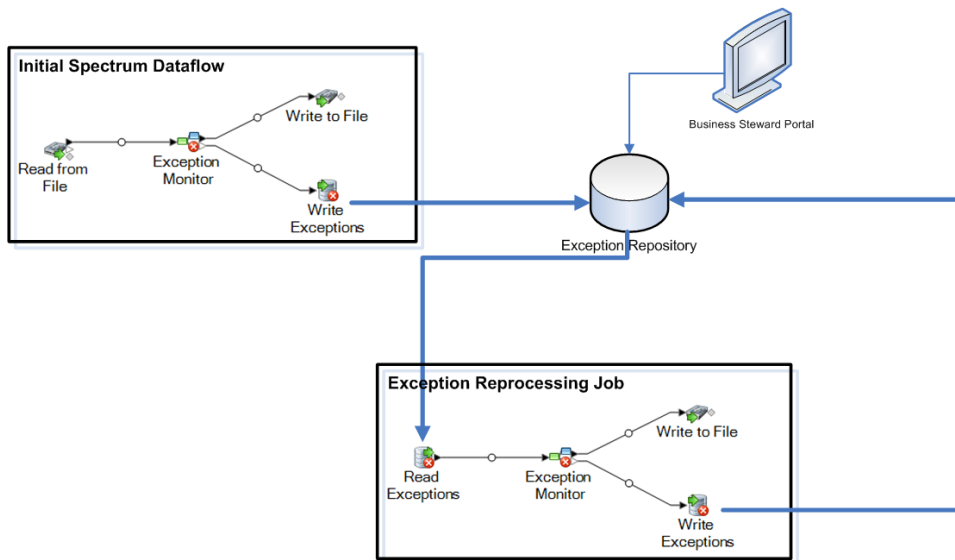
| | |
|---|-----|
| Entwerfen eines Datenflusses zur Handhabung von Ausnahmen | 158 |
| Entwerfen eines Datenflusses zur erneuten Überprüfung in Echtzeit | 160 |

Entwerfen eines Datenflusses zur Handhabung von Ausnahmen

Wenn Sie das Business Steward-Modul lizenziert haben, können Sie einen Ausnahmeverwaltungsprozess in Ihre Datenflüsse einschließen. Die Standardbausteine eines Ausnahmeverwaltungsprozesses umfassen:

- Einen ursprünglichen Datenfluss, über den ein Datenqualitätsprozess durchgeführt wird, wie z. B. die Deduplizierung eines Datenflusses, eine Adressenüberprüfung oder Geocoding.
- Ein Schritt vom Typ „Exception Monitor“, in dem Datensätze identifiziert werden, die nicht verarbeitet werden konnten.
- Ein Schritt vom Typ „Write Exceptions“, in dem die im „Exception Monitor“-Schritt identifizierten Ausnahmedatensätze genommen und zur manuellen Überprüfung in die Ausnahme-Datenbank geschrieben werden.
- Das Business Steward Portal, ein browserbasiertes Tool, mit dem Sie Ausnahmedatensätze überprüfen und bearbeiten können. Sobald die Datensätze bearbeitet wurden, werden sie als „Genehmigt“ markiert und stehen dann für die erneute Verarbeitung zur Verfügung.
- Einen Auftrag zur erneuten Verarbeitung von Ausnahmen, der mithilfe des „Read Exceptions“-Schrittes genehmigte Datensätze aus der Ausnahme-Datenbank in den Auftrag liest. Anschließend wird in dem Auftrag versucht, die korrigierten Datensätze erneut zu verarbeiten. Normalerweise wird hierbei die gleiche Logik verwendet wie beim ursprünglichen Datenfluss. Im „Exception Monitor“-Schritt wird erneut überprüft, ob Ausnahmen vorliegen. Anschließend werden im „Write Exceptions“-Schritt Ausnahmen für eine zusätzliche Überprüfung zurück zur Ausnahme-Datenbank gesendet.

Im Folgenden wird ein Beispielszenario dargestellt, das die grundlegende Implementierung einer Ausnahmenverwaltung verdeutlicht:



In diesem Beispiel gibt es zwei Datenflüsse: den ursprünglichen Datenfluss, in dem die Postleitzahlendaten der Eingabedatensätze ausgewertet werden, und der Auftrag zur erneuten Verarbeitung von Ausnahmen, in dem anhand bearbeiteter Ausnahmen überprüft wird, ob die Datensätze jetzt gültige Postleitzahlendaten enthalten.

In beiden Datenflüssen gibt es den „Exception Monitor“-Schritt. Dieser Schritt enthält die Bedingungen, mit denen Sie bestimmen möchten, ob ein Datensatz zur manuellen Überprüfung weitergeleitet werden soll. Diese Bedingungen bestehen aus mindestens einem Ausdruck, z. B.: `PostalCode is empty`. Dies bedeutet, dass alle Datensätze, die keine Postleitzahl enthalten, als Ausnahme gelten, im „Write Exceptions“-Schritt weitergeleitet und in die Ausnahme-Datenbank geschrieben würden. Weitere Informationen finden Sie unter [Exception Monitor](#) auf Seite 255.

Alle von Exception Monitor als Ausnahmen identifizierten Datensätze werden mithilfe des „Write Exceptions“-Schrittes an eine Ausnahme-Datenbank weitergeleitet. Data Stewards überprüfen die Ausnahmen in der Datenbank über das Business Steward Portal, einem browserbasierten Tool zum Anzeigen und Ändern von Ausnahmedatensätzen. In unserem Beispiel könnte der Data Steward mit dem Ausnahmen-Editor im Business Steward Portal manuell Postleitzahlen zu den Ausnahmedatensätzen hinzufügen und diese als „Genehmigt“ markieren.

Sobald ein Datensatz im Business Steward Portal als „Genehmigt“ markiert wurde, steht der Datensatz zum Einlesen in einen Spectrum™ Technology Platform-Datenfluss zur Verfügung. Dies geschieht durch Verwendung des „Read Exceptions“-Schrittes. Wenn bestimmte Datensätze weiterhin zu einer Ausnahme führen, werden sie erneut in die Ausnahme-Datenbank geschrieben, um von einem Data Steward überprüft zu werden.

Die folgenden Fragen sollen Ihnen dabei helfen, den besten Ansatz für Ihre Situation zu finden:

- **Wie sollen Ausnahmedatensätze identifiziert werden?** Im „Exception Monitor“-Schritt kann der Wert eines beliebigen Feldes oder eine beliebige Kombination aus Feldern ausgewertet werden, um zu ermitteln, ob ein Datensatz eine Ausnahme darstellt. Sie sollten die Ergebnisse analysieren, die Sie derzeit über Ihren Datenfluss abrufen, um zu ermitteln, wie Ausnahmen identifiziert werden

sollen. Sie sollten Datensätze im Mittelbereich der gleich bleibend hohen Datenqualität identifizieren und nicht Datensätze, die genau überprüft wurden oder deutlich fehlgeschlagen sind.

- **Sollen bearbeitete und genehmigte Ausnahmedatensätze mit der gleichen Logik erneut verarbeitet werden, die auch im ursprünglichen Datenfluss verwendet wurde?** Wenn ja, dann sollten Sie einen Unterfluss verwenden, um eine wiederverwendbar Geschäftslogik zu erstellen. Der Unterfluss könnte beispielsweise in einem ursprünglichen Datenfluss verwendet werden, in dem die Adressenüberprüfung durchgeführt wird, und in einem Auftrag zur erneuten Verarbeitung von Ausnahmen, in dem die korrigierten Datensätze zur Überprüfung der Korrekturen erneut verarbeitet werden. Anschließend können Sie unterschiedliche Quellen- und Datenladeschritte für den Datenfluss und den Auftrag verwenden. Der ursprüngliche Datenfluss enthält möglicherweise den „Read from DB“-Schritt, in dem Daten zur Verarbeitung aus Ihrer Kundendatenbank entnommen werden. Der Auftrag zur erneuten Verarbeitung von Ausnahmen enthält möglicherweise den „Read Exceptions“-Schritt, in dem bearbeiteten und genehmigten Ausnahmedatensätze aus der Ausnahme-Datenbank entnommen werden.
- **Sollen korrigierte und genehmigte Ausnahmen in einem vordefinierten Zeitplan erneut verarbeitet werden?** Wenn ja, können Sie Ihren Auftrag zur erneuten Verarbeitung über „Planung“ in der Management Console planen.

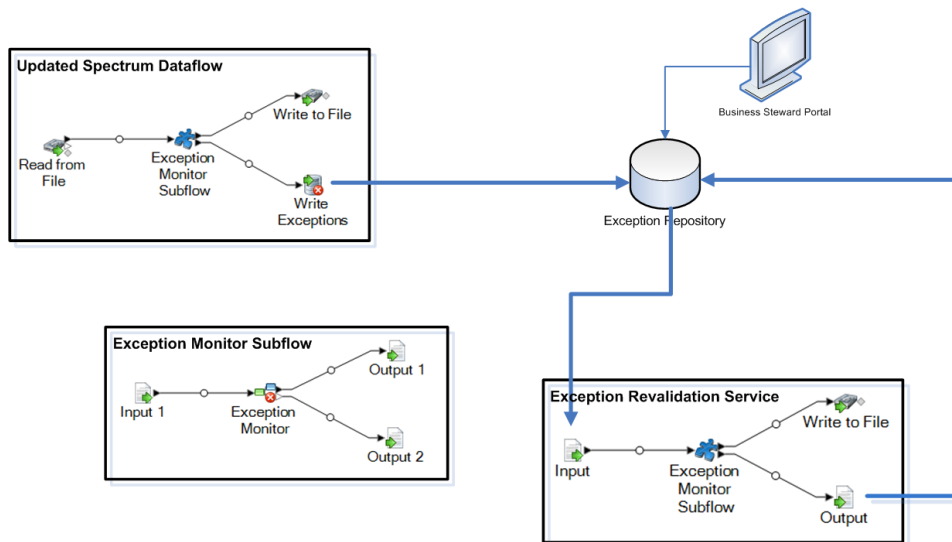
Entwerfen eines Datenflusses zur erneuten Überprüfung in Echtzeit

Wenn Sie in Ihrem Datenfluss die Ausnahmeverwaltung verwenden, können Sie mithilfe des Feature zur erneuten Überprüfung Ausnahmedatensätze über den Überprüfungsprozess erneut ausführen, nachdem diese im Business Steward Portal korrigiert wurden. Dadurch können Sie bestimmen, ob der Datensatz durch die von Ihnen vorgenommene Änderung erfolgreich in Echtzeit verarbeitet werden kann. Um das Ergebnis zu sehen, müssen Sie nicht warten, bis der Read Exceptions-Batchauftrag erneut ausgeführt wird.

Die Standardbausteine in einer Umgebung zur erneuten Überprüfung umfassen:

- Einen Auftrag oder einen Dienst, der einen verfügbar gemachten Unterfluss wiederverwendet oder enthält. Er muss zudem eine Eingabequelle, den Unterflussschritt, in dem die Eingabe verarbeitet wird, einen Schritt vom Typ „Write Exceptions“ und ein Laden von Daten für erfolgreich verarbeitete Datensätze enthalten.
- Einen verfügbar gemachten Unterfluss, der einen Schritt vom Typ „Exception Monitor“ enthält, der auf einen Dienst zur erneuten Überprüfung verweist und für die erneute Überprüfung konfiguriert ist. Darin inbegriffen ist die Angabe, ob erneut überprüfte Datensätze erneut verarbeitet oder genehmigt werden sollen.
- Einen verfügbar gemachten Dienst, der auch den verfügbar gemachten Unterfluss wiederverwendet oder enthält. Er verarbeitet Datensätze, die bearbeitet, gespeichert und für eine erneute Überprüfung im Business Steward Portal gesendet wurden.

Im Folgenden wird ein Beispielszenario dargestellt, das die Implementierung einer erneuten Überprüfung verdeutlicht:



In diesem Beispiel gibt es drei Datenflüsse: einen Auftrag, einen Unterfluss und einen Dienst. Der Auftrag führt Eingabedaten über den Unterfluss aus. Der Unterfluss enthält einen Schritt vom Typ „Exception Monitor“, in dem bestimmt wird, ob ein Datensatz zur manuellen Überprüfung weitergeleitet werden soll. In unserem Beispiel bedeutet dies, dass alle Datensätze, die im Feld „Postleitzahl“ keine Daten enthalten, als Ausnahme gelten würden und im „Write Exceptions“-Schritt weitergeleitet werden würden. Diese Ausnahmen werden im Business Steward Portal angezeigt. Datensätze, die in diesem Feld etwas anderes enthalten, würden im „Write to File“-Schritt weitergeleitet werden.

Der von Ihnen bei der Konfiguration des „Exception Monitor“-Schrittes angegebene Dienst für die erneute Überprüfung von Ausnahmen wird aufgerufen, wenn Sie mindestens einen Ausnahmedatensatz im Exception Editor des Business Steward Portal bearbeiten und auf **Erneut überprüfen und speichern** klicken. Der Dienst enthält, wie der Auftrag, den Exception Monitor-Unterfluss, der für die erneute Verarbeitung des Datensatzes bzw. der Datensätze die gleiche Geschäftslogik verwendet. Wenn die Datensätze mindestens eine der im „Exception Monitor“-Schritt festgelegten Bedingungen nicht erfüllen, werden die Ausnahmen in der Datenbank aktualisiert. Wenn die Datensätze die im „Exception Monitor“-Schritt festgelegten Bedingungen erfüllen, wird abhängig von der im Feld „Aktion nach erneuter Überprüfung“ getroffenen Auswahl eine von zwei Aktionen ausgeführt:

- **Datensätze erneut verarbeiten** – Datensätze werden aus der Datenbank gelöscht und erneut verarbeitet.
- **Datensätze genehmigen** – Datensätze werden als genehmigt markiert und zurück zur Datenbank gesendet.

Führen Sie die folgenden Schritte aus, um ein Szenario für die erneute Überprüfung in Echtzeit zu erstellen und zu verwenden:

1. Öffnen oder erstellen Sie einen Auftrag oder Dienstdatenfluss, der den „Exception Monitor“-Schritt, eine Eingabequelle (z. B. der Schritt „Read from File“ oder „Input“), ein Laden

von Daten (z. B. der Schritt „Write to File“ oder „Output“) und den Schritt „Write Exceptions“ enthält.

2. Wandeln Sie den „Exception Monitor“-Schritt in einen Unterfluss um und ordnen Sie die Eingabe- und Ausgabefelder so zu, dass sie mit den Feldern im ursprünglichen Datenfluss übereinstimmen. Stellen Sie sicher, dass das Feld „Metadaten für Ausnahme“ für die Eingabequelle und der Output-Schritt eingeschlossen wird, in dem der „Write Exceptions“-Schritt des Auftrags aufgefüllt wird. Machen Sie den Unterfluss verfügbar, damit er in dem Auftrag und dem Dienst verwendet werden kann.
3. Erstellen Sie einen Dienst, der den Input-Schritt enthält, den in Schritt 2 erstellten Unterfluss, den Output-Schritt und ein Laden von Daten (z. B. der „Write to File“- oder „Write to DB“-Schritt). Ordnen Sie die Eingabe- und Ausgabefelder so zu, dass sie mit den Feldern im ursprünglichen Datenfluss übereinstimmen. Stellen Sie dabei sicher, dass das Feld „Metadaten für Ausnahme“ für den Input- und Output-Schritt eingeschlossen wird. Machen Sie den Dienst verfügbar, damit er im Unterfluss verwendet werden kann.
4. Kehren Sie zum Unterfluss zurück und öffnen Sie die Registerkarte „Konfiguration“ im „Exception Monitor“-Schritt. Wählen Sie den in Schritt 3 erstellten Dienst für die erneute Überprüfung aus und geben Sie an, welche Maßnahmen nach der erneuten Überprüfung ergriffen werden sollen. Speichern Sie dies und machen Sie den Unterfluss erneut verfügbar.
5. Kehren Sie zum Dienst zurück. Dort wird eine Nachricht angezeigt, die Sie über Änderungen am Datenfluss informiert und mitteilt, dass der Dienst aktualisiert wird. Klicken Sie auf **OK**. Speichern Sie anschließend den Dienst und machen Sie ihn erneut verfügbar.
6. Kehren Sie zum ursprünglichen Auftrag oder Dienst zurück. Dort wird eine Nachricht angezeigt, die Sie über Änderungen am Unterfluss informiert und mitteilt, dass der Datenfluss aktualisiert wird. Klicken Sie auf **OK** und speichern Sie den Datenfluss anschließend.
7. Führen Sie den Auftrag aus.

Anmerkung: Nachdem Sie das Szenario für die erneute Überprüfung erstellt haben, um die Datenbank mit Datensätzen aufzufüllen, die für die erneute Überprüfung geeignet sind, müssen Sie den ursprünglichen Auftrag oder Dienst selbst dann erneut ausführen, wenn Sie diesen bereits vorher ausgeführt haben. Ob Datensätze im Exception Editor für eine erneute Überprüfung geeignet sind, können Sie daran erkennen, dass die Schaltfläche „Erneut überprüfen und speichern“ bei diesen Datensätzen aktiv ist.

7 - Suchtabellen

In this section

| | |
|---|-----|
| Einführung in Suchtabellen | 164 |
| Tabellen des Data Normalization-Moduls | 164 |
| Tabellen des Universal Name-Moduls | 170 |
| Anzeigen des Inhalts einer Nachschlagetabelle | 171 |
| Hinzufügen eines Begriffs zu einer Nachschlagetabelle | 172 |
| Entfernen eines Begriffs aus einer Nachschlagetabelle | 173 |
| Ändern der standardisierten Form eines Begriffs | 173 |
| Rückgängigmachen von Tabellenanpassungen | 173 |
| Erstellen einer Nachschlagetabelle | 174 |
| Importieren von Daten | 175 |

Einführung in Suchtabellen

Eine Suchtabelle ist eine Tabelle mit Schlüssel/Wert-Paaren, die von Spectrum™ Technology Platform-Schritten verwendet werden, um Daten über Token-Ersetzungen zu standardisieren. Um die Inhalte der im Advanced Transformer, Open Parser und Table Lookup verwendeten Suchtabellen zu ändern, verwenden Sie das Tabellenverwaltungstool im Enterprise Designer.

Tabellen des Data Normalization-Moduls

Advanced Transformer-Tabellen

Der Advanced Transformer verwendet die folgenden Tabellen, um Begriffe zu identifizieren. Verwenden Sie die Tabellenverwaltung, um neue Tabellen zu erstellen oder vorhandene zu ändern. Weitere Informationen finden Sie unter [Einführung in Suchtabellen](#) auf Seite 164.

- Abkürzungen der Luftfahrt
- Alle Akronymabkürzungen
- Abkürzungen für Geschäftsnamen
- Abkürzungen für kanadische Territorien
- IT-Abkürzungen
- Begrenzungszeichen
- Deutsche Unternehmen
- Fortune 1000
- Abkürzungen für geografische Richtungen
- „Global Sentry Noise“-Begriffe
- Länder mit „Global Sentry“-Sanktionierung
- Abkürzungen für Regierungsbehörden
- IATA-Bezeichner für Fluggesellschaften
- Länderspezifische IATA-Bezeichner für Fluggesellschaften
- Rechtliche Abkürzungen
- Medizinische Abkürzungen
- Abkürzungen für medizinische Organisationen
- Militärische Abkürzungen
- Spitznamen

- Abkürzungen für Sekundäreinheiten
- Umkehrung der Sekundäreinheit
- Singapur-Abkürzungen
- Spanische Abkürzungen
- Spanische Richtungsabkürzungen
- Spanische Abkürzungen für Straßensuffixe
- Abkürzungen für Bundeslandnamen
- Umkehrung von Bundeslandnamen
- Abkürzungen für Straßensuffixe
- Umkehrung von Straßensuffixen
- Unterordnung des übergeordneten Elements
- US-amerikanische Armeeabkürzungen
- US-amerikanische Navy-Abkürzungen

Open Parser-Tabellen

Open Parser verwendet die folgenden Tabellen, um Begriffe zu identifizieren. Verwenden Sie die Tabellenverwaltung, um neue Tabellen zu erstellen oder vorhandene Tabellen zu ändern. Weitere Informationen finden Sie unter [Einführung in Suchtabellen](#) auf Seite 164.

Basistabellen

Basistabellen werden mit dem Installationspaket des Data Normalization-Moduls bereitgestellt.

- Kontobeschreibungen
- Unternehmen
- Unternehmenszusammenschlüsse
- Präposition im Unternehmensnamen
- Unternehmenssuffixe
- Unternehmensbegriffe
- Konjunktionen
- Familiennamen-Präfixe
- Familiennamen
- Allgemeine Suffixe
- Deutsche Unternehmen
- Vornamen
- Generations-Suffixe
- Spanische Vornamen
- Spanische Familiennamen
- Titel

Core Names-Tabellen

Die Core Names-Tabellen werden nicht mit dem Installationspaket des Data Normalization-Moduls bereitgestellt und erfordern deshalb eine zusätzliche Lizenz. Wenden Sie sich für weitere Informationen an Ihren Kundenbetreuer.

Die Core Names-Tabellen müssen mit der Database Load Utility des Data Normalization-Moduls geladen werden. Anweisungen dazu finden Sie im *Spectrum™ Technology Platform-Installationshandbuch*.

- Erweiterte Familiennamen
- Erweiterte Vornamen

Company Names-Tabellen

Die Company Names-Tabellen werden nicht mit dem Installationspaket des Data Normalization-Moduls bereitgestellt und erfordern deshalb eine zusätzliche Lizenz. Wenden Sie sich für weitere Informationen an Ihren Kundenbetreuer.

Die Company Names-Tabellen müssen mit der Database Load Utility des Data Normalization-Moduls geladen werden. Anweisungen dazu finden Sie im *Spectrum™ Technology Platform-Installationshandbuch*.

- Unternehmen – Nord- und Südamerika
- Unternehmen – Asien-Pazifik
- Unternehmen – EMEA
- Unternehmensartikel
- Unternehmenszusammenschlüsse

Arabic Plus Pack-Tabellen

Arabic Plus Pack-Tabellen werden nicht mit dem Installationspaket des Data Normalization-Moduls bereitgestellt und erfordern deshalb eine zusätzliche Lizenz. Wenden Sie sich für weitere Informationen an Ihren Kundenbetreuer.

Arabic Plus Pack-Tabellen müssen mit der Database Load Utility des Data Normalization-Moduls geladen werden. Anweisungen dazu finden Sie im *Spectrum™ Technology Platform-Installationshandbuch*.

- Arabische Familiennamen (Arabisch)
- Arabische Familiennamen (Romanisiert)
- Arabische Vornamen (Arabisch)
- Arabisch Vornamen (Romanisiert)

Asian Plus Pack-Tabellen

Asian Plus Pack-Tabellen werden nicht mit dem Installationspaket des Data Normalization-Moduls bereitgestellt und erfordern deshalb eine zusätzliche Lizenz. Wenden Sie sich für weitere Informationen an Ihren Kundenbetreuer.

Asian Plus Pack-Tabellen müssen mit der Database Load Utility des Data Normalization-Moduls geladen werden. Anweisungen dazu finden Sie im *Spectrum™ Technology Platform-Installationshandbuch*.

- Chinesische Familiennamen (Muttersprache)
- Chinesische Familiennamen (Romanisiert)
- Chinesische Vornamen (Muttersprache)
- Chinesische Vornamen (Romanisiert)
- Koreanische Familiennamen (Muttersprache)
- Koreanische Familiennamen (Romanisiert)
- Koreanische Vornamen (Muttersprache)
- Koreanische Vornamen (Romanisiert)
- Japanische Familiennamen (Kana)
- Japanische Familiennamen (Kanji)
- Japanische Familiennamen (Romanisiert)
- Japanische Vornamen (Kana)
- Japanische Vornamen (Kanji)
- Japanische Namen (Romanisiert)

„Table Lookup“-Tabellen

Beim Table Lookup werden die folgenden Tabellen zum Identifizieren von Begriffen verwendet. Verwenden Sie die Tabellenverwaltung, um neue Tabellen zu erstellen oder vorhandene Tabellen zu ändern. Weitere Informationen finden Sie unter [Einführung in Suchtabellen](#) auf Seite 164.

Basistabellen

Basistabellen werden mit dem Installationspaket des Data Normalization-Moduls bereitgestellt.

- Aeronautical Abbreviations
- All Acronyms Initialism
- Business Names Abbreviations
- Canadian Territory Abbreviations
- Computing/IT Abbreviations
- EU Acronyms
- Fortune 1000
- French Abbreviations
- French Arrondissement to Department Number

- French Commune to Postal Code
- French Department to Region
- French Department Number to Department
- Gender Codes
- Geographic Directional Abbreviations
- German Acronyms
- German City to State Code
- German Area Code to City
- German District to State Code
- German State Abbreviations
- Global Sentry Sanctioned Countries
- Government Agencies Abbreviations
- IATA Airline Designator
- IATA Airline Designator Country
- Legal Abbreviations
- Medical Abbreviations
- Medical Organizations Acronyms
- Military Abbreviations
- Nicknames
- Secondary Unit Abbreviations
- Secondary Unit Reverse
- Singapore Abbreviations
- Spanish Abbreviations
- Spanish Directional Abbreviations
- Spanish Street Suffix Abbreviations
- State Name Abbreviations
- State Name Reverse
- Street Suffix Abbreviations
- Street Suffix Reverse
- Subsidiary to Parent
- U.K. Town to Postcode Area
- U.K. Dialing Code Prefixes
- U.K. Dialing Codes to Town
- U.K. Postcode Area to Town
- US-amerikanische Army Acronyms
- US-amerikanische Navy Acronyms
- ZREPLACE (wird vom SAP-Modul für die französische Adressvalidierung verwendet)

Core Names

Für „Core Names“-Tabellen ist eine zusätzliche Lizenz erforderlich. Wenden Sie sich für weitere Informationen an Ihren Kundenbetreuer.

Die Core Names-Tabellen müssen mit der Database Load Utility des Data Normalization-Moduls geladen werden. Anweisungen dazu finden Sie im *Spectrum™ Technology Platform-Installationshandbuch*.

- Erweiterte Familiennamen, Ethnizität
- Erweiterte Geschlechtscodes
- Erweiterte Vornamen, Ethnizität

Arabic Plus Pack

Für „Arabic Plus Pack“-Tabellen ist eine zusätzliche Lizenz erforderlich. Wenden Sie sich für weitere Informationen an Ihren Kundenbetreuer.

Arabic Plus Pack-Tabellen müssen mit der Database Load Utility des Data Normalization-Moduls geladen werden. Anweisungen dazu finden Sie im *Spectrum™ Technology Platform-Installationshandbuch*.

- Arabische Familiennamen, Ethnizität (Arabisch)
- Arabische Familiennamen, Ethnizität (Romanisiert)
- Arabische Geschlechtscodes (Arabisch)
- Arabische Geschlechtscodes (Romanisiert)
- Arabische Vornamen, Ethnizität (Arabisch)
- Arabische Vornamen, Ethnizität (Romanisiert)

Asian Plus Pack

Für „Asian Plus Pack“-Tabellen ist eine zusätzliche Lizenz erforderlich. Wenden Sie sich für weitere Informationen an Ihren Kundenbetreuer.

Asian Plus Pack-Tabellen müssen mit der Database Load Utility des Data Normalization-Moduls geladen werden. Anweisungen dazu finden Sie im *Spectrum™ Technology Platform-Installationshandbuch*.

- CJK Familiennamen, Ethnizität (Muttersprache)
- CJK Familiennamen, Ethnizität (Romanisiert)
- CJK Vornamen, Ethnizität (Muttersprache)
- CJK Vornamen, Ethnizität (Romanisiert)
- Japanische Geschlechtscodes (Kana)
- Japanische Geschlechtscodes (Kanji)
- Japanische Geschlechtscodes (Romanisiert)

Tabellen des Universal Name-Moduls

„Name Variant Finder“-Tabellen

Im „Name Variant Finder“-Schritt werden folgende Tabellen verwendet. Für jede Tabelle ist eine separate Lizenz erforderlich.

- Arabic Plus Pack: `g1-cdq-cjki-arabic-<date>.jar`
- Asian Plus Pack – Chinesisch: `g1-cdq-cjki-chinese-<date>.jar`
- Asian Plus Pack – Japanisch: `g1-cdq-cjki-japanese-<date>.jar`
- Asian Plus Pack – Koreanisch: `g1-cdq-cjki-korean-<date>.jar`
- Core Names-Datenbank: `g1-cdq-nomino-base-<date>.jar`

Open Name Parser-Tabellen

Open Name Parser verwendet die folgenden Tabellen, um Begriffe zu identifizieren. Verwenden Sie die Tabellenverwaltung, um neue Tabellen zu erstellen oder vorhandene Tabellen zu ändern. Weitere Informationen finden Sie unter [Einführung in Suchtabellen](#) auf Seite 164.

Basistabellen

Basistabellen werden mit dem Installationspaket des Universal Name-Moduls bereitgestellt.

- Kontobeschreibungen
- Unternehmenszusammenschlüsse
- Konjunktionen
- Familiennamen-Präfixe
- Familiennamen
- Allgemeine Suffixe
- Vornamen
- Generations-Suffixe
- Spanische Vornamen
- Spanische Familiennamen
- Titel

Core Names-Tabellen

Die Core Name-Tabellen werden nicht mit dem Installationspaket des Universal Name-Moduls bereitgestellt und erfordern deshalb eine zusätzliche Lizenz.

- Erweiterte Familiennamen
- Erweiterte Vornamen

Company Names-Tabellen

Die folgenden Company Names-Tabellen werden mit dem Installationspaket des Universal Name-Moduls bereitgestellt.

- Kontobeschreibungen
- Unternehmen
- Unternehmensartikel
- Unternehmenszusammenschlüsse
- Präposition im Unternehmensnamen
- Unternehmenssuffixe
- Unternehmensbegriffe
- Konjunktionen

Die folgenden Company Names-Tabellen werden nicht mit dem Installationspaket des Universal Name-Moduls bereitgestellt und erfordern deshalb eine zusätzliche Lizenz.

- Unternehmen – Nord- und Südamerika
- Unternehmen – Asien-Pazifik
- Unternehmen – EMEA

Asian Plus Pack-Tabellen

Die Asian Plus Pack-Tabellen werden nicht mit dem Installationspaket des Universal Name-Moduls bereitgestellt und erfordern deshalb eine zusätzliche Lizenz.

- Japanische Familiennamen (Kana)
- Japanische Familiennamen (Kanji)
- Japanische Familiennamen (Romanisiert)
- Japanische Vornamen (Kana)
- Japanische Vornamen (Kanji)
- Japanische Namen (Romanisiert)
- Japanische Titel

Anzeigen des Inhalts einer Nachschlagetabelle

Sie können den Inhalt einer Nachschlagetabelle mithilfe der Tabellenverwaltung im Enterprise Designer anzeigen.

1. Wählen Sie im Enterprise Designer **Tools** > **Tabellenverwaltung** aus.
2. Wählen Sie im Feld **Typ** den Schritt aus, dessen Nachschlagetabelle Sie anzeigen möchten.
3. Wählen Sie im Feld **Name** die Tabelle aus, die Sie anzeigen möchten.

4. Sie können die folgenden Optionen verwenden, um die Anzeige der Tabelle zu ändern:

| Option | Bezeichnung |
|--|---|
| Suchen Sie einen bestimmten Begriff | Geben Sie im Feld Beginnt mit den Begriff ein, den Sie suchen möchten, und klicken Sie dann auf Aktualisieren . |
| Blättern Sie durch die Tabelle | Klicken Sie rechts neben der Schaltfläche Aktualisieren auf die Schaltflächen „Vorwärts“ und „Zurück“. |
| Ändern Sie die Anzahl der pro Seite angezeigten Begriffe | Ändern Sie den Wert im Feld Elemente pro Seite . |
| Zeigen Sie alle Suchbegriffe für die einzelnen standardisierten Begriffe in einer „Table Lookup“-Tabelle an | Wählen Sie im Feld Anzeigen nach den Eintrag Standardisierter Begriff (Gruppierung) aus. Diese Option ist nur für „Table Lookup“-Tabellen verfügbar |

Hinzufügen eines Begriffs zu einer Nachschlagetabelle

Gehen Sie wie folgt vor, wenn Sie feststellen, dass Ihre Daten über Begriffe verfügen, die nicht in der Nachschlagetabelle enthalten sind, und Sie den Begriff zu einer Nachschlagetabelle hinzufügen möchten.

1. Wählen Sie im Enterprise Designer **Tools > Tabellenverwaltung** aus.
2. Wählen Sie im Feld **Typ** den Schritt aus, dessen Nachschlagetabelle Sie ändern möchten.
3. Wählen Sie im Feld **Name** die Tabelle aus, zu der Sie einen Begriff hinzufügen möchten.
4. Klicken Sie auf **Hinzufügen**.
5. Geben Sie im Feld **Suchbegriff** den Begriff ein, der in Ihren Daten vorhanden ist. Dies ist der Suchschlüssel, der verwendet wird.
6. Geben Sie bei „Table Lookup“-Tabellen im Feld **Standardisierter Begriff** den Begriff ein, den Sie als Ersatz für den Suchbegriff in Ihrem Datenfluss verwenden möchten.

Wenn Sie beispielsweise den Begriff „PB“ in „Pitney Bowes“ ändern möchten, müssen Sie „PB“ als Suchbegriff und „Pitney Bowes“ als standardisierten Begriff eingeben.

7. Aktivieren Sie bei „Table Lookup“-Tabellen das Kästchen **Vorhandenen Begriff außer Kraft setzen**, wenn dieser Begriff bereits in der Tabelle vorhanden ist und Sie ihn durch den in Schritt 5 eingegebenen Wert ersetzen möchten.

8. Klicken Sie auf **Hinzufügen**.

Entfernen eines Begriffs aus einer Nachschlagetabelle

So entfernen Sie einen Begriff aus einer Nachschlagetabelle:

1. Wählen Sie im Enterprise Designer **Tools** > **Tabellenverwaltung** aus.
2. Wählen Sie den Begriff aus und klicken Sie auf **Entfernen**.
3. Klicken Sie auf **Ja**, um den Tabellenbegriff zu entfernen.

Ändern der standardisierten Form eines Begriffs

Bei Tabellen, die von Table Lookup zum Standardisieren von Begriffen verwendet werden, können Sie die standardisierte Form eines Begriffs ändern. Wenn Sie beispielsweise in einer Tabelle die Suchbegriffe „PB“ und „PB Software“ haben, der standardisierte Begriff „Pitney Bowes“ lautet und Sie die standardisierte Form in „Pitney Bowes Inc“ ändern möchten, ist dies möglich, indem Sie wie folgt vorgehen.

1. Wählen Sie im Enterprise Designer **Tools** > **Tabellenverwaltung** aus.
2. Wählen Sie im Feld **Typ** die Option **Table Lookup** aus.
3. Wählen Sie im Feld **Name** die Tabelle aus, die Sie ändern möchten.
4. Wählen Sie den Begriff aus, den Sie ändern möchten, und klicken Sie auf **Ändern**.

Tip: Wenn es mehrere Suchbegriffe für einen standardisierten Begriff gibt, können Sie alle Suchbegriffe ohne großen Aufwand ändern, um den neuen standardisierten Begriff zu verwenden, indem Sie im Feld **Anzeigen nach** den Eintrag **Anzeigen nach standardisiertem Begriff (Gruppierung)** auswählen, die Gruppe auswählen und auf **Ändern** klicken.

5. Geben Sie im Feld **Standardisierter Begriff** einen neuen Wert ein.
6. Klicken Sie auf **OK**.

Rückgängigmachen von Tabellenanpassungen

Wenn Sie an einer Tabelle Änderungen vornehmen, können Sie den ursprünglichen Zustand der Tabelle wiederherstellen. So machen Sie Tabellenanpassungen rückgängig:

1. Wählen Sie im Enterprise Designer **Tools > Tabellenverwaltung** aus.
2. Wählen Sie die Tabelle aus, die Sie wiederherstellen möchten.
3. Klicken Sie auf **Wiederherstellen**.

Das Fenster **Wiederherstellen** wird angezeigt. Hier sind alle Begriffe aufgelistet, die hinzugefügt, entfernt oder geändert wurden.

4. Aktivieren Sie das Kästchen **Wiederherstellen** bei jedem Tabelleneintrag, den Sie wiederherstellen möchten. Sie können auch auf **Alle auswählen** oder **Gesamte Auswahl aufheben** klicken, um alle Kästchen **Wiederherstellen** zu aktivieren bzw. zu deaktivieren.
5. Klicken Sie auf **OK**.

Erstellen einer Nachschlagetabelle

Das Advanced Matching-Modul, das Data Normalization-Modul und das Universal Name-Modul enthalten eine Vielzahl von Tabellen, die für eine Vielzahl von Prozessen zum Ersetzen von Begriffen und Standardisierungsprozessen verwendet werden können. Wenn diese Tabellen jedoch nicht Ihren Anforderungen entsprechen, können Sie Ihre eigene Tabelle mit Suchbegriffen erstellen, die mit dem Advanced Transformer, Open Parser oder Table Lookup verwendet werden kann. Gehen Sie wie folgt vor, um eine Tabelle zu erstellen.

1. Wählen Sie im Enterprise Designer **Tools > Tabellenverwaltung** aus.
2. Wählen Sie im Feld **Typ** den Schritt aus, für den Sie eine Nachschlagetabelle erstellen möchten.
3. Klicken Sie auf **Neu**. Das Dialogfeld „**Tabelle hinzufügen**“ wird angezeigt.
4. Geben Sie im Feld **Tabellename** einen Namen für die neue Tabelle ein.
5. Wenn Sie eine neue, leere Tabelle des ausgewählten Typs haben möchten, lassen Sie **Kopieren von** auf **Keine** festgelegt. Wählen Sie einen Tabellennamen aus der Liste **Kopieren von** aus, wenn die neue Tabelle aus einer vorhandenen Tabelle aufgefüllt werden soll.
6. Klicken Sie auf **OK**.

Weitere Informationen zum Hinzufügen von Tabellenelementen zu Ihrer neuen Tabelle finden Sie unter [Hinzufügen eines Begriffs zu einer Nachschlagetabelle](#) auf Seite 172.

Importieren von Daten

Importieren von Daten in eine Nachschlagetabelle

Sie können Daten aus einer Datei für die Verwendung im Advanced Transformer, Open Parser oder Table Lookup in eine Nachschlagetabelle importieren. Um Daten aus einer Datei in eine Nachschlagetabelle importieren zu können, muss die Datei die folgenden Anforderungen erfüllen:

- Sie muss UTF-8-codiert sein.
- Es muss eine Datei mit Begrenzungszeichen sein. Unterstützte Trennzeichen sind Komma (,), Semikolon (;), Pipe (|) und Tabulator (\t).
- Felder mit eingebetteten Begrenzern müssen mit doppelten Anführungszeichen beginnen und enden, zum Beispiel „1, a“, „2, b“, „3, c“.
- Ein wörtliches Zitat in einem Feld, das mit doppelten Anführungszeichen beginnt und endet, muss zwei Anführungszeichen haben, zum Beispiel „2““ Füße“.

So importieren Sie Daten aus einer Datei in eine Nachschlagetabelle:

1. Wählen Sie im Enterprise Designer **Tools > Tabellenverwaltung** aus.
2. Wählen Sie die Tabelle aus, in die Sie die Daten importieren möchten. Alternativ können Sie eine neue Tabelle erstellen. Anweisungen zum Erstellen einer Tabelle finden Sie unter **Erstellen einer Nachschlagetabelle** auf Seite 174.
3. Klicken Sie auf **Importieren**.
4. Klicken Sie auf **Durchsuchen** und wählen Sie die Datei aus, die die Daten enthält, die Sie importieren möchten.
5. Klicken Sie auf **Öffnen**. Eine Vorschau der Daten in der zu importierenden Datei wird in der Dateivorschau angezeigt.
6. Sie können Spalten aus einer benutzerdefinierten Tabelle auswählen und der vorhandenen Tabelle zuordnen. Angenommen, Sie möchten zwei Spalten aus der benutzerdefinierten Tabelle importieren. Sie enthält Spalte1 und Spalte2. In der Spaltenliste würden Spalte1 und Spalte2 angezeigt. Sie können Spalte2 für die Zuordnung zu einem Suchbegriff auswählen und Spalte1 für die Zuordnung zu einem standardisierten Begriff.
7. Wählen Sie **Nur neue Begriffe importieren** aus, um nur neue Datensätze aus der benutzerdefinierten Tabelle zu importieren, oder wählen Sie **Vorhandene Begriffe überschreiben** aus, um alle Datensätze der ausgewählten Spalten zu importieren.
8. Klicken Sie auf **OK**.

Verwenden des erweiterten Imports



Mit der Funktion „Erweiterter Import“ können Sie selektiv Daten in Nachschlagetabellen importieren, die vom Advanced Transformer, Table Lookup und Open Parser verwendet werden. Verwenden Sie den erweiterten Import für folgende Zwecke:

- Extrahieren von Ausdrücken aus einer ausgewählten Spalte in eine benutzerdefinierte Datei mit Begrenzungszeichen.
- Extrahieren von Einzelbegriffen (Token) aus einer ausgewählten Spalte in eine benutzerdefinierte Datei mit Begrenzungszeichen. Wenn Sie Token extrahieren, können Sie identifizieren, wie oft die Begriffe in einer bestimmten Spalte in der Datei auftreten, und Gruppierungen für zugehörige Begriffe erstellen und diese zur Tabelle hinzufügen.

Die Datei, die die zu importierenden Daten enthält, muss die folgenden Anforderungen erfüllen:

- Sie muss UTF-8-codiert sein.
- Es muss eine Datei mit Begrenzungszeichen sein. Unterstützte Trennzeichen sind Komma (,), Semikolon (;), Pipe (|) und Tabulator (\t).
- Felder mit eingebetteten Begrenzern müssen mit doppelten Anführungszeichen beginnen und enden, zum Beispiel „1, a“, „2, b“, „3, c“.
- Ein wörtliches Zitat in einem Feld, das mit doppelten Anführungszeichen beginnt und endet, muss zwei Anführungszeichen haben, zum Beispiel „2“ Füße“.

1. Wählen Sie im Enterprise Designer **Tools** > **Tabellenverwaltung** aus.
2. Wählen Sie die Tabelle aus, in die Sie Daten importieren möchten.
3. Klicken Sie auf **Erw. Import**.
4. Klicken Sie auf **Durchsuchen** und wählen Sie die Datei aus, die Sie importieren möchten.
5. Klicken Sie auf **Öffnen**.
6. Wählen Sie aus der Liste **Spalte** eine Tabellenspalte aus. Die Beispieldaten zeigen, wie häufig die einzelnen in der benutzerdefinierten Tabelle aufgeführten Begriffe auftreten. Die Häufigkeit wird nur für Begriffe angezeigt, die noch nicht in der vorhandenen Tabelle enthalten sind.
7. Wählen Sie **In einzelne Wörter aufteilen** aus, um Begriffe als einzelne Wörter anzuzeigen.
8. Bei „Advanced Transformer“- und „Open Parser“-Tabellen:
 - a) Wählen Sie aus der Liste auf der linken Seite einen Begriff aus.
 - b) Klicken Sie auf den Pfeil nach rechts, um den Begriff zu der Liste auf der rechten Seite hinzuzufügen. Klicken Sie auf den Pfeil nach links, um einen ausgewählten Begriff aus der Tabellenliste zu löschen.
 - c) Klicken Sie auf **OK**, um die an der Tabelle vorgenommenen Änderungen zu speichern.
9. Bei „Table Lookup“-Tabellen:
 - a) Klicken Sie auf , um eine Tabellengruppierung hinzuzufügen.
 - b) Klicken Sie auf **Neu**.

- c) Geben Sie einen neuen Begriff ein und klicken Sie dann auf **Hinzufügen**. Fügen Sie so lange weitere Begriffe hinzu, bis Sie fertig sind, und klicken Sie anschließend auf **Schließen**.
- d) Wählen Sie einen Begriff aus der Liste aus, und klicken Sie dann auf „Hinzufügen“. Fügen Sie so lange weitere Begriffe hinzu, bis Sie fertig sind, und klicken Sie anschließend auf **Schließen**. Die neuen Begriffe werden zu der Begriffsliste auf der rechten Seite hinzugefügt.
- e) Wählen Sie einen Begriff auf der linken Seite aus und klicken Sie dann auf den Pfeil nach rechts, um den Begriff zur ausgewählten Gruppierung hinzuzufügen. Klicken Sie auf den Pfeil nach links, um einen Begriff aus einer der Gruppierungen zu löschen.
- f) Um einen Begriff zu ändern, wählen Sie ihn aus der Liste auf der rechten Seite aus und klicken Sie dann auf .
- g) Um einen Begriff zu löschen, wählen Sie ihn aus der Liste auf der rechten Seite aus und klicken Sie dann auf .
- h) Klicken Sie auf **OK**, um die an der Tabelle vorgenommenen Änderungen zu speichern.

8 - Schrittreferenz

In this section

| | |
|--------------------------|-----|
| Advanced Matching-Modul | 179 |
| Business Steward-Modul | 254 |
| Data Normalization-Modul | 302 |
| Universal Name-Modul | 322 |

Advanced Matching-Modul

Advanced Matching-Modul

Das Advanced Matching-Modul vergleicht Datensätze zwischen und/oder innerhalb einer beliebigen Anzahl von Eingabedateien. Mithilfe des Advanced Matching-Moduls können Sie Vergleiche auf vielen verschiedenen Feldern durchführen, einschließlich „Name“ und „Adresse“. Selbiges gilt für Felder, die nicht der Angabe von Namen oder Adressen dienen, z. B. für die Sozialversicherungsnummer oder das Geburtsdatum.

Komponenten

Das Advanced Matching-Modul setzt sich aus folgenden Komponenten zusammen:

- **Best of Breed:** Bei diesem Schritt wird ein „Best of Breed“-Datensatz aus dem Dublettencluster ausgewählt, indem ein Vorlagendatensatz ausgewählt wird. Anschließend wird mithilfe dieses Datensatzes der beste zusammengesetzte Datensatz erstellt. Dieser Datensatz wird anschließend als Survivor-Datensatz zurückgegeben.
- **Candidate Finder:** Bei diesem Schritt werden die Kandidatendatensätze abgerufen, welche die Gruppe potenzieller Übereinstimmungen bilden werden, die während des „Transactional Match“-Schrittes ausgewertet wird. Candidate Finder wird zusammen mit „Transactional Match“ verwendet.
- **Duplicate Synchronization:** Mithilfe diesem Schritt können Sie bestimmen, welche Felder aus einer Sammlung von Datensätzen in die entsprechenden Felder aller Datensätze in der Sammlung kopiert (veröffentlicht) werden.
- **Filter:** Über diesen Schritt können Sie die Kriterien festlegen, die Datensätze erfüllen müssen, um in einer Sammlung von Datensatzdubletten beibehalten oder daraus entfernt zu werden, entweder zur weiteren Downstream-Verarbeitung oder zum Schreiben in Ihre Ausgabedatei.
- **Interflow Match:** Anhand dieses Schrittes werden Übereinstimmungen zwischen ähnlichen Datensätzen unter mehreren Eingabestreams ermittelt.
- **Intraflow Match:** Anhand dieses Schrittes werden Übereinstimmungen zwischen ähnlichen Datensätzen innerhalb eines einzelnen Eingabestreams ermittelt.
- **Vergleichsanalyse:** Mithilfe der Vergleichsanalyse können Sie Ergebnisse aus Vergleichsschritten analysieren und vergleichen, um ein besseres Verständnis über das Resultat des Vergleichsprozesses zu erhalten und Ihre Vergleichsergebnisse zu verbessern.
- **Match Key Generator:** Bei diesem Schritt wird ein nicht eindeutiger Schlüssel erstellt, den alle ähnlichen Datensätze aufweisen. Datensätze mit demselben Schlüssel können für einen Vergleich gruppiert werden.
- **Private Match:** Bei diesem Schritt werden zwei Entitäten aktiviert, die Datensätze vergleichen und gemeinsame Datensätze ermitteln, ohne sensible Informationen zu gefährden. Mithilfe eines

Verschlüsselungs-Feature wird die Sicherheit aufrechterhalten, während eine Vergleichsfunktion ausgeführt wird. Ein Entschlüsselungs-Feature zeigt anschließend die Ausgabe der verglichenen Daten an.

- **Transactional Match:** Bei diesem Schritt werden Kopfdublettentransaktionen über den „Candidate Finder“-Schritt mit einer Datenbank abgeglichen, um potenzielle Kandidatendatensätze abzufragen und zurückzugeben.
- **Write to Search Index:** Mithilfe dieses Schrittes können Sie einen Volltextindex basierend auf den im Schritt eingehenden Daten erstellen. Wenn Sie diese Daten in einem fest zugeordneten Suchindex haben, erhalten Sie schnellere Antwortzeiten, wenn Sie über Candidate Finder Suchvorgänge im Index durchführen.

Best of Breed

Best of Breed konsolidiert Datensatzdubletten, indem die besten Daten in einer Sammlung von Datensatzdubletten ausgewählt werden und unter Verwendung der besten Daten ein neuer konsolidierter Datensatz erstellt wird. Dieser „Super“-Datensatz wird auch als „Best of Breed“-Datensatz bezeichnet. Die Regeln, die bei der Auswahl der zu verarbeitenden Datensätze angewendet werden, werden von Ihnen definiert. Wenn die Verarbeitung abgeschlossen ist, wird der „Best of Breed“-Datensatz im System behalten.

Optionen

In der folgenden Tabelle werden die Optionen für „Best of Breed“ aufgeführt.

| Name der Option | Beschreibung/gültige Werte |
|-----------------|---|
| Gruppieren nach | Gibt das Feld an, anhand dessen Gruppen von Datensätzen erstellt werden sollen, die in einen einzelnen „Best of Breed“-Datensatz zusammengeführt werden sollen, sodass ein „Best of Breed“-Datensatz aus jeder Gruppe erstellt wird. Wenn Sie den Vergleichsschritt früher im Datenfluss verwendet haben, sollten Sie das Feld „CollectionNumber“ auswählen, um die vom Vergleichsschritt als Gruppen erstellten Sammlungen zu verwenden. Wenn Sie jedoch Datensätze nach einem anderen Feld gruppieren möchten, wählen Sie das Feld hier aus. Wenn Sie beispielsweise alle Datensätze, die denselben Wert im Feld „AccountNumber“ aufweisen, in einem „Best of Breed“-Datensatz zusammenführen möchten, müssten Sie „AccountNumber“ auswählen. |
| Sortieren | Wenn Sie ein Feld im Gruppierungsfeld angeben, müssen Sie dieses Kästchen aktivieren, um die Datensätze nach dem Wert im ausgewählten Feld zu sortieren. Diese Option ist standardmäßig aktiviert. |

| Name der Option | Beschreibung/gültige Werte |
|-----------------|----------------------------|
|-----------------|----------------------------|

| | |
|-----------|--|
| Erweitert | |
|-----------|--|

| Name der Option | Beschreibung/gültige Werte |
|-----------------|----------------------------|
|-----------------|----------------------------|

Klicken Sie auf diese Schaltfläche, um die Optionen für die Sortierleistung anzugeben. Standardmäßig werden die Optionen für die Sortierleistung verwendet, die in der Management Console festgelegt sind. Es handelt sich dabei um die Standardleistungsoptionen für Ihr System. Wenn Sie die Standardleistungsoptionen Ihres Systems überschreiben möchten, müssen Sie das Kästchen **Optionen für Sortierleistung außer Kraft setzen** aktivieren und anschließend die für diese Felder gewünschten Werte angeben:

Beschränkung für Datensätze im Arbeitsspeicher Gibt die maximale Anzahl von Datenzeilen an, die ein Sorter im Arbeitsspeicher belässt, bevor er sie auf die Festplatte auslagert. Standardmäßig erfolgt eine Sortierung von bis zu 10.000 Datensätze in einem Arbeitsspeicher und eine Sortierung von mehr als 10.000 Datensätzen auf einem Datenträger. Der maximale Wert umfasst 100.000 Datensätze. In der Regel erfolgt eine Sortierung im Arbeitsspeicher deutlich schneller als auf einem Datenträger. Daher sollte hierfür ein Wert festgelegt werden, der hoch genug ist, damit ein Großteil der Sortierungen im Arbeitsspeicher erfolgen kann und nur umfangreiche Datasets auf einen Datenträger geschrieben werden.

Anmerkung: Vorsicht bei Umgebungen, in denen Aufträge gleichzeitig ausgeführt werden: Durch das Erhöhen der Einstellung **Beschränkung für Datensätze im Arbeitsspeicher** wird die Wahrscheinlichkeit erhöht, dass nicht mehr genügend Arbeitsspeicher verfügbar ist.

Maximale Anzahl temporärer Dateien Gibt die maximale Anzahl temporärer Dateien an, die von einem Sortiervorgang verwendet werden können. Die Verwendung einer größeren Anzahl temporärer Dateien kann zu einer besseren Leistung führen. Die optimale Anzahl hängt jedoch stark von der Konfiguration des Servers ab, auf dem Spectrum™ Technology Platform ausgeführt wird. Sie sollten mit verschiedenen Einstellungen experimentieren und dabei die Auswirkungen auf die Leistung beobachten, wenn mehr oder weniger temporäre Dateien verwendet werden. Verwenden Sie für die Berechnung der ungefähren Anzahl temporärer Dateien die folgende Gleichung:

$$\frac{(NumberOfRecords \times 2)}{InMemoryRecordLimit} = NumberOfTempFiles$$

Beachten Sie, dass die maximale Anzahl der temporären Dateien 1.000 nicht überschreiten darf.

Komprimierung aktivieren Gibt an, dass temporäre Dateien komprimiert werden, wenn sie auf die Festplatte geschrieben werden.

Anmerkung: Welche Einstellungen für die Sortierleistung optimal sind, hängt von der Hardwarekonfiguration Ihres Servers ab. Dennoch wird mit der folgenden Gleichung im Allgemeinen eine gute Sortierleistung erzielt:

| Name der Option | Beschreibung/gültige Werte |
|-----------------|----------------------------|
|-----------------|----------------------------|

$$(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2) \geq TotalNumberOfRecords$$

| | |
|--------------------------------------|--|
| Ursprüngliche Datensätze beibehalten | Wählen Sie diese Option aus, um alle Datensätze in der Sammlung zusammen mit dem „Best of Breed“-Datensatz beizubehalten. Deaktivieren Sie diese Option, wenn Sie lediglich den „Best of Breed“-Datensatz benötigen. |
|--------------------------------------|--|

| | |
|----------------------------|--|
| Ersten Datensatz verwenden | Wählen Sie diese Option aus, wenn „Best of Breed“ automatisch den ersten Datensatz in der Sammlung als den Vorlagendatensatz auswählen soll. Der Vorlagendatensatz ist der Datensatz, auf dem der „Best of Breed“-Datensatz basiert. |
|----------------------------|--|

| | |
|------------------------------|---|
| Vorlagendatensatz definieren | Wählen Sie diese Option aus, um Regeln für die Auswahl des Vorlagendatensatzes zu definieren. Weitere Informationen finden Sie unter Regeln für den Vorlagendatensatz definieren auf Seite 183. |
|------------------------------|---|

Regeln für den Vorlagendatensatz definieren

Bei der „Best of Breed“-Verarbeitung ist der Vorlagendatensatz der Datensatz in einer Sammlung, anhand dessen der „Best of Breed“-Datensatz erstellt wird. Der Vorlagendatensatz wird als Startpunkt für den Aufbau des „Best of Breed“-Datensatzes verwendet und basierend auf den von Ihnen definierten „Best of Breed“-Einstellungen geändert. Der „Best of Breed“-Schritt kann den Vorlagendatensatz automatisch auswählen. Alternativ können Sie Regeln für die Auswahl des Vorlagendatensatzes definieren. In diesem Thema wird beschrieben, wie Sie Regeln für die Auswahl des Vorlagendatensatzes definieren können.

Vorlagenregeln werden geschrieben, indem der Feldname, ein Operator, ein Werttyp und ein Wert angegeben werden. Es folgt ein Beispiel der Optionen für den Vorlagendatensatz:

Feldname: MatchScore
 Feldtyp: Numerisch
 Operator: Gleich
 Werttyp: Zeichenfolge
 Wert: 100

Diese Vorlagenregel wählt den Datensatz in der Sammlung aus, bei dem die Punktzahl dem Wert 100 entspricht.

Im folgenden Verfahren wird beschrieben, wie eine Vorlagendatensatzregel im „Best of Breed“-Schritt definiert wird.

1. Wählen Sie im „Best of Breed“-Schritt unter **Einstellungen für Vorlagendatensatz** die Option **Vorlagendatensatz definieren** aus.
2. Klicken Sie im Baum auf **Regeln**.
3. Klicken Sie auf **Regel hinzufügen**.
4. Füllen Sie die folgenden Felder aus.

| Option | Beschreibung |
|----------|--|
| Feldname | Gibt den Namen des Datenflussfeldes mit dem auszuwertenden Wert an, um zu bestimmen, ob der Datensatz der Vorlagendatensatz sein soll. |
| Feldtyp | Gibt den Datentyp im Feld an. Zur Auswahl stehen: <ul style="list-style-type: none"> Nicht numerisch Wählen Sie diese Option aus, wenn das Feld nicht numerische Daten enthält (z. B. string-Daten). Numerisch Wählen Sie diese Option aus, wenn das Feld numerische Daten enthält (z. B. double, float usw.). |

| Option | Beschreibung |
|----------------------------|---|
| Operator | Gibt den Vergleichstyp an, den Sie zur Auswertung des Feldes verwenden möchten. Zur Auswahl stehen: |
| Enthält | Bestimmt, ob das Feld den angegebenen Wert enthält. Zum Beispiel enthält „Segelboot“ den Wert „boot“. |
| Gleich | Bestimmt, ob das Feld exakt den angegebenen Wert enthält. |
| Größer als | Bestimmt, ob der Feldwert größer als der angegebene Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Größer oder gleich | Bestimmt, ob der Feldwert größer als oder gleich dem angegebenen Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Höchste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den höchsten Wert im Feld aufweist. Wenn die Felder in der Gruppe beispielsweise Werte von 10, 20, 30 und 100 enthalten, würde der Datensatz mit dem Feldwert 100 ausgewählt werden. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Ist leer | Bestimmt, ob das Feld keinen Wert enthält. |
| Ist nicht leer | Bestimmt, ob das Feld einen Wert enthält. |
| Kleiner als | Bestimmt, ob der Feldwert kleiner als der angegebene Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Kleiner oder gleich | Bestimmt, ob der Feldwert kleiner als oder gleich dem angegebenen Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Längste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den längsten Wert (in Bytes) im Feld aufweist. Wenn die Gruppe beispielsweise die Werte „Mike“ und „Michael“ enthält, wird der Datensatz mit dem Wert „Michael“ ausgewählt. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Niedrigste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den niedrigsten Wert im Feld aufweist. Wenn die Felder in der Gruppe beispielsweise Werte von 10, 20, 30 und 100 enthalten, würde der Datensatz mit dem Feldwert 10 ausgewählt werden. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Häufigste | Bestimmt, ob der Feldwert den Wert enthält, der in diesem Feld unter den Datensätzen in der Gruppe am häufigsten auftritt. Wenn zwei oder mehr Werte am häufigsten auftreten, werden keine Aktionen durchgeführt. |
| Ungleich | Bestimmt, ob der Feldwert nicht mit dem angegebenen Wert übereinstimmt. |

| Option | Beschreibung |
|---------|---|
| Werttyp | <p>Gibt den Werttyp an, den Sie mit dem Wert des Feldes vergleichen möchten. Zur Auswahl stehen:</p> <p>Anmerkung: Diese Option ist nicht verfügbar, wenn Sie den Operator „Höchste“, „Niedrigste“ oder „Längste“ auswählen.</p> <p>Feld Wählen Sie diese Option, wenn Sie den Wert eines anderen Datenflussfeldes mit dem Feld vergleichen möchten.</p> <p>Zeichenfolge Wählen Sie diese Option, wenn Sie das Feld mit einem spezifischen Wert vergleichen möchten.</p> |
| Wert | <p>Gibt den Wert an, der mit dem Wert des Feldes verglichen werden soll. Wenn Sie im Feld Feldtyp den Eintrag „Feld“ ausgewählt haben, müssen Sie ein Datenflussfeld auswählen. Wenn Sie im Feld Werttyp den Eintrag „Zeichenfolge“ ausgewählt haben, müssen Sie den bei dem Vergleich zu verwendenden Wert eingeben.</p> <p>Anmerkung: Diese Option ist nicht verfügbar, wenn Sie den Operator „Höchste“, „Niedrigste“ oder „Längste“ auswählen.</p> |

- Klicken Sie auf **OK**.
- Wenn Sie zusätzliche Regeln festlegen möchten, müssen Sie auf **Regel hinzufügen** klicken.

Wenn Sie zusätzliche Regeln hinzufügen, müssen Sie einen logischen Operator auswählen, der zwischen jeder Regel zu verwenden ist. Wählen Sie **And** aus, wenn die neue Regel und die vorherige Regel beide erfüllt sein müssen, damit der Datensatz als Vorlagendatensatz ausgewählt wird. Wählen Sie **Or** aus, wenn entweder die vorherige Regel oder die neue Regel erfüllt sein muss, damit der Datensatz als Vorlagendatensatz ausgewählt wird.

Sie haben nun Regeln konfiguriert, die zur Auswahl des Vorlagendatensatzes verwendet werden sollen. Konfigurieren Sie die Best-of-Breed-Einstellungen, um die Konfiguration des „Best of Breed“-Schrittes abzuschließen.

Definieren von „Best of Breed“-Regeln und -Aktionen

„Best of Breed“-Regeln und -Aktionen interagieren, um zu bestimmen, welche Felder aus Datensatzdubletten in einer Sammlung in den „Best of Breed“-Datensatz kopiert werden. Die Regeln testen Werte in einem Datensatz. Wenn der Datensatz die Regeln erfüllt, werden die Daten aus dem Datensatz in den Vorlagendatensatz kopiert. Aktionen definieren die zu kopierenden Daten und das Feld in dem Vorlagendatensatz, das die Daten erhalten soll. Nachdem alle Regeln und Aktionen ausgeführt wurden, wird der Vorlagendatensatz zum „Best of Breed“-Datensatz.

Regeln und Aktionen können zusammen als Bedingungen gruppiert werden. Sie können mehrere Bedingungen haben. Dadurch können Sie Folgendes ausführen:

1. Klicken Sie im „Best of Breed“-Schritt im Baum unter **Best-of-Breed-Einstellungen** auf den Knoten **Regeln**.
2. Klicken Sie auf **Regel hinzufügen**.
3. Füllen Sie die folgenden Felder aus:

| Option | Beschreibung |
|----------|---|
| Feldname | Gibt den Namen des Datenflussfeldes mit dem auszuwertenden Wert an, um zu bestimmen, ob die Bedingung erfüllt ist und die dazugehörigen Aktionen ausgeführt werden sollen. |
| Feldtyp | Gibt den Datentyp im Feld an. Zur Auswahl stehen: Nicht numerisch Wählen Sie diese Option aus, wenn das Feld nicht numerische Daten enthält (z. B. string-Daten). Numerisch Wählen Sie diese Option aus, wenn das Feld numerische Daten enthält (z. B. double, float usw.). |

| Option | Beschreibung |
|----------------------------|---|
| Operator | Gibt den Vergleichstyp an, den Sie zur Auswertung des Feldes verwenden möchten. Zur Auswahl stehen: |
| Enthält | Bestimmt, ob das Feld den angegebenen Wert enthält. Zum Beispiel enthält „Segelboot“ den Wert „boot“. |
| Gleich | Bestimmt, ob das Feld exakt den angegebenen Wert enthält. |
| Größer als | Bestimmt, ob der Feldwert größer als der angegebene Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Größer oder gleich | Bestimmt, ob der Feldwert größer als oder gleich dem angegebenen Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Höchste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den höchsten Wert im Feld aufweist. Wenn die Felder in der Gruppe beispielsweise Werte von 10, 20, 30 und 100 enthalten, würde der Datensatz mit dem Feldwert 100 ausgewählt werden. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Ist leer | Bestimmt, ob das Feld keinen Wert enthält. |
| Ist nicht leer | Bestimmt, ob das Feld einen Wert enthält. |
| Kleiner als | Bestimmt, ob der Feldwert kleiner als der angegebene Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Kleiner oder gleich | Bestimmt, ob der Feldwert kleiner als oder gleich dem angegebenen Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Längste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den längsten Wert (in Bytes) im Feld aufweist. Wenn die Gruppe beispielsweise die Werte „Mike“ und „Michael“ enthält, wird der Datensatz mit dem Wert „Michael“ ausgewählt. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Niedrigste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den niedrigsten Wert im Feld aufweist. Wenn die Felder in der Gruppe beispielsweise Werte von 10, 20, 30 und 100 enthalten, würde der Datensatz mit dem Feldwert 10 ausgewählt werden. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Häufigste | Bestimmt, ob der Feldwert den Wert enthält, der in diesem Feld unter den Datensätzen in der Gruppe am häufigsten auftritt. Wenn zwei oder mehr Werte am häufigsten auftreten, werden keine Aktionen durchgeführt. |
| Ungleich | Bestimmt, ob der Feldwert nicht mit dem angegebenen Wert übereinstimmt. |

| Option | Beschreibung |
|---------|---|
| Werttyp | <p>Gibt den Werttyp an, den Sie mit dem Wert des Feldes vergleichen möchten. Zur Auswahl stehen:</p> <p>Anmerkung: Diese Option ist nicht verfügbar, wenn Sie den Operator „Höchste“, „Niedrigste“ oder „Längste“ auswählen.</p> <p>Feld Wählen Sie diese Option, wenn Sie den Wert eines anderen Datenflussfeldes mit dem Feld vergleichen möchten.</p> <p>Zeichenfolge Wählen Sie diese Option, wenn Sie das Feld mit einem spezifischen Wert vergleichen möchten.</p> |

| | |
|------|--|
| Wert | <p>Gibt den Wert an, der mit dem Wert des Feldes verglichen werden soll. Wenn Sie im Feld Feldtyp den Eintrag „Feld“ ausgewählt haben, müssen Sie ein Datenflussfeld auswählen. Wenn Sie im Feld Werttyp den Eintrag „Zeichenfolge“ ausgewählt haben, müssen Sie den bei dem Vergleich zu verwendenden Wert eingeben.</p> <p>Anmerkung: Diese Option ist nicht verfügbar, wenn Sie den Operator „Höchste“, „Niedrigste“ oder „Längste“ auswählen.</p> |
|------|--|

- Klicken Sie auf **OK**.
- Wenn Sie zusätzliche Regeln für diese Bedingung festlegen möchten, müssen Sie auf **Regel hinzufügen** klicken.

Wenn Sie zusätzliche Regeln hinzufügen, müssen Sie einen logischen Operator auswählen, der zwischen jeder Regel zu verwenden ist. Wählen Sie **And** aus, wenn die neue Regel und die vorherige Regel beide erfüllt sein müssen, damit die Bedingung erfüllt ist und die dazugehörigen Aktionen ausgeführt werden. Wählen Sie **Or** aus, wenn entweder die vorherige Regel oder die neue Regel erfüllt sein muss, damit die Bedingung erfüllt ist.

- Klicken Sie im Baum auf den Knoten **Aktionen**.
- Klicken Sie auf **Aktion hinzufügen**.
- Füllen Sie die folgenden Felder aus.

| Option | Beschreibung |
|------------|---|
| Quellentyp | <p>Gibt den Datentyp an, der in den „Best of Breed“-Datensatz kopiert werden soll. Die folgenden Optionen stehen zur Auswahl:</p> <p>Feld Wählen Sie diese Option aus, wenn Sie einen Wert aus einem Feld in den „Best of Breed“-Datensatz kopieren möchten.</p> <p>Zeichenfolge Wählen Sie diese Option aus, wenn Sie einen konstanten Wert in den „Best of Breed“-Datensatz kopieren möchten.</p> |

| Option | Beschreibung |
|-----------------------|--|
| Quelldaten | Gibt die Daten an, die in den „Best of Breed“-Datensatz kopiert werden sollen. Wenn der Quelltyp „Feld“ ist, müssen Sie das Feld auswählen, dessen Wert in das Zielfeld kopiert werden soll. Wenn der Quelltyp „Zeichenfolge“ ist, müssen Sie einen konstanten Wert angeben, der in das Zielfeld kopiert werden soll. |
| Ziel | Gibt das Feld im „Best of Breed“-Datensatz an, in das die im Feld Quelldaten angegebenen Daten kopiert werden sollen. |
| Quelldaten kumulieren | <p>Wenn es sich bei den Daten im Feld Quelldaten um numerische Daten handelt, können Sie diese Option aktivieren, um die Quelldaten für alle Datensatzdubletten zu kombinieren und den Gesamtwert in den „Best of Breed“-Datensatz einzufügen.</p> <p>Beispiel: Es sind drei Datensatzdubletten in der Gruppe vorhanden und weisen die folgenden Werte im Feld „Einzahlungen“ auf:</p> <p>100,00 20,00 5,00</p> <p>Alle drei Werte werden addiert und die Summe (125,00) wird in das Feld „Einzahlungen“ des „Best of Breed“-Datensatzes eingefügt.</p> |

9. Klicken Sie auf **OK**.
10. Wenn Sie weitere Aktionen angeben möchten, die für diese Bedingung durchgeführt werden sollen, müssen Sie auf **Aktion hinzufügen** klicken und die Schritte oben wiederholen.
11. Klicken Sie zum Hinzufügen einer weiteren Bedingung auf die Stammbedingung im Baum, und klicken Sie anschließend auf **Bedingung hinzufügen**.

Beispiel einer „Best of Breed“-Regel und -Aktion

Anhand dieser „Best of Breed“-Regel wird der Datensatz ausgewählt, bei dem die Punktzahl dem Wert 100 entspricht. Die Kontonummerndaten, die mit den ausgewählten Feldern übereinstimmen, werden anschließend in das Feld „AccountNumber“ im „Best of Breed“-Datensatz kopiert.

Regel

Feldname: MatchScore
 Feldtyp: Numerisch
 Operator: Gleich
 Werttyp: Zeichenfolge
 Wert: 100

Aktion

Quelltyp: Feld

| |
|--|
| Quelldaten: AccountNumber Ziel: AccountNumber |
|--|

Ausgabe

Tabelle 8: „Best of Breed“-Ausgabe

| Feldname | Format | Beschreibung/gültige Werte |
|----------------------|--------------------|---|
| CollectionRecordType | Zeichenfolge | Gibt die Vorlage- und „Best of Breed“-Datensätze in einer Sammlung von Datensatzdubletten an. Die folgenden Werte sind möglich: |
| | Primary | Der Datensatz ist der ausgewählte Vorlagendatensatz in einer Sammlung. |
| | Secondary | Der Datensatz ist nicht der ausgewählte Vorlagendatensatz in einer Sammlung. |
| | BestOfBreed | Der Datensatz ist der neu erstellte „Best of Breed“-Datensatz in der Sammlung. |

Candidate Finder

Candidate Finder ruft die Kandidatendatensätze ab, welche die Gruppe potenzieller Übereinstimmungen bilden werden. Datenbanksuchen interagieren mit Transactional Match, während Suchindexsuchen unabhängig von Transactional Match arbeiten. Je nach dem Format Ihrer Daten muss Candidate Finder gegebenenfalls auch den Namen oder die Adresse des Kopfdubletten-Datensatzes, der Kandidatendatensätze oder den Namen/die Adresse von beiden parsen.

Candidate Finder ermöglicht außerdem Volltext-Indexsuchen und unterstützt Sie bei der Definition von Suchkriterien für Zeichen und Texte anhand von Suchtypen (Jedes Wort, das beginnt mit, Enthält, Enthält alle, Enthält beliebiges, Enthält keines, Unschärf, Muster, Nähe, Bereich, Platzhalter) und Bedingungen (Alle Übereinstimmungen, Beliebige Übereinstimmung, Keine Übereinstimmung).

Datenbankoptionen

Mithilfe des Dialogfeldes „Candidate Finder“ können Sie SQL-Anweisungen definieren, die potenzielle Übereinstimmungskandidaten aus einer Datenbank abfragen, und die aus der Datenbank ausgewählten Spalten den Feldnamen zuordnen, die in Ihrem Datenfluss definiert sind.

Tabelle 9: Datenbankoptionen für Candidate Finder

| Name der Option | Beschreibung/gültige Werte |
|-------------------------------|--|
| Finder-Typ | Wählen Sie die Datenbank aus. |
| Verbindung | <p>Wählen Sie die Datenbank mit den Kandidatendatensätzen aus. Sie können eine beliebige, auf der Management Console konfigurierte Verbindung auswählen. Um eine nicht aufgeführte Datenbank zu verbinden, müssen Sie in der Management Console eine Verbindung zu dieser Datenbank konfigurieren, und anschließend „Candidate Finder“ schließen und erneut öffnen, um die Liste der Verbindungen zu aktualisieren.</p> <p>Anmerkung: Das Feature „Datenflussoptionen“ im Enterprise Designer ermöglicht es, den Verbindungsnamen für die Konfiguration zur Laufzeit verfügbar zu machen.</p> |
| SQL-Anweisung | Geben Sie wie in Definieren der SQL-Abfrage auf Seite 192 beschrieben eine SQL-Anweisung in das Textfeld ein. |
| Registerkarte „Feldzuordnung“ | Wählen Sie die Einstellungen für das Feld-Mapping wie in Mapping von Schrittfeldern zu Datenbankspalten auf Seite 193 beschrieben aus. |
| Registerkarte „Vorschau“ | Klicken Sie auf diese Registerkarte, um einen Test-Vergleichsschlüssel einzugeben, um Ihre SQL-Anweisung <code>SELECT</code> auf Ihrer Indexabfrage zu testen. |

Definieren der SQL-Abfrage

Sie können eine beliebige gültige SQL `SELECT`-Anweisung in das Textfeld im Dialogfeld „**Candidate Finder**“-Optionen eingeben.

Anmerkung: `Select *` ist nicht zulässig.

Angenommen, Sie haben eine Tabelle mit Namen „Customer_Table“ in Ihrer Datenbank, die die folgenden Spalten aufweist:

- Customer_Table
- Cust_Name
- Cust_Address
- Cust_City
- Cust_State
- Cust_Zip

Um alle Zeilen aus der Datenbank abzurufen, könnten Sie eine Abfrage ähnlich der folgenden erstellen:

```
SELECT Cust_Name, Cust_Address, Cust_City, Cust_State, Cust_Zip from
Customer_Table;
```

Sie werden nur selten Ihre Transaktion mit allen Zeilen in der Datenbank vergleichen wollen. Um nur die relevanten Kandidatendatensätze zurückzugeben, fügen Sie eine `WHERE`-Klausel hinzu und nutzen Sie die Variablenersetzung. Mit „Variablenersetzung“ ist eine spezielle Syntax gemeint, die Sie verwenden, um die Candidate Selection-Engine zu veranlassen, die Variable durch die eigentlichen Daten aus Ihrem Kopfdubletten-Datensatz zu ersetzen.

Um die Variablenersetzung zu verwenden, schließen Sie den Feldnamen in geschweiften Klammern ein, denen Sie ein Dollarzeichen voranstellen: `${FieldName}`. Beispiel: Die folgende Abfrage gibt nur Datensätze mit einem Wert in „Cust_Zip“ zurück, der mit dem Wert in „PostalCode“ im Kopfdubletten-Datensatz übereinstimmt.

```
SELECT Cust_Name, Cust_Address, Cust_City, Cust_State, Cust_Zip
FROM Customer_Table
WHERE Cust_Zip = ${PostalCode};
```

Für SQL 2000 muss der Datentyp mit dem Datentyp für Candidate Finder identisch sein. Der JDBC-Treiber legt die Candidate Finder-Eingabevariable (z. B. `${MatchKey}`), die in der `WHERE`-Klausel verwendet wird, auf den Datentyp „`nVarChar(4000)`“ fest. Wenn die Daten in der Datenbank auf den Datentyp „`VarChar`“ festgelegt werden, ignoriert der SQL-Server den Index in der Datenbank. Wenn der Index ignoriert wird, sinkt die Leistung. Verwenden Sie deshalb die folgende Abfrage für SQL 2000:

```
SELECT Cust_Name, Cust_Address, Cust_City, Cust_State, Cust_Zip
FROM Customer_Table
WHERE Cust_Zip = CAST(${PostalCode} AS VARCHAR(255));
```

Mapping von Schrittfeldern zu Datenbankspalten

Wenn die Spaltennamen in Ihrer Datenbank genau mit den Komponentenfeldnamen übereinstimmen, werden sie automatisch den entsprechenden Schrittfeldern zugeordnet. Wenn sie nicht genau gleich benannt sind, müssen Sie „Ausgewählte Felder“ (Spalten aus der Datenbank) verwenden, um sie den Schrittfeldern (im Datenfluss definierte Feldnamen) zuzuordnen.

Nehmen wir zum Beispiel eine Tabelle mit Namen „Customer_Table“ mit den folgenden Spalten:

- Cust_Name
- Cust_Address
- Cust_City
- Cust_State
- Cust_Zip

Wenn Sie diese Datensätze aus der Datenbank abrufen, müssen Sie die Spaltennamen den Feldnamen zuordnen, die durch „Transactional Match“ und andere Komponenten in Ihrem Datenfluss

verwendet werden. So könnte beispielsweise „Cust_Address“ „AddressLine1“ und „Cust_Zip“ „PostalCode“ zugeordnet werden.

1. Wählen Sie im Dialogfeld **Candidate Finder-Optionen** die Dropdown-Liste unter **Ausgewählte Felder** aus. Wählen Sie danach die Datenbankspalte „Cust_Zip“ aus.
2. Wählen Sie die Dropdown-Liste unter **Schrittfelder** aus. Wählen Sie danach das Feld aus, das Sie zuordnen möchten.

Wenn Sie beispielsweise „Cust_Zip“ „Postal Code“ zuordnen möchten, wählen Sie zuerst „Cust_Zip“ unter „Ausgewählte Felder“ und dann „PostalCode“ in der entsprechenden Zeile „Schrittfeld“ aus.

Alternative Methode zum Mapping von Feldern

Sie können eine spezielle Syntax in Ihrer SQL-Abfrage verwenden, um das Mapping vorzunehmen. Schließen Sie dazu den Feldnamen, dem Sie einen Namen zuordnen möchten, nach dem Spaltennamen in Ihrer Abfrage in geschweifte Klammern ein. Wenn Sie das tun, werden die ausgewählten Felder automatisch den entsprechenden Schrittfeldern zugeordnet.

Beispiel:

```
select Cust_Name {Name}, Cust_Address {AddressLine1},
       Cust_City {City}, Cust_State {StateProvince},
       Cust_Zip {PostalCode}
from Customer
where Cust_Zip = ${PostalCode};
```

Konfigurieren von Verbindungsnamen zur Laufzeit

Der Verbindungsname kann zur Laufzeit konfiguriert und weitergeleitet werden, wenn er als eine Datenflussoption verfügbar gemacht wird. Dadurch können Sie Ihren Datenfluss ausführen, während Sie einen anderen Verbindungsnamen verwenden.

1. Öffnen Sie im Enterprise Designer einen Datenfluss, der den „Candidate Finder“-Schritt verwendet.
2. Speichern Sie diesen Datenfluss und machen Sie ihn verfügbar.
3. Öffnen Sie **Bearbeiten > Datenflussoptionen**.
4. Erweitern Sie Candidate Finder in der Tabelle **Datenflussoptionen zu Schritten zuordnen**, und bearbeiten Sie die Optionen nach Bedarf. Aktivieren Sie das Kästchen für die zu bearbeitende Option, und ändern Sie anschließend den Wert in der Dropdown-Liste **Standardwert**.
5. Optional: Ändern Sie den Namen der Optionen im Feld **Optionsbeschriftung**.
6. Klicken Sie zweimal auf **OK**.

Suchindexoptionen

Mithilfe des Dialogfeldes „Candidate Finder“ können Sie eine einfache Suche, bei der Eingabefeldwerte innerhalb von Suchindizes verglichen werden, oder eine erweiterte Suche

durchführen, um Vergleichsregeln zu erstellen, die potenzielle Übereinstimmungskandidaten aus Suchindizes abrufen.

Indexoptionen für die einfache Suche

Tabelle 10: „Candidate Finder“-Optionen

| Name der Option | Beschreibung/gültige Werte |
|--|---|
| Finder-Typ | Wählen Sie den Suchindex aus. |
| Name | Wählen Sie den entsprechenden Index aus, der mithilfe des Schrittes Write to Search Index auf Seite 247 im Rahmen der bereitgestellten „Advanced Matching“-Schritte im Enterprise Designer erstellt wurde. |
| Startdatensatz | Geben Sie die Datensatznummer ein, bei der die Suchergebnisse beginnen sollen. Der Standardwert ist 1. |
| Maximale Anzahl Ergebnisse | Geben Sie die maximale Anzahl an Ergebnissen ein, die vom Suchindex zurückgegeben werden sollen. Der Standardwert ist 10. Anmerkung: Wenn „Maximale Anzahl Ergebnisse“ beliebig groß ist, verarbeiten Sie sie in Batches mithilfe des Feldes Batch-Größe abrufen . |
| Batch-Größe abrufen | Wenn Maximale Anzahl Ergebnisse beliebig groß ist, geben Sie die Größe der Batches ein, in denen die Ergebnisse verarbeitet werden sollen. Dies optimiert die Verarbeitung einer großen Anzahl von Datensätzen. Der Standardwert ist 10000. Der empfohlene Wert für eine Batch-Größe ist ein Wert kleiner als der Wert Maximale Anzahl Ergebnisse , und wenn der Wert der Batch-Größe größer als der Wert von Maximale Anzahl Ergebnisse ist, werden die Datensätze in einem einzigen Batch verarbeitet. Anmerkung: Dieses Feld gilt nur für die vom Cluster unterstützte Suchmaschine und nicht für die Legacy-Suchmaschine. |
| Anzahl der Übereinstimmungen zurückgeben | Gibt die Summe der ermittelten Übereinstimmungen zurück. Wenn Sie beispielsweise den Standardwert von 10 für das obige Feld „Maximale Anzahl Ergebnisse“ verwenden, werden nur 10 Ergebnisse zurückgegeben. Wenn Sie dieses Kästchen allerdings aktivieren, wird im Ausgabefeld „Gesamtanzahl Übereinstimmungen“ angezeigt, wie viele Übereinstimmungen während der Verarbeitung ermittelt wurden. |

| Name der Option | Beschreibung/gültige Werte |
|-----------------|---|
| Indexsuchtyp | Bestimmt den Typ der Indexsuche, die Sie durchführen möchten. Wählen Sie Einfache Suche aus. |
| Indexfelder | Wählen Sie die Indexfelder aus, die Sie für den Vergleich in der einfachen Suche verwenden möchten. |
| Eingabefeld | Wählen Sie das Eingabefeld aus, das Sie für den Vergleich in der einfachen Suche verwenden möchten. |

| Name der Option | Beschreibung/gültige Werte |
|-----------------|----------------------------|
|-----------------|----------------------------|

| | |
|-------------------|--|
| Eingabeanalysator | |
|-------------------|--|

Name der Option

Beschreibung/gültige Werte

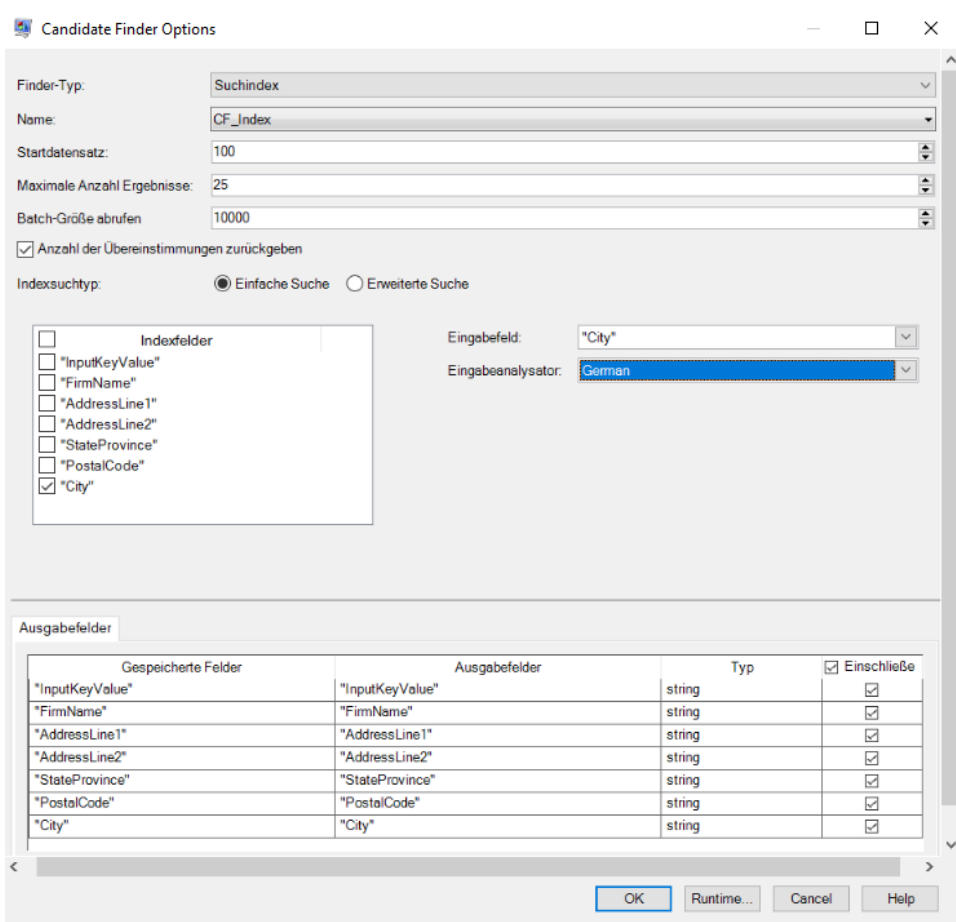
Geben Sie an, welcher Analysator zum Tokenisieren der Eingabezeichenfolge verwendet werden soll. Eine von diesen:

- Standard: Bietet einen grammatikbasierten Token-Ersteller, der ein Superset der Analysatoren „Leerzeichen“ und „Stoppwort“ enthält. Erkennt die englische Interpunktion zum Aufschlüsseln von Wörtern, kennt die zu ignorierenden Wörter (über den Stoppwort-Analysator) und führt Suchen durch, welche die Groß-/Kleinschreibung nicht beachten, indem er Vergleiche in Kleinschreibung durchführt. Die Zeichenfolge „Pitney Bowes Software“ würde beispielsweise als drei Token zurückgegeben werden: „Pitney“, „Bowes“ und „Software“.
- Leerzeichen: Trennt Token mit einem Leerzeichen. Eine Art Untergruppe des standardmäßigen Analysators, da Wortumbrüche in englischen Texten basierend auf Leerzeichen und Zeilenumbrüchen erkannt werden.
- StopWord: Entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
- Schlüsselwort: Entfernt ein Token aus einem Datenstream. Die Zeichenfolge „Pitney Bowes Software“ würde beispielsweise als ein einzelner Token zurückgegeben werden: Pitney Bowes Software.
- Russisch: Unterstützt Indizes und Type-Ahead-Dienste in russischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „und“, „ich“ und „du“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
- Deutsch: Unterstützt Indizes und Type-Ahead-Dienste in deutscher Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
- Dänisch: Unterstützt Indizes und Type-Ahead-Dienste in dänischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „um“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
- Niederländisch: Unterstützt Indizes und Type-Ahead-Dienste in niederländischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
- Finnisch: Unterstützt Indizes und Type-Ahead-Dienste in finnischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „ist“, „und“ und „von“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
- Französisch: Unterstützt Indizes und Type-Ahead-Dienste in französischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
- Ungarisch: Unterstützt Indizes und Type-Ahead-Dienste in ungarischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
- Italienisch: Unterstützt Indizes und Type-Ahead-Dienste in italienischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
- Norwegisch: Unterstützt Indizes und Type-Ahead-Dienste in norwegischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“,

| Name der Option | Beschreibung/gültige Werte |
|-------------------------------|--|
| | <p>„und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.</p> <ul style="list-style-type: none"> • Portugiesisch: Unterstützt Indizes und Type-Ahead-Dienste in portugiesischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern. • Spanisch: Unterstützt Indizes und Type-Ahead-Dienste in spanischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern. • Schwedisch: Unterstützt Indizes und Type-Ahead-Dienste in schwedischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern. • Hindi: Unterstützt Indizes und Type-Ahead-Dienste in Hindi. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „von“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern. |
| Registerkarte „Ausgabefelder“ | <p>Aktivieren Sie das Kästchen Einschließen, um auszuwählen, welche gespeicherten Felder in der Ausgabe enthalten sein sollen.</p> <p>Anmerkung: Wenn das Eingabefeld aus einem früheren Schritt im Datenfluss stammt und denselben Namen wie der gespeicherte Feldname aus dem Suchindex aufweist, überschreiben die Werte aus dem Eingabefeld die Werte im Ausgabefeld.</p> |

Im Bildschirm unten wird ein Beispiel eines abgeschlossenen „Candidate Finder-Optionen“-Schrittes mithilfe einer einfachen Indexsuche aufgeführt:

- Ein Suchindex mit dem **Namen** „CF_Index“
- Ein **Startdatensatz**: von 100, d. h., dass die Suchergebnisse beim 100. Datensatz beginnen werden.
- **Maximale Anzahl Ergebnisse**, auf 25 eingestellt, d. h., dass nur 25 Ergebnisse zurückgegeben werden sollen
- **Batch-Größe abrufen**, um Ergebnisse in Batches zu verarbeiten
- Die Option **Anzahl der Übereinstimmungen zurückgeben** ist aktiviert, sodass alle Datensätze zurückgegeben werden, statt nur die 25, in die wir diese Ansicht eingrenzen
- Ein einfacher **Indexsuchtyp**
- Ein **Indexfeld** von „City“ zum Abgleich mit dem Eingabefeld „City“
- Ein **Eingabefeld** von „City“ zum Abgleich mit dem Indexfeld „City“
- Ein Eingabeanalysator auf Deutsch zum Vergleich von Feldern
- Eine Feldzuordnung, mit der angezeigt wird, dass alle Felder in der Ausgabe zurückgegeben werden



Indexoptionen für die erweiterte Suche

Tabelle 11: „Candidate Finder“-Optionen

| Name der Option | Beschreibung/gültige Werte |
|-----------------|--|
| Finder-Typ | Wählen Sie den Suchindex aus. |
| Name | Wählen Sie den entsprechenden Index aus, der mithilfe des Schrittes Write to Search Index auf Seite 247 im Rahmen der bereitgestellten „Advanced Matching“-Schritte im Enterprise Designer erstellt wurde. |
| Startdatensatz | Geben Sie die Datensatznummer ein, bei der die Suchergebnisse beginnen sollen. Der Standardwert ist 1. |

| Name der Option | Beschreibung/gültige Werte |
|--|---|
| Maximale Anzahl Ergebnisse | <p>Geben Sie die maximale Anzahl an Antworten ein, die von der Indexsuche zurückgegeben werden sollen. Der Standardwert ist 10.</p> <p>Anmerkung: Wenn „Maximale Anzahl Ergebnisse“ beliebig groß ist, verarbeiten Sie sie in Batches mithilfe des Feldes Batch-Größe abrufen.</p> |
| Batch-Größe abrufen | <p>Wenn Maximale Anzahl Ergebnisse beliebig groß ist, geben Sie die Größe der Batches ein, in denen die Ergebnisse verarbeitet werden sollen. Dies optimiert die Verarbeitung einer großen Anzahl von Datensätzen. Der Standardwert ist 10000.</p> <p>Der empfohlene Wert für eine Batch-Größe ist ein Wert kleiner als der Wert Maximale Anzahl Ergebnisse, und wenn der Wert der Batch-Größe größer als der Wert von Maximale Anzahl Ergebnisse ist, werden die Datensätze in einem einzigen Batch verarbeitet.</p> <p>Anmerkung: Dieses Feld gilt nur für die vom Cluster unterstützte Suchmaschine und nicht für die Legacy-Suchmaschine.</p> |
| Anzahl der Übereinstimmungen zurückgeben | <p>Gibt die Summe der ermittelten Übereinstimmungen zurück. Wenn Sie beispielsweise den Standardwert von 10 für das obige Feld „Maximale Anzahl Ergebnisse“ verwenden, werden nur 10 Ergebnisse zurückgegeben. Wenn Sie dieses Kästchen allerdings aktivieren, wird im Ausgabefeld „Gesamtanzahl Übereinstimmungen“ angezeigt, wie viele Übereinstimmungen während der Verarbeitung ermittelt wurden.</p> |
| Indexsuchtyp | <p>Bestimmt den Typ der Indexsuche, die Sie durchführen möchten. Wählen Sie Erweiterte Suche aus.</p> |
| Schaltfläche „Übergeordnetes Element hinzufügen“ | <p>Greifen Sie auf übergeordnete Optionen zu.</p> |
| Übergeordnete Optionen – Name | <p>Geben Sie einen Namen für das übergeordnete Element ein.</p> |

| Name der Option | Beschreibung/gültige Werte |
|---|---|
| Übergeordnete Optionen – Suchmethode | <p>Geben Sie an, wie bestimmt werden soll, ob es sich bei einem übergeordneten Element um eine Übereinstimmung handelt. Eine von diesen:</p> <p>Alle Übereinstimmungen: Ein übergeordnetes Element wird als Übereinstimmung erkannt, wenn alle untergeordneten Elemente übereinstimmen. Über diese Methode wird eine AND-Verbindungsstraße zwischen den untergeordneten Elementen erstellt.</p> <p>Beliebige Übereinstimmung: Ein übergeordnetes Element wird als Übereinstimmung erkannt, wenn mindestens ein untergeordnetes Element übereinstimmt. Über diese Methode wird eine OR-Verbindungsstraße zwischen den untergeordneten Elementen erstellt.</p> <p>Keine Übereinstimmung: Ein übergeordnetes Element wird als Übereinstimmung erkannt, wenn keines der untergeordneten Elemente übereinstimmt. Über diese Methode wird eine NOT-Verbindungsstraße zwischen den untergeordneten Elementen erstellt.</p> |
| Schaltfläche „Untergeordnetes Element hinzufügen“ | Greifen Sie auf untergeordnete Optionen zu. |
| Untergeordnete Optionen – Indexfeld | Wählen Sie das Indexfeld aus, das Sie für den Vergleich in der erweiterten Suche verwenden möchten. |
| Untergeordnete Optionen – Suchtyp | Gibt die Such-/Übereinstimmungskriterien an, die bestimmen, ob die Eingabedaten anhand der indizierten Daten gesucht/abgeglichen werden. Bei allen Suchvorgängen wird die Groß-/Kleinschreibung beachtet. |
| Untergeordnete Optionen – Eingabefeld | Wählen Sie das Eingabefeld aus, das Sie für den Vergleich in der erweiterten Suche verwenden möchten. |
| Jedes Wort/jeder Ausdruck, der/das beginnt mit | <p>Bestimmt, ob der Text im Suchindexfeld mit dem Text aus dem Eingabefeld beginnt.</p> <p>Der Text im Eingabefeld „tech“ würde beispielsweise als Übereinstimmung für Suchindexfelder gelten, die „Technisch“, „Technologie“, „Technologien“, „Techniker“ oder sogar „National University of Technical Sciences“ enthalten. Ebenso würde ein Ausdruck im Eingabefeld „DEF Sof“ als Übereinstimmung für Suchindexfelder mit dem Inhalt „ABC DEF Software“, „DEF Software“ und „DEF Software India“ gelten, jedoch nicht für Suchindexfelder mit dem Inhalt „Software DEF“ oder „DEF ABC Software“.</p> |

| Name der Option | Beschreibung/gültige Werte |
|--------------------|---|
| Enthält | Bestimmt, ob das Suchindexfeld Daten aus dem Eingabefeld enthält. Bei diesem Suchtyp wird die Reihenfolge der Wörter im Eingabefeld bei der Suche im Suchindexfeld beachtet. Die Eingabefelddaten „Pitney“ und „Pitney Bowes“ würden beispielsweise in einem Suchindexfeld von „Pitney Bowes Software Inc.“ enthalten sein. |
| Enthält alle | Bestimmt, ob das Suchindexfeld alle alphanumerischen Wörter aus dem Eingabefeld enthält. Bei diesem Suchtyp wird die Reihenfolge der Wörter im Eingabefeld bei der Suche im Suchindexfeld nicht beachtet. |
| Enthält beliebiges | Bestimmt, ob das Suchindexfeld eines der alphanumerischen Wörter aus dem Eingabefeld enthält. |
| Enthält keines | Bestimmt, ob das Suchindexfeld keines der alphanumerischen Wörter aus dem Eingabefeld enthält. |
| Unscharf | <p>Bestimmt die Ähnlichkeit zwischen zwei alphanumerischen Wörtern anhand der Anzahl der Löschungen, Einfügungen oder Ersetzungen, die zum Umwandeln eines Wortes in ein anderes erforderlich sind.</p> <p>Legen Sie mithilfe des Parameters Maximale Anzahl Bearbeitungen einen Grenzwert für die Anzahl an Bearbeitungen fest, die zugelassen sind, um als erfolgreiche Übereinstimmung zu gelten:</p> <ul style="list-style-type: none"> • 0: Lässt keine Löschungen, Ergänzungen oder Ersetzungen zu. Die Daten im Eingabefeld und im Suchindexfeld müssen identisch sein. • 1: Lässt höchstens eine Löschung, Ergänzung oder Ersetzung zu. Ein Eingabefeld mit dem Inhalt „Barton“ wird mit einem Suchindexfeld mit dem Inhalt „Carton“ übereinstimmen. • 2: Lässt höchstens zwei Löschungen, Ergänzungen oder Ersetzungen zu. Ein Eingabefeld mit dem Inhalt „Barton“ wird mit einem Suchindexfeld mit dem Inhalt „Martin“ übereinstimmen. <p>Der Suchtyp „Unscharf“ wird ausschließlich für die Suche nach Einzelwörtern verwendet. Klicken Sie auf Zusätzliche Wörter ignorieren, damit Candidate Finder beim Vergleich des Eingabefeldes mit dem Indexfeld lediglich das erste Wort im Feld berücksichtigt. Wenn das Indexfeld beispielsweise „Pitney“ und das Eingabefeld „Pitney Bowes“ enthält, würde keine Übereinstimmung aufgrund des Wortes „Bowes“ erzielt werden. Wenn Sie jedoch dieses Kästchen aktivieren, wird „Bowes“ ignoriert. Da „Pitney“ das erste Wort ist, würden die beiden Wörter als Übereinstimmung gelten.</p> |

| Name der Option | Beschreibung/gültige Werte |
|-----------------|--|
| Numerisch | <p>Bestimmt, ob das Suchindexfeld Nummern aus dem Eingabefeld enthält.</p> <p>Der Suchtyp „Numerisch“ wird ausschließlich für die Suche nach Einzelwörtern verwendet.</p> <p>Klicken Sie auf Zusätzliche Wörter ignorieren, damit Candidate Finder beim Vergleich des Eingabefeldes mit dem Indexfeld lediglich das erste Wort im Feld berücksichtigt.</p> |
| Muster | <p>Bestimmt, ob das Textmuster des Eingabefeldes mit dem Textmuster der Suchkriterien übereinstimmt. Sie können das Textmuster weiter im Feld Musterzeichenfolge verfeinern. Wenn das Eingabefeld beispielsweise „nlm“ enthält und das definierte Muster „a*b?c“ lautet, wird es mit den folgenden Wörtern übereinstimmen: „Neelam“, „nelam“, „neelum“, „nilam“ usw.</p> <p>Der Suchtyp „Muster“ wird ausschließlich für die Suche nach Einzelwörtern verwendet. Klicken Sie auf Zusätzliche Wörter ignorieren, damit Candidate Finder beim Vergleich des Eingabefeldes mit dem Indexfeld lediglich das erste Wort im Feld berücksichtigt.</p> |
| Nähe | <p>Bestimmt, ob zwischen Wörtern in den Eingabefeldern ein bestimmter Abstand besteht.</p> <ul style="list-style-type: none"> • Definieren Sie die Eingabefelder Erste Eingabe und Zweite Eingabe, nach denen Sie im Index suchen möchten. • Bestimmen Sie anhand des Parameters Entfernung die maximal zugelassene Entfernung zwischen den Wörtern, die in den Feldern „Erste“ und „Zweite“ eingegeben wurden, um als Übereinstimmung zu gelten. <p>Sie könnten diesen Suchtyp beispielsweise verwenden, um nach dem ersten Feld „Spectrum“ und dem zweiten Feld „Pitney“ innerhalb von zehn Wörtern Entfernung voneinander in einem Suchindexfeld zu suchen, das den Satz „Spectrum Technology Platform ist ein Produkt von Pitney Bowes Software Inc.“ enthält.</p> <p>Der Suchtyp „Nähe“ wird ausschließlich für die Suche nach Einzelwörtern verwendet. Klicken Sie auf Zusätzliche Wörter ignorieren, damit Candidate Finder beim Vergleich des Eingabefeldes mit dem Indexfeld lediglich das erste Wort im Feld berücksichtigt.</p> |

| Name der Option | Beschreibung/gültige Werte |
|--|--|
| Bereich | <p>Führt eine inklusive Suche nach Begriffen in einem Bereich durch. Dieser wird mit einem Feld für die Untergrenze (Anfangsbegriff) und einem Feld für die Obergrenze (Endbegriff) definiert. Alle alphanumerischen Wörter sind im Suchindexfeld lexikografisch angeordnet.</p> <ul style="list-style-type: none"> • Wählen Sie anhand des Parameters Feld Untergrenze das Feld aus, das als Startbegriff verwendet werden soll. • Wählen Sie anhand des Parameters Feld Obergrenze das Feld aus, das als Endbegriff verwendet werden soll. <p>Wenn Sie beispielsweise nach Postleitzahlen zwischen 20001 (unter „Feld Untergrenze“ definiert) und 20009 (unter „Feld Obergrenze“ definiert) gesucht haben, würde die Suche alle Adressen mit Postleitzahlen in diesem Bereich zurückgeben.</p> <p>Der Suchtyp „Bereich“ wird ausschließlich für die Suche nach Einzelwörtern verwendet. Klicken Sie auf Zusätzliche Wörter ignorieren, damit Candidate Finder beim Vergleich des Eingabefeldes mit dem Indexfeld lediglich das erste Wort im Feld berücksichtigt.</p> |
| Platzhalter | <p>Sucht mit Platzhaltern für ein oder mehrere Zeichen.</p> <p>Wählen Sie die Position in Ihrer Eingabedatei aus, in die Sie das Platzhalterzeichen einfügen möchten.</p> <p>Der Suchtyp „Platzhalter“ wird ausschließlich für die Suche nach Einzelwörtern verwendet. Klicken Sie auf Zusätzliche Wörter ignorieren, damit Candidate Finder beim Vergleich des Eingabefeldes mit dem Indexfeld lediglich das erste Wort im Feld berücksichtigt.</p> |
| Untergeordnete Optionen – Relevanzfaktor | <p>Steuern Sie die Relevanz eines untergeordneten Feldes, indem Sie eine Zahl bis 100 hier eingeben. Je höher der Boost-Faktor ist, desto relevanter wird ein Feld. Wenn beispielsweise Ergebnisse aus dem Feld „Firmenname“ relevanter als die Ergebnisse aus anderen Feldern sein sollen, müssen Sie „Firmenname“ aus dem Indexfeldnamen auswählen und hier eine „5“ eingeben.</p> <p>Anmerkung: Hier eingegebene Zahlen müssen positiv sein, können jedoch unter „1“ liegen, z. B. wäre „0,05“ gültig.</p> |
| Leerzeichen ignorieren | <p>Deaktivieren Sie dieses Kästchen, wenn die Abfrage die leeren Felder der Eingabedatei berücksichtigen soll.</p> <p>Anmerkung: Standardmäßig ignoriert die Abfrage leere Felder.</p> |

| Name der Option | Beschreibung/gültige Werte |
|-------------------------------|---|
| Registerkarte „Ausgabefelder“ | <p>Aktivieren Sie das Kästchen Einschließen, um auszuwählen, welche gespeicherten Felder in der Ausgabe enthalten sein sollen.</p> <p>Anmerkung: Wenn das Eingabefeld aus einem früheren Schritt im Datenfluss stammt und denselben Namen wie der gespeicherte Feldname aus dem Suchindex aufweist, überschreiben die Werte aus dem Eingabefeld die Werte im Ausgabefeld.</p> |

Im Bildschirm unten ist ein Beispiel des abgeschlossenen Schrittes „Candidate Finder-Optionen“ mithilfe einer erweiterten Indexsuche aufgeführt:

- Ein Suchindex mit dem **Namen** „CF_Index“
- Ein **Startdatensatz** von 26, d. h., dass die Suchergebnisse beim 26. Datensatz beginnen werden.
- **Maximale Anzahl Ergebnisse**, auf 10 eingestellt, d. h., dass nur 10 Ergebnisse zurückgegeben werden sollen
- **Batch-Größe abrufen**, um Ergebnisse in Batches zu verarbeiten
- Die Option **Anzahl der Übereinstimmungen zurückgeben** ist aktiviert, sodass alle Datensätze zurückgegeben werden, statt nur die 10, in die wir diese Ansicht eingrenzen
- Ein erweiterter **Indexsuchtyp**
- Ein **Typ des übergeordneten Elements** mit der Bezeichnung „Neuer Typ“
- Ein **Typ des untergeordneten Elements** mit der Bezeichnung „StateProvince“, basierend auf dem Indexfeldnamen
- Ein **Indexfeld** von „StateProvince“ zum Vergleich mit dem Eingabefeld „StateProvince“
- Ein Suchtyp **Unschärf**, bei dem **Maximale Anzahl Bearbeitungen** auf 2 eingestellt ist, sodass bei einer erfolgreichen Übereinstimmung bis zu zwei Bearbeitungen möglich sind
- Ein **Eingabefeld** von „StateProvince“ zum Abgleich mit dem Indexfeld „StateProvince“
- Ein **Relevanzfaktor** von 1,0, um die Relevanz der Daten zum Bundesland/Kanton zu erhöhen
- Eine Feldzuordnung, mit der angezeigt wird, dass alle Felder in der Ausgabe zurückgegeben werden

Konfigurieren von Optionen zur Laufzeit

Einige „Candidate Finder“-Optionen können zur Laufzeit konfiguriert und weitergeleitet werden, wenn sie als Datenflussoptionen verfügbar gemacht werden. Dadurch können Sie Ihren Datenfluss ausführen, während Sie unterschiedliche Konfigurationen verwenden. Die folgenden Datenflussoptionen sind für Candidate Finder verfügbar:

- **Verbindungsname:** Der Name der Datenbank mit den Kandidatendatensätzen.
- **Suchindexname:** Der Name des Suchindex, der im Datenfluss von Candidate Finder verwendet wird.
- **Startdatensatz:** Die Datensatznummer, bei der die Suchergebnisse beginnen sollen.
- **Maximale Anzahl Ergebnisse:** Die maximale Anzahl an Antworten, die von der Indexsuche zurückgegeben werden sollen.
- **Anzahl der Übereinstimmungen zurückgeben:** Die Summe der ermittelten Übereinstimmungen. Dieses Feld ist hilfreich, wenn Sie eine niedrigere Zahl in das Feld „Maximale Anzahl Ergebnisse“ eingeben, jedoch die Summe der ermittelten Übereinstimmungen erhalten möchten.

So definieren Sie die „Candidate Finder“-Optionen zur Laufzeit:

1. Öffnen Sie im Enterprise Designer einen Datenfluss, der den „Candidate Finder“-Schritt verwendet.
2. Speichern Sie diesen Datenfluss und machen Sie ihn verfügbar.
3. Öffnen Sie **Bearbeiten > Datenflussoptionen**.
4. Erweitern Sie Candidate Finder in der Tabelle **Datenflussoptionen zu Schritten zuordnen**, und bearbeiten Sie die Optionen nach Bedarf. Aktivieren Sie das Kästchen für die zu

bearbeitende Option, und ändern Sie anschließend den Wert in der Dropdown-Liste **Standardwert**.

5. Optional: Ändern Sie den Namen der Optionen im Feld **Optionsbeschriftung**.
6. Klicken Sie zweimal auf **OK**.

Ausgabe

Tabelle 12: Ausgaben von Candidate Finder

| Feldname | Format | Beschreibung/gültige Werte | | | | |
|-----------------------|---|--|----------------|--|------------------|---|
| CandidateGroup | Zeichenfolge | Über dieses Feld wird die Gruppierung eines Kopfdubletten-Datensatzes und seiner Kandidaten identifiziert. Jedem Kopfdubletten-Datensatz wird eine Kandidatengruppennummer zugewiesen. Den Kandidaten dieser Kopfdublette wird dieselbe Kandidatengruppennummer zugewiesen. Wenn „John Smith“ beispielsweise ein Kopfdubletten-Datensatz ist und seine Kandidatendatensätze „John Smith“ und „Jon Smth“ sind, hätten alle drei Datensätze denselben Kandidatengruppenwert. | | | | |
| TotalMatchCount | Zeichenfolge | In diesem Feld wird die Gesamtzahl der Übereinstimmungen angegeben, die während der Verarbeitung ermittelt wurden. | | | | |
| TransactionRecordType | Zeichenfolge | Zur Auswahl stehen: <table border="0" style="margin-left: 20px;"> <tr> <td style="padding-right: 10px;">Suspect</td> <td>Ein Kopfdubletten-Datensatz wird als Eingabe für eine Abfrage verwendet.</td> </tr> <tr> <td>Candidate</td> <td>Ein Kandidatendatensatz ist ein Ergebnis, das aus einer Abfrage zurückgegeben wird.</td> </tr> </table> | Suspect | Ein Kopfdubletten-Datensatz wird als Eingabe für eine Abfrage verwendet. | Candidate | Ein Kandidatendatensatz ist ein Ergebnis, das aus einer Abfrage zurückgegeben wird. |
| Suspect | Ein Kopfdubletten-Datensatz wird als Eingabe für eine Abfrage verwendet. | | | | | |
| Candidate | Ein Kandidatendatensatz ist ein Ergebnis, das aus einer Abfrage zurückgegeben wird. | | | | | |

Duplicate Synchronization

„Duplicate Synchronization“ bestimmt, welche Felder aus einer Sammlung von Datensätzen in die entsprechenden Felder aller Datensätze in der Sammlung kopiert werden sollen. Sie können die Regeln festlegen, die Datensätze erfüllen müssen, um die Felddaten in andere Datensätze in der Sammlung zu kopieren. Nach Abschluss der Verarbeitung werden alle Datensätze in der Sammlung beibehalten.

Optionen

In der folgenden Tabelle sind die Optionen für den **Duplicate Synchronization**-Schritt aufgeführt.

| Name der Option | Beschreibung/gültige Werte |
|-----------------|--|
| Gruppieren nach | Gibt das Feld an, das zum Erstellen von zu synchronisierenden Datensatzgruppen verwendet werden soll. Wenn Sie den Vergleichsschritt früher im Datenfluss verwendet haben, z. B. bei Interflow Match, Intraflow Match oder Transactional Match, sollten Sie das Feld „CollectionNumber“ auswählen, um die vom Vergleichsschritt als Gruppen erstellten Sammlungen zu verwenden. Wenn Sie jedoch Datensätze nach einem anderen Feld gruppieren möchten, wählen Sie das Feld hier aus. Wenn Sie beispielsweise Datensätze mit demselben Wert im Feld „AccountNumber“ synchronisieren möchten, müssten Sie „AccountNumber“ auswählen. |
| Sortieren | Wenn Sie ein Feld im Gruppierungsfeld angeben, müssen Sie dieses Kästchen aktivieren, um die Datensätze nach dem Wert im ausgewählten Feld zu sortieren. Diese Option ist standardmäßig aktiviert. |

| Name der Option | Beschreibung/gültige Werte |
|-----------------|----------------------------|
|-----------------|----------------------------|

| | |
|-----------|--|
| Erweitert | |
|-----------|--|

Name der Option

Beschreibung/gültige Werte

Klicken Sie auf diese Schaltfläche, um die Optionen für die Sortierleistung anzugeben. Standardmäßig werden die Optionen für die Sortierleistung verwendet, die in der Management Console festgelegt sind. Es handelt sich dabei um die Standardleistungsoptionen für Ihr System. Wenn Sie die Standardleistungsoptionen Ihres Systems überschreiben möchten, müssen Sie das Kästchen **Optionen für Sortierleistung außer Kraft setzen** aktivieren und anschließend die für diese Felder gewünschten Werte angeben:

Beschränkung für Datensätze im Arbeitsspeicher Gibt die maximale Anzahl von Datenzeilen an, die ein Sorter im Arbeitsspeicher belässt, bevor er sie auf die Festplatte auslagert. Standardmäßig erfolgt eine Sortierung von bis zu 10.000 Datensätzen in einem Arbeitsspeicher und eine Sortierung von mehr als 10.000 Datensätzen auf einem Datenträger. Der maximale Wert umfasst 100.000 Datensätze. In der Regel erfolgt eine Sortierung im Arbeitsspeicher deutlich schneller als auf einem Datenträger. Daher sollte hierfür ein Wert festgelegt werden, der hoch genug ist, damit ein Großteil der Sortierungen im Arbeitsspeicher erfolgen kann und nur umfangreiche Datasets auf einen Datenträger geschrieben werden.

Anmerkung: Vorsicht bei Umgebungen, in denen Aufträge gleichzeitig ausgeführt werden: Durch das Erhöhen der Einstellung **Beschränkung für Datensätze im Arbeitsspeicher** wird die Wahrscheinlichkeit erhöht, dass nicht mehr genügend Arbeitsspeicher verfügbar ist.

Maximale Anzahl temporärer Dateien Gibt die maximale Anzahl temporärer Dateien an, die von einem Sortiervorgang verwendet werden können. Die Verwendung einer größeren Anzahl temporärer Dateien kann zu einer besseren Leistung führen. Die optimale Anzahl hängt jedoch stark von der Konfiguration des Servers ab, auf dem Spectrum™ Technology Platform ausgeführt wird. Sie sollten mit verschiedenen Einstellungen experimentieren und dabei die Auswirkungen auf die Leistung beobachten, wenn mehr oder weniger temporäre Dateien verwendet werden. Verwenden Sie für die Berechnung der ungefähren Anzahl temporärer Dateien die folgende Gleichung:

$$\frac{(NumberOfRecords \times 2)}{InMemoryRecordLimit} = NumberOfTempFiles$$

Beachten Sie, dass die maximale Anzahl der temporären Dateien 1.000 nicht überschreiten darf.

Komprimierung aktivieren Gibt an, dass temporäre Dateien komprimiert werden, wenn sie auf die Festplatte geschrieben werden.

Anmerkung: Welche Einstellungen für die Sortierleistung optimal sind,

| Name der Option | Beschreibung/gültige Werte |
|-----------------|----------------------------|
|-----------------|----------------------------|

hängt von der Hardwarekonfiguration Ihres Servers ab. Dennoch wird mit der folgenden Gleichung im Allgemeinen eine gute Sortierleistung erzielt:

$$(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2) \geq TotalNumberOfRecords$$

Regeln

Die „Duplicate Synchronization“-Regeln bestimmen, aus welchen Datensätzen die Daten in alle anderen Datensätze in der Sammlung kopiert werden sollen.

Wählen Sie zum Hinzufügen einer Regel „Regeln“ in der Regelhierarchie aus, und klicken Sie auf **Regel hinzufügen**.

Wenn Sie mehrere Regeln angeben, müssen Sie einen logischen Operator auswählen, der zwischen jeder Regel zu verwenden ist. Wählen Sie **And** aus, wenn die neue Regel und die vorherige Regel beide erfüllt sein müssen, damit die Bedingung erfüllt ist. Wählen Sie **Or** aus, wenn entweder die vorherige Regel oder die neue Regel erfüllt sein muss, damit die Bedingung erfüllt ist.

| Option | Beschreibung |
|--------|--------------|
|--------|--------------|

| | |
|----------|---|
| Feldname | Gibt den Namen des Datenflussfeldes an, dessen Wert Sie auswerten möchten, um zu bestimmen, ob der Datensatz gefiltert werden soll. |
|----------|---|

| | |
|---------|--|
| Feldtyp | Gibt den Datentyp im Feld an. Zur Auswahl stehen: <ul style="list-style-type: none"> Nicht numerisch Wählen Sie diese Option aus, wenn das Feld nicht numerische Daten enthält (z. B. string-Daten). Numerisch Wählen Sie diese Option aus, wenn das Feld numerische Daten enthält (z. B. double, float usw.). |
|---------|--|

| Option | Beschreibung |
|--------|--------------|
|--------|--------------|

| | |
|----------|--|
| Operator | |
|----------|--|

Option

Beschreibung

Gibt den Vergleichstyp an, den Sie zur Auswertung des Feldes verwenden möchten. Zur Auswahl stehen:

| | |
|----------------------------|---|
| Enthält | Bestimmt, ob das Feld den angegebenen Wert enthält. Zum Beispiel enthält „Segelboot“ den Wert „boot“. |
| Gleich | Bestimmt, ob das Feld exakt den angegebenen Wert enthält. |
| Größer als | Bestimmt, ob der Feldwert größer als der angegebene Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Größer oder gleich | Bestimmt, ob der Feldwert größer als oder gleich dem angegebenen Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Höchste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den höchsten Wert im Feld aufweist. Wenn die Felder in der Gruppe beispielsweise Werte von 10, 20, 30 und 100 enthalten, würde der Datensatz mit dem Feldwert 100 ausgewählt werden. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Ist leer | Bestimmt, ob das Feld keinen Wert enthält. |
| Ist nicht leer | Bestimmt, ob das Feld einen Wert enthält. |
| Kleiner als | Bestimmt, ob der Feldwert kleiner als der angegebene Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Kleiner oder gleich | Bestimmt, ob der Feldwert kleiner als oder gleich dem angegebenen Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Längste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den längsten Wert (in Bytes) im Feld aufweist. Wenn die Gruppe beispielsweise die Werte „Mike“ und „Michael“ enthält, wird der Datensatz mit dem Wert „Michael“ ausgewählt. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Niedrigste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den niedrigsten Wert im Feld aufweist. Wenn die Felder in der Gruppe beispielsweise Werte von 10, 20, 30 und 100 enthalten, würde der Datensatz mit dem Feldwert 10 ausgewählt werden. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Häufigste | Bestimmt, ob der Feldwert den Wert enthält, der in diesem Feld unter den Datensätzen in der Gruppe am häufigsten auftritt. Wenn zwei oder mehr Werte am häufigsten auftreten, werden keine Aktionen durchgeführt. |
| Ungleich | Bestimmt, ob der Feldwert nicht mit dem angegebenen Wert |

| Option | Beschreibung |
|---------|---|
| | übereinstimmt. |
| Werttyp | <p>Gibt den Werttyp an, den Sie mit dem Wert des Feldes vergleichen möchten. Zur Auswahl stehen:</p> <p>Anmerkung: Diese Option ist nicht verfügbar, wenn Sie den Operator „Höchste“, „Niedrigste“ oder „Längste“ auswählen.</p> <p>Feld Wählen Sie diese Option, wenn Sie den Wert eines anderen Datenflussfeldes mit dem Feld vergleichen möchten.</p> <p>Zeichenfolge Wählen Sie diese Option, wenn Sie das Feld mit einem spezifischen Wert vergleichen möchten.</p> |
| Wert | <p>Gibt den Wert an, der mit dem Wert des Feldes verglichen werden soll. Wenn Sie im Feld Feldtyp den Eintrag „Feld“ ausgewählt haben, müssen Sie ein Datenflussfeld auswählen. Wenn Sie im Feld Werttyp den Eintrag „Zeichenfolge“ ausgewählt haben, müssen Sie den bei dem Vergleich zu verwendenden Wert eingeben.</p> <p>Anmerkung: Diese Option ist nicht verfügbar, wenn Sie den Operator „Höchste“, „Niedrigste“ oder „Längste“ auswählen.</p> |

Aktionen

Aktionen bestimmen, welches Feld in andere Datensätze in der Gruppe kopiert werden soll. Um eine Aktion hinzuzufügen, müssen Sie „Aktion“ im Bedingungsbaum „Duplicate Synchronization“ auswählen und anschließend auf **Aktion hinzufügen** klicken. Definieren Sie die Aktion anhand der folgenden Optionen.

| Option | Beschreibung |
|----------|---|
| Quelltyp | <p>Gibt den Datentyp an, der in andere Datensätze in der Gruppe kopiert werden soll. Die folgenden Optionen stehen zur Auswahl:</p> <p>Feld Wählen Sie diese Option aus, wenn Sie einen Wert aus einem Feld in die anderen Datensätze in der Gruppe kopieren möchten.</p> <p>Zeichenfolge Wählen Sie diese Option aus, wenn Sie einen konstanten Wert in die anderen Datensätze in der Gruppe kopieren möchten.</p> |

| Option | Beschreibung |
|------------|--|
| Quelldaten | <p>Gibt die Daten an, die in die anderen Datensätze in der Gruppe kopiert werden sollen. Wenn der Quelltyp „Field“ ist, müssen Sie das Feld auswählen, dessen Wert in die anderen Datensätze in der Gruppe kopiert werden soll. Wenn der Quelltyp „String“ ist, müssen Sie einen konstanten Wert angeben, der in die anderen Datensätze in der Gruppe kopiert werden soll.</p> <p>Anmerkung: Wenn die Quelldaten Nullwerte aufweisen, werden sie nicht zu den anderen Datensätzen in der Gruppe kopiert. Die anderen Datensätze behalten ihre ursprünglichen Werte.</p> |
| Ziel | <p>Gibt das Feld in den anderen Datensätzen an, in das die im Feld Quelldaten angegebenen Daten kopiert werden sollen. Wenn Sie beispielsweise die Daten in das Feld „AccountBalance“ in allen anderen Datensätzen in der Gruppe kopieren möchten, müssten Sie „AccountBalance“ angeben.</p> |

Beispiel einer „Duplicate Synchronization“-Regel und -Aktion

Diese „Duplicate Synchronization“-Regel und -Aktion wählen den Datensatz aus, bei dem die Punktzahl 100 beträgt, und kopieren das Feld „AccountNumber“ in alle anderen Datensätze in der Gruppe.

Regel

Feldname: MatchScore
 Feldtyp: Numerisch
 Operator: Gleich
 Werttyp: Zeichenfolge
 Wert: 100

Aktion

Quelltyp: Feld
 Quelldaten: AccountNumber
 Ziel: NewAccountNumber

Filter

Der Filter-Schritt entfernt oder behält Datensätze aus einer Gruppe von Datensätzen basierend auf den von Ihnen festgelegten Regeln bei.

Optionen

In der folgenden Tabelle sind die Optionen für den Filter-Schritt aufgeführt.

| Name der Option | Beschreibung/gültige Werte |
|-----------------|--|
| Gruppieren nach | <p>Gibt das Feld an, das zum Erstellen von zu filternden Datensatzgruppen verwendet werden soll. Der Filter-Schritt wird einen oder mehrere Datensätze aus jeder Gruppe beibehalten, je nachdem, wie Sie den Schritt konfigurieren. Wenn Sie den Vergleichsschritt früher im Datenfluss verwendet haben, z. B. bei Interflow Match, Intraflow Match oder Transactional Match, sollten Sie das Feld „CollectionNumber“ auswählen, um die vom Vergleichsschritt als Gruppen erstellten Sammlungen zu verwenden. Wenn Sie jedoch Datensätze nach einem anderen Feld gruppieren möchten, wählen Sie das Feld hier aus. Wenn Sie beispielsweise alle bis auf einen Datensatz aus den Datensätzen mit demselben Wert im Feld „AccountNumber“ herausfiltern möchten, müssten Sie „AccountNumber“ auswählen.</p> |
| Sortieren | <p>Wenn Sie ein Feld im Gruppierungsfeld angeben, müssen Sie dieses Kästchen aktivieren, um die Datensätze nach dem Wert im ausgewählten Feld zu sortieren. Diese Option ist standardmäßig aktiviert.</p> |

| Name der Option | Beschreibung/gültige Werte |
|-----------------|----------------------------|
|-----------------|----------------------------|

| | |
|-----------|--|
| Erweitert | |
|-----------|--|

Name der Option

Beschreibung/gültige Werte

Klicken Sie auf diese Schaltfläche, um die Optionen für die Sortierleistung anzugeben. Standardmäßig werden die Optionen für die Sortierleistung verwendet, die in der Management Console festgelegt sind. Es handelt sich dabei um die Standardleistungsoptionen für Ihr System. Wenn Sie die Standardleistungsoptionen Ihres Systems überschreiben möchten, müssen Sie das Kästchen **Optionen für Sortierleistung außer Kraft setzen** aktivieren und anschließend die für diese Felder gewünschten Werte angeben:

Beschränkung für Datensätze im Arbeitsspeicher Gibt die maximale Anzahl von Datenzeilen an, die ein Sorter im Arbeitsspeicher belässt, bevor er sie auf die Festplatte auslagert. Standardmäßig erfolgt eine Sortierung von bis zu 10.000 Datensätzen in einem Arbeitsspeicher und eine Sortierung von mehr als 10.000 Datensätzen auf einem Datenträger. Der maximale Wert umfasst 100.000 Datensätze. In der Regel erfolgt eine Sortierung im Arbeitsspeicher deutlich schneller als auf einem Datenträger. Daher sollte hierfür ein Wert festgelegt werden, der hoch genug ist, damit ein Großteil der Sortierungen im Arbeitsspeicher erfolgen kann und nur umfangreiche Datasets auf einen Datenträger geschrieben werden.

Anmerkung: Vorsicht bei Umgebungen, in denen Aufträge gleichzeitig ausgeführt werden: Durch das Erhöhen der Einstellung **Beschränkung für Datensätze im Arbeitsspeicher** wird die Wahrscheinlichkeit erhöht, dass nicht mehr genügend Arbeitsspeicher verfügbar ist.

Maximale Anzahl temporärer Dateien Gibt die maximale Anzahl temporärer Dateien an, die von einem Sortiervorgang verwendet werden können. Die Verwendung einer größeren Anzahl temporärer Dateien kann zu einer besseren Leistung führen. Die optimale Anzahl hängt jedoch stark von der Konfiguration des Servers ab, auf dem Spectrum™ Technology Platform ausgeführt wird. Sie sollten mit verschiedenen Einstellungen experimentieren und dabei die Auswirkungen auf die Leistung beobachten, wenn mehr oder weniger temporäre Dateien verwendet werden. Verwenden Sie für die Berechnung der ungefähren Anzahl temporärer Dateien die folgende Gleichung:

$$\frac{(NumberOfRecords \times 2)}{InMemoryRecordLimit} = NumberOfTempFiles$$

Beachten Sie, dass die maximale Anzahl der temporären Dateien 1.000 nicht überschreiten darf.

Komprimierung aktivieren Gibt an, dass temporäre Dateien komprimiert werden, wenn sie auf die Festplatte geschrieben werden.

Anmerkung: Welche Einstellungen für die Sortierleistung optimal sind,

| Name der Option | Beschreibung/gültige Werte |
|-----------------|----------------------------|
|-----------------|----------------------------|

hängt von der Hardwarekonfiguration Ihres Servers ab. Dennoch wird mit der folgenden Gleichung im Allgemeinen eine gute Sortierleistung erzielt:

$$(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2) \geq TotalNumberOfRecords$$

| | |
|---|---|
| Anzahl zurückgegebener doppelter Datensätze beschränken | Gibt die maximale Anzahl der Datensätze an, die aus jeder Gruppe zurückgegeben werden. Wenn Sie diese Option auf 1 einstellen, können Sie Filterregeln definieren, um zu bestimmen, welcher Datensatz in jeder Gruppe zurückgegeben werden soll. Wenn keine Regeln definiert sind, wird der erste Datensatz in jeder Sammlung zurückgegeben und der Rest wird verworfen. In diesem Modus definieren die Filterregeln, welcher Datensatz beibehalten wird. |
|---|---|

Wenn Sie beispielsweise eine Regel definieren, bei der der Datensatz mit der höchsten Punktzahl in einer Gruppe beibehalten wird, und Sie diese Option auf 1 einstellen, wird der Datensatz mit der höchsten Punktzahl in jeder Gruppe beibehalten und die anderen Datensätze in der Gruppe werden verworfen.

Wenn Sie diese Option auf einen Wert über 1 einstellen, können Sie keine Filterregeln festlegen.

Anmerkung: Wenn keiner der Datensätze in der Sammlung den definierten Regelkriterien entspricht, werden keine Datensätze aus der Gruppe zurückgegeben.

| | |
|----------------------------------|---|
| Duplikate aus Sammlung entfernen | Gibt an, dass Filterregeln verwendet werden sollen, um zu bestimmen, welche Datensätze aus der Sammlung entfernt werden. Die restlichen Datensätze in der Sammlung werden beibehalten. Wenn diese Option aktiviert ist, müssen Sie eine Regel definieren. |
|----------------------------------|---|

Anmerkung: Wenn lediglich ein Datensatz in einer Gruppe enthalten ist, werden die Filterregeln ignoriert und der Datensatz wird beibehalten.

Regeloptionen

Filterregeln bestimmen, welche Datensätze in einer Gruppe beibehalten oder entfernt werden sollen. Wenn Sie die Option **Anzahl zurückgegebener doppelter Datensätze beschränken** auswählen, bestimmen die Regeln, welche Datensätze nach dem Filtern erhalten bleiben. Wenn Sie die Option **Duplikate aus Sammlung entfernen** auswählen, bestimmen die Regeln, welche Datensätze aus dem Datenfluss entfernt werden.

Wählen Sie zum Hinzufügen einer Regel „Regeln“ in der Regelhierarchie aus, und klicken Sie auf **Regel hinzufügen**.

Wenn Sie mehrere Regeln angeben, müssen Sie einen logischen Operator auswählen, der zwischen jeder Regel zu verwenden ist. Wählen Sie **And** aus, wenn die neue Regel und die vorherige Regel

beide erfüllt sein müssen, damit die Bedingung erfüllt ist. Wählen Sie **Or** aus, wenn entweder die vorherige Regel oder die neue Regel erfüllt sein muss, damit die Bedingung erfüllt ist.

Anmerkung: Es kann lediglich eine Bedingung in einem Filter-Schritt vorhanden sein. Wenn Sie „Bedingung“ in der Regelhierarchie auswählen, werden die Schaltflächen abgeblendet.

| Option | Beschreibung |
|----------|---|
| Feldname | Gibt den Namen des Datenflussfeldes an, dessen Wert Sie auswerten möchten, um zu bestimmen, ob der Datensatz gefiltert werden soll. |
| Feldtyp | Gibt den Datentyp im Feld an. Zur Auswahl stehen: Nicht numerisch Wählen Sie diese Option aus, wenn das Feld nicht numerische Daten enthält (z. B. string-Daten). Numerisch Wählen Sie diese Option aus, wenn das Feld numerische Daten enthält (z. B. double, float usw.). |

| Option | Beschreibung |
|--------|--------------|
|--------|--------------|

| | |
|----------|--|
| Operator | |
|----------|--|

Option

Beschreibung

Gibt den Vergleichstyp an, den Sie zur Auswertung des Feldes verwenden möchten. Zur Auswahl stehen:

| | |
|----------------------------|---|
| Enthält | Bestimmt, ob das Feld den angegebenen Wert enthält. Zum Beispiel enthält „Segelboot“ den Wert „boot“. |
| Gleich | Bestimmt, ob das Feld exakt den angegebenen Wert enthält. |
| Größer als | Bestimmt, ob der Feldwert größer als der angegebene Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Größer oder gleich | Bestimmt, ob der Feldwert größer als oder gleich dem angegebenen Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Höchste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den höchsten Wert im Feld aufweist. Wenn die Felder in der Gruppe beispielsweise Werte von 10, 20, 30 und 100 enthalten, würde der Datensatz mit dem Feldwert 100 ausgewählt werden. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Ist leer | Bestimmt, ob das Feld keinen Wert enthält. |
| Ist nicht leer | Bestimmt, ob das Feld einen Wert enthält. |
| Kleiner als | Bestimmt, ob der Feldwert kleiner als der angegebene Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Kleiner oder gleich | Bestimmt, ob der Feldwert kleiner als oder gleich dem angegebenen Wert ist. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. |
| Längste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den längsten Wert (in Bytes) im Feld aufweist. Wenn die Gruppe beispielsweise die Werte „Mike“ und „Michael“ enthält, wird der Datensatz mit dem Wert „Michael“ ausgewählt. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Niedrigste | Vergleicht den Wert des Feldes für die gesamte Datensatzgruppe und bestimmt, welcher Datensatz den niedrigsten Wert im Feld aufweist. Wenn die Felder in der Gruppe beispielsweise Werte von 10, 20, 30 und 100 enthalten, würde der Datensatz mit dem Feldwert 10 ausgewählt werden. Dieser Vorgang kann nur auf numerischen Feldern ausgeführt werden. Wenn mehrere Datensätze gleichzeitig den längsten Wert aufweisen, wird ein Datensatz ausgewählt. |
| Häufigste | Bestimmt, ob der Feldwert den Wert enthält, der in diesem Feld unter den Datensätzen in der Gruppe am häufigsten auftritt. Wenn zwei oder mehr Werte am häufigsten auftreten, werden keine Aktionen durchgeführt. |
| Ungleich | Bestimmt, ob der Feldwert nicht mit dem angegebenen Wert |

| Option | Beschreibung |
|---------|---|
| | übereinstimmt. |
| Werttyp | <p>Gibt den Werttyp an, den Sie mit dem Wert des Feldes vergleichen möchten. Zur Auswahl stehen:</p> <p>Anmerkung: Diese Option ist nicht verfügbar, wenn Sie den Operator „Höchste“, „Niedrigste“ oder „Längste“ auswählen.</p> <p>Feld Wählen Sie diese Option, wenn Sie den Wert eines anderen Datenflussfeldes mit dem Feld vergleichen möchten.</p> <p>Zeichenfolge Wählen Sie diese Option, wenn Sie das Feld mit einem spezifischen Wert vergleichen möchten.</p> |
| Wert | <p>Gibt den Wert an, der mit dem Wert des Feldes verglichen werden soll. Wenn Sie im Feld Feldtyp den Eintrag „Feld“ ausgewählt haben, müssen Sie ein Datenflussfeld auswählen. Wenn Sie im Feld Werttyp den Eintrag „Zeichenfolge“ ausgewählt haben, müssen Sie den bei dem Vergleich zu verwendenden Wert eingeben.</p> <p>Anmerkung: Diese Option ist nicht verfügbar, wenn Sie den Operator „Höchste“, „Niedrigste“ oder „Längste“ auswählen.</p> |

Beispiel einer Filterregel

Diese Regel behält in jeder Gruppe den Datensatz mit dem höchsten Wert im Feld „MatchScore“ bei. Hinweis: Die Optionen **Wert** und **Werttyp** gelten nicht, wenn der Operator „Höchste“ oder „Niedrigste“ eingestellt ist.

Feldname = MatchScore
 Feldtyp = Numerisch
 Operator = Höchste

Die Regel behält den Datensatz mit dem Wert „12345“ im Feld „AccountNumber“ bei.

Feldname = AccountNumber
 Feldtyp = Numerisch
 Operator = Gleich
 Werttyp = Zeichenfolge
 Wert = 12345

Interflow Match

Interflow Match ermittelt Übereinstimmungen zwischen ähnlichen Datensätzen in zwei Eingabedatensatzstreams. Der erste Datensatzstream bildet eine Quelle für Kopfdubletten-Datensätze und der zweite Stream ist eine Quelle für Kandidatendatensätze.

Mithilfe von Übereinstimmungsgruppenkriterien (z. B. ein Vergleichsschlüssel) identifiziert Interflow Match eine Gruppe von Datensätzen, die möglicherweise Dubletten eines bestimmten Kopfdubletten-Datensatzes sind.

Jeder Kandidat wird einzeln mit der Kopfdublette abgeglichen und entsprechend Ihrer Vergleichsregeln mit einer Punktzahl versehen. Wenn es sich bei dem Kandidaten um eine Dublette handelt, wird ihm eine Sammlungsnummer zugewiesen, der übereinstimmende Datensatztyp wird als eine Dublette gekennzeichnet und ausgeschrieben. Die nicht übereinstimmenden eindeutigen Kandidaten können auf Wunsch des Benutzers ausgeschrieben werden. Wenn Interflow Match alle Kandidatendatensätze in der aktuellen Übereinstimmungsgruppe durchlaufen hat, wird dem übereinstimmenden Kopfdubletten-Datensatz eine Sammlungsnummer zugewiesen, die der Datensatzdublette entspricht. Wenn alternativ keine Übereinstimmungen ermittelt wurden, erhält die Kopfdublette die Sammlungsnummer 0 und wird als eindeutiger Datensatz gekennzeichnet.

Anmerkung: Interflow Match gleicht lediglich Kopfdubletten-Datensätze mit Kandidatendatensätzen ab. Es vergleicht keine Kopfdubletten-Datensätze mit anderen Kopfdubletten-Datensätzen, wie es bei Intraflow Match der Fall ist.

Der Vergleichsprozess für eine bestimmte Kopfdublette wird eventuell beendet, bevor alle möglichen Kandidaten verglichen werden, wenn Sie ein Limit für Dubletten festgelegt haben und dieses Limit für die aktuelle Kopfdublette überschritten wurde.

Der Vergleichstyp (Intraflow oder Interflow) bestimmt, wie Ergebnisse aus dem Schnellschlüsselvergleich in die Kandidatenpunktzahl umgewandelt werden. Beim Interflow-Vergleich überträgt ein erfolgreicher Schnellschlüsselvergleich stets ein MatchScore von 100 auf den Kandidaten. Auf der anderen Seite hängt beim Intraflow-Vergleich die Punktzahl, die ein Kandidat infolge eines Schnellschlüsselvergleichs erzielt, davon ab, ob der Datensatz, mit dem der Kandidat übereinstimmte, die Übereinstimmung einer anderen Kopfdublette war. Schnellschlüssel-Dubletten einer Kopfdublette erzielen immer ein MatchScore von 100, während Schnellschlüssel-Dubletten eines anderen Kandidaten (bei dem es sich um eine Dublette einer Kopfdublette handelte) den MatchScore (nicht unbedingt 100) dieses Kandidaten übernehmen werden.

Optionen

1. Wählen Sie im Feld **Vergleichsregel laden** eine der vordefinierten Vergleichsregeln aus, die Sie entweder unverändert verwenden oder nach Bedarf ändern können. Wenn Sie eine neue Vergleichsregel erstellen möchten, ohne eine der vordefinierten Vergleichsregeln als Startpunkt zu verwenden, klicken Sie auf **Neu**. Im Datenfluss darf lediglich eine benutzerdefinierte Regel enthalten sein.

Anmerkung: Das Feature „Datenflussoptionen“ im Enterprise Designer ermöglicht es, die Vergleichsregel für die Konfiguration zur Laufzeit verfügbar zu machen.

2. Klicken Sie auf **Gruppieren nach**, um ein Feld auszuwählen, das zum Gruppieren von Datensätzen in der Vergleichswarteschlange verwendet werden soll. Intraflow Match führt lediglich Vergleichsvorgänge zwischen Datensätzen in derselben Vergleichswarteschlange durch.
3. Aktivieren Sie das Kästchen **Sortieren**, um eine Vorabvergleichssortierung Ihrer Eingabe basierend auf dem im Feld **Gruppieren nach** ausgewählten Feld durchzuführen.
4. Klicken Sie auf **Erweitert**, um zusätzliche Optionen der Sortierleistung festzulegen.

Beschränkung für Datensätze im Arbeitsspeicher Gibt die maximale Anzahl von Datenzeilen an, die ein Sorter im Arbeitsspeicher belässt, bevor er sie auf die Festplatte auslagert. Standardmäßig erfolgt eine Sortierung von bis zu 10.000 Datensätze in einem Arbeitsspeicher und eine Sortierung von mehr als 10.000 Datensätzen auf einem Datenträger. Der maximale Wert umfasst 100.000 Datensätze. In der Regel erfolgt eine Sortierung im Arbeitsspeicher deutlich schneller als auf einem Datenträger. Daher sollte hierfür ein Wert festgelegt werden, der hoch genug ist, damit ein Großteil der Sortierungen im Arbeitsspeicher erfolgen kann und nur umfangreiche Datasets auf einen Datenträger geschrieben werden.

Anmerkung: Vorsicht bei Umgebungen, in denen Aufträge gleichzeitig ausgeführt werden: Durch das Erhöhen der Einstellung **Beschränkung für Datensätze im Arbeitsspeicher** wird die Wahrscheinlichkeit erhöht, dass nicht mehr genügend Arbeitsspeicher verfügbar ist.

Maximale Anzahl temporärer Dateien Gibt die maximale Anzahl temporärer Dateien an, die von einem Sortiervorgang verwendet werden können. Die Verwendung einer größeren Anzahl temporärer Dateien kann zu einer besseren Leistung führen. Die optimale Anzahl hängt jedoch stark von der Konfiguration des Servers ab, auf dem Spectrum™ Technology Platform ausgeführt wird. Sie sollten mit verschiedenen Einstellungen experimentieren und dabei die Auswirkungen auf die Leistung beobachten, wenn mehr oder weniger temporäre Dateien verwendet werden. Verwenden Sie für die Berechnung der ungefähren Anzahl temporärer Dateien die folgende Gleichung:

$$\frac{(NumberOfRecords \times 2)}{InMemoryRecordLimit} = NumberOfTempFiles$$

Beachten Sie, dass die maximale Anzahl der temporären Dateien 1.000 nicht überschreiten darf.

Komprimierung aktivieren Gibt an, dass temporäre Dateien komprimiert werden, wenn sie auf die Festplatte geschrieben werden.

Anmerkung: Welche Einstellungen für die Sortierleistung optimal sind, hängt von der Hardwarekonfiguration Ihres Servers ab. Dennoch wird mit der folgenden Gleichung im Allgemeinen eine gute Sortierleistung erzielt:

$$(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2) \geq TotalNumberOfRecords$$

5. Klicken Sie auf **Schnellvergleich aktiviert**, um einen Erstvergleich zwischen Schnellschlüsselwerten durchzuführen, was zur Bestimmung dient, ob zwei Datensätze als Übereinstimmung gelten.

Der Schnellschlüsselvergleich kann ein hilfreiches Tool zur Reduzierung der Anzahl an durchgeführten Vergleichen sein und somit die Ausführungsgeschwindigkeit verbessern. Ein zu weit gefasster Schnellschlüssel hat viele falsch-positive Übereinstimmungen zur Folge. Sie können einen Schnellschlüssel als Teil der Generierung eines Vergleichsschlüssels über den MatchKeyGenerator generieren. Weitere Informationen finden Sie unter [Match Key Generator](#) auf Seite 236.

Wenn zwei Datensätze eine genaue Übereinstimmung im Schnellschlüssel aufweisen, wird der Kandidat als eine 100 %-ige Dublette eingeordnet. Wenn der Schnellvergleichsschlüssel in zwei Datensätzen nicht übereinstimmt, werden diese anhand der regelbasierten Methode verglichen.

Um zu ermitteln, ob ein Kandidat mithilfe eines Schnellvergleichsschlüssels verglichen wurde, sehen Sie nach, ob das Feld **ExpressKeyIdentified** „Y“ (ja) oder „N“ (nein) enthält. Beachten Sie, dass Datensatzkopfdubletten immer den Wert „N“ in **ExpressKeyIdentified** enthalten.

6. Geben Sie im Textfeld **Anfängliche Sammlungsnummer** die Startnummer ein, die dem Sammlungsnummernfeld für Datensatzdubletten zugewiesen werden soll.

Die Sammlungsnummer identifiziert jede Datensatzdublette in einer Vergleichswarteschlange. Eindeutige Datensätze erhalten die Sammlungsnummer 0. Jeder Datensatzdublette wird eine Sammlungsnummer zugewiesen, die mit dem im Textfeld **Anfängliche Sammlungsnummer** angegebenen Wert beginnt.

7. Wählen Sie eine der folgenden Optionen aus:

| Option | Bezeichnung |
|--|---|
| Kopfdublette mit allen Kandidaten vergleichen | Bei dieser Option wird die Kopfdublette mit allen Kandidaten in derselben Übereinstimmungsgruppe (Option „Gruppieren nach“) verglichen, auch wenn bereits eine Dublette innerhalb der Übereinstimmungsgruppe ermittelt wurde. Beispiel: Kopfdublette: John Smith Kandidat: Bill Jones Kandidat: John Smith Kandidat: John Smith |

| Option | Bezeichnung |
|--|--|
| | <p>Im Beispiel würde die Kopfdublette „John Smith“ mit den beiden Kandidaten „John Smith“ verglichen werden.</p> <p>Aktivieren Sie das Kästchen Eindeutige Kandidaten zurückgeben, um Datensätze innerhalb einer Übereinstimmungsgruppe aus dem Kandidatenport zurückzugeben, die als eindeutige Datensätze identifiziert wurden.</p> |
| Kopfdublette nicht mehr mit Kandidaten vergleichen, wenn n Dubletten gefunden wurden | <p>Bei dieser Option wird die Kopfdublette mit allen Kandidaten in derselben Übereinstimmungsgruppe (Option „Gruppieren nach“) verglichen. Der Vergleich wird jedoch beendet, wenn die benutzerdefinierte Anzahl an Dubletten ermittelt wurde. Wenn Sie beispielsweise wählen, Kandidaten nicht mehr zu vergleichen, sobald eine Dublette ermittelt wird, und Ihnen die folgenden Daten vorliegen:</p> <p>Kopfdublette: John Smith Kandidat: Bill Jones Kandidat: John Smith Kandidat: John Smith</p> <p>Im Beispiel würde der Vergleich des Kopfdubletten-Datensatzes „John Smith“ in der Übereinstimmungsgruppe beendet werden, wenn der erste Kandidat „John Smith“ als Dublette identifiziert werden würde.</p> |

- Klicken Sie auf **Daten für Analyse generieren**, um Vergleichsergebnisse zu generieren. Weitere Informationen finden Sie unter [Analysieren von Vergleichsergebnissen](#) auf Seite 121.
- Über **Eindeutigen Datensätzen Sammlungsnummer 0 zuweisen** (standardmäßig aktiviert) werden Nullen als Sammlungsnummern eindeutigen Datensätzen zugewiesen. Deaktivieren Sie diese Option, um statt der Null andere Sammlungsnummern für eindeutige Datensätze zu generieren. Die Sammlungsnummern für eindeutige Datensätze werden gemäß der Reihenfolge der anderen Sammlungsnummern nummeriert. Wenn Ihr Vergleichsdatenfluss beispielsweise fünf Datensätze ermittelt und die ersten drei Datensätze eindeutig sind, werden die Sammlungsnummern wie in der ersten Gruppe unten dargestellt zugewiesen. Wenn Ihr Vergleichsdatenfluss fünf Datensätze ermittelt und die letzten beiden Datensätze eindeutig sind, werden die Sammlungsnummern wie in der zweiten Gruppe unten dargestellt zugewiesen.

| Option | Bezeichnung |
|------------------------|---------------------|
| Sammlungsnummer | Datensatztyp |
| 1 | Eindeutig |

| Option | Bezeichnung |
|-----------------|-----------------------|
| 2 | Eindeutig |
| 3 | Eindeutig |
| 4 | Dublette/Kopfdublette |
| 4 | Dublette/Kopfdublette |
| | |
| Sammlungsnummer | Datensatztyp |
| 1 | Dublette/Kopfdublette |
| 1 | Dublette/Kopfdublette |
| 2 | Eindeutig |
| 3 | Eindeutig |
| 4 | Eindeutig |

Wenn dieses Kästchen aktiviert bleibt, wird allen in Ihrem Datenfluss ermittelten eindeutigen Datensätzen standardmäßig die Sammlungsnummer Null zugewiesen.

10. Wenn Sie eine neue benutzerdefinierte Vergleichsregel erstellen, finden Sie unter [Erstellen einer Vergleichsregel](#) auf Seite 72 weitere Informationen.
11. Klicken Sie auf **Auswerten**, um zu prüfen, wie ein Kopfdubletten-Datensatz im Vergleich zu Kandidatendatensätzen ausgewertet wurde. Weitere Informationen finden Sie unter [Interflow Match](#) auf Seite 225.

Ausgabe

Tabelle 13: Ausgabefelder für Interflow Match

| Feldname | Beschreibung/gültige Werte |
|------------------|--|
| CollectionNumber | Identifiziert eine Sammlung von Datensatzdubletten. Möglich sind Werte ab 1. |

| Feldname | Beschreibung/gültige Werte |
|------------------------|---|
| ExpressMatchIdentified | Gibt an, ob die Übereinstimmung anhand des Schnellvergleichsschlüssels ermittelt wurde. Möglich sind die Werte „Yes“ oder „No“. |
| HasDuplicates | <p>Identifiziert, ob der Datensatz eine Dublette eines anderen Datensatzes ist. Zur Auswahl stehen:</p> <p>Y Bei dem Datensatz handelt es sich um einen Kopfdubletten-Datensatz mit Dubletten.</p> <p>N Bei dem Datensatz handelt es sich um einen Kopfdubletten-Datensatz ohne Dubletten.</p> <p>D Der Datensatz ist ein Kandidatendatensatz und eine Dublette des Kopfdubletten-Datensatzes.</p> <p>U Der Datensatz ist ein Kandidatendatensatz, aber keine Dublette des Kopfdubletten-Datensatzes.</p> |
| InterflowSourceType | Möglich sind die Werte input_port_0 oder input_port_1. |
| MatchRecordType | <p>Bestimmt den Typ des Vergleichsdatensatzes in einer Sammlung. Die folgenden Werte sind möglich:</p> <p>suspect Der ursprüngliche Eingabedatensatz, bei dem markiert wurde, dass er möglicherweise Datensatzdubletten enthält.</p> <p>duplicate Ein Datensatz, der eine Dublette des Eingabedatensatzes ist.</p> <p>unique Ein Datensatz ohne Dubletten.</p> |
| MatchScore | Identifiziert die Gesamtpunktzahl zwischen zwei Datensätzen. Möglich sind Werte zwischen 0 und 100, wobei 0 eine niedrige Übereinstimmung und 100 eine genaue Übereinstimmung angibt. |

Anmerkung: Die Schritte „Validate Address“ und „Advanced Matching-Modul“ verwenden beide das Feld „MatchScore“. Der Wert des Feldes „MatchScore“ in der Ausgabe eines Datenflusses wird von dem letzten Schritt bestimmt, der den Wert verändert, bevor er an einen Ausgabeschritt gesendet wird. Wenn Sie einen Datenfluss mit den Schritten „Validate Address“ und „Advanced Matching-Modul“ haben und das Ausgabefeld „MatchScore“ für jeden Schritt anzeigen möchten, kopieren Sie den MatchScore-Wert mithilfe eines Transformer-Schrittes in ein anderes Feld. „Validate Address“ erzeugt beispielsweise ein Ausgabefeld mit dem Namen „MatchScore“. Ein Transformer-Schritt kopiert anschließend

das Feld „MatchScore“ aus „Validate Address“ in ein Feld mit dem Namen „AddressMatchScore“. Wenn ein Vergleichsschritt ausgeführt wird, füllt er das Feld „MatchScore“ mit dem Wert aus dem Vergleich aus und gibt den Wert für „AddressMatchScore“ aus „Validate Address“ weiter.

Intraflow Match

Intraflow Match ermittelt Übereinstimmungen zwischen ähnlichen Datensätzen innerhalb eines einzelnen Eingabestreams. Sie können hierarchische Regeln basierend auf beliebigen Felder erstellen, die in anderen Schritten des Datenflusses definiert oder erstellt wurden.

Optionen

1. Wählen Sie im Feld **Vergleichsregel laden** eine der vordefinierten Vergleichsregeln aus, die Sie entweder unverändert verwenden oder nach Bedarf ändern können. Wenn Sie eine neue Vergleichsregel erstellen möchten, ohne eine der vordefinierten Vergleichsregeln als Startpunkt zu verwenden, klicken Sie auf **Neu**. Im Datenfluss darf lediglich eine benutzerdefinierte Regel enthalten sein.

Anmerkung: Das Feature „Datenflussoptionen“ im Enterprise Designer ermöglicht es, die Vergleichsregel für die Konfiguration zur Laufzeit verfügbar zu machen.

2. Klicken Sie auf **Gruppieren nach**, um ein Feld auszuwählen, das zum Gruppieren von Datensätzen in der Vergleichswarteschlange verwendet werden soll. Intraflow Match führt lediglich Vergleichsvorgänge zwischen Datensätzen in derselben Vergleichswarteschlange durch.
3. Aktivieren Sie das Kästchen **Sortieren**, um eine Vorabvergleichssortierung Ihrer Eingabe basierend auf dem im Feld **Gruppieren nach** ausgewählten Feld durchzuführen.
4. Klicken Sie auf **Erweitert**, um zusätzliche Optionen der Sortierleistung festzulegen.

Beschränkung für Datensätze im Arbeitsspeicher Gibt die maximale Anzahl von Datenzeilen an, die ein Sorter im Arbeitsspeicher belässt, bevor er sie auf die Festplatte auslagert. Standardmäßig erfolgt eine Sortierung von bis zu 10.000 Datensätze in einem Arbeitsspeicher und eine Sortierung von mehr als 10.000 Datensätzen auf einem Datenträger. Der maximale Wert umfasst 100.000 Datensätze. In der Regel erfolgt eine Sortierung im Arbeitsspeicher deutlich schneller als auf einem Datenträger. Daher sollte hierfür ein Wert festgelegt werden, der hoch genug ist, damit ein Großteil der Sortierungen im Arbeitsspeicher erfolgen kann und nur umfangreiche Datasets auf einen Datenträger geschrieben werden.

Anmerkung: Vorsicht bei Umgebungen, in denen Aufträge gleichzeitig ausgeführt werden: Durch das Erhöhen der Einstellung **Beschränkung für Datensätze im Arbeitsspeicher** wird die

Wahrscheinlichkeit erhöht, dass nicht mehr genügend Arbeitsspeicher verfügbar ist.

Maximale Anzahl temporärer Dateien Gibt die maximale Anzahl temporärer Dateien an, die von einem Sortiervorgang verwendet werden können. Die Verwendung einer größeren Anzahl temporärer Dateien kann zu einer besseren Leistung führen. Die optimale Anzahl hängt jedoch stark von der Konfiguration des Servers ab, auf dem Spectrum™ Technology Platform ausgeführt wird. Sie sollten mit verschiedenen Einstellungen experimentieren und dabei die Auswirkungen auf die Leistung beobachten, wenn mehr oder weniger temporäre Dateien verwendet werden. Verwenden Sie für die Berechnung der ungefähren Anzahl temporärer Dateien die folgende Gleichung:

$$\text{NumberOfRecords} \times 2 \div \text{InMemoryRecordLimit} = \text{NumberOfTempFiles}$$

Beachten Sie, dass die maximale Anzahl der temporären Dateien 1.000 nicht überschreiten darf.

Komprimierung aktivieren Gibt an, dass temporäre Dateien komprimiert werden, wenn sie auf die Festplatte geschrieben werden.

Anmerkung: Welche Einstellungen für die Sortierleistung optimal sind, hängt von der Hardwarekonfiguration Ihres Servers ab. Dennoch wird mit der folgenden Gleichung im Allgemeinen eine gute Sortierleistung erzielt:

$$(\text{InMemoryRecordLimit} \times \text{MaxNumberOfTempFiles} \div 2) \geq \text{TotalNumberOfRecords}$$

5. Klicken Sie auf **Schnellvergleich aktiviert**, um einen Erstvergleich zwischen Schnellschlüsselwerten durchzuführen, was zur Bestimmung dient, ob zwei Datensätze als Übereinstimmung gelten.
Sie können einen Schnellschlüssel als Teil der Generierung eines Vergleichsschlüssels über den MatchKeyGenerator generieren. Weitere Informationen finden Sie unter [Match Key Generator](#) auf Seite 236.
6. Geben Sie im Textfeld **Anfängliche Sammlungsnummer** die Startnummer ein, die dem Sammlungsnummernfeld für Datensatzdubletten zugewiesen werden soll.
Die Sammlungsnummer identifiziert jede Datensatzdublette in einer Vergleichswarteschlange. Eindeutige Datensätze erhalten die Sammlungsnummer 0. Jeder Datensatzdublette wird eine Sammlungsnummer zugewiesen, die mit dem im Textfeld **Anfängliche Sammlungsnummer** angegebenen Wert beginnt.
7. Klicken Sie auf **Gleitendes Fenster**, um diese Vergleichsmethode zu aktivieren. Weitere Informationen zu „Gleitendes Fenster“ finden Sie unter [Vergleichsmethode „Gleitendes Fenster“](#) auf Seite 234.

8. Klicken Sie auf **Daten für Analyse generieren**, um Vergleichsergebnisse zu generieren. Weitere Informationen finden Sie unter [Analysieren von Vergleichsergebnissen](#) auf Seite 121.
9. Über **Eindeutigen Datensätzen Sammlungsnummer 0 zuweisen** (standardmäßig aktiviert) werden Nullen als Sammlungsnummern eindeutigen Datensätzen zugewiesen. Deaktivieren Sie diese Option, um statt der Null andere Sammlungsnummern für eindeutige Datensätze zu generieren. Die Sammlungsnummern für eindeutige Datensätze werden gemäß der Reihenfolge der anderen Sammlungsnummern nummeriert. Wenn Ihr Vergleichsdatenfluss beispielsweise fünf Datensätze ermittelt und die ersten drei Datensätze eindeutig sind, werden die Sammlungsnummern wie in der ersten Gruppe unten dargestellt zugewiesen. Wenn Ihr Vergleichsdatenfluss fünf Datensätze ermittelt und die letzten beiden Datensätze eindeutig sind, werden die Sammlungsnummern wie in der zweiten Gruppe unten dargestellt zugewiesen.

| Option | Bezeichnung |
|------------------------|-----------------------|
| Sammlungsnummer | Datensatztyp |
| 1 | Eindeutig |
| 2 | Eindeutig |
| 3 | Eindeutig |
| 4 | Dublette/Kopfdublette |
| 4 | Dublette/Kopfdublette |
| Sammlungsnummer | Datensatztyp |
| 1 | Dublette/Kopfdublette |
| 1 | Dublette/Kopfdublette |
| 2 | Eindeutig |
| 3 | Eindeutig |
| 4 | Eindeutig |

Wenn dieses Kästchen aktiviert bleibt, wird allen in Ihrem Datenfluss ermittelten eindeutigen Datensätzen standardmäßig die Sammlungsnummer Null zugewiesen.

10. Weitere Informationen zum Ändern der anderen Optionen finden Sie unter [Erstellen einer Vergleichsregel](#) auf Seite 72.
11. Klicken Sie auf **Auswerten**, um zu prüfen, wie ein Kopfdubletten-Datensatz im Vergleich zu Kandidatendatensätzen ausgewertet wurde. Weitere Informationen finden Sie unter [Interflow Match](#) auf Seite 225.

Standardvergleichsmethode

Mithilfe der vom Benutzer festgelegten Option „Gruppieren nach“ werden beim Vergleich Gruppen von Datensätzen identifiziert, die eventuell Dubletten voneinander sind. Beim Vergleich wird anschließend jeder Datensatz in der Gruppe untersucht. Wenn der Datensatz mit einer vorhandenen Kopfdublette übereinstimmt, wird dieser Datensatz als Dublette dieser Kopfdublette eingeordnet, erhält eine Punktzahl, Sammlungsnummer und einen Vergleichsdatensatztyp (Dublette) und wird anschließend aus dem Vergleich entfernt. Wenn der Datensatz jedoch mit keiner vorhandenen Kopfdublette innerhalb der Übereinstimmungsgruppe übereinstimmt, wird der Datensatz als eine neue Kopfdublette eingeordnet, indem er zur aktuellen Übereinstimmungsgruppe hinzugefügt wird, damit er mit nachfolgenden Datensätzen abgeglichen werden kann. Wenn beim Vergleich alle Datensätze in der aktuellen Übereinstimmungsgruppe untersucht wurden, werden alle Kopfdubletten aus dem Vergleich entfernt, ihr Vergleichsdatensatztyp wird als „Eindeutig“ markiert und sie erhalten die Sammlungsnummer 0. Diese Kopfdubletten mit mindestens einer Dublette bleiben als Vergleichsdatensatztyp „Kopfdublette“ erhalten und ihnen wird dieselbe Sammlungsnummer wie ihrer übereinstimmenden Datensatzdublette zugewiesen. Wenn schließlich alle Datensätze innerhalb einer Übereinstimmungsgruppe in die Ausgabe geschrieben wurden, wird eine neue Übereinstimmungsgruppe verglichen.

Anmerkung: Die standardmäßige Vergleichsmethode vergleicht nur Datensätze, die sich innerhalb derselben Übereinstimmungsgruppe befinden.

Der Vergleichstyp (Intraflow oder Interflow) bestimmt, wie Ergebnisse aus dem Schnellschlüsselvergleich in die Kandidatenpunktzahl umgewandelt werden. Beim Interflow-Vergleich überträgt ein erfolgreicher Schnellschlüsselvergleich stets ein MatchScore von 100 auf den Kandidaten. Auf der anderen Seite hängt beim Intraflow-Vergleich die Punktzahl, die ein Kandidat infolge eines Schnellschlüsselvergleichs erzielt, davon ab, ob der Datensatz, mit dem der Kandidat übereinstimmte, die Übereinstimmung einer anderen Kopfdublette war. Schnellschlüssel-Dubletten einer Kopfdublette erzielen immer ein MatchScore von 100, während Schnellschlüssel-Dubletten eines anderen Kandidaten (bei dem es sich um eine Dublette einer Kopfdublette handelte) den MatchScore (nicht unbedingt 100) dieses Kandidaten übernehmen werden.

Vergleichsmethode „Gleitendes Fenster“

Der Algorithmus „Gleitendes Fenster“ ist ein Algorithmus, der der Reihe nach eine vorbestimmte Puffergröße, Fenster genannt, mit der entsprechenden Menge an Datenzeilen füllt. Bei jedem Hinzufügen einer Zeile in das Fenster wird sie mit jedem Element verglichen, das sich bereits im Fenster befindet. Wenn eine Übereinstimmung mit einem Element ermittelt wird, erhalten sowohl der Treiberdatensatz (das neue Element, das dem Fenster hinzugefügt werden soll) als auch die

Kandidaten (Elemente, die sich bereits im Fenster befinden) dieselbe Gruppen-ID. Dieser Vergleich wird fortgesetzt, bis der Treiberdatensatz mit allen Elementen im Fenster verglichen wurde.

Da neue Treiber hinzugefügt werden, wird das Fenster irgendwann seine vorbestimmte Kapazität erreichen. Zu diesem Zeitpunkt beginnt das Fenster zu gleiten und wird deswegen „Gleitendes Fenster“ genannt. Mit Gleiten wird ausgedrückt, dass der Fensterpuffer das älteste Element im Fenster entfernt und ausschreibt, während er den neuen Treiberdatensatz dem Fenster hinzufügt.

Ausgabe

Tabelle 14: Ausgabe von Intraflow Match

| Feldname | Beschreibung/gültige Werte |
|------------------------|--|
| CollectionNumber | Identifiziert eine Sammlung von Datensatzdubletten. Möglich sind Werte ab 1. |
| ExpressMatchIdentified | Gibt an, ob die Übereinstimmung anhand des Schnellvergleichsschlüssels ermittelt wurde. Möglich sind die Werte „Yes“ oder „No“. |
| MatchRecordType | Bestimmt den Typ des Vergleichsdatensatzes in einer Sammlung. Die folgenden Werte sind möglich: <ul style="list-style-type: none"> suspect Ein Datensatz, mit dem andere Datensätze verglichen werden, um zu bestimmen, ob sie Dubletten voneinander sind. In jeder Sammlung ist lediglich ein Kopfdubletten-Datensatz vorhanden. duplicate Ein Datensatz, der eine Dublette des Kopfdubletten-Datensatzes ist. unique Ein Datensatz ohne Dubletten. |
| MatchScore | Identifiziert die Gesamtpunktzahl zwischen zwei Datensätzen. Möglich sind Werte zwischen 0 und 100, wobei 0 eine niedrige Übereinstimmung und 100 eine genaue Übereinstimmung angibt. |

Anmerkung: Die Schritte „Validate Address“ und „Advanced Matching-Modul“ verwenden beide das Feld „MatchScore“. Der Wert des Feldes „MatchScore“ in der Ausgabe eines Datenflusses wird von dem letzten Schritt bestimmt, der den Wert verändert, bevor er an einen Ausgabeschritt gesendet wird. Wenn Sie einen Datenfluss mit den Schritten „Validate Address“ und „Advanced Matching-Modul“ haben und das Ausgabefeld „MatchScore“ für jeden Schritt anzeigen möchten, kopieren Sie den MatchScore-Wert mithilfe eines Transformer-Schrittes in ein anderes Feld. „Validate Address“ erzeugt beispielsweise ein Ausgabefeld mit dem Namen „MatchScore“. Ein Transformer-Schritt kopiert anschließend

das Feld „MatchScore“ aus „Validate Address“ in ein Feld mit dem Namen „AddressMatchScore“. Wenn ein Vergleichsschritt ausgeführt wird, füllt er das Feld „MatchScore“ mit dem Wert aus dem Vergleich aus und gibt den Wert für „AddressMatchScore“ aus „Validate Address“ weiter.

Match Key Generator

Match Key Generator erstellt einen nicht eindeutigen Schlüssel für jeden Datensatz, der dann von Vergleichsschritten verwendet werden kann, um Gruppen potenzieller Datensatzdubletten zu identifizieren. Vergleichsschlüssel erleichtern den Vergleichsprozess, da sie Ihnen erlauben, Datensätze nach Vergleichsschlüssel zu gruppieren und dann nur Datensätze innerhalb dieser Gruppen zu vergleichen.

Der Vergleichsschlüssel wird mithilfe von Regeln erstellt, die Sie definieren, und besteht aus Eingabefeldern. Jedes angegebene Eingabefeld verfügt über einen ausgewählten Algorithmus, der dafür ausgeführt wird. Das Ergebnis eines jeden Algorithmus wird dann verkettet, um ein einzelnes Vergleichsschlüsselfeld zu erstellen.

Zusätzlich zum Erstellen von Vergleichsschlüsseln können Sie auch Schnellvergleichsschlüssel erstellen, die später im Datenfluss durch einen „Intraflow Match“- oder „Interflow Match“-Schritt verwendet werden.

Sie können mehrere Vergleichsschlüssel und Schnellvergleichsschlüssel erstellen.

Beispiel: Wenn der eingehende Datensatz wie folgt lautet:

Vorname – Fred
 Nachname – Mertz
 Postleitzahl – 21114-1687
 Geschlechtscode – M

Und Sie eine Vergleichsschlüsselregel definieren, die einen Vergleichsschlüssel generiert, indem Daten aus dem Datensatz wie folgt kombiniert werden:

| Eingabefeld | Startposition | Länge |
|--------------|---------------|-------|
| Postleitzahl | 1 | 5 |
| Postleitzahl | 7 | 4 |
| Nachname | 1 | 5 |
| Vorname | 1 | 5 |

| Eingabefeld | Startposition | Länge |
|-----------------|---------------|-------|
| Geschlechtscode | 1 | 1 |

Dann ergäbe sich der Schlüssel:

211141687MertzFredM

Eingabe

Die Eingabe ist ein beliebiges Feld in den Quelldaten.

Optionen

Klicken Sie zum Definieren von Match Key Generator-Optionen auf die Schaltfläche **Hinzufügen**. Das Dialogfeld **Vergleichsschlüsselfeld** wird angezeigt.

Anmerkung: Das Feature „Datenflussoptionen“ im Enterprise Designer ermöglicht es, Match Key Generator zur Laufzeit zur Konfiguration verfügbar zu machen.

Tabelle 15: Match Key Generator-Optionen

| Name der Option | Beschreibung/gültige Werte |
|-----------------|----------------------------|
|-----------------|----------------------------|

| | |
|-------------|--|
| Algorithmus | |
|-------------|--|

Name der Option

Beschreibung/gültige Werte

Gibt den Algorithmus an, der verwendet werden soll, um den Vergleichsschlüssel zu generieren. Zur Auswahl stehen:

| | |
|-----------------------------|--|
| Consonant | Gibt die angegebenen Felder mit entfernten Konsonanten zurück. |
| Double Metaphone | Gibt einen auf einer phonetischen Darstellung ihrer Zeichen basierenden Code zurück. „Double Metaphone“ ist eine verbesserte Version des Algorithmus „Metaphone“ und versucht, viele der in unterschiedlichen Sprachen vorkommenden Unregelmäßigkeiten zu berücksichtigen. |
| Köln | Indiziert Namen nach Klang, so wie sie auf Deutsch ausgesprochen werden. Ermöglicht die Codierung von Namen mit derselben Aussprache in derselben Darstellung, sodass sie trotz kleiner Unterschiede bei der Schreibweise verglichen werden können. Das Ergebnis ist immer eine Folge von Zahlen; Sonderzeichen und Leerzeichen werden ignoriert. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren. |
| MD5 | Ein Message Digest-Algorithmus, der einen 128-Bit-Hashwert erzeugt. Dieser Algorithmus wird im Allgemeinen verwendet, um die Datenintegrität zu überprüfen. |
| Metaphone | Gibt einen per Metaphone codierten Schlüssel ausgewählter Felder zurück. Metaphone ist ein Algorithmus, um Wörter anhand ihrer englischen Aussprache zu codieren. |
| Metaphone (Spanisch) | Gibt einen per Metaphone codierten Schlüssel ausgewählter Felder für die spanische Sprache zurück. Dieser Metaphone-Algorithmus codiert Wörter anhand ihrer spanischen Aussprache. |
| Metaphone 3 | Übertrifft die Algorithmen „Metaphone“ und „Double Metaphone“ mit genaueren Einstellungen für Konsonanten und inneren Vokalen, was Ihnen ermöglicht, Wörter oder Namen auf phonetischer Basis mehr oder weniger eng mit Suchbegriffen abgeglichen zu erzeugen. Metaphone 3 erhöht die Genauigkeit der phonetischen Codierung auf 98 %. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren. |
| NYSIIS | Phonetischer Codealgorithmus, der eine ungefähr übereinstimmende Aussprache mit der exakten Schreibweise vergleicht und Wörter indiziert, deren Aussprache ähnlich ist. Bestandteil des New York State Identification and Intelligence System. Angenommen, Sie suchen die Daten einer Person in einer Personendatenbank. Sie meinen, dass der Name der Person wie „John Smith“ klingt, er wird aber „Jon Smyth“ geschrieben. Würden Sie eine Suche durchführen, in der nach einer genauen Übereinstimmung mit „John Smith“ gesucht wird, würden keine Ergebnisse zurückgegeben. Wenn Sie jedoch die Datenbank mit dem Algorithmus NYSIIS indizieren und eine |

| Name der Option | Beschreibung/gültige Werte |
|--------------------------------|--|
| | erneute Suche mit dem Algorithmus NYSIIS durchführen, wird die richtige Übereinstimmung zurückgegeben, weil vom Algorithmus sowohl „John Smith“ als auch „Jon Smyth“ als „JAN SNATH“ indiziert werden. |
| Phonix | Vorverarbeitet Namenszeichenfolgen, indem mehr als 100 Transformationsregeln auf einzelne Zeichen oder auf Zeichenfolgen angewendet werden. 19 dieser Regeln werden nur angewendet, wenn das bzw. die Zeichen am Anfang der Zeichenfolge stehen, während 12 der Regeln nur angewendet werden, wenn sie in der Mitte der Zeichenfolge stehen. 28 der Regeln werden nur angewendet, wenn sie am Ende der Zeichenfolge stehen. Die transformierte Namenszeichenfolge wird als Code codiert, der aus einem Anfangsbuchstaben gefolgt von drei Stellen besteht (Nullen und doppelt vorhandene Zahlen werden entfernt). Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren; sie ist komplexer und deshalb langsamer als Soundex. |
| Soundex | Gibt einen Soundex-Code ausgewählter Felder zurück. Soundex erzeugt einen auf der englischen Aussprache eines Wortes basierenden Code mit fester Länge. |
| Teilzeichenfolge | Gibt einen angegebenen Teil des ausgewählten Feldes zurück. |
| Feldname | Gibt das Feld an, auf das Sie den ausgewählten Algorithmus anwenden möchten, um den Vergleichsschlüssel zu generieren. Wenn Sie beispielsweise ein Feld mit Namen „LastName“ auswählen und den Soundex-Algorithmus wählen, würde der Soundex-Algorithmus auf die Daten im Feld „LastName“ angewendet, um einen Vergleichsschlüssel zu erzeugen. |
| Startposition | Gibt die Startposition innerhalb des angegebenen Feldes an. Nicht alle Algorithmen erlauben Ihnen, eine Startposition anzugeben. |
| Länge | Gibt die Länge der Zeichen an, die ab der Startposition eingeschlossen werden sollen. Nicht alle Algorithmen erlauben Ihnen, eine Länge anzugeben. |
| Sonderzeichen entfernen | Entfernt alle nicht numerischen und nicht alphabetischen Zeichen wie Bindestriche, Leerzeichen und andere Sonderzeichen aus einem Eingabefeld. |

| Name der Option | Beschreibung/gültige Werte |
|-------------------|---|
| Eingabe sortieren | Sortiert alle Zeichen in einem Eingabefeld oder alle Begriffe in einem Eingabefeld in alphabetischer Reihenfolge. |
| | Zeichen Sortiert die Zeichenwerte aus einem Eingabefeld vor dem Erstellen einer eindeutigen ID. |
| | Begriffe Sortiert jeden Begriffswert aus einem Eingabefeld vor dem Erstellen einer eindeutigen ID. |

Wenn Sie mehrere Algorithmen zur Vergleichsschlüsselgenerierung hinzufügen, können Sie die Schaltflächen **Nach oben** und **Nach unten** verwenden, um die Reihenfolge zu ändern, in der die Algorithmen angewendet werden.

Schnellvergleichsschlüssel generieren

Aktivieren Sie die Option **Schnellvergleichsschlüssel generieren** und klicken Sie auf **Hinzufügen**, um einen Schnellvergleichsschlüssel zu definieren, der später im Datenfluss durch „Intraflow Match“- oder „Interflow Match“-Schritt verwendet wird.

Wenn die Option **Schnellvergleichsschlüssel generieren** aktiviert ist und die Option **Schnellvergleich aktiviert** in einem nachgestellten „Interflow Match“- oder „Intraflow Match“-Schritt ausgewählt wird, wird der erste Vergleichsversuch mit dem hier erstellten Schnellvergleichsschlüssel unternommen. Wenn die Schnellvergleichsschlüssel von zwei Datensätzen übereinstimmen, dann wird der Datensatz als Übereinstimmung gewertet und keine weitere Verarbeitung versucht. Wenn die Schnellvergleichsschlüssel der Datensätze nicht übereinstimmen, dann werden die in „Interflow Match“ oder „Intraflow Match“ definierten Vergleichsregeln verwendet, um zu bestimmen, ob die Datensätze übereinstimmen.

Ausgabe

Tabelle 16: Match Key Generator-Ausgabe

| Feldname | Beschreibung/gültige Werte |
|-----------------|--|
| ExpressMatchKey | Ein Wert, der die Vergleichsebene angibt. Wenn der Schnellvergleichsschlüssel übereinstimmt, ist die Punktzahl 100. Wenn der Schnellvergleichsschlüssel nicht übereinstimmt, wird die Punktzahl 0 zurückgegeben. |
| MatchKey | Der Schlüssel, der generiert wird, um Datensätze zu identifizieren. |

Private Match

„Private Match“ ermöglicht zwei Entitäten, Datasets zu vergleichen und gemeinsame Datensätze zu identifizieren, ohne sensible Informationen zu kompromittieren. So könnten beispielsweise zwei Unternehmen daran interessiert sein, eine gemeinsame Marketingkampagne zu starten. Jedes Unternehmen hat eine eigene Datenbank mit Kundendaten, und die Unternehmen möchten für einen zielgerichteteren Kampagnenansatz erfahren, welche Kunden bei beiden Unternehmen kaufen. Um jedoch die Datensicherheit zu gewährleisten und Datenschutzbestimmungen einzuhalten, möchten die Unternehmen diese Datenbanken nicht gemeinsam nutzen oder Dritten überlassen, um diesen Vergleich durchzuführen. Das Feature „Private Match“ ermöglicht es, die beiden Datenbanken miteinander zu vergleichen, ohne die Sicherheit zu beeinträchtigen oder gegen Datenschutzgesetze zu verstoßen.

„Private Match“ wird in einem von drei Modi verwendet:

- Modus „Verschlüsseln“ – Der erste Benutzer gibt seine Daten ein, und ein Indexfeld und Vergleichsfeld werden extrahiert und verschlüsselt. Ein öffentlicher Schlüssel und eine Ersetzungstabelle, die die Daten des ersten Benutzers enthält, werden für den zweiten Benutzer generiert, und ein privater Schlüssel wird für den ersten Benutzer zur späteren Verwendung generiert.
- Modus „Private Match“ – Der zweite Benutzer gibt seine Daten und die verschlüsselten Daten des ersten Benutzers ein, stellt den öffentlichen Schlüssel und die Ersetzungstabelle bereit und führt einen Vergleich durch. Es wird eine Datei generiert, die die abgeglichenen Daten enthält, die dem ersten Benutzer gesendet werden sollen.
- Modus „Entschlüsseln“ – Der erste Benutzer gibt die verschlüsselten Daten des zweiten Benutzers ein, stellt den privaten Schlüssel bereit und generiert die Ausgabe, die einen abgeglichenen Index der Daten beider Benutzer enthält.

Durch Verwendung der Verschlüsselungsfunktion (Modus „Verschlüsseln“) bleibt die Sicherheit gewahrt, während eine Vergleichsfunktion ausgeführt wird (Modus „Private Match“). Danach zeigt eine Entschlüsselungsfunktion die Ausgabe der verglichenen Daten (Modus „Entschlüsseln“). Alle generierten und von Benutzern gemeinsam genutzten Dateien sind verschlüsselt und nicht lesbar.

Eingabe

Die Eingabeanforderungen für den „Private Match“-Schritt hängen von der durchgeführten Aufgabe ab:

- Modus „Verschlüsseln“ – Eine Datei, die die Daten des ersten Benutzers enthält, muss an den Eingabeport angefügt werden. Das kann eine Textdatei, Datenbank oder beinahe jede Art von Quellendatei sein.
- Modus „Private Match“ – Eine Datei, die die Daten des zweiten Benutzers enthält, muss an den ersten Eingabeport angefügt werden. Dies kann ebenfalls eine Textdatei, Datenbank oder beinahe jede Art von Quellendatei sein. Die vom ersten Benutzer generierten verschlüsselten Daten müssen an den zweiten Eingabeport angefügt werden.

- Modus „Entschlüsseln“ – Die vom zweiten Benutzer generierte Ausgabedatei.

Optionen

Die Optionen für den „Private Match“-Schritt hängen von der durchgeführten Aufgabe ab.

Anmerkung: Wenn Sie ein Upgrade von Version 10.x auf Version 11.0 oder neuer durchführen und private Schlüssel erzeugt haben, müssen Sie diese Schlüssel erneut generieren, weil in Version 10.x generierte Schlüssel nicht mit Version 11.0 und neuer kompatibel sind.

Modus „Verschlüsseln“

1. Wählen Sie den Vorgang **Verschlüsseln** aus.
2. Wählen Sie das **Indexfeld** aus, das eine eindeutige ID für jeden Datensatz in der Datei bereitstellt. Die eindeutige ID muss ein numerischer Wert sein.
3. Wählen Sie das **Vergleichsfeld** aus, das mit den Daten des zweiten Benutzers verglichen werden soll.
4. Geben Sie den Pfad und Namen der Datei mit dem **öffentlichen Schlüssel** an, die beim Ausführen des Auftrags erstellt wird.
5. Geben Sie den Pfad und Namen der Datei mit dem **Schlüsselspeicher** an, die beim Ausführen des Auftrags erstellt wird.
6. Geben Sie den Pfad und Namen der Datei mit der **Ersetzungstabelle** an, die beim Ausführen des Auftrags erstellt wird.
7. Geben Sie einen Namen für die **Ausgabespalte** an, die die verschlüsselten Daten in der Ausgabedatei enthalten soll, die an den zweiten Benutzer gesendet wird.
8. Klicken Sie auf **OK**.

Modus „Private Match“

1. Wählen Sie den Vorgang **Private Match** aus.
2. Wählen Sie das **Indexfeld** aus, das eine eindeutige ID für jeden Datensatz in der Datei bereitstellt. Die eindeutige ID muss ein numerischer Wert sein.
3. Wählen Sie das **Vergleichsfeld** aus, das mit den Daten des ersten Benutzers verglichen werden soll.
4. Geben Sie den Namen des **Feldes für verschlüsselte Daten** ein, den der erste Benutzer in Schritt 7 der Anleitung zum Modus „Verschlüsseln“ eingegeben hat.
5. Navigieren Sie zur Datei **Öffentlicher Schlüssel**.
6. Navigieren Sie zur Datei **Ersetzungstabelle**.
7. Geben Sie einen Namen für die **Ausgabespalte** an, die die verschlüsselten Daten in der Ausgabedatei enthalten soll, die an den ersten Benutzer gesendet wird.

Modus „Entschlüsseln“

1. Wählen Sie den Vorgang **Entschlüsseln** aus.

2. Wählen Sie das **Feld für verschlüsselte Daten** aus, das in Schritt 7 der Anleitung zum Modus „Private Match“ eingegeben wurde.
3. Navigieren Sie zur Datei **Schlüsselspeicher**.
4. Wählen Sie die **Ausgabespalte** aus, die die entschlüsselten Daten in der Ausgabedatei enthalten soll. Die Daten in diesem Feld haben das Format abgeglicherer Index der Daten des ersten Benutzers und abgeglicherer Index der Daten des zweiten Benutzers, getrennt durch ein Komma (,). Siehe dazu:

```
User1Data,User2Data
```

Ausgabe

Die Ausgabeanforderungen für den „Private Match“-Schritt hängen von der durchgeführten Aufgabe ab:

- Modus „Verschlüsseln“ – Eine durch den „Write to File“-Schritt generierte Datei, die die verschlüsselten Daten des ersten Benutzers enthält.
- Modus „Private Match“ – Eine durch den „Write to File“-Schritt generierte Datei, die verschlüsselte Informationen zu den Vergleichsergebnissen enthält.
- Modus „Entschlüsseln“ – Eine durch den „Write to File“-Schritt generierte Datei, die den abgeglichenen Index der Daten beider Benutzer enthält.

Transactional Match

Transactional Match vergleicht Kopfdubletten-Datensätze mit Kandidatendatensätzen, die vom „Candidate Finder“-Schritt zurückgegeben wurden. „Transactional Match“ verwendet Vergleichsregeln, um den Kopfdubletten-Datensatz mit allen Kandidatendatensätzen mit derselben (durch Candidate Finder zugewiesenen) Kandidatengruppennummer zu vergleichen, um Dubletten zu finden. Wenn es sich bei dem Kandidatendatensatz um eine Dublette handelt, wird ihm eine Sammlungsnummer zugewiesen, der übereinstimmende Datensatztyp wird als eine Dublette gekennzeichnet und der Datensatz wird anschließend ausgeschrieben. Allen Kandidaten in der Gruppe ohne Übereinstimmung wird die Sammlungsnummer 0 zugewiesen und sie werden als eindeutig markiert und ebenfalls ausgeschrieben.

Anmerkung: Transactional Match gleicht lediglich Kopfdubletten-Datensätze mit Kandidaten ab. Es vergleicht keine Kopfdubletten-Datensätze mit anderen Kopfdubletten-Datensätzen, wie es bei Intraflow Match der Fall ist.

Optionen

1. Wählen Sie im Feld **Vergleichsregel laden** eine der vordefinierten Vergleichsregeln aus, die Sie entweder unverändert verwenden oder nach Bedarf ändern können. Wenn Sie eine neue Vergleichsregel erstellen möchten, ohne eine der vordefinierten Vergleichsregeln als Startpunkt

zu verwenden, klicken Sie auf **Neu**. Im Datenfluss darf lediglich eine benutzerdefinierte Regel enthalten sein.

Anmerkung: Das Feature „Datenflussoptionen“ im Enterprise Designer ermöglicht es, die Vergleichsregel für die Konfiguration zur Laufzeit verfügbar zu machen.

2. Wählen Sie **Eindeutige Kandidaten zurückgeben** aus, wenn eindeutige Kandidatendatensätze in der Ausgabe des Schrittes enthalten sein sollen.
3. Wählen Sie **Daten für Analyse generieren** aus, wenn Sie das Tool „Vergleichsanalyse“ zur Analyse der Ergebnisse des Datenflusses verwenden möchten. Weitere Informationen finden Sie unter [Analysieren von Vergleichsergebnissen](#) auf Seite 121.
4. Weitere Informationen zum Ändern der anderen Optionen finden Sie unter [Erstellen einer Vergleichsregel](#) auf Seite 72.
5. Klicken Sie auf **Auswerten**, um zu prüfen, wie ein Kopfdubletten-Datensatz im Vergleich zu Kandidatendatensätzen ausgewertet wurde. Weitere Informationen finden Sie unter [Interflow Match](#) auf Seite 225.

Ausgabe

Tabelle 17: Ausgabe von Transactional Match

| Feldname | Beschreibung/gültige Werte |
|---------------|---|
| HasDuplicates | <p>Identifiziert, ob der Datensatz eine Dublette eines anderen Datensatzes ist. Zur Auswahl stehen:</p> <p>Y Bei dem Datensatz handelt es sich um einen Kopfdubletten-Datensatz mit Dubletten.</p> <p>N Bei dem Datensatz handelt es sich um einen Kopfdubletten-Datensatz ohne Dubletten.</p> <p>D Der Datensatz ist ein Kandidatendatensatz und eine Dublette des Kopfdubletten-Datensatzes.</p> <p>U Der Datensatz ist ein Kandidatendatensatz, aber keine Dublette des Kopfdubletten-Datensatzes.</p> |

| Feldname | Beschreibung/gültige Werte |
|------------------------------|---|
| MatchRecordType | <p>Bestimmt den Typ des Vergleichsdatensatzes in einer Sammlung. Die folgenden Werte sind möglich:</p> <p>Suspect Der ursprüngliche Eingabedatensatz, bei dem markiert wurde, dass er möglicherweise Datensatzdubletten enthält.</p> <p>Duplicate Ein Datensatz, der eine Dublette des Eingabedatensatzes ist.</p> <p>Unique Ein Datensatz ohne Dubletten.</p> |
| MatchScore | <p>Identifiziert die Gesamtpunktzahl zwischen zwei Datensätzen. Möglich sind Werte zwischen 0 und 100, wobei 0 eine niedrige Übereinstimmung und 100 eine genaue Übereinstimmung angibt.</p> |
| MatchInfo.<RuleName>.IsMatch | <p>Zeigt den Vergleichsstatus für die Regel als Wahr oder Falsch an, wobei Wahr für eine Übereinstimmung und Falsch für keine Übereinstimmung steht. Diese Information wird angezeigt, wenn Sie die Option Ausführliche Vergleichsinformationen zurückgeben im Bildschirm Transactional Match-Optionen auswählen.</p> <p>Anmerkung: Der Wert in diesem Feld wird aus dem Vergleichsstatus der untergeordneten Knoten basierend auf der für übergeordnete Elemente definierten Regelkonfiguration abgeleitet.</p> |
| MatchInfo.<RuleName>.Score | <p>Zeigt den Übereinstimmungswert für die Regel an. Möglich sind Werte zwischen 0 und 100, wobei 0 eine niedrige Übereinstimmung und 100 eine genaue Übereinstimmung angibt. Diese Information wird angezeigt, wenn Sie die Option Ausführliche Vergleichsinformationen zurückgeben im Bildschirm Transactional Match-Optionen auswählen.</p> <p>Anmerkung: Die Punktzahl in diesem Feld wird aus den Punktzahlen der untergeordneten Knoten basierend auf der für übergeordnete Elemente definierten Regelkonfiguration abgeleitet.</p> |

| Feldname | Beschreibung/gültige Werte |
|---|--|
| MatchInfo.<RuleName>.<RuleNodeName>.IsMatch | <p>Zeigt den Vergleichsstatus jedes Knotens in der Regelhierarchie als Wahr oder Falsch an, wobei Wahr für eine Übereinstimmung und Falsch für keine Übereinstimmung steht. RuleNodeName ist eine Variable, die in Ihrer Vergleichsregel durch die hierarchischen Knotennamen ersetzt wird.</p> <p>Dieses Feld wird für jeden Knoten in der Regelhierarchie ausgegeben, wenn Sie die Option Ausführliche Vergleichsinformationen zurückgeben im Bildschirm Transactional Match-Optionen ausgewählt haben.</p> |
| MatchInfo.<RuleName>.<RuleNodeName>.Score | <p>Zeigt den Übereinstimmungswert für jeden Knoten in der Regelhierarchie an. RuleNodeName ist eine Variable, die in Ihrer Vergleichsregel durch die hierarchischen Knotennamen ersetzt wird.</p> <p>Dieses Feld wird für jeden Knoten in der Regelhierarchie ausgegeben, wenn Sie die Option Ausführliche Vergleichsinformationen zurückgeben im Bildschirm Transactional Match-Optionen ausgewählt haben.</p> <p>Möglich sind Werte zwischen 0 und 100, wobei 0 eine niedrige Übereinstimmung und 100 eine genaue Übereinstimmung angibt.</p> |

Anmerkung: Die Schritte „Validate Address“ und „Advanced Matching-Modul“ verwenden beide das Feld „MatchScore“. Der Wert des Feldes „MatchScore“ in der Ausgabe eines Datenflusses wird von dem letzten Schritt bestimmt, der den Wert verändert, bevor er an einen Ausgabeschritt gesendet wird. Wenn Sie einen Datenfluss mit den Schritten „Validate Address“ und „Advanced Matching-Modul“ haben und das Ausgabefeld „MatchScore“ für jeden Schritt anzeigen möchten, kopieren Sie den MatchScore-Wert mithilfe eines Transformer-Schrittes in ein anderes Feld. „Validate Address“ erzeugt beispielsweise ein Ausgabefeld mit dem Namen „MatchScore“. Ein Transformer-Schritt kopiert anschließend das Feld „MatchScore“ aus „Validate Address“ in ein Feld mit dem Namen „AddressMatchScore“. Wenn ein Vergleichsschritt ausgeführt wird, füllt er das Feld „MatchScore“ mit dem Wert aus dem Vergleich aus und gibt den Wert für „AddressMatchScore“ aus „Validate Address“ weiter.

Write to Search Index

„Write to Search Index“ ermöglicht Ihnen, einen Volltextindex auf Basis der im Schritt eingehenden Daten zu erstellen. Diese Daten in einem dedizierten Suchindex führen zu einer schnelleren Antwortzeit, wenn Sie Suchen anhand des Index aus anderen Spectrum™ Technology

Platform-Schritten durchführen. Volltext-Suchindizes sind relationalen Datenbanken vorzuziehen, wenn Sie ziemlich viele unformatierte Textdaten haben, die durchsucht oder kategorisiert werden müssen, oder wenn Sie eine große Menge interaktiver, textbasierter Abfragen unterstützen.

„Write to Search Index“ nutzt einen Analysator, um Eingabetext in kleine Indizierungselemente, als Token bezeichnet, zu unterteilen. Anschließend werden Suchindexbegriffe aus diesen Token extrahiert. Der Typ des verwendeten Analysators – die Art und Weise, auf die Eingabetext in Token unterteilt wird –, bestimmt, wie Sie später nach diesem Text suchen können. Beispiel: Der Analysator *Schlüsselwort* führt immer einen genauen Vergleich durch und verwendet die gesamte Zeichenfolge als ein Token, während der Analysator *Standard* die Zeichenfolge unterteilt, um Token zu erstellen. Der Analysator *Stoppwort* ist besonders ausgeklügelt und entfernt vor der Tokenbildung Artikel wie „a“ oder „the“ aus der Zeichenfolge, was zu einer kleineren Indexgröße führt.

Suchindizes unterstützen ein Echtzeit-Feature, sodass Indizes fast unverzüglich aktualisiert werden – ohne die Notwendigkeit, den Schritt zu schließen und mithilfe des Suchindex neu zu erstellen.

Weitere Informationen zu den Suchindexaufgaben, die Sie über die Befehlszeile ausführen können, finden Sie unter „Suchindizes“ im Abschnitt [Administrationsumgebung](#) im Administratorhandbuch.

Optionen

1. Doppelklicken Sie im Enterprise Designer auf den „Write to Search Index“-Schritt auf der Arbeitsfläche.
2. Geben Sie einen **Namen** für den Index ein.
3. Wählen Sie einen **Schreibmodus** aus. Wenn Sie einen Index erneut generieren, haben Sie verschiedene Optionen, wie die neuen Daten sich auf die vorhandenen Daten auswirken.
 - **Erstellen oder überschreiben**: Neue Daten überschreiben die vorhandene Daten. Die vorher vorhandenen Daten sind nicht mehr im Index enthalten.
 - **Aktualisieren oder anfügen**: Neue Daten überschreiben vorhandene Daten. Alle neuen Daten, die vorher nicht vorhanden waren, werden dem Index hinzugefügt.
 - **Anfügen**: Neue Daten werden zu den vorhandenen Daten hinzugefügt. Die vorhandenen Daten bleiben intakt.
 - **Löschen**: Daten im ausgewählten Feld werden aus dem Suchindex gelöscht.
4. Wählen Sie das **Schlüsselfeld** danach aus, ob Sie die Datensätze **Aktualisieren oder anfügen** oder **Löschen** möchten.
 - Im Modus **Erstellen oder überschreiben** muss das **Schlüsselfeld** für (in einer verteilten Umgebung verwendete) Elasticsearch-Indizes eindeutig sein. Wenn Sie das Feld leer lassen, werden alle Datensätze im Index gespeichert, unabhängig davon, ob Dubletten entstehen. Sie können dann jedoch bei diesem Index keine Schreibvorgänge wie Aktualisieren, Anfügen und Löschen durchführen. In der folgenden Tabelle wird das Indizierungsverhalten für den Fall erklärt, dass das **Schlüsselfeld** für Lucene- und Elasticsearch-Indizes nicht eindeutig ist.

| Schreibmodus | Schlüsselfeld | Lucene-Index | Elasticsearch-Index |
|----------------------------|---|-------------------------------------|---|
| Create or Overwrite | Datensatzdubletten mit demselben Schlüsselfeld | Alle Datensätze werden gespeichert. | Alle Datensatzdubletten mit demselben Schlüsselfeld werden überschrieben . |
| Aktualisieren oder anfügen | Datensatzdubletten mit demselben Schlüsselfeld | Dubletten werden überschrieben. | Dubletten werden überschrieben. |

5. Aktivieren Sie das Kästchen **Übergabe als Batch**, wenn Sie die Anzahl der Datensätze angeben möchten, die beim Erstellen des Suchindex in einem Batch übergeben werden sollen. Geben Sie dann die Anzahl im Feld **Batchgröße** ein. Der Standardwert ist 5000.
6. Wählen Sie einen **Analysator** für die Erstellung aus:
 - Standard: Bietet einen grammatikbasierten Token-Ersteller, der ein Superset der Analysatoren „Leerzeichen“ und „Stoppwort“ enthält. Erkennt die englische Interpunktion zum Aufschlüsseln von Wörtern, kennt die zu ignorierenden Wörter (über den Stoppwort-Analysator) und führt Suchen durch, welche die Groß-/Kleinschreibung nicht beachten, indem er Vergleiche in Kleinschreibung durchführt. Die Zeichenfolge „Pitney Bowes Software“ würde beispielsweise als drei Token zurückgegeben werden: „pitney“, „bowes“ und „software“. Einen Vergleich zwischen Standard- und Schlüsselwort-Analysator finden Sie unter **Standard- und Schlüsselwort-Analysator** auf Seite 253.
 - Leerzeichen: Trennt Token mit einem Leerzeichen. Eine Art Untergruppe des standardmäßigen Analysators, da Wortumbrüche in englischen Texten basierend auf Leerzeichen und Zeilenumbrüchen erkannt werden.
 - Stoppwort: Entfernt Wörter wie „the“, „and“ und „a“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
 - Schlüsselwort: Erstellt ein einzelnes Token unverändert aus einem Datenstream. Die Zeichenfolge „Pitney Bowes Software“ würde beispielsweise als ein einzelner Token zurückgegeben werden: Pitney Bowes Software. Einen Vergleich zwischen Standard- und Schlüsselwort-Analysator finden Sie unter **Standard- und Schlüsselwort-Analysator** auf Seite 253.
 - Russisch: Unterstützt Indizes und Type-Ahead-Dienste in russischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „und“, „ich“ und „du“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
 - Deutsch: Unterstützt Indizes und Type-Ahead-Dienste in deutscher Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.

- Dänisch: Unterstützt Indizes und Type-Ahead-Dienste in dänischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „um“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
 - Niederländisch: Unterstützt Indizes und Type-Ahead-Dienste in niederländischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
 - Finnisch: Unterstützt Indizes und Type-Ahead-Dienste in finnischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „ist“, „und“ und „von“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
 - Französisch: Unterstützt Indizes und Type-Ahead-Dienste in französischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
 - Ungarisch: Unterstützt Indizes und Type-Ahead-Dienste in ungarischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
 - Italienisch: Unterstützt Indizes und Type-Ahead-Dienste in italienischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
 - Norwegisch: Unterstützt Indizes und Type-Ahead-Dienste in norwegischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
 - Portugiesisch: Unterstützt Indizes und Type-Ahead-Dienste in portugiesischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
 - Spanisch: Unterstützt Indizes und Type-Ahead-Dienste in spanischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
 - Schwedisch: Unterstützt Indizes und Type-Ahead-Dienste in schwedischer Sprache. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „der/die/das“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
 - Hindi: Unterstützt Indizes und Type-Ahead-Dienste in Hindi. Unterstützt ebenfalls viele Stoppwörter und entfernt Wörter wie „von“, „und“ und „ein/e/er/es“, um die Indexgröße zu reduzieren und die Leistung zu verbessern.
7. Um den Analysator aller in der Liste vorhandenen Felder zu aktualisieren, wählen Sie einen Analysator von **Analysatoren aktualisieren auf...** aus.
 8. Um das Schema neu vom Server zu laden, klicken Sie auf **Schema neu laden**.

Anmerkung: Sie können die Feldnamen ändern, indem Sie den neuen Namen direkt in der Spalte **Felder** eingeben. Sie können die **Schrittfelder** oder den **Typ** jedoch nicht ändern.
 9. Um einzelne Felder zu ihrer Eingabequelle hinzuzufügen oder aus ihr zu entfernen, klicken Sie auf **Schnell hinzufügen**. Im Dialogfenster **Schnell hinzufügen** wird eine Liste aller Felder der

Eingabequelle angezeigt. Wählen Sie die Felder aus, die hinzugefügt werden sollen, und klicken Sie auf **OK**.

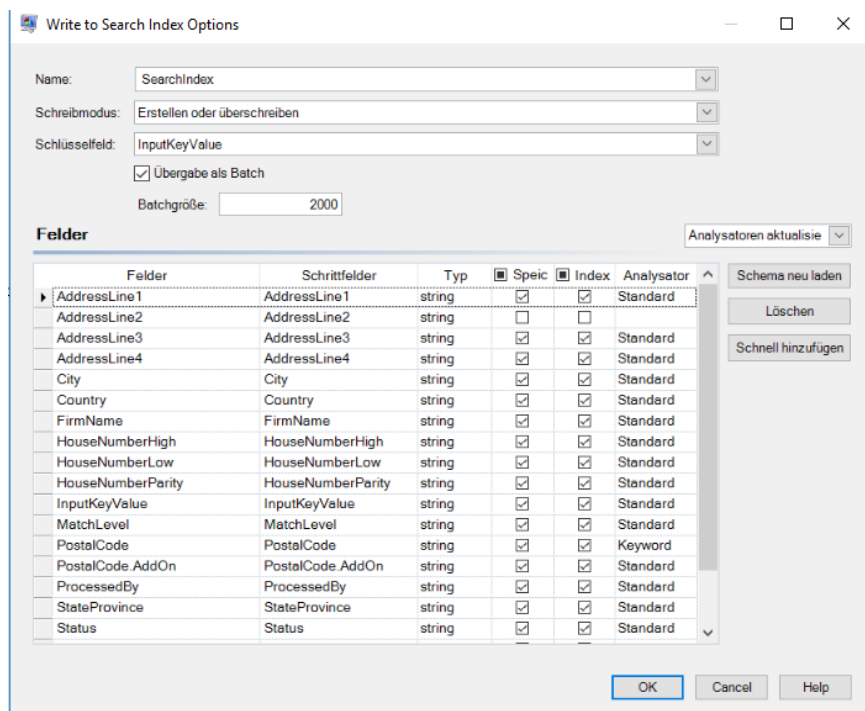
10. Wählen Sie die Felder aus, deren Daten Sie speichern möchten. Bei einer Eingabedatei mit Adressen können Sie beispielsweise nur das Feld „Postleitzahl“ indizieren und die restlichen Felder (wie Adresszeile 1, Stadt, Bundesland/Kanton) speichern, damit die gesamte Adresse zurückgegeben wird, wenn über die Indexsuche eine Übereinstimmung gefunden wird.
11. Wählen Sie die Felder aus, deren Daten zum Index für eine Suchabfrage hinzugefügt werden sollen.

Anmerkung: Wenn Sie einzelne Felder löschen möchten, wählen Sie diese aus und klicken Sie auf **Löschen**.

12. Falls erforderlich ändern Sie den Analysator für jedes Feld, das einen anderen Analysator verwenden soll, als den Analysator, den Sie im Feld „Analysator“ ausgewählt haben.
13. Klicken Sie auf **OK**.

Unten finden Sie ein Beispiel eines abgeschlossenen „Write to Search Index Options“-Schrittes:

- Name des Suchindex
- Modus „Erstellen oder überschreiben“
- ein Schlüsselfeld „InputKeyValue“
- eine Batchgröße bei einer Übergabe als Batch von 2000 Datensätzen
- Verwendung des Standard-Analysators
- eine Liste von Feldern in der Eingabedatei
- eine Liste von Feldern, die zusammen mit den Indexdaten gespeichert werden; in diesem Fall wird nur „AddressLine2“ gespeichert.
- eine Liste von Feldern, die den Index bilden; in diesem Fall wird nur „AddressLine2“ indiziert.
- Verwendung des Schlüsselwort-Analysators für das Feld „PostalCode“



Ausgabe

„Write to Search Index“ hat einen Ausgabeport namens **Fehlerport**, der Daten für Datensätze sammelt, die nicht verarbeitet und dem Suchindex hinzugefügt werden konnten. Datensätze, die durch den Fehlerport in die Datenladung gelangen, werden als falsch formatiert gewertet.

Datensätze werden zum Fehlerport weitergeleitet, wenn der Wert von **KeyField** im Eingabedatensatz leer oder null ist. Bei Suchindex-Engines, die eigenständige Suchvorgänge (die alte Index-Engine) unterstützen, funktioniert der Fehlerport nur für die Aktualisierungs- und Löschvorgänge. Bei der Index-Engine mit verteilter Unterstützung sammelt er jedoch falsch formatierte Datensätze für alle vier Vorgänge: Erstellen, Aktualisieren, Löschen und Anfügen.

Suchindexverwaltung

Mithilfe des Tools **Suchindexverwaltung** können Sie folgende Aufgaben erledigen:

- Felder zu einem vorhandenen Suchindex hinzufügen oder aus ihm entfernen
- einen oder mehrere Suchindizes löschen

So fügen Sie Felder zu einem vorhandenen Suchindex hinzu oder entfernen sie aus ihm:

1. Wählen Sie **Tools > Suchindexverwaltung** aus. Im Dialogfenster **Suchindexverwaltung** werden die vorhandenen Suchindizes angezeigt.
2. Wählen Sie den zu ändernden Index aus und klicken Sie auf **Ändern**. Die Seite **Index ändern** wird angezeigt. Sie listet Liste aller Felder im Suchindex mit ihren Details auf.
3. Um ein Feld zu löschen, wählen Sie es aus der Liste aus und klicken Sie auf **Feld entfernen**.

4. Um ein neues Feld hinzuzufügen, klicken Sie auf **Feld hinzufügen**. Im Dialogfenster **Feld hinzufügen** können Sie die folgenden Details hinzufügen:

- **Feldname:** Geben Sie den Namen des Feldes ein.
- **Typ:** Wählen Sie den Feldtyp aus. Folgende Optionen sind verfügbar: string, integer, long, float und double.
- **Index:** Aktivieren Sie das Kästchen, um das Feld zum Index für Suchabfragen hinzuzufügen.
- **Speichern:** Aktivieren Sie das Kästchen, um das Feld zu speichern.
- **Analysator:** Wählen Sie den Analysator aus der Liste aus, den Sie zur Indizierung des Feldes verwenden möchten.

Anmerkung: Diese Option ist nur aktiviert, wenn Sie die Indizierung dieses Feldes aktiviert haben.

So löschen Sie einen vorhandenen Suchindex:

1. Wählen Sie **Tools > Suchindexverwaltung** aus.
2. Wählen Sie die zu löschenden Suchindizes aus.
3. Klicken Sie auf **Löschen**.
4. Klicken Sie auf **Schließen**.

Anmerkung: Sie können einen Suchindex auch über die Administrationsumgebung löschen. Der Befehl lautet `index delete --n Indexname`, wobei „Indexname“ der Name des zu löschenden Index ist.

Standard- und Schlüsselwort-Analysator

Bevor Sie sich zwischen dem Standard- und dem Schlüsselwort-Analysator entscheiden, sollten Sie folgende Unterschiede in der Arbeitsweise dieser beiden Analysatoren berücksichtigen:

- **Standard-Analysator:** Er unterteilt die Zeichenfolge, um sie zu tokenisieren, und konvertiert alle Token in Kleinbuchstaben.
- **Schlüsselwort-Analysator:** Er tokenisiert die gesamte Zeichenfolge unverändert (ändert nicht die Groß-/Kleinschreibung).

Beispiel:

Das folgende Beispiel verwendet den Algorithmus *Enthält beliebiges* und die Eingabe „P O“. In diesem Fall gibt der **Standard**-Analysator sowohl Datensätze mit „P O“ als auch mit „P“ zurück, während der **Schlüsselwort**-Analysator nur Datensätze mit „P O“ zurückgibt (siehe unten).

Suchergebnis mit Standard-Analysator:

| Point | | Point 2 | | | | | | |
|------------|----------------|---------------|-------|------------|--------|-----------------------|-----------|--|
| SearchTerm | CandidateGroup | HasDuplicates | Index | IndexField | Term | TransactionRecordType | | |
| P O | 1 | Y | | | | | Suspect | |
| P O | 1 | D | P O | 4 | Test4 | | Candidate | |
| P O | 1 | D | P | 1 | Test 1 | | Candidate | |

Suchergebnis mit Schlüsselwort-Analysator:

| Point | | Point 2 | | | | | | |
|------------|----------------|---------------|-------|------------|-------|-----------------------|--|--|
| SearchTerm | CandidateGroup | HasDuplicates | Index | IndexField | Term | TransactionRecordType | | |
| PO | 1 | Y | | | | Suspect | | |
| PO | 1 | D | PO | 4 | Test4 | Candidate | | |

Business Steward-Modul

Business Steward-Modul

Das Business Steward-Modul besteht aus einer Reihe von Features, mit denen Sie Ausnahmedatensätze identifizieren und auflösen können. Bei diesen Datensätzen handelt es sich um Datensätze, die von Spectrum™ Technology Platform nicht zuverlässig verarbeitet werden konnten und von einem Dateneigner manuell überprüft werden müssen. Einige Beispiele für diese Datensätze lauten wie folgt:

- Fehler bei Adressenüberprüfungen
- Fehler beim Geocoding
- Übereinstimmungen mit niedrigem Zuverlässigkeitsniveau
- Zusammenführungs-/Konsolidierungsentscheidungen

Das Business Steward-Modul stellt eine browserbasierte Lösung für die manuelle Überprüfung von Datensätzen bereit. Sobald Datensätze manuell korrigiert und genehmigt wurden, können sie wieder in Ihren Spectrum™ Technology Platform-Datenqualitätsprozess eingegliedert werden.

Viele der Aktionen, die innerhalb des Business Steward-Moduls ausgeführt werden, werden im Überwachungsprotokoll widerspiegelt, einer Funktion der Management Console, mit der Benutzeraktivität aufgezeichnet wird. Die folgenden Aktionen sind im Protokoll enthalten:

- Hinzufügen von Ausnahmen im Schritt „Write Exceptions“
- Löschen von Ausnahmen im Schritt „Read Exceptions“ und auf der Seite „Ausnahmen verwalten“ im Business Steward Portal
- Zuweisen von Ausnahmen im Exception Manager des Business Steward Portal
- Abrufen von Ausnahmen im Ausnahmen-Editor des Business Steward Portal
- Aktualisieren von Ausnahmen im Ausnahmen-Editor des Business Steward Portal
- Erneutes Überprüfen von Ausnahmen durch Klicken auf „Speichern“ und „Erneut überprüfen“ im Editor des Business Steward Portal

Weitere Informationen zum Prüfprotokoll finden Sie im Spectrum™ Technology Platform-Administratorhandbuch für die Webbenutzeroberfläche.

Komponenten

Das Business Steward-Modul setzt sich aus folgenden Komponenten zusammen:

- **Exception Monitor:** Ein Schritt, in dem Datensätze auf Basis von einer Reihe von Bedingungen ausgewertet werden, um zu ermitteln, ob eine manuelle Überprüfung durch einen Dateneigner erforderlich ist. Wenn Datensätze diese Bedingungen erfüllen, kann in diesem Schritt eine E-Mail versendet werden, in der Empfänger über die Ausnahmen benachrichtigt werden.
- **Write Exceptions:** Ein Schritt, in dem Ausnahmedatensätze in die Ausnahme-Datenbank geschrieben werden. Sobald die Ausnahmedatensätze in der Datenbank enthalten sind, stehen sie dem Dateneigner zur Überprüfung zur Verfügung.
- **Exception Dashboard und Editor:** Eine browserbasierte Lösung, in der ein Dashboard mit Zusammenfassungsstatistiken und Diagrammen angezeigt wird, durch die Sie die in Ihren Daten auftretenden Ausnahmearten nachvollziehen können. Es dient zudem als Editor, mit dem Sie Ausnahmedatensätze bearbeiten und für die erneute Verarbeitung genehmigen können.
- **Manage Exceptions:** Eine browserbasierte Lösung, mit der Sie die Ausnahmedatensatzaktivität für alle Benutzer überprüfen und verwalten können.
- **Data Quality Performance:** Eine browserbasierte Lösung, die Informationen zu Trends innerhalb Ihrer Ausnahmedatensätze enthält und mit dem Sie wichtige Leistungsindikatoren ermitteln und Benachrichtigungen senden können, wenn bestimmte Bedingungen erfüllt wurden.
- **Read Exceptions:** Ein Schritt, bei dem Datensätze aus der Ausnahme-Datenbank in einen Datenfluss eingelesen werden. In diesem Schritt können Sie von einem Dateneigner korrigierte Ausnahmedatensätze erneut verarbeiten.

Exception Monitor

Im „Exception Monitor“-Schritt werden Datensätze auf Basis von einer Reihe von Bedingungen ausgewertet, um zu ermitteln, ob eine manuelle Überprüfung des Datensatzes durch einen Data Steward erforderlich ist. Mit Exception Monitor können Sie Datensätze weiterleiten, die Spectrum™ Technology Platform nicht erfolgreich für ein Tool zur manuellen Überprüfung (das Business Steward Portal) verarbeiten konnte.

Einige Beispiele für Ausnahmen lauten wie folgt:

- Fehler bei Adressenüberprüfungen
- Fehler beim Geocoding
- Übereinstimmungen mit niedrigem Zuverlässigkeitsniveau
- Zusammenführungs-/Konsolidierungsentscheidungen

Sie können nicht nur Bedingungen festlegen, durch die bestimmt wird, ob Datensätze manuell überprüft werden müssen, sondern auch Exception Monitor so konfigurieren, dass eine Benachrichtigung an mindestens eine E-Mail-Adresse gesendet wird, wenn diese Bedingungen in einer bestimmten Häufigkeit erfüllt wurden.

Weitere Informationen zur Verarbeitung von Ausnahmen finden Sie unter [Business Steward-Modul](#) auf Seite 254.

Eingabe

Exception Monitor wählt einen beliebigen Datensatz als Eingabe. Wenn die Eingabedaten kein Feld mit dem Namen „Sammlungsnummer“ enthalten, wird die Option **Alle Datensätze in der Gruppe der Ausnahme zurückgeben** deaktiviert.

Anmerkung: Exception Monitor kann keine Felder überwachen, die komplexe Daten enthalten, wie z. B. Listen oder Geometrie-Objekte.

Optionen

Registerkarte „Bedingungen“

Tabelle 18: Exception Monitor-Optionen

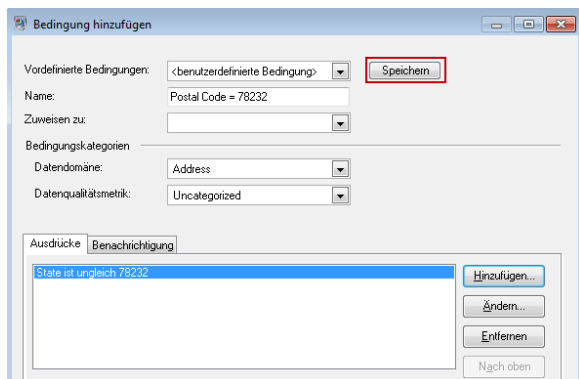
| Name der Option | Beschreibung |
|---|---|
| Auswertung beenden, wenn eine Bedingung erfüllt ist | Gibt an, ob nach der Erfüllung einer Bedingung mit der Auswertung eines Datensatzes auf Basis der verbleibenden Bedingungen fortgefahren werden soll. Durch Aktivierung dieser Option kann die Leistung verbessert werden, da die Anzahl der vom System durchzuführenden Auswertungen möglicherweise reduziert wird. Wenn nicht alle Bedingungen ausgewertet werden, geht jedoch ein gewisser Grad der Vollständigkeit in den im Business Steward Portal angezeigten Ausnahmeberichten verloren. Beispiel: Wenn Sie drei Bedingungen („Vollständigkeit der Adresse“, „Namenszuverlässigkeit“ Und „Geocodezuverlässigkeit“) definieren, ein Datensatz die unter „Vollständigkeit der Adresse“ definierten Kriterien erfüllt und Sie diese Option aktivieren, würde der Datensatz nicht auf Basis von „Namenszuverlässigkeit“ Und „Geocodezuverlässigkeit“ ausgewertet werden. Wenn sich der Datensatz zudem für eine Ausnahme qualifiziert, da er die Bedingung „Namenszuverlässigkeit“ erfüllt, würden diese Informationen nicht erfasst werden. Stattdessen würde gemeldet werden, dass bei dem Datensatz statt unter „Vollständigkeit der Adresse“ und „Namenszuverlässigkeit“ nur unter „Vollständigkeit der Adresse“ ein Problem aufgetreten ist. |

Hinzufügen oder Ändern von Bedingungen und Ausdrücken

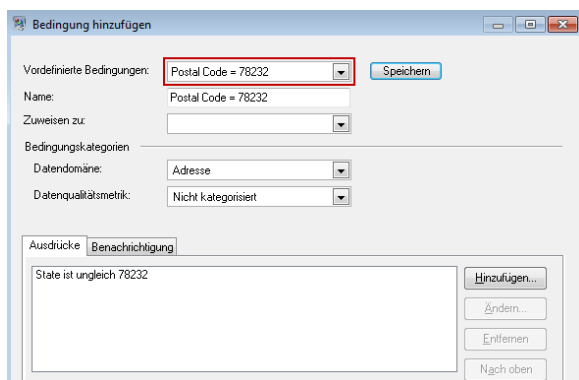
Eine Bedingung definiert die Kriterien, mit denen bestimmt wird, ob ein Datensatz eine „Ausnahme“ darstellt und zur manuellen Überprüfung weitergeleitet werden muss. In der Regel bedeutet dies, dass Sie Bedingungen definieren sollten, über die permanent Datensätze identifiziert werden können, bei denen entweder zuvor im Datenfluss die automatisierte Verarbeitung fehlgeschlagen ist oder die einen geringen Zuverlässigkeitsgrad aufweisen und daher manuell überprüft werden müssen.

Im „Exception Monitor“-Schritt können Sie über das Dialogfeld „Bedingung hinzufügen“ vordefinierte Bedingungen und benutzerdefinierte Bedingungen erstellen. Vordefinierte Bedingungen sind für alle Datenflüsse verfügbar, während benutzerdefinierte Bedingungen nur für die Datenflüsse verfügbar sind, für die sie erstellt wurden. Der Konfigurationsprozess ist bei beiden Typen fast identisch. Zur

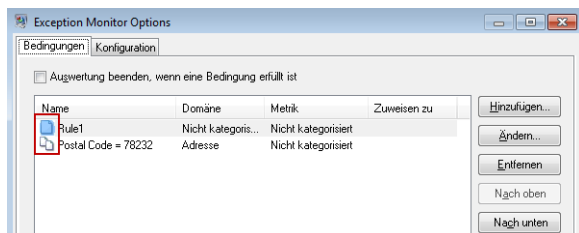
Erstellung einer vordefinierten Bedingung müssen Sie diese jedoch speichern, indem Sie die Felder ausfüllen und auf **Speichern** klicken, wie nachstehend im roten Feld abgebildet.



Nachdem Sie eine benutzerdefinierte Bedingung gespeichert haben, ändert sich das Feld **Vorgefertigte Bedingungen**, damit statt „<benutzerdefinierte Bedingung>“ der Name der Bedingung angezeigt wird.



Nach der Erstellung vordefinierter oder benutzerdefinierter Bedingungen werden diese auf der Registerkarte „Bedingungen“ im Dialogfeld „Exception Monitor-Optionen“ angezeigt. Wie im folgenden Bild dargestellt, gibt das Symbol neben dem Namen der Bedingung an, ob es sich um eine vordefinierte oder um eine benutzerdefinierte Bedingung handelt. Das Symbol mit zwei Dokumenten steht für eine vordefinierte Bedingung und das Symbol mit einem Dokument steht für eine benutzerdefinierte Bedingung.



1. Klicken Sie auf der Registerkarte „Bedingungen“ im Fenster „**Exception Monitor-Optionen**“ auf **Hinzufügen**, um eine neue Bedingung zu erstellen, oder auf **Ändern**, um eine vorhandene Bedingung zu bearbeiten. Füllen Sie die folgenden Felder aus:

- **Vordefinierte Bedingungen** – Wählen Sie eine vordefinierte Bedingung aus oder behalten Sie „<benutzerdefinierte Bedingung>“ im Dropdown-Menü bei, um eine neue Bedingung zu erstellen.
- **Name** – Ein Name für die Bedingung. Sie können einen beliebigen Namen auswählen. Da der Bedingungsname im Business Steward Portal angezeigt wird, sollten Sie einen beschreibenden Namen verwenden. Beispiel: „MatchScore < 80“ oder „FailedDPV“. Wenn Sie versuchen, einer neuen Bedingung einen Namen zu geben, der mit dem Namen einer vorhandenen Bedingung identisch ist, jedoch am Ende andere Zeichen enthält (z. B. „FailedDPV“ und „FailedDPV2“), werden Sie gefragt, ob Sie die vorhandene Bedingung überschreiben möchten, wenn Sie das letzte Zeichen eingegeben haben, das mit dem zugehörigen Namen identisch ist (in unserem Beispiel „n“). Beantworten Sie dies mit „Ja“ und geben Sie den Namen der Bedingung fertig ein. Wenn Sie auf **OK** oder auf **Speichern** klicken, sind im Dialogfeld „Exception Monitor-Optionen“ beide Bedingungen sichtbar. Die neue Bedingung überschreibt nur dann die vorhandene Bedingung, wenn ihre Namen hundertprozentig identisch sind.
- **Zuweisen zu** – Wählen Sie einen Benutzer aus, dem die Ausnahmedatensätze zugewiesen werden sollten, die diese Bedingung erfüllen. Wenn Sie in diesem Feld keine Auswahl treffen, werden die Ausnahmedatensätze automatisch dem Benutzer zugewiesen, der den Auftrag ausgeführt hat.
- **Datendomäne** – (Optional) Gibt die Art von Daten an, die von der Bedingung ausgewertet werden. Dies wird lediglich für die Berichterstellung verwendet, um zu zeigen, welche Ausnahmetypen in Ihren Daten auftreten. Wenn die Bedingung beispielsweise den Erfolg oder Misserfolg einer Adressenüberprüfung auswertet, könnte die Datendomäne „Adresse“ lauten. Wenn die Bedingungen den Erfolg oder Misserfolg eines Geocoding-Vorgangs auswertet, könnte die Datendomäne „Räumlich“ usw. lauten. Sie können Ihre eigene Datendomäne angeben oder eine der vordefinierten Domänen auswählen:
 - Account – Die Bedingung überprüft einen Unternehmens- oder Organisationsnamen, der einem Verkaufskonto zugeordnet ist.
 - Address – Die Bedingung überprüft Adressdaten, z. B. eine vollständige Postanschrift oder eine Postleitzahl.
 - Asset – Die Bedingung überprüft Daten zum Eigentum eines Unternehmens, z. B. physisches Eigentum, Immobilien, Personal oder andere Objekte.
 - Date – Die Bedingung überprüft Daten zu Datumsangaben.
 - Email – Die Bedingung überprüft E-Mail-Daten.
 - Financial – Die Bedingung überprüft Daten zu Währung, Sicherheiten usw.
 - Name – Die Bedingung überprüft Daten zu Personennamen, z. B. einen Vor- oder Nachnamen.
 - Phone – Die Bedingung überprüft Daten zu Telefonnummern.
 - Product – Die Bedingung überprüft Daten zu Materialien, Artikeln, Warenhandel usw.
 - Spatial – Die Bedingung überprüft Daten zu Punkten, Vielecken oder Linien, die eine definierte geografische Besonderheit darstellen, z. B. Überschwemmungsgebiete, Küstenlinien, Häuser, Absatzgebiete usw.
 - SSN – Die Bedingung überprüft Daten zur US- Sozialversicherungsnummer.

- Uncategorized – Wählen Sie diese Option aus, wenn diese Bedingung nicht kategorisiert werden soll.
 - **Datenqualitätsmetrik** – (Optional) Gibt die Metrik an, die von dieser Bedingung erfasst wird. Dies wird lediglich für die Berichterstellung verwendet, um zu zeigen, welche Ausnahmetypen in Ihren Daten auftreten. Wenn die Bedingung beispielsweise die Vollständigkeit eines Datensatzes auswerten soll (also z. B. ob alle Adressen eine Postleitzahl enthalten), könnten Sie „Vollständigkeit“ als Datenqualitätsmetrik angeben. Sie können Ihre eigene Metrik angeben oder eine der vordefinierten Metriken auswählen:
 - Accuracy – Die Bedingung erfasst, ob die Daten auf Basis einer vertrauenswürdigen Quelle überprüft werden konnten. Wenn eine Adresse beispielsweise nicht anhand der Daten von der Postbehörde überprüft werden konnte, könnte sie als Ausnahme gelten, da sie nicht korrekt ist.
 - Completeness – Die Bedingung erfasst, ob in den Daten wichtige Attribute fehlen. Beispiel: eine Adresse, bei der die Postleitzahl fehlt, oder ein Konto ohne Namen der Kontaktperson.
 - Consistency – Die Bedingung erfasst, ob die Daten zwischen mehreren Systemen konsistent sind. Wenn Ihr Kundendatensystem beispielsweise die Geschlechtscodes M und F verwendet, die von Ihnen verarbeiteten Daten jedoch die Geschlechtscodes 0 und 1 aufweisen, könnte bei den Daten von einem Konsistenzproblem ausgegangen werden.
 - Interpretability – Die Bedingung erfasst, ob Daten korrekt in eine Datenstruktur geparkt wurden, die von einem anderen System interpretiert werden kann. Sozialversicherungsnummern sollten beispielsweise nur numerische Daten enthalten. Wenn die Daten Buchstaben enthalten, z. B. xxx-xx-xxxx, könnte bei ihnen von Auswertbarkeitsproblemen ausgegangen werden.
 - Recency – Die Bedingung erfasst, ob die Daten auf dem neuesten Stand sind. Wenn beispielsweise eine Person umzieht, in Ihrem System jedoch die alte Adresse dieser Person gespeichert ist, könnte bei den Daten von einem Aktualitätsproblem ausgegangen werden.
 - Uncategorized – Wählen Sie diese Option aus, wenn diese Bedingung nicht kategorisiert werden soll.
 - Uniqueness – Die Bedingung erfasst, ob Daten doppelt vorliegen. Wenn der Datenfluss keine doppelten Daten konsolidieren konnte, könnte davon ausgegangen werden, dass es sich bei den Datensätzen um Ausnahmen handelt.
2. Sie müssen mindestens eine Ausnahme zu der Bedingung hinzufügen. Ein Ausdruck ist eine logische Anweisung, die den Wert eines Feldes überprüft. Klicken Sie auf **Hinzufügen**, um einen Ausdruck hinzuzufügen. Klicken Sie auf **Ändern**, um einen vorhandenen Ausdruck zu ändern. Füllen Sie die folgenden Felder aus:
- **Mit Ausdrucks-Generator erstellter Ausdruck** – Wählen Sie diese Option aus, um einen grundlegenden Ausdruck zu erstellen.
 - **Benutzerdefinierter Ausdruck** – Wählen Sie diese Option aus, um über Groovy-Scripting einen Ausdruck zu schreiben. Verwenden Sie einen benutzerdefinierten Ausdruck, wenn Sie eine komplexere Logik verwenden müssen, z. B. verschachtelte Auswertungen. Weitere Informationen finden Sie unter [Verwenden benutzerdefinierter Ausdrücke in Exception Monitor](#) auf Seite 261.

- Wenn für diese Bedingung bereits andere Ausdrücke definiert wurden, können Sie im Feld **Logischer Operator** einen Operator auswählen. Zur Auswahl stehen:
 - **And** – Dieser Ausdruck muss zusätzlich zu dem vorherigen Ausdruck wahr sein, damit die Bedingung wahr ist.
 - **Or**: – Wenn dieser Ausdruck wahr ist, kann die Bedingung selbst dann wahr sein, wenn der vorherige Ausdruck nicht wahr ist.
 - Wenn Sie einen Ausdruck mit dem Ausdrucks-Generator erstellen möchten, sind folgende Felder verfügbar:
 - **Feldname** – Wählen Sie das Feld aus, das von diesem Ausdruck ausgewertet werden soll. Die Liste der verfügbaren Felder wird basierend auf den Schritten vor dem „Exception Monitor“-Schritt aufgefüllt.
 - **Operator** – Wählen Sie den Operator aus, den Sie bei der Auswertung verwenden möchten.
 - **Wert** – Geben Sie den Wert an, der mit dem im Feld „Operator“ ausgewählten Operator von dem Ausdruck überprüft werden soll.
3. Klicken Sie auf **Hinzufügen**, um den Ausdruck hinzuzufügen. Klicken Sie auf **Schließen**, wenn Sie mit dem Hinzufügen von Ausdrücken fertig sind.
 4. Verwenden Sie die Schaltflächen **Nach oben** und **Nach unten**, um die Reihenfolge der Auswertung von Ausdrücken zu ändern.
 5. Klicken Sie auf die Registerkarte **Benachrichtigung**, wenn Exception Monitor eine Nachricht an eine oder mehrere E-Mail-Adressen senden soll, sobald die Bedingung in einer bestimmten Häufigkeit erfüllt wurde. Diese E-Mail enthält einen Link zu den fehlgeschlagenen Datensätzen im Ausnahmen-Editor des Business Steward Portal, wo Sie manuell die korrekten Daten eingeben können. Wenn Sie keine Benachrichtigungen mehr einrichten möchten, fahren Sie mit Schritt 11 fort. Um den Empfang von Benachrichtigungen an einer bestimmten E-Mail-Adresse zu stoppen, entfernen Sie diese Adresse in der Zeile **Benachrichtigungen senden an** auf der Registerkarte „Benachrichtigungen“ im Dialogfeld „Bedingung ändern“ aus der Liste der Empfänger.

Anmerkung: Bevor Sie erfolgreich eine Benachrichtigung aus Exception Monitor verwenden können, müssen Sie die Benachrichtigungen in der Management Console einrichten. Weitere Informationen zur Konfiguration von Benachrichtigungen finden Sie im *Administratorhandbuch*.
 6. Geben Sie die E-Mail-Adresse(n) ein, an die die Benachrichtigung gesendet werden soll. Trennen Sie mehrere Adressen mit Kommas, Leerzeichen oder Semikolons voneinander.
 7. Geben Sie an, wann eine Benachrichtigung gesendet werden soll. Sie können festlegen, dass sie beim ersten Eintreten der Bedingung gesendet wird, oder dass sie gesendet wird, wenn die Bedingung in einer bestimmten Häufigkeit eingetreten ist. Der maximale Wert beträgt 1.000.000 Vorkommen.
 8. Aktivieren Sie das Feld **Erinnerung senden nach**, wenn nach der ersten E-Mail Erinnerungsnachrichten an die angegebene(n) E-Mail-Adresse(n) gesendet werden sollen.
 9. Geben Sie die Anzahl der Tage nach der ersten E-Mail an, nach denen die E-Mail-Erinnerungsnachricht gesendet werden soll.

10. Klicken Sie auf **Täglich erinnern**, wenn nach der ersten E-Mail-Erinnerung täglich Erinnerungsnachrichten gesendet werden sollen.
11. Klicken Sie auf **Speichern**, wenn Sie diese Bedingung zur Wiederverwendung als vordefinierte Bedingung speichern möchten. Wenn Sie eine vorhandene Bedingung ändern und auf „Speichern“ klicken, werden Sie gefragt, ob Sie die vorhandene Bedingung überschreiben möchten. Beachten Sie, dass die durch das Überschreiben einer vordefinierten Bedingung entstehenden Änderungen auf alle Datenflüsse Auswirkungen haben, die diese Bedingung verwenden.
12. Klicken Sie auf **OK**, wenn Sie mit dem Arbeiten mit Ausdrücken fertig sind.
13. Fügen Sie nach Bedarf zusätzliche Bedingungen hinzu oder ändern Sie diese.
14. Verwenden Sie die Schaltflächen **Nach oben** und **Nach unten**, um die Reihenfolge der Auswertung von Bedingungen zu ändern. Die Reihenfolge der Bedingungen ist nur dann wichtig, wenn Sie die Option **Auswertung beenden, wenn eine Bedingung erfüllt ist** aktiviert haben. Weitere Informationen finden Sie unter **Registerkarte „Konfiguration“** auf Seite 263.
15. Klicken Sie nach Abschluss des Vorgangs auf **OK**.

Entfernen einer Bedingung oder eines Ausdrucks

- Um eine Bedingung zu entfernen, öffnen Sie Exception Monitor, wählen Sie die Bedingung aus, die Sie entfernen möchten, und klicken Sie dann auf **Entfernen**. Beachten Sie, dass beim Entfernen einer Bedingung alle Ausdrücke in der Bedingung entfernt werden.
- Um einen Ausdruck zu entfernen, öffnen Sie die Bedingung, die den Ausdruck enthält, und klicken Sie dann auf **Entfernen**.

Verwenden benutzerdefinierter Ausdrücke in Exception Monitor

Sie können Ihre eigenen benutzerdefinierten Ausdrücke schreiben, um zu steuern, wie Exception Monitor zum Erstellen eines Ausdrucks über Groovy-Scripting Datensätze weiterleitet.

Verwenden von Groovy-Scripting

Informationen zu Groovy finden Sie unter groovy-lang.org.

Die Auswertung der im „Exception Monitor“-Schritt verwendeten Groovy-Ausdrücke muss einen booleschen Wert (wahr oder falsch) ergeben, der darauf hindeutet, ob der Datensatz als Ausnahme gilt und zur manuellen Überprüfung weitergeleitet werden soll. Ausnahmedatensätze werden an den Ausnahmeport weitergeleitet.

Wenn Sie beispielsweise Datensätze mit einem Zuverlässigkeitsniveau für Überprüfungen in Höhe von < 85 überprüfen müssen, würde Ihr Script folgendermaßen aussehen:

```
data['Confidence']<85
```

Bei der Überwachung würde der Wert im Feld „Zuverlässigkeit“ auf Basis Ihrer Kriterien ausgewertet werden, damit ermittelt werden kann, an welchen Ausgabeport er gesendet werden soll.

Überprüfen eines Feldes auf einen einzelnen Wert

Bei diesem Beispiel ergibt die Auswertung „wahr“, wenn das Statusfeld „F“ enthält. Dies müsste eine genaue Übereinstimmung sein, die Auswertung von „f“ würde folglich nicht „wahr“ ergeben.

```
return data['Status'] == 'F';
```

Überprüfen eines Feldes auf mehrere Werte

Bei diesem Beispiel ergibt die Auswertung „wahr“, wenn das Statusfeld „F“ oder „f“ enthält.

```
boolean returnValue = false;
if (data['Status'] == 'F' || data['Status'] == 'f')
{
    returnValue = true;
}
return returnValue;
```

Auswerten der Feldlänge

Bei diesem Beispiel ergibt die Auswertung „wahr“, wenn das Feld „PostalCode“ mehr als 5 Zeichen enthält.

```
return data['PostalCode'].length() > 5;
```

Prüfen auf ein Zeichen innerhalb eines Feldwertes

Bei diesem Beispiel ergibt die Auswertung „wahr“, wenn das Feld „PostalCode“ einen Gedankenstrich enthält.

```
boolean returnValue = false;
if (data['PostalCode'].indexOf('-') != -1)
{
    returnValue = true;
}
return returnValue;
```

Häufige Fehler

Im Folgenden werden häufige Fehler bei der Verwendung des Skriptings aufgeführt.

Das folgende Beispiel ist falsch, da „PostalCode“ (der Spaltenname) in einfachen oder doppelten Anführungszeichen stehen muss

```
return data[PostalCode];
```

Das folgende Beispiel ist falsch, da keine Spalte angegeben ist

```
return data[];
```

Das folgende Beispiel ist falsch, da „row.set()“ keinen booleschen Wert zurückgibt. Die Auswertung wird immer „falsch“ ergeben und das Feld „PostalCode“ wird immer in 88989 geändert.

```
return row.set('PostalCode', '88989');
```

Verwenden Sie ein einfaches Gleichheitszeichen, um den Wert eines Feldes festzulegen, und ein doppeltes Gleichheitszeichen, um den Wert eines Feldes zu überprüfen.

Registerkarte „Konfiguration“

Tabelle 19: Exception Monitor-Optionen

| Name der Option | Beschreibung |
|--|--|
| Exception Monitor deaktivieren | Schaltet Exception Monitor ein oder aus. Wenn Sie Exception Monitor deaktivieren, durchlaufen Datensätze einfach den Schritt, ohne dass Maßnahmen ergriffen werden müssen. Dies ist vergleichbar mit dem Entfernen von Exception Monitor aus dem Datenfluss. |
| Auftrag nach Erreichen der Ausnahmengrenze beenden | Gibt an, ob die Ausführung des Auftrags abgebrochen werden soll, wenn die angegebene Anzahl der Datensätze die Ausnahmebedingungen erfüllt. |
| Maximale Anzahl von Ausnahmedatensätzen | Wenn die Option Auftrag nach Erreichen der Ausnahmengrenze beenden ausgewählt wurde, können Sie in diesem Feld die maximal zulässige Anzahl der Ausnahmedatensätze angeben, bevor die Ausführung des Auftrags abgebrochen wird. Wenn Sie beispielsweise 100 angeben, wird der Auftrag beendet, sobald der 101. Datensatz aufgetreten ist. |
| Nur Bericht (keine Ausnahmen erstellen) | Ermöglicht es Ihnen, Datensätze nachzuverfolgen, die Ausnahmebedingungen erfüllen und meldet diese Statistiken auf der Seite „Datenqualitäts-Leistung“ im Business Steward Portal. Mit der Option werden jedoch keine Ausnahmen für diese Datensätze erstellt. |

| Name der Option | Beschreibung |
|--|---|
| Alle Datensätze in der Gruppe der Ausnahme zurückgeben | <p>Gibt an, ob alle Datensätze zurückgegeben werden sollen, die zu der Gruppe des Ausnahmedatensatzes gehören, statt nur den Datensatz zurückzugeben. Beispiel: Eine Übereinstimmungsgruppe (basierend auf einem Übereinstimmungsschlüssel) enthält vier Datensätze. Bei einem handelt es sich um die Datensatzkopfdublette, bei dem anderen um eine Dublette mit einer Punktzahl von 90 und bei den zwei letzten Datensätzen handelt es sich um eindeutige Datensätze mit einer Punktzahl von 80 und 83. Wenn Sie über eine Bedingung verfügen, in der angegeben wird, dass jeder Datensatz mit einem MatchScore zwischen 80 und 89 eine Ausnahme darstellt, würden standardmäßig nur die Datensätze mit einem MatchScore von 80 und 83 an den Ausnahmeport gesendet werden. Wenn Sie diese Option aktivieren, würden jedoch alle vier Datensätze an den Ausnahmeport gesendet werden.</p> <p>Aktivieren Sie diese Option, wenn Sie möchten, dass Dateneigner den Ausnahmedatensatz mit anderen Datensätzen in der Gruppe vergleichen können. Durch den Vergleich aller Datensätze in der Gruppe können Dateneigner möglicherweise fundiertere Entscheidungen bzgl. der Handhabung eines Ausnahmedatensatzes treffen. In einer Vergleichssituation könnte ein Dateneigner beispielsweise alle Kandidaten sehen, um zu ermitteln, ob es sich bei der Ausnahme um eine Dublette handelt.</p> <p>Anmerkung:</p> <p>Wenn die Eingabedaten kein Feld mit dem Namen „Sammlungsnummer“ enthalten, wird diese Option deaktiviert.</p> |
| Gruppieren nach | <p>Wenn Sie Alle Datensätze in der Gruppe der Ausnahme zurückgeben ausgewählt haben, wählen Sie das Feld aus, nach dem die Datensätze gruppiert werden sollen.</p> <p>Anmerkung: Das Eingabefeld „Sammlungsnummer“ wird in dieser Liste nicht angezeigt, da es bei dem Feature „Gruppieren nach“ keine gültige Auswahl darstellt.</p> |
| Dienst für erneute Überprüfung | Wählen Sie den Dienst aus, der ausgeführt werden soll, wenn Sie Datensätze aus diesem Datenfluss erneut überprüfen. |
| Aktion nach erneuter Überprüfung | Gibt an, ob Sie Datensätze erneut verarbeiten möchten oder ob Sie Datensätze genehmigen möchten, die erneut überprüft worden sind. |

| Name der Option | Beschreibung |
|--|--|
| Ausnahmedatensätze anhand von Vergleichsfeld vergleichen | <p>Verwendet Vergleichsfelder, um Eingabedatensätze mit Ausnahmedatensätzen in der Datenbank zu vergleichen. Aktivieren Sie diese Option, wenn Ihre Eingabe Datensätze enthält, durch die zuvor Ausnahmen generiert wurden, die jedoch jetzt in der Eingabe korrigiert wurden.</p> <p>Die Eingabedatensätze werden auf Basis der Bedingung(en) ausgewertet und anschließend mit den vorhandenen Ausnahmedatensätzen in der Datenbank verglichen. Wenn ein Eingabedatensatz die Bedingung(en) erfüllt und mit einem Ausnahmedatensatz übereinstimmt, wird dieser Datensatz aus der Datenbank entfernt. Wenn ein Eingabedatensatz die Bedingung(en) nicht erfüllt und mit einem Ausnahmedatensatz übereinstimmt, wird dieser Datensatz aktualisiert und bleibt in der Datenbank erhalten. Wenn in der Datenbank Dubletten vorhanden sind, wird zudem nur eine übereinstimmende Ausnahme pro Datenfluss aktualisiert. Alle anderen Dubletten dieses Datenflusses werden gelöscht.</p> |
| Vergleichsfelder | <p>Enthält eine Liste mit allen Eingabefeldern, die zum Erstellen eines Schlüssels für den Vergleich eines Ausnahmedatensatzes in der Datenbank verwendet werden. Sie müssen mindestens ein Vergleichsfeld definieren, wenn Sie das Feld Datensätze anhand von Vergleichsfeld vergleichen aktiviert haben.</p> |

Ausgabe

Exception Monitor gibt Datensätze an zwei Ports zurück. Ein Port enthält Datensätze, die keine der im „Exception Monitor“-Schritt definierten Bedingungen erfüllen. Der andere Port, der Ausnahmeport, enthält alle Datensätze, die mindestens eine Ausnahmebedingung erfüllen. Der Ausnahmeport kann auch Nicht-Ausnahmedatensätze enthalten, wenn Sie die Option **Alle Datensätze in der Gruppe der Ausnahme zurückgeben** aktivieren. Exception Monitor fügt innerhalb eines Datensatzes keine Felder hinzu oder ändert diese.

Read Exceptions

„Read Exceptions“ ist ein Schritt, in dem Datensätze aus der Ausnahme-Datenbank als Eingabe in einen Datenfluss eingelesen werden. (Weitere Informationen zur Ausnahme-Datenbank finden Sie unter [Business Steward-Modul](#) auf Seite 254.)

Anmerkung: Sobald ein Datensatz im „Read Exceptions“-Schritt in einen Datenfluss eingelesen wurde, wird er aus der Datenbank gelöscht.

Eingabe

Im „Read Exceptions“-Schritt werden Daten aus einer Ausnahme-Datenbank eingelesen. Dabei wird keine Eingabe aus einem anderen Schritt im Datenfluss einbezogen.

Anmerkung: Es werden nur die Datensätze in den Datenfluss eingelesen, die im Business Steward Portal als „Genehmigt“ markiert sind.

Optionen

Der „Read Exceptions“-Schritt verfügt über die folgenden Optionen.

Registerkarte „Allgemein“

Die Optionen auf der Registerkarte **Allgemein** geben an, welche Ausnahmedatensätze Sie in den Datenfluss einlesen möchten.

Mit den **Filteroptionen** können Sie mithilfe der folgenden Kriterien eine Teilmenge der Datensätze aus der Ausnahme-Datenbank auswählen:

- **Benutzer:** Der Benutzer, der den Datenfluss ausgeführt hat, über den die Ausnahmen generiert wurden, die Sie in den Datenfluss einlesen möchten.
- **Datenflussname:** Der Name des Datenflusses, über den die Ausnahmen generiert wurden, die Sie in den Datenfluss einlesen möchten.
- **Schrittbeschriftung:** Die Beschriftung des „Exception Monitor“-Schrittes, wie sie im Datenfluss im Enterprise Designer angezeigt wird. Diese Kriterien sind hilfreich, wenn der Datenfluss, über den die Ausnahmen generiert wurden, mehrere „Exception Monitor“-Schritte enthält und Sie nur die Ausnahmen aus einen dieser „Exception Monitor“-Schritte einlesen möchten.
- **Anfangsdatum:** Das Datum und die Uhrzeit der ältesten Datensätze, die Sie in den Datenfluss einlesen möchten. Als Datum eines Ausnahmedatensatzes gilt das Datum, an dem dieser zuletzt geändert wurde.
- **Enddatum:** Das Datum und die Uhrzeit der neuesten Datensätze, die Sie in den Datenfluss einlesen möchten. Als Datum eines Ausnahmedatensatzes gilt das Datum, an dem dieser zuletzt geändert wurde.

In der Auflistung der **Felder** werden die Felder angezeigt, die in den Datenfluss eingelesen werden. Standardmäßig werden alle Felder eingeschlossen. Sie können jedoch Felder ausschließen, indem Sie das Kästchen in der Spalte **Einschließen** deaktivieren.

In der Auflistung der **Vorschau** werden die Datensätze angezeigt, die die von Ihnen unter **Filter** angegebenen Kriterien erfüllen.

Anmerkung: In der Vorschau werden nur Datensätze angezeigt, die im Business Steward Portal als „Genehmigt“ markiert wurden und die Filterkriterien erfüllen.

Registerkarte „Sortieren“

Über die Registerkarte „Sortieren“ können Sie die Eingabedatensätze basierend auf Feldwerten sortieren.

- **Hinzufügen:** Fügt ein Feld hinzu, nach dem sortiert werden soll.
- Spalte **Feldname:** Zeigt den Namen des Feldes an, nach dem sortiert werden soll. Sie können ein Feld auswählen, indem Sie auf die Dropdown-Schaltfläche klicken.

- Spalte **Sortierung**: Gibt an, ob die Sortierreihenfolge absteigend oder aufsteigend sein soll.
- **Nach oben** und **Nach unten**: Ändert die Sortierreihenfolge. Datensätze werden zuerst nach dem ersten Feld in der Liste, dann nach dem zweiten Feld usw. sortiert.
- **Entfernen**: Entfernt ein Sortierfeld.

Registerkarte „Laufzeit“

- **Startdatensatz**: Geben Sie die Position des ersten Datensatzes in der Datenbank an, den Sie in den Datenfluss einlesen möchten. Wenn Sie beispielsweise die ersten 99 Datensätze in der Datenbank überspringen möchten, würden Sie 100 angeben. Der 100. Datensatz würde zuerst in die Datenbank eingelesen werden, wenn er die auf der Registerkarte **Allgemein** angegebenen Kriterien erfüllt. Die Position eines Datensatzes wird durch die Reihenfolge der Datensätze im Business Steward Portal bestimmt.
- **Alle Datensätze**: Wählen Sie diese Option aus, wenn Sie alle Datensätze einlesen möchten, die die auf der Registerkarte **Allgemein** angegebenen Suchkriterien erfüllen.
- **Max. Anzahl Datensätze**: Wählen Sie diese Option aus, wenn Sie die Anzahl der Datensätze beschränken möchten, die in den Datenfluss eingelesen werden. Wenn Sie beispielsweise nur die ersten 1.000 Datensätze einlesen möchten, die die Auswahlkriterien erfüllen, wählen Sie diese Option aus und geben Sie 1000 an.

Ausgabe

Im „Read Exceptions“-Schritt werden Datensätze aus der Ausnahme-Datenbank zurückgegeben, die genehmigt wurden und den in den „Read Exceptions“-Optionen angegebenen Auswahlkriterien entsprechen. Zusätzlich zu den Feldern der Datensätze werden im „Read Exceptions“-Schritt die folgenden Felder zurückgegeben, in denen die letzten, am Datensatz im Business Steward Portal vorgenommenen Änderungen beschrieben werden.

Tabelle 20: Read Exceptions-Ausgabe

| Feldname | Beschreibung |
|--------------------------|---|
| Exception.Comment | Kommentare, die von der Person eingegeben wurden, die die Ausnahme behoben hat. In Kommentaren könnten beispielsweise die Änderungen beschrieben werden, die der Business Steward am Datensatz vorgenommen hat. |
| Exception.LastModifiedBy | Der Benutzer, der den Datensatz zuletzt im Business Steward Portal geändert hat |

| Feldname | Beschreibung |
|------------------------------------|--|
| Exception.LastModifiedMilliseconds | Die Zeit, zu der der Datensatz im Business Steward Portal zuletzt geändert wurde. Die Zeit wird seit dem 01. Januar 1970 00:00 GMT in Millisekunden angegeben. Dieses Verfahren wird in der Java-Programmiersprache standardmäßig für die Berechnung der Zeit verwendet. Sie können mit diesem Wert Datumsangaben vergleichen oder eine Transformation erstellen, um diesen Wert in das von Ihnen gewünschte Datumsformat umzuwandeln. |
| Exception.LastModifiedString | Die Zeit, zu der der Datensatz im Business Steward Portal zuletzt geändert wurde. Das folgende Feld bietet eine verständlichere Darstellung des Datums als das Feld „Ausnahme.Letzte Änderung in Millisekunden“. Die Zeit wird im folgenden Format angegeben: Thu Feb 17 13:34:32 CST 2011 |

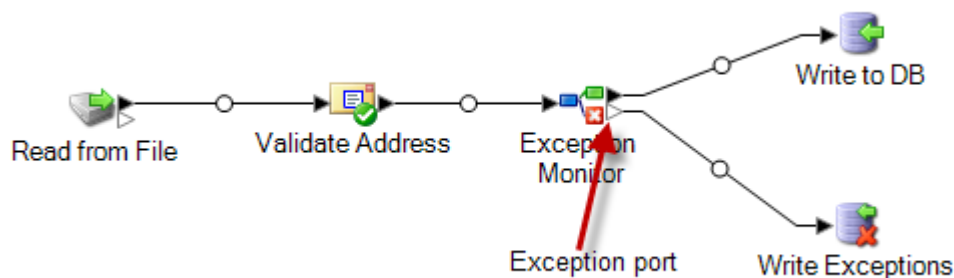
Write Exceptions

Der Schritt „Write Exceptions“ schreibt Datensätze, die durch den „Exception Monitor“-Schritt als Ausnahmen erkannt wurden, in das Ausnahme-Repository. Sobald sie sich im Ausnahme-Repository befinden, können die Datensätze mithilfe des Business Steward Portal erneut angezeigt und bearbeitet werden.

Anmerkung: Ausnahmedatensätze werden nicht zum Business Steward Portal geschrieben, wenn die Aufträge oder Dienste im Enterprise Designer im Überprüfungsmodus oder in der Management Console im Vorschaumodus ausgeführt werden.

Eingabe

Im „Write Exceptions“-Schritt werden Datensätze aus dem Ausnahmeport im „Exception Monitor“-Schritt entnommen, die anschließend in die Ausnahme-Datenbank geschrieben werden. Der „Write Exceptions“-Schritt sollte hinter dem Ausnahmeport des „Exception Monitor“-Schrittes platziert werden. Der Ausnahmeport ist der untere Ausgabeport im „Exception Monitor“-Schritt:



Optionen

Mithilfe des Schritts „Write Exceptions“ können Sie auswählen, von welchem Feld die Daten in das Ausnahme-Repository zurückgegeben werden sollen. Welche Felder angezeigt werden, hängt von den Schritten ab, die nachgeschaltet im Datenfluss auftreten. Wenn in Ihrem Datenfluss beispielsweise ein „Validate Address“-Schritt vorhanden ist, würden Felder wie AddressLine1, AddressLine2, City, PostalCode usw. im „Write Exceptions“-Schritt angezeigt werden. Standardmäßig sind alle diese Felder aktiviert. Deaktivieren Sie die Kästchen für alle Felder, die nicht in das Ausnahme-Repository zurückgegeben werden sollen. Die Reihenfolge der Felder wird durch ihre Reihenfolge beim Übergang in den „Write Exceptions“-Schritt bestimmt. Sie können die Felder neu anordnen, indem Sie eine Zeile auswählen und diese mithilfe der Pfeile auf der rechten Seite des Bildschirms nach oben oder unten verschieben. Die hier von Ihnen ausgewählte Reihenfolge bleibt für alle Benutzer im Business Steward Portal erhalten. Allerdings kann jeder Benutzer die Felder innerhalb des Portals nach Bedarf neu anordnen.

Wählen Sie die Option **Benutzern das Erstellen von „Best of Breed“-Datensätzen im Portal erlauben** aus, um eine manuelle Datensatzkonsolidierung durchzuführen, wenn keine Datenvergleiche mithilfe der Standardregeln möglich sind. Diese Funktion kopiert einen ausgewählten Datensatz innerhalb einer Gruppe und verwendet ihn anstelle der Dubletten für die Verarbeitung. Diese Option ist nur für gruppierte Ausnahmedatensätze verfügbar, die aus einem Vergleichsauftrag erstellt wurden, der durch das Vorhandensein des CandidateGroup-Feldes oder des CollectionNumber-Feldes identifiziert wird. Bei Verwendung dieser Option wird dem Ausnahmedatensatz ein schreibgeschütztes Feld mit dem Namen „CollectionRecordType“ hinzugefügt. Sie können dieses Feld unten in der Liste sehen. Beachten Sie, dass alle Optionen für dieses Feld deaktiviert sind.

Anmerkung: Wenn Ihr Datenfluss ebenfalls für eine **erneute Überprüfung** konfiguriert wird, müssen Sie das CollectionRecordType-Feld manuell in den „Exception Monitor“-Schritt/Unterfluss einfügen und verfügbar machen. Das Gleiche gilt für den entsprechenden Dienst.

Nach dem Hinzufügen eines **Best of Breed**-Datensatzes in der Ansicht „Dubletten auflösen“ des Ausnahmen-Editors im Business Steward Portal wird dieses Feld auf „BestOfBreed“ eingestellt. Wenn Sie „Best of Breed“-Datensätze erstellen möchten, verwenden Sie für diese Datensätze im Business Steward Portal die Option **Allen zustimmen**. Weitere Informationen über „Best of Breed“-Datensätze im Business Steward Portal finden Sie [hier](#).

Es gibt gegebenenfalls Eingabefelder, die in das Repository aufgenommen, jedoch nicht im Business Steward Portal angezeigt werden sollen. Der Grund dafür könnte sein, dass die Felder sensible Daten enthalten oder dass Sie die Anzeige des Portals optimieren möchten. Aktivieren Sie das Kästchen **Anzeigen zulassen**, um festzulegen, welches der ausgewählten Felder angezeigt werden soll, sobald es in das Ausnahme-Repository aufgenommen wurde. Standardmäßig können alle Felder angezeigt werden. Deaktivieren Sie das Kästchen für jedes Feld, das nicht im Portal angezeigt werden soll.

Sie können außerdem festlegen, welches der ausgewählten Felder im Portal bearbeitbar sein soll, sobald es in das Ausnahme-Repository aufgenommen wurde. Standardmäßig ist die Spalte **Bearbeiten zulassen** für alle Felder aktiviert, die in den „Write Exceptions“-Schritt übergehen. Deaktivieren Sie das Kästchen für alle Felder, die schreibgeschützt in das Ausnahme-Repository zurückgegeben werden sollen.

Sie können schließlich mithilfe des Feature **Suche** eine Suche nach einem Feld mit problematischen Daten durchführen. Sie können aus der Liste der Suchvorgänge auswählen, die im Business Steward-Konfigurationstool definiert wurden, oder den Namen des Suchvorgangs manuell eingeben. Weitere Informationen zu Suchvorgängen finden Sie unter [Suchvorgänge](#) auf Seite 295.

Anmerkung: Suchvorgänge können nur Feldern zugewiesen werden, die vom string-Typ sind.

Ausgabe

„Write Exceptions“ gibt keine Ausgaben im Datenfluss zurück. Stattdessen werden Ausnahmedatensätze in die Ausnahmen-Datenbank geschrieben.

Business Steward Portal

Einführung in das Business Steward Portal

Was ist das Business Steward Portal?

Das Business Steward Portal ist ein Tool zum Überprüfen, Ändern und Genehmigen von Datensätzen, deren automatische Verarbeitung fehlgeschlagen ist bzw. die nicht mit ausreichendem Zuverlässigkeitsniveau verarbeitet wurden. Verwenden Sie das Business Steward Portal, um richtige oder zusätzliche Daten in einen Datensatz einzugeben. Wenn beispielsweise bei einem Kundendatensatz der Adressenüberprüfungsprozess fehlschlägt, könnten Sie mithilfe der Suchtools Recherche betreiben und die Adresse des Kunden bestimmen, um den Datensatz anschließend so zu ändern, dass die richtige Adresse darin enthalten ist. Abhängig von Ihrer Konfiguration könnte der geänderte Datensatz dann von Spectrum™ Technology Platform genehmigt und erneut verarbeitet werden, an einen anderen Datenüberprüfungs- oder Datenanreicherungsprozess gesendet werden oder in eine Datenbank geschrieben werden. Sie könnten über das Portal auch Informationen hinzufügen, die im ursprünglichen Datensatz nicht enthalten waren.

Das Business Steward Portal enthält zudem zusammenfassende Diagramme, die Einblicke in die Art von Daten geben, die eine Verarbeitung von Ausnahmen auslösen. Darin inbegriffen sind die Datendomäne (Name, Adressen, Geodaten usw.) und die Datenqualitätsmetrik (Vollständigkeit, Genauigkeit, Konsistenz usw.), bei denen die Daten Fehler aufweisen.

Zusätzlich können Sie auf der Seite „Ausnahmen verwalten“ im Business Steward Portal die Ausnahmedatensatzaktivität überprüfen und verwalten. Darin inbegriffen ist die Neuzuweisung von Datensätzen eines Benutzers zu einem anderen Benutzer. Schließlich enthält die Seite „Datenqualität“ im Business Steward Portal Informationen zu Trends bei Datenflüssen und Schritten.

Weitere Informationen zur Verarbeitung von Ausnahmen finden Sie unter [Business Steward-Modul](#) auf Seite 254.

Zugriff auf das Business Steward Portal

Klicken Sie zum Öffnen des Business Steward Portal auf **Start > Alle Programme > Pitney Bowes > Spectrum Technology Platform > Server > Begrüßungsseite**, und wählen Sie **Spectrum Data Quality** und anschließend **Business Steward Portal** aus. Klicken Sie daraufhin auf **Business Steward Portal öffnen**.

Alternativ können Sie wie folgt vorgehen:

1. Öffnen Sie einen Webbrowser, und laden Sie die Seite `http://<servername>:8080/bsmportal`.

Beispiel:

`http://myserver:8080/bsmportal`

Kontaktieren Sie Ihren Spectrum™ Technology Platform-Administrator, wenn Sie den Servernamen und/oder den Port nicht kennen.

2. Melden Sie sich bei Spectrum™ Technology Platform an. Kontaktieren Sie Ihren Spectrum™ Technology Platform-Administrator, wenn Sie Probleme bei der Anmeldung haben.

Das Menü „Business Steward Portal“

Das Menü „Business Steward Portal“ besteht aus vier Optionen und ermöglicht den Zugriff auf das Hilfesystem, wie unten dargestellt:

- **Dashboard:** Anzeige des Status der Ihnen zugewiesenen Ausnahmen. Wenn Sie Anzeigeberechtigungen besitzen, können Sie auch den Status der Ausnahmen anzeigen, die anderen Benutzern zugewiesen sind.
- **Editor:** Prüfen, Bearbeiten und Genehmigen von Ausnahmedatensätzen für die erneute Verarbeitung.
- **Verwalten:** Wenn Sie Anzeigeberechtigungen besitzen, können Sie sich selbst und anderen Ausnahmen zuweisen. Wenn Sie Löschberechtigungen besitzen, können Sie Ausnahmedatensätze aus der Datenbank löschen.
- **Datenqualität:** Wenn Sie Anzeigeberechtigungen besitzen, können Sie auf diese Seite zugreifen, um statistische Informationen anzuzeigen und wichtige Leistungskennzahlen für

Ausnahmedatensätze zu konfigurieren. Wenn Sie keine Anzeigeberechtigungen besitzen, haben Sie keinen Zugriff auf diese Seite.

- Dropdown-Liste „Benutzer“: Zugriff auf das Hilfesystem von Business Steward Portal, die gesamte Dokumentation zu Spectrum Technology Platform und die Seite „Profil“, auf der Sie eine Sprachpräferenz einstellen und ein Land angeben können, um die Kulturpräferenz anzugeben. Außerdem können Sie eine E-Mail-Adresse für Benachrichtigungen bereitstellen und Ihr Kennwort ändern.

Die Seite „Dashboard“

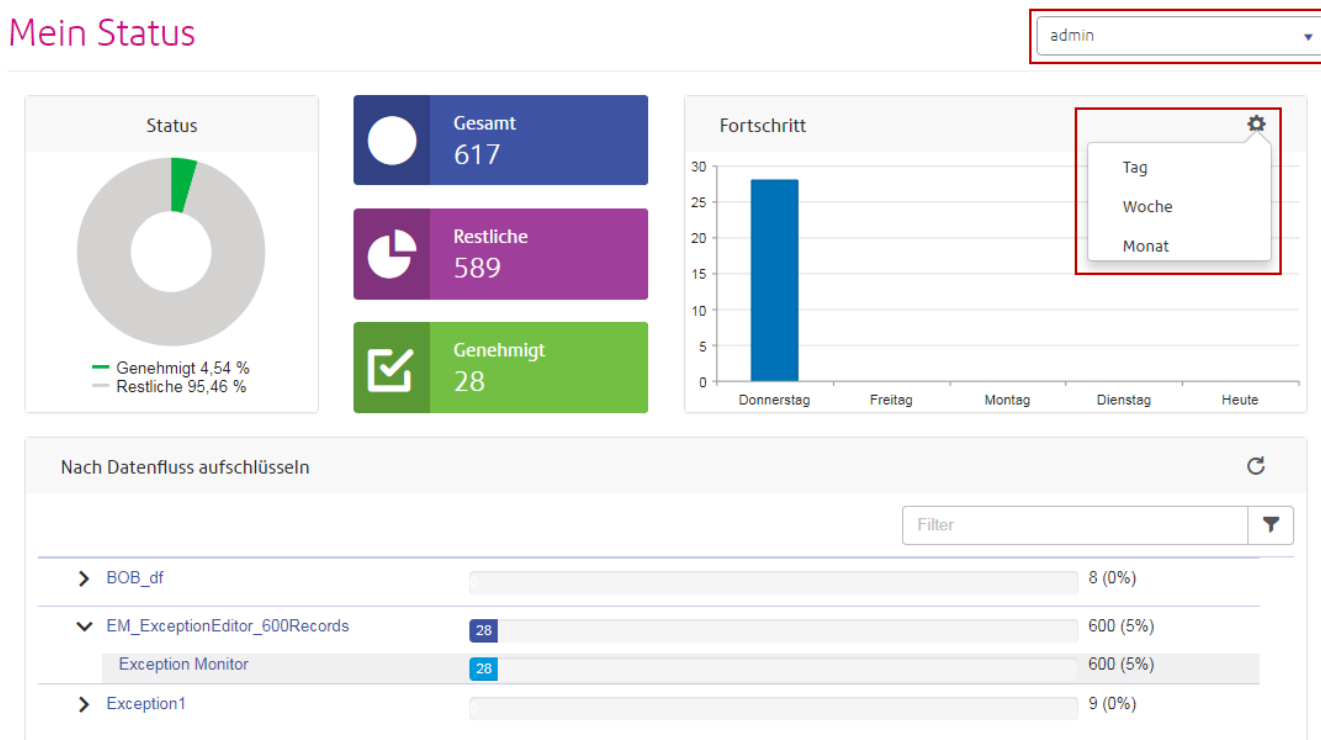
Im Exceptions-Dashboard werden Daten angezeigt, die den Status der Ausnahmedatensätze, die Ihnen und anderen Benutzern gehören, zusammenfassen. (Beachten Sie, dass Sie nur Daten anderer Benutzer anzeigen können, wenn Sie über Berechtigungen zum Ändern verfügen.) Dies beinhaltet Folgendes:

- Die Anzahl und den Prozentsatz von genehmigten Ausnahmen gegenüber nicht genehmigten Ausnahmen
- Den Fortschritt des ausgewählten Benutzers pro Tag, Woche oder Monat
- Den Status der Genehmigung von Ausnahmedatensätzen nach Datenfluss

Anmerkung: Die Fortschrittsdiagramme für Benutzer und Datenfluss werden nicht angezeigt, wenn Sie die Option „Fortschritt nachverfolgen“ auf der Seite „Business Steward-Einstellungen“ der Management Console ausgeschaltet haben.

1. Wählen Sie auf der Registerkarte „Dashboard“ im ersten Dropdown-Feld, das nachstehend hervorgehoben wird, den Benutzer aus, dessen Ausnahmeaktivität Sie anzeigen möchten. Beachten Sie, dass nur Benutzer in dieser Liste angezeigt werden, denen aktuell Ausnahmen zugewiesen sind.

Mein Status



- Klicken Sie auf das Zahnrad **Einstellungen** (das im zweiten Dropdown-Feld angezeigt und vorstehend hervorgehoben wird), um einen Zeitraum in Tagen, Wochen oder Monaten auszuwählen.
- Geben Sie in das Feld **Filter** Suchkriterien ein, um die Liste der Datenflüsse oder die Teilmenge der Datenflüsse einzuschränken. Der Suchbegriff muss im Datenflussnamen enthalten sein.

Die Seite „Editor“

Mit dem Exception Editor können Sie Ausnahmedatensätze manuell überprüfen, ändern und genehmigen. Die manuelle Überprüfung dient zum Ermitteln falscher oder fehlender Daten und zum Aktualisieren und Genehmigen dieser Daten, insbesondere dann, wenn Spectrum™ Technology Platform sie im Rahmen eines automatisierten Datenflussprozesses nicht korrigieren konnte. Anschließend können Sie Ausnahmedatensätze für eine Genehmigung oder erneute Verarbeitung erneut überprüfen.

Mit dem Exception Editor können Sie zudem Ausnahmedatensatzdubletten auflösen und über Suchtools nach Informationen zum Bearbeiten, Genehmigen und erneuten Ausführen von Datensätzen suchen.

Anpassen der Inhalte im Ausnahmen-Editor

Es gibt mehrere Wege, um die Anzeige im Ausnahmen-Editor anzupassen. Sie können die **Auswahloptionen** verwenden, um Datensätze für einen Benutzer, einen Datenfluss, eine Auftrags-ID usw. zurückzugeben. Sie können das Tool **Feldfilter** verwenden, um Datensätze anzuzeigen, die

bestimmte Kriterien erfüllen. Dabei werden die Datensätze ausgeblendet, die nicht den Kriterien entsprechen.

Wenn Sie das Tool **Elemente pro Seite** verwenden und anschließend Auswahloptionen oder einen Feldfilter anwenden, um die Liste der angezeigten Ausnahmedatensätze einzuschränken, denken Sie daran, dass mehrere Ergebnisseiten zurückgegeben werden können und nicht nur der Anfangsbildschirm alle Ergebnisse enthält. Legen Sie beispielsweise für Elemente pro Seite 10 fest und wenden Sie dann einen Feldfilter an, um nur die Datensätze anzuzeigen die eine bestimmte Postleitzahl enthalten. Es könnte am Anfang so aussehen, als würden nur 10 Datensätze zurückgegeben werden, aber es könnte mehrere Ergebnisseiten geben, da Sie die Grenze der anzuzeigenden Datensätze auf 10 beschränkt haben.

Verwenden von Auswahloptionen

Wählen Sie im Auswahloptionsbereich Kriterien in den Dropdown-Feldern aus, um die im Ausnahmeraster angezeigten Datensätze einzuschränken.

| | |
|----------------------------|--|
| Benutzer | Erforderlich. Die ID des Benutzers, der dem Datenfluss zugewiesen wurde. Diese Informationen sind nur sichtbar, wenn Sie über Berechtigungen zum Ändern verfügen. |
| Datenflussname | Erforderlich. Der Name des Datenflusses, in dem die Ausnahmedatensätze generiert wurden. |
| Schrittbeschriftung | Erforderlich. Der benutzerdefinierte Name, der dem „Exception Monitor“-Schritt im Datenfluss gegeben wurde. Diese Information ist besonders in den Fällen hilfreich, in denen ein Datenfluss mehrere „Exception Monitor“-Schritte enthält. Wenn die Person, die den Datenfluss erstellt hat, jedem „Exception Monitor“-Schritt eine aussagekräftige Bezeichnung gegeben hat, können Sie erkennen, welcher Exception Monitor den Ausnahmedatensatz generiert hat. Die Standardbezeichnung lautet „Exception Monitor“. |
| Auftrags-ID | Erforderlich. Ein numerischer Identifikator, der vom System einem Auftrag zugewiesen wird. Bei jeder Ausführung eines Auftrags wird ihm eine neue Auftrags-ID zugewiesen. |
| Status | Optional: Der Genehmigungsstatus des Datensatzes. |
| Datum | Optional: Das Datum (und optional die Uhrzeit), zu dem der Datenfluss ausgeführt wurde. Wenn Sie die Uhrzeit eingeben, geben Sie diese nach dem Datum ein. |
| Datendomäne | Optional: Die Art von Daten, die zu einer Ausnahme geführt haben. Beispiele für Datendomänen umfassen den Namen, die Adresse und die Telefonnummer. Anhand dieser Informationen können Sie erkennen, welche Felder im Datensatz bearbeitet werden müssen. |
| Qualitätsmetriken | Optional: Die Qualitätsmessung, dass der Datensatz fehlerhaft ist. Beispiele für die Qualitätsmetrik sind Genauigkeit, Vollständigkeit und Eindeutigkeit. Mithilfe dieser Informationen können Sie ermitteln, weshalb der Datensatz als Ausnahme identifiziert wurde. |

Verwenden des Feldfilters

Nachdem Sie Auswahloptionen festgelegt haben und die Ausnahmedatensätze geladen wurden, können Sie Felder filtern, um nur die Datensätze anzuzeigen, die für Sie von Interesse sind. Das

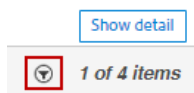
Business Steward Portal zeigt standardmäßig nur Datensätze aus je einem Spectrum™ Technology Platform-Datenfluss an. Sie können die Datensatzliste filtern, um nur die Datensätze anzuzeigen, die innerhalb eines bestimmten Feldes bestimmte Kriterien erfüllen. Sobald ein Filter erstellt wurde, wird dieser automatisch gespeichert. Wenn Sie den Datenfluss das nächste Mal im Ausnahmen-Editor öffnen, wird der Filter angewendet.

Anmerkung: Bevor Sie das Tool zum Filtern von Feldern verwenden können, müssen Sie Auswahloptionen anwenden.

Sie können Filter für mehrere Felder erstellen, sie sollten jedoch für jedes Feld nur einen Filter erstellen. Wenn auf ein Feld bereits ein Filter angewendet wird, ist der Hintergrund des Pfeils blau:

| City | FirstName |
|-------|-----------|
| KEENE | LAKSHMI |
| KEENE | LAKSHMI |

Filter können nur in der Tabellenansicht erstellt werden. Filter, die in der Tabellenansicht definiert wurden, werden jedoch auch in der Formularansicht wiederspiegelt. Wie bei der Tabellenansicht ist unten im Formular ein Indikator vorhanden, der angibt, dass ein Datensatz gefiltert wurde:



Gehen Sie wie folgt vor, um die Liste der Datensätze zu filtern:

1. Klicken Sie auf die Schaltfläche **Filter**. Neben den einzelnen Spaltenüberschriften werden Ihnen Filtersymbole angezeigt.
2. Klicken Sie bei dem Feld, dessen Daten Sie filtern möchten, auf das Filtersymbol.
3. Wählen Sie einen Operator aus, der dem Datentyp des Feldes entspricht, gefolgt von einem Wert.

ist gleich Sucht nach Datensätzen mit genau dem von Ihnen angegebenen Wert. Dies kann ein numerischer Wert oder ein Textwert sein. Sie können beispielsweise nach Datensätzen mit genau 82 als Wert für „MatchScore“ oder nach Datensätzen mit „Smith“ als Wert für „Nachname“ suchen.

ist ungleich Sucht nach Datensätzen mit einem anderen Wert als dem von Ihnen angegebenen Wert. Dies kann ein numerischer Wert oder ein Textwert sein. Sie können beispielsweise nach Datensätzen mit einem beliebigen Wert für „MatchScore“ mit Ausnahme von 100 oder nach Datensätzen mit einem beliebigen Wert für „LastName“ mit Ausnahme von „Smith“ suchen.

beginnt mit Sucht nach Datensätzen, die im ausgewählten Feld mit einem bestimmten Wert beginnen. Wenn Sie beispielsweise im Feld „LastName“ nach „Van“ filtern, würden Ihnen Datensätze mit „Van Buren“, „Vandenburg“ oder „Van Dyck“ angezeigt werden.

enthält Sucht nach Datensätzen mit dem von Ihnen angegebenen Wert an einer beliebigen Stelle innerhalb des ausgewählten Feldes. Wenn Sie

beispielsweise im Feld „AddressLine1“ nach „South“ filtern, würden Ihnen Datensätze mit „12 South Ave.“, „9889 Southport St.“, „600 South Shore Dr.“ und „4089 5th St. South“ angezeigt werden.

enthält nicht Sucht nach Datensätzen ohne den von Ihnen angegebenen Wert an einer beliebigen Stelle innerhalb des ausgewählten Feldes. Wenn Sie beispielsweise im Feld „AddressLine1“ nach „South“ filtern, würden Ihnen keine Datensätze mit „12 South Ave.“, „9889 Southport St.“, „600 South Shore Dr.“ und „4089 5th St.“ angezeigt werden. „South“ angezeigt werden.

endet mit Sucht nach Datensätzen, die im ausgewählten Feld mit einem bestimmten Wert enden. Wenn Sie beispielsweise nach Datensätzen filtern, die im Feld „City“ mit „burg“ enden, würden Ihnen Datensätze mit „Gettysburg“, „Fredricksburg“ und „Blacksburg“ angezeigt werden.

ist größer als Sucht nach Datensätzen, die einen numerischen Wert enthalten, der größer als der von Ihnen angegebene Wert ist.

ist größer oder gleich Sucht nach Datensätzen, die einen numerischen Wert enthalten, der größer oder gleich dem von Ihnen angegebenen Wert ist. Wenn Sie beispielsweise 50 angeben, würden Ihnen Datensätze mit einem Wert von 50 oder mehr im ausgewählten Feld angezeigt werden.

ist kleiner als Sucht nach Datensätzen, die einen numerischen Wert enthalten, der kleiner als der von Ihnen angegebene Wert ist.

ist kleiner oder gleich Sucht nach Datensätzen, die einen numerischen Wert enthalten, der kleiner oder gleich dem von Ihnen angegebenen Wert ist. Wenn Sie beispielsweise 50 angeben, würden Ihnen Datensätze mit einem Wert von 50 oder weniger im ausgewählten Feld angezeigt werden.

ist später Sucht nach Datensätzen, die einen Datums- oder Zeitwert enthalten, der später als der von Ihnen angegebene Wert ist.

ist später oder gleich Sucht nach Datensätzen, die einen Datums- oder Zeitwert enthalten, der später oder gleich dem von Ihnen angegebenen Wert ist.

ist früher Sucht nach Datensätzen, die einen Datums- oder Zeitwert enthalten, der früher als der von Ihnen angegebene Wert ist.

ist früher oder gleich Sucht nach Datensätzen, die einen Datums- oder Zeitwert enthalten, der früher oder gleich dem von Ihnen angegebenen Wert ist.

4. Klicken Sie auf die Schaltfläche **Filter**, um die Kriterien anzuwenden. Es werden nur die Datensätze angezeigt, deren Daten den Kriterien entsprechen.
5. Klicken Sie erneut auf das Filtersymbol und klicken Sie dann auf **Löschen**, um den Filter zu entfernen. Alternativ können Sie auf die Schaltfläche **Filter** klicken, um alle Filter zu entfernen. Diese Aktion kann in der Tabellenansicht oder in der Formularansicht ausgeführt werden.

Anzeigen von Datensätzen

Sie können Ausnahmedatensätze in zwei verschiedenen Formaten anzeigen. Die Standardansicht ist die Tabellenansicht, in der Sie bis zu 100 Ausnahmedatensätze pro Seite laden können. Sie können durch die Liste scrollen und die Datensätze in einer beliebigen Reihenfolge bearbeiten. Wenn Sie mehrere Datensätze bearbeiten, können Sie alle Änderungen gleichzeitig speichern. Sie müssen die Änderungen nicht für jeden einzelnen Datensatz speichern. Wenn Sie diese Ansicht verwenden, können Sie im Dropdown-Menü im unteren Bereich des Bildschirms angeben, wie viele Datensätze pro Seite angezeigt werden sollen.

Die andere Methode zum Anzeigen von Ausnahmedatensätzen stellt die Formularansicht dar, in der jeweils ein Datensatz angezeigt und bearbeitet wird. In dieser Ansicht können Sie nicht mehrere Datensätze gleichzeitig bearbeiten. Ebenso müssen Sie die Änderungen für jeden einzelnen Datensatz speichern. Die Änderungen können nicht für mehrere Datensätze gleichzeitig gespeichert werden.

Anzeigen von Datensatzdetails

Unabhängig von der von Ihnen verwendeten Ansicht werden im Ausnahmeraster alle Felder eines Datensatzes sowie der Genehmigungsstatus, der Ausnahmetyp und zum Datensatz hinzugefügte Kommentare angezeigt. In der Tabellenansicht können Sie weitere Details zu einem Datensatz anzeigen, indem Sie auf den Pfeil links neben einem Datensatz klicken. Klicken Sie in der Formularansicht auf **Details anzeigen**. Durch diese Aktionen wird die Registerkarte **Details** angezeigt, auf der folgende Informationen angezeigt werden.

| | |
|------------------------|---|
| Auftrags-ID | Ein numerischer Identifikator, der vom System einem Auftrag zugewiesen wird. Bei jeder Ausführung eines Auftrags wird ihm eine neue Auftrags-ID zugewiesen. |
| Gruppieren nach | Wenn der Datenfluss so konfiguriert wurde, dass er alle Datensätze in der Gruppe der Ausnahmedatensätze zurückgeben soll, wird hier das Feld angezeigt, nach dem die Datensätze gruppiert werden. Dies gilt nur für Datensätze, die Vergleiche durchführen, z. B. Datenflüsse, die Datensatzdubletten erkennen, oder Datenflüsse, die Datensätze in Haushalte gruppieren. |
| Ausnahmezeit | Das Datum und die Uhrzeit, an dem bzw. zu der der Exception Monitor den Datensatz als Ausnahme identifiziert hat. |
| Datensatztyp | Der von Ihnen ausgewählte Datensatztyp. Zur Auswahl stehen: <ul style="list-style-type: none"> • E—Exception (Ausnahme) • GE—Group Exception (Ausnahme gruppieren) |
| Status | Der Status des von Ihnen ausgewählten Datensatzes. Zur Auswahl stehen: <ul style="list-style-type: none"> • Neu • Behoben (nur möglich, wenn der Datensatz seit seiner Erstellung mindestens einmal bearbeitet wurde) |

Klicken Sie auf die Registerkarte **Bedingungen**, um folgende Informationen anzuzeigen:

| | |
|------------------|---|
| Bedingung | Der Name der Bedingung, die den Datensatz als Ausnahme identifiziert hat. Bedingungsnamen werden von der Person definiert, die den Datenfluss eingerichtet hat. |
| Domäne | Die Art von Daten, die zu einer Ausnahme geführt haben. Beispiele für Datendomänen umfassen den Namen, die Adresse und die Telefonnummer. Anhand dieser Informationen können Sie erkennen, welche Felder im Datensatz bearbeitet werden müssen. |
| Metrik | Die Qualitätsmessung, dass der Datensatz fehlerhaft ist. Beispiele für die Qualitätsmetrik sind Genauigkeit, Vollständigkeit und Eindeutigkeit. Mithilfe dieser Informationen können Sie ermitteln, weshalb der Datensatz als Ausnahme identifiziert wurde. |

Bei jeder Datensatzänderung bewahrt das System bestimmte Informationen auf. Diese umfassen die Angabe, wer den Datensatz geändert hat und wann er geändert wurde sowie den Namen des Benutzers, dem der Datensatz zugewiesen wurde, und alle Daten, die bei diesem Datensatz in das Kommentarfeld eingegeben wurden. Der Datensatz, der im Ausnahmeraster angezeigt wird, spiegelt die letzten Änderungen und den letzten Kommentar (sofern vorhanden) wider; auf der Registerkarte **Verlauf** werden jedoch folgende Informationen zu der gesamten Bestandszeit des Datensatzes angezeigt – von dem Zeitpunkt, zu dem der Datensatz erstmalig als Ausnahme zur Datenbank hinzugefügt wurde, bis hin zu dem Zeitpunkt, zu dem Sie den Datensatz anzeigen:

| | |
|-----------------------------|---|
| Version | Die Revisionsnummer der Änderung. |
| Zuletzt geändert von | Der Benutzer, der die Änderung vorgenommen hat. |
| Zugewiesen zu | Der Benutzer, dem der Ausnahmedatensatz zum Zeitpunkt der Revision zugewiesen wurde. |
| Wann | Das Datum und die Uhrzeit, an dem bzw. zu der die Änderung gespeichert wurde. |
| Kommentare | Die Kommentare (falls vorhanden), die von der Person eingegeben wurden, die die Änderung vorgenommen hat. |

Sortierfelder

Wenn Sie die Tabellenansicht verwenden, können Sie die Reihenfolge der angezeigten Datensätze basierend auf einem bestimmten Feld durch Klicken auf eine beliebige Stelle in der Spaltenüberschrift sortieren. Wenn Sie beispielsweise Datensätze in alphabetischer Reihenfolge nach dem Status sortieren möchten, klicken Sie einfach auf die Spaltenüberschrift „Status“. Durch den ersten Klick werden die Datensätze in absteigender Reihenfolge und durch einen zweiten Klick in aufsteigender Reihenfolge sortiert. Ein dritter Klick führt dazu, dass die Sortierreihenfolge gelöscht wird und die Datensätze wieder in der ursprünglichen Reihenfolge angezeigt werden.

Konfigurieren von Feldern

Sie können auswählen, welche Felder angezeigt werden, und die Reihenfolge ändern, in der sie angezeigt werden. Klicken Sie dazu auf die Schaltfläche **Ansicht konfigurieren** (das Zahnrad auf der rechten Seite des Bildschirms unter der Dropdown-Liste „Benutzer“), und nehmen Sie die entsprechenden Änderungen vor. Diese Änderungen werden auf dem Server mit dem Benutzernamen und dem Datenflussname gespeichert. Deshalb wird die Konfiguration auch angewendet, wenn Sie den Datenfluss zu einem späteren Zeitpunkt öffnen. Gleichmaßen werden hier vorgenommene

Änderungen auf die Anzeige angewendet, wenn Sie Ausnahmedatensätze mithilfe der Formularansicht bearbeiten. Verwenden Sie diese „Ansicht konfigurieren“-Features, um die im Ausnahmen-Editor angezeigten Felder anzupassen:

Suchen nach Feldern in „Ansicht konfigurieren“

Geben Sie einen vollständigen Feldnamen oder einen Teil davon in das Feld „Suchen“ ein, und die Liste der verfügbaren Felder wird dynamisch aktualisiert. Bei der Suche wird die Groß-/Kleinschreibung nicht beachtet.

Ausblenden von Feldern in der Ansicht

Wenn Sie nicht jedes Feld eines Ausnahmedatensatzes anzeigen möchten, klicken Sie auf **Ansicht konfigurieren** und wählen Sie die auszublendenden Felder ab. Die angezeigte Liste hat die gleiche Reihenfolge wie die Anzeige im Grid „Ausnahmen“.

Ändern der Feldreihenfolge

Sie können die Reihenfolge der Felder ändern, indem Sie die Felder in die gewünschte Reihenfolge ziehen. Bei der Anzeige der Suchergebnisse können Sie die Felder jedoch nicht erneut anordnen. Sie müssen das Suchfenster löschen und „Alle“ auswählen, um die Felder erneut anordnen zu können.


Anmerkung: Wenn Sie die Tabellenansicht verwenden, können Sie die Spaltenköpfe direkt aus dem Ausnahmen-Editor ziehen, um die Reihenfolge der Felder zu ändern; Sie müssen dies nicht unter „Ansicht konfigurieren“ konfigurieren tun.

Konfigurieren des Layouts der Formularansicht

Sie können das Standardlayout der Felder in der Formularansicht nicht nur durch Auswahl der anzuzeigenden Spalten ändern, sondern auch durch die Angabe der Spaltenbreite. Dabei gilt als einzige Regel, dass die Spaltenbreiten für alle sichtbaren Felder in der Summe 12 ergeben müssen. Wenn Sie beispielsweise vier Felder nebeneinander in einer Zeile platzieren möchten, könnten Sie jedem Feld eine Spaltenbreite von 3 zuweisen. Wenn Sie nur zwei Felder einschließen möchten, könnten Sie 6 als Spaltenbreite für jedes Feld angeben. Wenn Sie alternativ fünf Adressfelder in einer Zeile platzieren möchten, müssen Sie für jedes dieser Felder eine Zahl angeben, sodass die Zahlen zusammen 12 ergeben. Die erste untere Abbildung zeigt, wie Sie die fünf Felder konfigurieren können. Die zweite Abbildung zeigt ein Beispiel, wie diese Konfiguration gerendert wird.


Anmerkung: Wenn Sie das Standardlayout für ein Feld oder eine Feldkombination überschreiben, reagieren diese Felder nicht mehr auf die Größe des Fensters. Felder mit Standardwerten reagieren jedoch nach wie vor auf die Größe des Browser-Fensters.

Ansicht konfigurieren




Filter 





Alle Sichtbar
Ausgeblendet

| Feldname | Formularspalten |
|--|-----------------|
| <input type="checkbox"/> Kommentieren | Standard |
| <input checked="" type="checkbox"/> AddressLine1 | 4 |
| <input checked="" type="checkbox"/> City | 4 |
| <input type="checkbox"/> FirstName | Standard |
| <input type="checkbox"/> LastName | Standard |
| <input checked="" type="checkbox"/> State | 1 |
| <input checked="" type="checkbox"/> PostalCode | 2 |
| <input checked="" type="checkbox"/> Status | 1 |

Auf Standard zurücksetzen 


OK Abbrechen





  

Genehmigt

| AddressLine1 | City | State | PostalCode | Status |
|--------------------|--------------|-------|------------|--------|
| 555.5522D BIRCH ST | PRESQUE ISLE | ME | | |

Details anzeigen 

1 von 572 Elementen

Bearbeiten von Ausnahmedatensätzen

Der Zweck der Bearbeitung eines Ausnahmedatensatzes ist die Korrektur oder Vergrößerung und Genehmigung des Datensatzes, damit er erfolgreich verarbeitet werden kann. Die Bearbeitung eines Ausnahmedatensatzes kann die Verwendung anderer Spectrum Technology Platform-Dienste oder die Konsultation externer Ressourcen wie Karten, das Internet oder andere Informationssysteme Ihres Unternehmens beinhalten.

Nach der Überprüfung der Datensätze können Sie sie direkt im Ausnahmen-Editor bearbeiten und genehmigen. In der Tabellenansicht können Sie einen oder mehrere Datensätze gleichzeitig bearbeiten, während in der Formularansicht nur jeweils ein Datensatz bearbeitet werden kann.

Beachten Sie, dass schreibgeschützte Felder nicht bearbeitet werden können. Wenn Sie ein schreibgeschütztes Feld bearbeiten möchten, müssen Sie alle Ausnahmedatensätze für diesen Datenfluss und diese Auftrags-ID löschen und den Datenfluss erneut ausführen, nachdem Sie die Felder entsprechend im „Write Exceptions“-Schritt konfiguriert haben. Dadurch würden neue Ausnahmedatensätze mit bearbeitbaren Feldern erzeugt werden. Des Weiteren können Sie ein Feld nicht bearbeiten, damit es einen Wert enthält, der nicht mit dem Datentyp übereinstimmt. Das heißt, Sie können beispielsweise kein Feld mit numerischem Datentyp bearbeiten, so dass es Buchstaben enthält.

Tabellenansicht

Gehen Sie wie folgt vor, um einen einzelnen Datensatz in der Tabellenansicht zu bearbeiten:

1. Klicken Sie auf das Feld, das Sie bearbeiten möchten, und ändern Sie den Feldwert entsprechend. Schreibgeschützte Felder werden ausgegraut dargestellt. Klicken Sie mit der rechten Maustaste auf das Feld, um auf Optionen für das Ausschneiden, Kopieren und Einfügen zuzugreifen. Wenn Sie ein Feld bearbeitet haben, werden Sie feststellen, dass in der oberen linken Ecke dieses Feldes ein grünes Dreieck erscheint. Dies ist ein visueller Hinweis, der Sie daran erinnern soll, dass der Wert des Feldes geändert, aber noch nicht gespeichert wurde.
2. Aktivieren Sie das Feld **Genehmigt** für den geänderten Datensatz. Dadurch wird der Datensatz als für die Verarbeitung durch Spectrum™ Technology Platform bereit markiert.
3. Wenn Sie eine von Ihnen vorgenommene Änderung rückgängig machen müssen, wählen Sie den Datensatz aus, den Sie ändern möchten, und klicken Sie auf die Schaltfläche **Änderungen rückgängig machen**.
4. Klicken Sie auf die Schaltfläche **Speichern**, wenn Sie mit dem Bearbeiten der Datensätze fertig sind.

Gehen Sie wie folgt vor, um ein Feld für mehrere Datensätze in der Tabellenansicht zu bearbeiten:

1. Klicken Sie bei gedrückter Strg-Taste oder Umschalttaste auf das Feld, das Sie für alle ausgewählten Datensätze ändern möchten. Wenn Sie beispielsweise alle Instanzen von „L.A.“ in „Los Angeles“ ändern möchten, drücken Sie die Strg-Taste und klicken Sie bei einem der Datensätze in das Feld „City“. Klicken Sie dann, während Sie die Strg-Taste oder die Umschalttaste weiter gedrückt halten, bei den anderen Datensätzen, die Sie ändern möchten, in das gleiche Feld.
2. Ändern Sie die Feldwerte entsprechend. Sie können diese Felder bearbeiten, Sie sollten sich aber darüber im Klaren sein, dass hier vorgenommene Änderungen für alle ausgewählten Datensätze gelten, auch wenn die Werte in diesen Feldern zuvor unterschiedlich waren. Ebenso werden Daten aus allen ausgewählten Datensätzen gelöscht, wenn Sie beim Bearbeiten mehrerer Datensätze die Daten aus einem Feld löschen.
3. Aktivieren Sie das Feld **Genehmigt** für die geänderten Datensätze. Dadurch wird der Datensatz als für die Verarbeitung durch Spectrum™ Technology Platform bereit markiert. Alternativ können Sie auf die Schaltfläche **Alle genehmigen** klicken; hierdurch wird jeder einzelne zu

genehmigende Ausnahmedatensatz im Editor angezeigt. (Das Feature „Alle genehmigen“ gilt nicht für übereinstimmende Ausnahmedatensätze.)

4. Wenn Sie eine von Ihnen vorgenommene Änderung rückgängig machen müssen, wählen Sie die Datensätze aus, die Sie ändern möchten, und klicken Sie auf die Schaltfläche **Änderungen rückgängig machen**.
5. Klicken Sie auf die Schaltfläche **Speichern**, wenn Sie mit dem Bearbeiten der Datensätze fertig sind. Die Änderungen der Datensätze werden in der Ausnahme-Datenbank gespeichert und die Ansicht wird aktualisiert. Wenn Sie keinen Workflow für die erneute Überprüfung definiert haben, wird die Liste der Ausnahmen erneut geladen. Jedoch wird mindestens einer der bearbeiteten Datensätze möglicherweise nicht in der aktualisierten Liste angezeigt, da dieser nicht mehr den Such-/Filterkriterien entspricht. Wenn Sie einen Workflow für die erneute Überprüfung *definiert haben*, sind die bearbeiteten Datensätze möglicherweise nicht in der aktualisierten Liste enthalten, wenn diese jetzt gültig sind und aus der Datenbank gelöscht wurden.
6. Rufen Sie über die Navigationsschaltflächen im unteren Bereich des Bildschirms die vorherige und die nächste Seite der Ausnahmedatensätze auf; über diese Schaltflächen können Sie auch direkt den ersten oder letzten Ausnahmedatensatz aufrufen.

Formularansicht

Gehen Sie wie folgt vor, um Datensätze über die Formularansicht zu bearbeiten:

1. Klicken Sie auf **Formularansicht**. Der erste Datensatz der Gruppe wird angezeigt.
2. Klicken Sie auf das Feld, das Sie bearbeiten möchten, und ändern Sie den Feldwert entsprechend. Schreibgeschützte Felder werden ausgegraut dargestellt. Klicken Sie mit der rechten Maustaste auf das Feld, um auf Optionen für das Ausschneiden, Kopieren und Einfügen zuzugreifen. Wenn Sie ein Feld bearbeitet haben, werden Sie feststellen, dass die Umrandung dieses Feldes grün wird. Dies ist ein visueller Hinweis, der Sie daran erinnern soll, dass der Wert des Feldes geändert, aber noch nicht gespeichert wurde.
3. Sie können Kommentare zu Ihren Änderungen in die Spalte **Kommentare** einfügen. Die Kommentare sind für andere Benutzer sichtbar und können zur Nachverfolgung von Änderungen am Datensatz verwendet werden.
4. Wenn Sie sicher sind, dass Sie alle notwendigen Änderungen vorgenommen haben, um eine Gültigkeit des Datensatzes zu erzielen, aktivieren Sie das Feld **Genehmigt**. Dadurch wird der Datensatz als für die Verarbeitung durch Spectrum™ Technology Plattform bereit markiert.
5. Wenn Sie eine von Ihnen vorgenommene Änderung rückgängig machen müssen, klicken Sie auf die Schaltfläche **Änderungen rückgängig machen**.
6. Klicken Sie auf **Speichern**. Die Änderungen des Datensatzes werden in der Ausnahme-Datenbank gespeichert und die Ansicht wird aktualisiert. Entweder wird Ihnen derselbe Datensatz mit Ihren Änderungen angezeigt oder Ihnen wird der nächste Datensatz in der Liste angezeigt, da der von Ihnen geänderte Datensatz nicht mehr verfügbar ist bzw. nicht mehr Ihren Such-/Filterkriterien entspricht.
7. Rufen Sie über die Navigationsschaltflächen im unteren Bereich des Bildschirms den vorherigen und den nächsten Ausnahmedatensatz auf; über diese Schaltflächen können Sie auch direkt den ersten oder letzten Ausnahmedatensatz aufrufen.

Beheben von Datensatzdubletten

Dublettenauflösungsausnahmen treten auf, wenn Spectrum™ Technology Platform nicht sicher bestimmen kann, ob ein Datensatz eine Dublette eines anderen Datensatzes ist. Es gibt zwei Wege, Datensatzdubletten zu beheben.

Anmerkung: Datensatzdubletten können nur mit der Funktion „Dubletten auflösen“ in der Tabellenansicht behoben werden. Sie können diese Datensätze jedoch weiterhin in der Formularansicht bearbeiten.

Ein Ansatz ist die Gruppierung der Datensatzdubletten in Sammlungen. Wenn Sie die Datensätze genehmigen, können sie anschließend mithilfe eines Konsolidierungsvorgangs verarbeitet werden, um die Datensatzdubletten in jeder Sammlung in Ihren Daten zu eliminieren.

Ein weiterer Ansatz ist die Bearbeitung der Datensätze, damit sie wahrscheinlicher als Dubletten erkannt werden, z. B. durch die Korrektur der Schreibweise eines Straßennamen. Wenn Sie die Datensätze genehmigen, verarbeitet Spectrum™ Technology Platform die Datensätze erneut mithilfe eines Vergleichs- und Konsolidierungsprozesses. Nach erfolgreicher Korrektur der Datensätze kann Spectrum™ Technology Platform einen Datensatz als Dublette identifizieren.

Erstellen einer Dublette aus einem anderen Datensatz

Datensatzdubletten werden als Gruppen von Datensätzen im Business Steward Portal angezeigt. Sie können einen Datensatz zu einer Dublette machen, indem Sie ihn in die gleiche Gruppe wie die Datensatzdublette verschieben.

So machen Sie einen Datensatz zu einer Dublette:

1. Wählen Sie den zu bearbeitenden Datensatz aus, und klicken Sie auf **Dubletten auflösen**.

In der Ansicht **Dublettenauflösung** werden Datensatzdubletten angezeigt. Die Datensätze sind in Sammlungen oder Kandidatengruppen gruppiert, die diese Übereinstimmungsdatensatztypen enthalten:

| | |
|------------------|--|
| suspect | Ein Datensatz, mit dem andere Datensätze verglichen werden, um zu bestimmen, ob sie Dubletten voneinander sind. In jeder Sammlung ist lediglich ein Kopfdubletten-Datensatz vorhanden. |
| duplicate | Ein Datensatz, der eine Dublette des Kopfdubletten-Datensatzes ist. |
| unique | Ein Datensatz ohne Dubletten. |

Sie können den Typ eines Datensatzes bestimmen, indem Sie die Spalte „MatchRecordType“ ansehen.

2. Korrigieren Sie bei Bedarf einzelne Datensätze. Weitere Informationen finden Sie unter [Bearbeiten von Ausnahmedatensätzen](#) auf Seite 280. Alternativ können Sie Datensätze zwischen Gruppen verschieben.
3. Geben Sie in das Feld „Sammlungsnummer“ oder „Kandidatengruppe“ die Zahl der Gruppe ein, in die Sie den Datensatz verschieben möchten. Der Datensatz wird zu einer Dublette der anderen Datensätze in der Gruppe.

In manchen Fällen können Sie einen Datensatz mit einem MatchRecordType-Wert eines „Verdächtigen“ nicht in eine andere Sammlung von Dubletten verschieben.

Anmerkung: Datensätze werden entweder nach dem Feld „Sammlungsnummer“ oder dem Feld „Kandidatengruppe“ gruppiert, je nachdem welcher Typ der Vergleichslogik im Datenfluss verwendet wird, der die Ausnahmen erzeugt hat. Kontaktieren Sie Ihren Spectrum™ Technology Platform-Administrator, wenn Sie zusätzliche Informationen über das Vergleichen benötigen.

4. Wenn Sie mit der Änderung der Datensätze fertig sind, aktivieren Sie das Kästchen **Genehmigt**. Dadurch wird angegeben, dass der Datensatz für die erneute Verarbeitung durch Spectrum™ Technology Platform bereit ist.
5. Klicken Sie zum Speichern der Änderungen auf **Speichern**.

Erstellen einer neuen Gruppe von Datensatzdubletten

In manchen Situationen können Sie eine neue Gruppe von Datensätzen erstellen, von denen Sie Dubletten erstellen möchten. In anderen Situationen können Sie keine neuen Gruppen erstellen. Ihre Fähigkeit, neue Gruppen erstellen zu können, wird durch den Typ der Spectrum™ Technology Platform-Verarbeitung bestimmt, der die Ausnahmedatensätze erstellt hat.

1. Wählen Sie den zu bearbeitenden Datensatz aus, und klicken Sie auf **Dubletten auflösen**.

In der Ansicht **Dublettenauflösung** werden Datensatzdubletten angezeigt. Die Datensätze sind in Sammlungen oder Kandidatengruppen gruppiert, die diese Übereinstimmungsdatensatztypen enthalten:

| | |
|------------------|--|
| suspect | Ein Datensatz, mit dem andere Datensätze verglichen werden, um zu bestimmen, ob sie Dubletten voneinander sind. In jeder Sammlung ist lediglich ein Kopfdubletten-Datensatz vorhanden. |
| duplicate | Ein Datensatz, der eine Dublette des Kopfdubletten-Datensatzes ist. |
| unique | Ein Datensatz ohne Dubletten. |

Sie können den Typ eines Datensatzes bestimmen, indem Sie die Spalte „MatchRecordType“ ansehen.

2. Korrigieren Sie bei Bedarf einzelne Datensätze. Weitere Informationen finden Sie unter [Bearbeiten von Ausnahmedatensätzen](#) auf Seite 280.
3. Wählen Sie einen Datensatz aus, den die neue Sammlung enthalten soll, und klicken Sie auf **Neue Sammlung**. Die neue Sammlung erhält automatisch eine einzigartige Sammlungsnummer, und der von Ihnen ausgewählte Datensatz wird zu einem verdächtigen Datensatz.

Anmerkung: Wenn die Schaltfläche **Neue Sammlung** nicht angezeigt wird, können Sie keine neue Sammlung für die Datensätze erstellen, mit denen Sie gerade arbeiten. Sie können nur neue Sammlungen erstellen, wenn der Datenfluss, der die Ausnahmen erzeugt hat, einen „Interflow Match“- oder „Intraflow Match“-Schritt enthält. Dies gilt nicht bei einem enthaltenen „Transactional Match“-Schritt. Kontaktieren Sie Ihren Spectrum™

Technology Platform-Administrator, wenn Sie zusätzliche Informationen über diesen Vergleichsschritt benötigen.

4. Sie nehmen zusätzliche Datensätze in die Sammlung auf, indem Sie die Nummer der neuen Sammlung in das Datensatzfeld „Sammlungsnummer“ eingeben.
5. Wenn Sie mit der Änderung der Datensätze fertig sind, aktivieren Sie das Kästchen **Genehmigt**. Dadurch wird angegeben, dass der Datensatz für die erneute Verarbeitung durch Spectrum™ Technology Platform bereit ist.
6. Klicken Sie zum Speichern der Änderungen auf **Speichern**.

Erstellen eines eindeutigen Datensatzes

So ändern Sie einen Datensatz von einer Dublette in einen eindeutigen Datensatz:

1. Geben Sie in das Feld „MatchRecordType“ das Wort „Eindeutig“ ein.
2. Wenn Sie mit der Änderung der Datensätze fertig sind, aktivieren Sie das Kästchen **Genehmigt**. Dadurch wird angegeben, dass der Datensatz für die erneute Verarbeitung durch Spectrum™ Technology Platform bereit ist.
3. Klicken Sie zum Speichern der Änderungen auf **Speichern**.

Während der Dublettenauflösung automatisch angepasste Felder

Wenn Sie in der Ansicht der Dublettenauflösung im Business Steward Portal Datensätze ändern, werden einige Felder automatisch angepasst, um die neue Disposition des Datensatzes widerzuspiegeln.

Tabelle 21: Von Interflow oder Intraflow Match verarbeitete Datensätze

| Aktion | Automatisch auf Felder angewendete Werte |
|--|---|
| Verschieben eines Datensatzes von einer Sammlung in die andere | <p>Wenn Sie einen Datensatz in eine Sammlung von Dubletten verschieben:</p> <ul style="list-style-type: none"> • MatchRecordType: Dublette • MatchScore: 100 • Enthält Dubletten: D (Dieses Feld ist nur vorhanden, wenn im Datenfluss ein „Interflow Match“-Schritt enthalten war.) <p>Wenn Sie eine Datensatzdublette in die Sammlung eindeutiger Datensätze verschieben (Sammlung 0):</p> <ul style="list-style-type: none"> • MatchRecordType: Eindeutig • MatchScore: Keine Änderung • Enthält Dubletten: U (Dieses Feld ist nur vorhanden, wenn im Datenfluss ein „Interflow Match“-Schritt enthalten war.) <p>Wenn Sie eine Datensatzkopfdublette in die Sammlung eindeutiger Datensätze verschieben (Sammlung 0):</p> <ul style="list-style-type: none"> • MatchRecordType: Eindeutig • MatchScore: 0 • HasDuplicates: N (Dieses Feld ist nur vorhanden, wenn im Datenfluss ein „Interflow Match“-Schritt enthalten war.) |
| Erstellen einer neuen Sammlung | <ul style="list-style-type: none"> • MatchRecordType: Kopfdublette • MatchScore: Kein Wert • HasDuplicates: Y (Dieses Feld ist nur vorhanden, wenn im Datenfluss ein „Interflow Match“-Schritt enthalten war.) <p>Anmerkung: Wenn der Datensatz aus einem Datenfluss stammt, in dem ein „Interflow Match“-Schritt enthalten war, können nur Datensätze mit dem Wert „input_port_0“ im Feld „Interflow-Quellentyp“ Datensatzkopfdubletten sein.</p> |

Tabelle 22: Von „Transactional Match“ verarbeitete Datensätze

| Aktion | Automatisch auf Felder angewendete Werte |
|---|--|
| „MatchRecordType“ in „Dublette“ ändern | <ul style="list-style-type: none"> • HasDuplicates: D • MatchScore: 100 |
| „MatchRecordType“ in „Eindeutig“ ändern | <ul style="list-style-type: none"> • HasDuplicates: U • MatchScore: Nicht geändert |

| Aktion | Automatisch auf Felder angewendete Werte |
|-------------------------------|--|
| „HasDuplicates“ in „D“ ändern | <ul style="list-style-type: none"> • MatchRecordType: Dublette • MatchScore: 100 |
| „HasDuplicates“ in „U“ ändern | <ul style="list-style-type: none"> • MatchRecordType: Eindeutig • MatchScore: Nicht geändert |
| „HasDuplicates“ in „Y“ ändern | <ul style="list-style-type: none"> • MatchRecordType: Kopfdublette • MatchScore: Leer |
| „HasDuplicates“ in „N“ ändern | <ul style="list-style-type: none"> • MatchRecordType: Kopfdublette • MatchScore: Leer |

Verwenden von Suchtools

Der Exception Editor des Business Steward Portal stellt Suchtools zur Verfügung, mit denen Sie nach Informationen suchen können, die beim Bearbeiten und Genehmigen sowie beim erfolgreichen erneuten Ausführen von Ausnahmedatensätzen hilfreich sein können. Die Tools umfassen Bing Maps und die Dienste, die Sie in Spectrum™ Technology Platform lizenziert haben. Bei Tools mit Symbolen, die ein Pluszeichen in der Mitte haben, handelt es sich um Premiumtools. Für die Verwendung dieser Tools fallen Gebühren an.

Anmerkung: Diese Suchtools müssen zunächst auf der Seite „Business Steward-Einstellungen“ der Management Console konfiguriert werden. Wenn auf dieser Seite keine Tools ausgewählt wurden, werden dort keine Tools angezeigt. Unabhängig davon, welche Tools in der Management Console ausgewählt werden, sind die Dienste, die dem Benutzer an dieser Stelle angezeigt werden, zudem auf die Dienste beschränkt, für die er über Berechtigungen zum Anzeigen verfügt.

Verwenden von Bing Maps

Bing Maps zeigt den Standort einer Adresse auf einer Karte an und stellt Steuerungen bereit, mit denen Sie die Karte zoomen sowie hin- und herbewegen können. Zusätzlich können Sie zum Abrufen von Adressen auf die Karte klicken.

Anmerkung: Das Bing Maps-Suchtool wird von Microsoft bereitgestellt. Für die Nutzung dieses Services ist eine Internetverbindung erforderlich.

1. Klicken Sie auf den Datensatz, zu dem Sie eine Recherche durchführen möchten.
2. Klicken Sie unter der Tabelle mit den Datensätzen auf **Suchtools**, um die Ansicht zu erweitern.
3. Wählen Sie im Feld **Service Bing Map** aus.

4. Wählen Sie aus der Spalte **Feldname** die Felder aus, die Sie für Ihre Suche verwenden möchten. Wenn Sie beispielsweise auf einer Karte nach einer Adresse suchen möchten, sollten Sie „AddressLine1“ und „City“ auswählen. Wenn Sie lediglich die Stadt auf einer Karte anzeigen möchten, sollten Sie nur „City“ und „StateProvince“ auswählen.
5. Wählen Sie für eine traditionelle Kartenansicht „Straße“ und für ein Luftbild „Vogelperspektive“ aus. Wenn Bing Maps entscheiden soll, welche Ansicht geeigneter ist, wählen Sie „Automatisch“ aus.
6. Klicken Sie auf **Los**. Die Ergebnisse, einschließlich Längen- und Breitengraden, werden im Feld **Ergebnisse** und auf der Karte in Form einer Stecknadel angezeigt. Klicken Sie auf die Schaltflächen zum Drehen, um die Perspektive um 90° zu verschieben. Klicken Sie auf die Pfeile auf dem Kompass, um den Fokus inkrementell in die ausgewählte Richtung zu verschieben. Verwenden Sie die Schaltflächen „Vergrößern“ und „Verkleinern“, um die Karte mehr oder weniger nah zu fokussieren. Sie können die Stecknadel auch verschieben, indem Sie sie auf eine neue Position ziehen. Die Karteninformationen werden dynamisch aktualisiert.
7. Klicken Sie auf die Karte, um die Adressen anderer Gebäude abzurufen. Ein Wechsel in die Vogelperspektive kann für die Suche nach Gebäuden hilfreich sein.

Nachdem Sie eine erste Kartensuche abgeschlossen haben, können Sie auf einen anderen Ausnahmedatensatz und auf die Schaltfläche „Los“ klicken. Die Karte wird dann entsprechend aktualisiert.

Verwenden von Spectrum-Suchtooldiensten

Pitney Bowes-Suchtooldienste umfassen alle Dienste, für die Sie über eine Lizenz verfügen, wie z. B. die Adressenüberprüfung, das Abrufen von Postleitzahlen usw. Sie können mit diesen Diensten im Exception Editor nach Ausnahmedaten, die Sie korrigieren oder vergrößern möchten, suchen und diese überprüfen.

Beachten Sie, dass Ihnen bei der Verwendung dieses Feature nur dann Dienste angezeigt werden, wenn Sie unter der Plattform-Gruppe für Rollensicherheit über Berechtigungen zum Anzeigen von Diensten verfügen. Ebenso benötigen Sie Berechtigungen zum Ausführen von Diensten, um diese ausführen zu können. Diese Berechtigungen können jedoch mit der Option „Außerkräftsetzungen gesicherter Entitäten“ geändert werden. Durch Verwendung einer Kombination aus Berechtigungen der obersten Ebene und Außerkräftsetzungen kann der Administrator über das Dropdown-Menü „BSM Portal-Dienst“ die Liste der Dienste verwalten, auf die ein bestimmter Benutzer bzw. eine bestimmte Rolle Zugriff hat.

1. Wählen Sie den Datensatz aus, der die Daten enthält, die Sie durchsuchen möchten.
2. Klicken Sie unter dem Exception Editor auf **Suchtools**.
3. Wählen Sie im Feld **Dienst** den Dienst aus, den Sie verwenden möchten, z. B. die Adressenüberprüfung oder das Abrufen von Kandidatenadressen.
4. Wenn der Ausnahmedatensatz Felder enthält, die in diesem Dienst verwendet werden, jedoch unter anderen Namen, ordnen Sie auf der Registerkarte **Dienstfelder** den Ausnahmefeldern die Felder „Diensteingabe“ zu. Wenn Sie beispielsweise die Adressenüberprüfung verwenden und Ihr Ausnahmedatensatz nicht das Feld „AddressLine1“, stattdessen aber das Feld „AddrLine1“ enthält, dann wählen Sie in der Spalte „Ausnahmefelder“ in der Zeile „AddressLine1“ das Feld „AddrLine1“ aus. Sie müssen mindestens ein Eingabefeld zugeordnet haben, bevor Sie den Dienst ausführen können.

Anmerkung: Das Business Steward Portal speichert die von Ihnen erstellten Mappings von Dienstfeldern zu Ausnahmefeldern, solange Sie Ausnahmedatensätze mit den gleichen Feldnamen, Schrittnamen und Datenflussnamen zuordnen. Wenn Ihr Ausnahmedatensatz beispielsweise ein Feld mit dem Namen „AddrLine1“ enthält und Sie dieses Feld dem Feld „AddressLine1“ zuordnen, wird diese Zuordnung gespeichert, solange Sie mit Datensätzen arbeiten, die „AddrLine1“ enthalten und im gleichen Schritt vom gleichen Datenfluss erstellt wurden.

5. Wiederholen Sie Schritt 4 für Dienstausgabefelder. Dieser Schritt ist optional. Sie müssen jedoch mindestens ein Ausgabefeld zugeordnet haben, bevor Sie die Dienstdaten anwenden können. Beachten Sie, dass Sie den Dienstausgabefeldern kein schreibgeschütztes Ausnahmefeld zuordnen können.
6. Klicken Sie auf die Registerkarte **Optionen**, um Datenbankressourcen auszuwählen und Dienstoptionen anzuzeigen und zu ändern, die in der Management Console festgelegt wurden. Wenn Ihnen der Zweck einer bestimmten Option nicht bekannt ist, zeigen Sie mit der Maus auf diese Option, um die zugehörige Beschreibung anzuzeigen. Änderungen, die an dieser Stelle von Ihnen vorgenommen werden, bleiben bestehen, wenn sie von demselben Benutzer, Datenfluss und Schritt des Ausnahmedatensatzes verwendet werden.

Anmerkung: Wenn für den Dienst, den Sie verwenden, eine Datenbank erforderlich ist, müssen Sie zunächst die Datenbankressource in der Management Console konfigurieren. Wenn Sie beispielsweise Datensätze der USA mithilfe von „Validate Address“ überprüfen, müssen Sie eine Datenbank für die USA in der Management Console konfiguriert haben.

7. Klicken Sie auf **Dienst ausführen**. Der aktualisierte Datensatz wird auf der Registerkarte **Ergebnisse** mit einem Statuscode angezeigt, der den Erfolg des Datensatzes angibt. Felder, die Ausnahmedatensätzen zugeordnet sind, werden mit einem Sternchen angegeben.
8. Wählen Sie den Ergebnisdatsatz aus und klicken Sie auf **Dienstdaten anwenden**, um die in die zugeordneten Felder des Ausnahmedatensatzes zu übertragen.
9. Wenn Sie den aktualisierten Datensatz erneut verarbeiten möchten, klicken Sie bei diesem Datensatz auf das Kästchen **Genehmigt** und anschließend auf **Speichern**.

Die Seite „Verwalten“

Auf der Seite „Verwalten“ kann ein Benutzer mit Berechtigungen zum Anzeigen und Ändern eine Ausnahmedatensatzaktivität für alle Personen, denen ein Ausnahmedatensatz zugewiesen ist, überprüfen und verwalten. Zudem bietet die Seite die Möglichkeit, Ausnahmedatensätze eines Benutzers einem anderen Benutzer neu zuzuweisen. Wenn Sie über Berechtigungen zum Löschen verfügen, können Sie die gesamte Gruppe von Ausnahmedatensätzen basierend auf dem Datenflussnamen und der Auftrags-ID von dem System löschen.

Überprüfen der Ausnahmedatensatzaktivität

Im Abschnitt **Status** auf der Seite „Ausnahmen verwalten“ wird eine Ausnahmedatensatzaktivität nach Zuweisung oder Datenflussname angezeigt. Sie können durch Klicken auf die Schaltfläche „Zuweisungen“ oder „Datenflüsse“ in der oberen rechten Ecke des Bildschirms angeben, welche

Ausnahmedatensatzaktivität angezeigt werden soll. „Zuweisungen“ enthält die Anzahl der Ausnahmedatensätze, die den einzelnen Benutzern zugewiesen wurden, sowie die Angabe, wie viele dieser Datensätze genehmigt wurden und wie viele übrig bleiben. „Datenflüsse“ enthält den Prozentsatz der Datenflüsse, die für die einzelnen Datenflüsse genehmigt wurden. Sie können die angezeigten Informationen filtern, indem Sie in der Zeile „Filter“ Suchkriterien eingeben. Die Liste wird dynamisch automatisch aufgefüllt, und zwar mit Datenflüssen oder Personen, denen die Datenflüsse zugewiesen sind, die mit den von Ihnen eingegebenen Buchstaben übereinstimmen.

Im Abschnitt **Fortschritt** auf dieser Seite wird der kumulative Fortschritt für alle Benutzer oder alle Datenflüsse angezeigt, deren individuelle Status auf dem Dashboard angezeigt werden. Wählen Sie im oberen Bereich des Bildschirms „Zuweisungen“ aus, um den Fortschritt der Benutzer anzuzeigen; wählen Sie „Datenflüsse“ aus, um den Fortschritt der Datenflüsse anzuzeigen. Verwenden Sie den Maßstab, um die Metrik für den Maßstab (Tag, Woche, Monat) und die Anzahl der anzuzeigenden Einheiten auszuwählen.

Zuweisen von Ausnahmedatensätzen

Im Abschnitt **Zuweisung** auf der Seite „Ausnahmen verwalten“ (unter **Wartung**) können Sie Ausnahmedatensätze eines Benutzers einem anderen Benutzer neu zuweisen.

1. Wählen Sie im Feld **Benutzer** einen Benutzer aus, dessen Ausnahmen Sie einem anderen Benutzer neu zuweisen möchten.
2. Wenn Sie alle Ausnahmedatensätze eines Benutzers neu zuweisen möchten, fahren Sie mit Schritt 4 fort. Füllen Sie mindestens eines der folgenden Felder aus, um einen Teil der Ausnahmedatensätze eines Benutzers neu zuzuweisen:
 - Datenflussname – Der Name des Datenflusses, der Ausnahmedatensätze erzeugt.
 - Schrittbeschriftung – Der Name des Schritts, in dem die Ausnahmedatensätze erzeugt werden.
 - Auftrags-ID – Die ID, die dem Auftrag mit den Ausnahmedatensätzen zugewiesen ist.
 - Datendomäne – Die Art von **Daten**, die in Exception Monitor zugewiesen werden.
 - Qualitätsmetriken – Die Art von **Metrik**, die in Exception Monitor zugewiesen wird.
 - Anfangsdatum/-uhrzeit – Das Startdatum in einem Datumsbereich, in dem die Ausnahmedatensätze erstellt wurden.
 - Enddatum/-uhrzeit – Das Enddatum in einem Datumsbereich, in dem die Ausnahmedatensätze erstellt wurden.
 - Genehmigungsstatus – Gibt an, ob die Ausnahmedatensätze genehmigt wurden.
3. Nachdem (mindestens) in den Feldern **Benutzer**, **Datenflussname** und **Schrittbeschriftung** eine Auswahl getroffen wurde, können Sie den Filter basierend auf Ausnahmefeldwerten weiter verfeinern.
 - a) Klicken Sie auf das Symbol zum Hinzufügen von Feldfiltern.
 - b) Wählen Sie in der Spalte **Feldname** das Feld aus, nach dem gefiltert werden soll.
 - c) Wählen Sie in der Spalte **Operator** eine der folgenden Optionen aus:

| | |
|-------------------|--|
| ist gleich | Sucht nach Datensätzen mit genau dem von Ihnen angegebenen Wert. Dies kann ein numerischer Wert oder ein Textwert sein. Sie können |
|-------------------|--|

beispielsweise nach Datensätzen mit genau 82 als Wert für „MatchScore“ oder nach Datensätzen mit „Smith“ als Wert für „Nachname“ suchen.

- ist ungleich** Sucht nach Datensätzen mit einem anderen Wert als dem von Ihnen angegebenen Wert. Dies kann ein numerischer Wert oder ein Textwert sein. Sie können beispielsweise nach Datensätzen mit einem beliebigen Wert für „MatchScore“ mit Ausnahme von 100 oder nach Datensätzen mit einem beliebigen Wert für „LastName“ mit Ausnahme von „Smith“ suchen.
- ist größer als** Sucht nach Datensätzen, die einen numerischen Wert enthalten, der größer als der von Ihnen angegebene Wert ist.
- ist größer oder gleich** Sucht nach Datensätzen, die einen numerischen Wert enthalten, der größer oder gleich dem von Ihnen angegebenen Wert ist. Wenn Sie beispielsweise 50 angeben, würden Ihnen Datensätze mit einem Wert von 50 oder mehr im ausgewählten Feld angezeigt werden.
- ist kleiner als** Sucht nach Datensätzen, die einen numerischen Wert enthalten, der kleiner als der von Ihnen angegebene Wert ist.
- ist kleiner oder gleich** Sucht nach Datensätzen, die einen numerischen Wert enthalten, der kleiner oder gleich dem von Ihnen angegebenen Wert ist. Wenn Sie beispielsweise 50 angeben, würden Ihnen Datensätze mit einem Wert von 50 oder weniger im ausgewählten Feld angezeigt werden.
- enthält** Sucht nach Datensätzen mit dem von Ihnen angegebenen Wert an einer beliebigen Stelle innerhalb des ausgewählten Feldes. Wenn Sie beispielsweise im Feld „AddressLine1“ nach „South“ filtern, würden Ihnen Datensätze mit „12 South Ave.“, „9889 Southport St.“, „600 South Shore Dr.“ und „4089 5th St. South“ angezeigt werden.
- enthält nicht** Sucht nach Datensätzen ohne den von Ihnen angegebenen Wert an einer beliebigen Stelle innerhalb des ausgewählten Feldes. Wenn Sie beispielsweise im Feld „AddressLine1“ nach „South“ filtern, jedoch nicht „enthält nicht“ auswählen, würden Ihnen keine Datensätze mit „12 South Ave.“, „9889 Southport St.“, „600 South Shore Dr.“ und „4089 5th St. South“ angezeigt werden.
- beginnt mit** Sucht nach Datensätzen, die im ausgewählten Feld mit einem bestimmten Wert beginnen. Wenn Sie beispielsweise im Feld „LastName“ nach „Van“ filtern, würden Ihnen Datensätze mit „Van Buren“, „Vandenburg“ oder „Van Dyck“ angezeigt werden.
- endet mit** Sucht nach Datensätzen, die im ausgewählten Feld mit einem bestimmten Wert enden. Wenn Sie beispielsweise nach Datensätzen filtern, die im Feld „City“ mit „burg“ enden, würden Ihnen Datensätze mit „Gettysburg“, „Fredricksburg“ und „Blacksburg“ angezeigt werden.

- d) Geben Sie in der Spalte **Feldwert** den Wert ein, der als Filterkriterium verwendet werden soll.

Anmerkung: Beim Suchwert muss die Groß-/Kleinschreibung beachtet werden. Das heißt, dass bei der Suche nach „SMITH“ nur Datensätze zurückgegeben werden, bei denen „SMITH“ in Großbuchstaben geschrieben wird, jedoch nicht „smith“ oder „Smith“.

- e) Um nach mehr als einem Feld zu filtern, müssen Sie mehrere Filter hinzufügen, indem Sie erneut auf das Symbol zum Hinzufügen von Feldfiltern klicken. Wenn Sie beispielsweise alle Datensätze mit „SMITH“ als Wert für den Nachnamen und „NY“ als Statuswert anzeigen möchten, könnten Sie zwei Werte verwenden: einen für das Feld „Nachname“ und einen für das Feld „Status“.

In dem folgenden Beispiel würden alle Datensätze mit dem Wert „FL“ im Feld „Status“ zurückgegeben werden:

| Feldname | Operator | Wert |
|----------|------------|------|
| State | Ist gleich | FL |

In dem folgenden Beispiel würden alle Datensätze zurückgegeben werden, die nicht „60510“ als Wert für „Postleitzahl“ enthalten:

| Feldname | Operator | Wert |
|------------|--------------|-------|
| PostalCode | Ist ungleich | 14226 |
| State | Ist gleich | NY |

In dem folgenden Beispiel würden alle Datensätze zurückgegeben werden, die „NY“ als Wert für „Bundesstaat/Provinz“ mit allen Postleitzahlen außer 14226 enthalten.

| Feldname | Operator | Wert |
|------------|--------------|-------|
| PostalCode | Ist ungleich | 60510 |

- Klicken Sie auf **Neu zuweisen**.
- Optional: Wählen Sie die Anzahl der Ausnahmen aus, die Sie neu zuweisen möchten. Sie können dem neuen Benutzer alle oder die resultierenden Ausnahmen zuweisen. Wenn Sie beispielsweise „10“ als Limit eingeben, werden nur die ersten 10 Datensätze neu zugewiesen, die die Kriterien erfüllen; die restlichen, die Kriterien erfüllenden Datensätze werden nicht neu zugewiesen.
- Wählen Sie im Dropdown-Menü **Neu zuweisen** einen anderen Benutzer aus.
- Klicken Sie auf **Bestätigen**.

Löschen von Ausnahmedatensätzen

Gelegentlich möchten Sie vielleicht Ausnahmedatensätze aus der Datenbank löschen. Sie könnten beispielsweise über Datensätze verfügen, die beim Testen des Systems übrig geblieben sind, oder über Datensätze, die fälschlicherweise nach der Verarbeitung als Ausnahmen galten. Möglicherweise möchten Sie auch erst genehmigte Datensätze verarbeiten und löschen und anschließend denselben

Auftrag erneut ausführen. Im Abschnitt **Löschen** auf der Seite „Verwalten“ (unter **Wartung**) ist all dies möglich.

Sie müssen in den Feldern **Datenflussname** und **Auftrags-ID** eine Auswahl treffen, bevor Sie auf **Entfernen** klicken können. Sie können jedoch im Feld „Auftrags-ID“ die Option „Alle“ auswählen, um Ausnahmedatensätze aus den einzelnen Aufträgen zu entfernen, die von dem ausgewählten Datenfluss ausgeführt werden. Klicken Sie auf **Datenqualitäts-Berichtsdaten entfernen**, um alle Leistungsdaten zu entfernen. Wenn diese Option nicht ausgewählt wird, werden die Ausnahmedatensätze des Auftrags aus der Datenbank entfernt, die Leistungsdaten werden jedoch weiterhin auf der Seite „Datenqualität“ angezeigt.

Die Seite „Datenqualität“

Die Seite „Datenqualität“ im Business Steward Portal enthält Informationen zu Trends innerhalb Ihrer Ausnahmedatensätze.

Anmerkung: Diese Seite wird deaktiviert, wenn Sie die Datenqualitäts-Berichterstellung auf der Seite „Business Steward-Einstellungen“ in der Management Console ausschalten.

Identifizieren von Trends

Im Abschnitt **Trends** auf der Seite „Datenqualitäts-Leistung“ werden die folgenden statistischen Informationen zu Ihren Datenflüssen dargestellt:

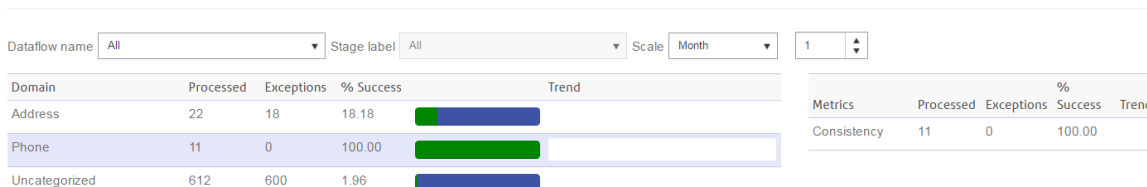
- Gesamtzahl der verarbeiteten Datensätze
- Gesamtzahl der Ausnahmedatensätze
- Prozentsatz der erfolgreich verarbeiteten Datensätze
- Prozentsatz der erfolgreich verarbeiteten Datensätze und Ausnahmedatensätze
- Der Trend Ihrer Daten in 30-Tage-Intervallen

Diese Informationen können innerhalb eines Datenflusses nach Datenflussname oder Schrittbeschriftung aufgeschlüsselt werden. Die hier angezeigten Werte werden durch die Einstellungen bestimmt, die Sie im „Exceptions Monitor“-Schritt Ihrer Datenflüsse ausgewählt haben.

1. Wählen Sie einen **Datenflussnamen** aus, wenn Sie Informationen zu einem bestimmten Datenfluss anzeigen möchten. Andernfalls werden Ihnen Daten zu allen Datenflüssen angezeigt.
2. Wählen Sie eine **Schrittbeschriftung** aus. Beachten Sie, dass Sie einen einzelnen Datenfluss auswählen müssen, wenn Sie die Ergebnisse basierend auf einem Schritt filtern möchten.
3. Wählen Sie für den **Zeitraum** eine Dauer aus, um anzugeben, aus welchem Zeitraum die Daten angezeigt werden sollen. Sie können eine beliebige Anzahl von Tagen, Wochen oder Monaten auswählen.

4. Wählen Sie die entsprechende Datenqualitätsmetrik aus, wenn Sie Ergebnisse nach Datendomäne filtern möchten. Die untere Abbildung zeigt eine erweiterte Metrik für Genauigkeit.

Trends



Business Steward-Einstellungen

Einführung

Einführung in die Business Steward-Einstellungen

Die Business Steward-Einstellungen enthalten die folgenden Tools für Benutzer mit Schreibberechtigungen:

- **Suchvorgänge:** Ermöglicht die Korrektur von Ausnahmedatensätzen mithilfe eines bestimmten Datensatzes für diese Korrekturen.
- **Domänen:** Ermöglicht die Festlegung der auszuwertenden Datenart.
- **Metriken:** Ermöglicht es festzulegen, auf welche Art die Daten gemessen werden.
- **Benachrichtigungen:** Ermöglicht das Verschicken einer Nachricht an eine oder mehrere E-Mail-Adressen, wenn eine bestimmte Anzahl von Ausnahmen mit einer bestimmten Domäne oder Metrik verknüpft ist.
- **Datenqualitäts-Berichterstellung:** Legt Einstellungen zur Nachverfolgung von Erfolgs-/Misserfolgsbedingungen und KPI fest.
- **Suchtooldienste:** Legt Einstellungen für Suchtools im Ausnahmen-Editor des Business Steward Portal fest.
- **Optionen:** Legt Einstellungen für Überwachungsprotokolle und die Nachverfolgung von Fortschritten fest.

Zugreifen auf Business Steward-Einstellungen

Gehen Sie wie folgt vor, um auf Business Steward-Einstellungen zuzugreifen:

1. Öffnen Sie die folgende URL in einem Webbrowser:

`http://Server.Port/managementconsole`

Dabei steht *Server* für den Servernamen oder die IP-Adresse Ihres Spectrum™ Technology Platform-Servers, und *Port* ist der HTTP-Port, der von Spectrum™ Technology Platform verwendet wird. Der HTTP-Port ist standardmäßig auf 8080 eingestellt.

2. Geben Sie einen gültigen Benutzernamen und das dazugehörige Kennwort ein.

3. Klicken Sie auf die Schaltfläche **Ressourcen**.
4. Wählen Sie **Business Steward-Einstellungen** aus.

Suchvorgänge

Das Tool „Suchvorgänge“ ermöglicht die Auswahl aus einer Liste von Werten für ein bestimmtes Feld, wenn die Datensätze im Ausnahmen-Editor aktualisiert werden. Dieses Feature ist besonders hilfreich, wenn Sie über mehrere Datensätze mit Daten im gleichen Feld verfügen, die Sie ändern möchten. So könnten Sie beispielsweise einen Satz von Ausnahmedatensätzen ändern wollen, der Bankdaten enthält. Eines der Felder in diesen Daten könnte aus Codes bestehen, die die Art des mit dem Datensatz verknüpften Kontos darstellen (1 = Girokonto, 2 = Sparkonto, 3 = Geldmarktkonto usw.). Die dazugehörigen Adressen beinhalten ISO-Codes anstelle von Namen im Feld „Land“, wodurch diese Adressen nicht überprüft werden können. Um dies zu korrigieren, könnten Sie einen Suchvorgang erstellen, der ISO-Codes mit ihren entsprechenden Ländernamen bereitstellt, und Korrekturen im Ausnahmen-Editor vornehmen. Dabei können Sie den ISO-Code aus einer Liste auswählen, der anschließend im Feld mit dem verknüpften Ländernamen geladen wird.

Ein weiterer Vorteil dieses Tools ist die Beschränkung der für die Korrekturen verfügbaren Optionen, wodurch die Möglichkeit weiterer Fehler verringert wird. Mit dem gleichen Beispiel von falschen oder fehlenden Ländernamen wird nun ein Suchvorgang erstellt, der eine Liste von Ländernamen bereitstellt. Anstatt diese Namen manuell für jeden Ausnahmedatensatz einzugeben, stellen Sie sicher, dass diese Namen korrekt geschrieben sind und sie mit größerer Wahrscheinlichkeit während der erneuten Verarbeitung überprüft werden.

Was ist der Suchvorgangsprozess?

Der Suchvorgangsprozess beinhaltet drei Schritte, nachdem Sie die Ausnahmen überprüft und ein wiederkehrendes Problem in diesen Ausnahmen erkannt haben (z. B. ungültige Daten in einem Länderfeld):

- Erstellen Sie den Suchvorgang mithilfe von Werten und/oder Beschriftungen von akkuraten Daten, die die falschen Daten überschreiben.
- Verknüpfen Sie mithilfe des „Write Exceptions“-Schrittes im Datenfluss, der die Ausnahmedatensätze erzeugt, das problematische Feld mit dem erstellten Suchvorgang und führen Sie den Datenfluss erneut aus.
- Korrigieren Sie die Ausnahmedatensätze im Ausnahmen-Editor des Business Steward Portal, indem Sie die falschen Daten im problematischen Feld mit korrekten Daten aus dem Suchvorgang überschreiben.

Erstellen von Suchvorgängen

Ein Suchvorgang besteht aus Werten oder Paaren aus Wert und Beschriftung, die Daten enthalten. Diese Daten sollen vorhandene problematische Daten in einem Datenfluss ersetzen, der Ausnahmedatensätze erzeugt. Der Wert ersetzt die problematischen Daten. Die Beschriftung wird in einer Liste angezeigt, aus der Sie auswählen können, wenn Sie mithilfe der Suchtabelle Datensätze im Ausnahmen-Editor korrigieren.

Anmerkung: Wenn Sie nur Werte in Ihre Suchtabelle einbeziehen, werden diese Werte auch als Beschriftungen verwendet.

Sie können einen Suchvorgang laden, indem Sie manuell die Informationen eingeben oder Sie aus einer externen Quelle kopieren und in das Dialogfeld **Viele hinzufügen** einfügen. Die externe Quelle kann eine Tabellenkalkulation, eine Textdatei oder praktisch jede andere Datei sein, solange die Informationen in einer oder zwei Spalten mit Komma, Tabstopp oder Semikolon als Trennzeichen dargestellt sind.

Anmerkung: Wenn Sie die Funktion **Viele hinzufügen** verwenden und auf **Speichern** klicken, werden zuvor vorhandene Werte oder Wert/Beschriftung-Paare für diesen Suchvorgang gelöscht. Nach Verwendung der Funktion „Viele hinzufügen“ können Sie manuell zusätzliche Werte oder Wert/Beschriftung-Paare hinzufügen.

1. Klicken Sie auf die Schaltfläche **Suche hinzufügen**.
2. Geben Sie einen Namen für den neuen Suchvorgang in das Textfeld ein.
3. So fügen Sie manuell ein Wert/Beschriftung-Paar hinzu:
 - Klicken Sie auf die Schaltfläche **Suchwert hinzufügen**.
 - Geben Sie einen Wert und/oder eine Beschriftung für den Suchvorgang ein.

So verwenden Sie die Funktion **Viele hinzufügen**:

- Klicken Sie auf die Schaltfläche **Viele hinzufügen**, um das Dialogfeld zu öffnen.
- Wählen Sie die Spalten und das entsprechende Trennzeichen aus. Wenn Sie Daten aus Microsoft Excel einfügen, verwenden Sie den Tabstopp als Trennzeichen. Wenn Sie das falsche Trennzeichen angeben, importiert das Tool die gesamte Zeile als Wert oder Beschriftung (je nachdem was als erste Spalte angegeben ist).
- Geben Sie die Werte, Trennzeichen und/oder Beschriftungen für alle Inhalte ein, oder fügen Sie die Inhalte aus einer anderen Anwendung ein.

Nachdem alle Werte oder Wert/Beschriftung-Paare hinzugefügt wurden, können Sie sie entweder in der Spalte „Wert“ oder „Beschriftung“ auf- oder absteigend sortieren. Beachten Sie, dass die Sortierung der Liste nicht aufgehoben werden kann (dies gilt jedoch nicht für die Rückwärtssortierung oder die Sortierung einer anderen Spalte).

4. Wiederholen Sie Schritt 3 nach Bedarf.
5. Klicken Sie auf **Speichern**.

Zuweisen von Suchvorgängen

Nach der Erstellung eines Suchvorgangs müssen Sie diesen Suchvorgang im „Write Exceptions“-Schritt des Datenflusses dem Feld mit den problematischen Daten zuweisen.

1. Öffnen Sie im Enterprise Designer den Datenfluss, der die Ausnahmedatensätze erzeugt.
2. Öffnen Sie den „Write Exceptions“-Schritt.
3. Wählen Sie in der Spalte **Name der Suche** für das Feld mit den problematischen Daten den Suchvorgang aus der Dropdown-Liste aus, der die neuen, korrekten Daten enthält, und klicken Sie auf **OK**.
4. Speichern Sie, und kehren Sie zum Datenfluss zurück.

Korrigieren von Datensätzen

Nach dem Erstellen eines Suchvorgangs und dem Zuweisen dieses Vorgangs zu einem Feld im Datenfluss müssen Sie die Ausnahmedatensätze im Business Steward Portal korrigieren.

1. Öffnen Sie im Ausnahmen-Editor den Datenfluss, der die Ausnahmedatensätze erzeugt.
2. Klicken Sie für den ersten problematischen Datensatz auf das Feld, dem Sie den Suchvorgang zugewiesen haben.
3. Klicken Sie auf die Dropdown-Schaltfläche in diesem Feld, und wählen Sie die korrekte Beschriftung für diesen Datensatz aus. Denken Sie daran, dass diese Beschriftung nicht unbedingt dem Wert entsprechen muss. Wenn ein Feld beispielsweise den Wert „Kalifornien“ enthalten soll, könnten Sie auf eine Beschriftung klicken, die „KA“ lautet.
4. Wiederholen Sie Schritt 3 für jeden problematischen Datensatz.
5. Speichern Sie die geänderten Ausnahmen.

Ändern oder Löschen von Suchvorgängen

1. Öffnen Sie die Seite „Suchvorgänge“.
2. Aktivieren Sie das Kästchen neben dem entsprechenden Suchvorgang.
3. Klicken Sie auf die Schaltfläche **Suche bearbeiten**, ändern Sie den Suchvorgang nach Wunsch, und klicken Sie auf **Speichern**.

oder

Klicken Sie auf die Schaltfläche **Suche löschen**.

Domänen

Domänen geben die Art der auszuwertenden Daten an. Dies wird für die Berichterstellung verwendet, um zu zeigen, welche Ausnahmetypen in Ihren Daten auftreten. Wenn die Bedingung beispielsweise den Erfolg oder Misserfolg einer Adressenüberprüfung auswertet, könnte die Datendomäne „Adresse“ lauten. Wenn die Bedingungen den Erfolg oder Misserfolg eines Geocoding-Vorgangs auswertet, könnte die Datendomäne „Räumlich“ usw. lauten.

Anmerkung: Die von Ihnen hier erstellten Domänen dienen als Standardoptionen für die Business Steward-Konfiguration und den „Exception Monitor“-Schritt.

Sie können eine der unten aufgeführten vordefinierten Domänen auswählen oder Ihre eigene Domäne durch Klicken auf **Element hinzufügen** und Ausfüllen der notwendigen Felder angeben. Des Weiteren können Sie Domänen bearbeiten, indem Sie eine Domäne auswählen, auf die Schaltfläche **Element bearbeiten** klicken und die notwendigen Änderungen vornehmen. Sie können auch die Liste der angezeigten Domänen filtern, indem Sie die Suchdaten in das Feld **Filter** eingeben. Die Ergebnisse werden dynamisch aktualisiert.

- Account – Die Bedingung überprüft einen Unternehmens- oder Organisationsnamen, der einem Verkaufskonto zugeordnet ist.
- Address – Die Bedingung überprüft Adressdaten, z. B. eine vollständige Postanschrift oder eine Postleitzahl.

- **Asset** – Die Bedingung überprüft Daten zum Eigentum eines Unternehmens, z. B. physisches Eigentum, Immobilien, Personal oder andere Objekte.
- **Date** – Die Bedingung überprüft Daten zu Datumsangaben.
- **Email** – Die Bedingung überprüft E-Mail-Daten.
- **Financial** – Die Bedingung überprüft Daten zu Währung, Sicherheiten usw.
- **Name** – Die Bedingung überprüft Daten zu Personennamen, z. B. einen Vor- oder Nachnamen.
- **Phone** – Die Bedingung überprüft Daten zu Telefonnummern.
- **Product** – Die Bedingung überprüft Daten zu Materialien, Artikeln, Warenhandel usw.
- **Spatial** – Die Bedingung überprüft Daten zu Punkten, Vielecken oder Linien, die eine definierte geografische Besonderheit darstellen, z. B. Überschwemmungsgebiete, Küstenlinien, Häuser, Absatzgebiete usw.
- **SSN** – Die Bedingung überprüft Daten zur US- Sozialversicherungsnummer.
- **Uncategorized** – Wählen Sie diese Option aus, wenn diese Bedingung nicht kategorisiert werden soll.

Metriken

Metriken geben die Art und Weise an, wie Daten gemessen werden. Dies wird für die Berichterstellung verwendet, um zu zeigen, welche Ausnahmetypen in Ihren Daten auftreten. Wenn die Bedingung beispielsweise die Vollständigkeit eines Datensatzes auswerten soll (also z. B. ob alle Adressen eine Postleitzahl enthalten), könnten Sie „Vollständigkeit“ als Datenqualitätsmetrik angeben.

Anmerkung: Die von Ihnen hier erstellten Metriken dienen als Standardoptionen für die Business Steward-Konfiguration und den „Exception Monitor“-Schritt.

Sie können eine der unten aufgeführten vordefinierten Metriken auswählen oder Ihre eigene Metrik durch Klicken auf **Element hinzufügen** und Ausfüllen der notwendigen Felder angeben. Des Weiteren können Sie Metriken bearbeiten, indem Sie eine Metrik auswählen, auf die Schaltfläche **Element bearbeiten** klicken und die notwendigen Änderungen vornehmen. Sie können auch die Liste der angezeigten Metriken filtern, indem Sie die Suchdaten in das Feld **Filter** eingeben. Die Ergebnisse werden dynamisch aktualisiert.

- **Accuracy** – Die Bedingung erfasst, ob die Daten auf Basis einer vertrauenswürdigen Quelle überprüft werden konnten. Wenn eine Adresse beispielsweise nicht anhand der Daten von der Postbehörde überprüft werden konnte, könnte sie als Ausnahme gelten, da sie nicht korrekt ist.
- **Completeness** – Die Bedingung erfasst, ob in den Daten wichtige Attribute fehlen. Beispiel: eine Adresse, bei der die Postleitzahl fehlt, oder ein Konto ohne Namen der Kontaktperson.
- **Consistency** – Die Bedingung erfasst, ob die Daten zwischen mehreren Systemen konsistent sind. Wenn Ihr Kundendatensystem beispielsweise die Geschlechtscodes M und F verwendet, die von Ihnen verarbeiteten Daten jedoch die Geschlechtscodes 0 und 1 aufweisen, könnte bei den Daten von einem Konsistenzproblem ausgegangen werden.
- **Interpretability** – Die Bedingung erfasst, ob Daten korrekt in eine Datenstruktur geparkt wurden, die von einem anderen System interpretiert werden kann. Sozialversicherungsnummern sollten beispielsweise nur numerische Daten enthalten. Wenn die Daten Buchstaben enthalten, z. B. xxx-xx-xxxx, könnte bei ihnen von Auswertbarkeitsproblemen ausgegangen werden.

- Recency – Die Bedingung erfasst, ob die Daten auf dem neuesten Stand sind. Wenn beispielsweise eine Person umzieht, in Ihrem System jedoch die alte Adresse dieser Person gespeichert ist, könnte bei den Daten von einem Aktualitätsproblem ausgegangen werden.
- Uncategorized – Wählen Sie diese Option aus, wenn diese Bedingung nicht kategorisiert werden soll.
- Uniqueness – Die Bedingung erfasst, ob Daten doppelt vorliegen. Wenn der Datenfluss keine doppelten Daten konsolidieren konnte, könnte davon ausgegangen werden, dass es sich bei den Datensätzen um Ausnahmen handelt.

Benachrichtigungen

Mit dem Feature „Benachrichtigungen“ können Sie das System eine Nachricht an eine oder mehrere E-Mail-Adressen schicken lassen, wenn eine bestimmte Anzahl von Ausnahmen mit einer bestimmten Domäne oder Metrik verknüpft ist. Diese E-Mail enthält einen Link zu den fehlgeschlagenen Datensätzen im Ausnahmen-Editor des Business Steward Portal, wo Sie manuell die korrekten Daten eingeben können. Um den Empfang von Benachrichtigungen an einer bestimmten E-Mail-Adresse zu stoppen, entfernen Sie diese Adresse aus der Liste der Empfänger in der Zeile „Benachrichtigung senden an“ auf der Seite „Domäne bearbeiten“.

Anmerkung: Bevor Sie erfolgreich eine Benachrichtigung aus der Business Steward-Konfiguration verwenden können, müssen Sie die Benachrichtigungen in der Management Console einrichten. Weitere Informationen zur Konfiguration von Benachrichtigungen finden Sie im Administratorhandbuch.

1. Öffnen Sie in der Business Steward-Konfiguration die Seite **Domänen** oder die Seite **Metriken**.
2. Wählen Sie die Domäne oder Metrik aus, für die Sie eine Benachrichtigung hinzufügen möchten, und klicken Sie auf die Schaltfläche **Element bearbeiten**.
3. Wählen Sie E-Mail-Adresse(n) aus der Dropdown-Liste aus (wie in der Management Console konfiguriert) oder geben Sie neue E-Mail-Adressen ein, an die die Benachrichtigungen geschickt werden sollen.
4. Wählen Sie die Anzahl der Ausnahmedatensätze aus, die eine Benachrichtigung auslösen sollen.
5. Geben Sie den Text ein, der als Betreff der Benachrichtigung geschickt werden soll.
6. Geben Sie die Nachricht ein, die als Text der Benachrichtigung erscheinen soll. Sie können Variable in der Nachricht verwenden, um wichtige Informationen über die Ausnahmen weiterzugeben. Zu den Variablen gehören:
 - `#{jobID}`: Die ID-Nummer des Auftrags, der die Ausnahmedatensätze erzeugt hat.
 - `#{jobName}`: Der Name des Auftrags, der die Ausnahmedatensätze erzeugt hat.
 - `#{userName}`: Der Name des Benutzers, dessen Auftrag die Ausnahmedatensätze erzeugt hat.
 - `#{stageLabel}`: Der Name des Datenflussschrittes, in dem die Ausnahmedatensätze erzeugt wurden.
 - `#{link}`: Ein Link zur Editor-Seite im Business Steward Portal, der Datensätze eines bestimmten Datenfluss zeigt.

7. Aktivieren Sie das Kästchen **Erinnerung senden**, wenn Sie eine Erinnerungsbenachrichtigung senden möchten. Wählen Sie die Anzahl der Tage aus, die vergehen sollen, bis die Erinnerung verschickt wird.
8. Geben Sie den Text ein, der als Betreff der Erinnerungsbenachrichtigung geschickt werden soll.
9. Geben Sie die Nachricht ein, die als Text der Erinnerungsbenachrichtigung erscheinen soll. Die Erinnerung verwendet eine zusätzliche Variable:
 - $\${Count}$: Die Anzahl der Ausnahmen für einen bestimmten Datenfluss oder Schritt, die noch gelöst werden müssen.
10. Aktivieren Sie das Kästchen **Täglich erinnern**, wenn jeden Tag eine Erinnerungsbenachrichtigung geschickt werden soll, bis die Ausnahmen gelöst wurden.

Datenqualitäts-Berichterstellung

1. Klicken Sie auf **Datenqualitäts-Berichterstellung**, um erfüllte oder nicht erfüllte Bedingungen im „Exception Monitor“-Schritt nachzuverfolgen. Wenn Sie diese Option deaktivieren, enthält die Seite „Datenqualität“ im Business Steward Portal keine Daten. Ebenso wird das Feld „Nur Bericht“ im „Exception Monitor“-Schritt für alle Datenflüsse deaktiviert.
2. Wählen Sie im Dropdown-Menü **Aufbewahrung** aus, wie lange (in Monaten) diese Daten aufbewahrt werden sollten.

Konfigurieren der wichtigen Leistungsindikatoren

Sie können im Abschnitt **KPI-Konfiguration** auf der Registerkarte „Datenqualitäts-Berichterstellung“ wichtige Leistungsindikatoren (Key Performance Indicators, kurz KPI) für Ihre Daten bestimmen und Benachrichtigungen für Fälle zuweisen, in denen diese KPIs bestimmte Bedingungen erfüllen.

1. Klicken Sie auf **KPI hinzufügen**.
2. Geben Sie einen **Namen** für den wichtigen Leistungsindikator ein. Dieser Name muss auf Ihrem Spectrum™ Technology Platform-Server eindeutig sein.
3. Wählen Sie eine der Datenqualitäts-**Metriken** für den wichtigen Leistungsindikator aus. Wenn Sie keine Auswahl treffen, wird dieser wichtige Leistungsindikator mit allen **Metriken** verknüpft.
4. Wählen Sie einen **Datenflussnamen** für den wichtigen Leistungsindikator aus. Wenn Sie keine Auswahl treffen, wird dieser wichtige Leistungsindikator mit allen Datenflüssen im Business Steward-Modul verknüpft.
5. Wählen Sie eine **Schrittbeschriftung** für den wichtigen Leistungsindikator aus. Wenn Sie keine Auswahl treffen, wird dieser wichtige Leistungsindikator mit allen Schritten in den Datenflüssen des Business Steward-Moduls verknüpft.
6. Wählen Sie eine **Datendomäne** für den wichtigen Leistungsindikator aus. Dieser wichtige Leistungsindikator wird mit allen **Domänen** verknüpft. Beachten Sie, dass das Feld „Bedingung“ durch die Auswahl einer Domäne an dieser Stelle deaktiviert wird.
7. Wählen Sie eine **Bedingung** für den wichtigen Leistungsindikator aus. Wenn Sie keine Auswahl treffen, wird für diesen wichtigen Leistungsindikator standardmäßig „Alle“ verwendet. Beachten Sie, dass Sie zunächst „Alle“ im Feld „Domäne“ aktiviert haben müssen, bevor Sie eine

Bedingung auswählen können. Sobald eine Bedingung ausgewählt wurde, wird das Feld „Domäne“ deaktiviert.

8. Wählen Sie einen **KPI-Zeitraum** aus, um die Intervalle zu bestimmen, in denen das Business Steward-Modul Ihre Daten überwachen und Benachrichtigungen senden soll. Wenn Sie beispielsweise „1“ und „Monatlich“ auswählen, wird eine KPI-Nachricht gesendet, wenn der Anteil der Ausnahmen pro Schwellenwert oder Streuung in einem Zeitraum auf monatlicher Basis zugenommen hat.
9. Geben Sie einen Prozentsatz für eine **Streuung** oder einen **Schwellenwert** an. Streuungswerte stellen den angestiegenen Anteil von Fehlern in Ausnahmedatensätzen seit dem letzten Zeitraum dar. Schwellenwerte stellen den Anteil von Fehlern dar, bei denen die Benachrichtigungen gesendet werden sollen. Der Wert muss 1 oder größer sein.
10. Wählen Sie E-Mail-Adressen aus der Liste aus oder geben Sie E-Mail-Adressen der **Empfänger** ein, die bei Erfüllung dieser Bedingungen benachrichtigt werden sollen. Wenn möglich, wird dieses Feld bei der Eingabe der E-Mail-Adressen automatisch vervollständigt. Sie müssen die Adressen nicht mit einem Komma, Semikolon oder einem anderen Satzzeichen voneinander trennen.
11. Geben Sie den **Betreff** ein, der in der Benachrichtigungs-E-Mail verwendet werden soll.
12. Geben Sie die **Nachricht** ein, die in der Benachrichtigung weitergegeben werden soll, wenn diese Bedingungen erfüllt sind.
13. Klicken Sie auf **OK**. Der neue KPI wird unter allen anderen bestehenden KPIs angezeigt. Sie können KPIs in allen Spalten sortieren, die Daten enthalten.
14. Klicken Sie auf **Speichern**.

Sie können KPIs ändern und entfernen, indem Sie einen KPI auswählen und auf **Ausgewählten KPI bearbeiten** oder auf **Ausgewählten KPI löschen** klicken.

Suchtooldienste

1. Wählen Sie aus, welche Suchtooldienste im Exception Editor des Business Steward Portal verfügbar sein sollen. Die Liste der verfügbaren Dienste basiert auf Benutzerberechtigungen und wird durch Ihre lizenzierten Module und Dienste innerhalb der Spectrum Technology Platform aufgefüllt. Mit der Option **Filter** können Sie die Liste der Dienste basierend auf Filterkriterien einschränken.
2. Klicken Sie auf **Premium**, um Benutzer darauf hinzuweisen, dass bei der Verwendung dieser Dienste (z. B. Dienste von Dun & Bradstreet) zusätzliche Gebühren entstehen.

Optionen

1. Klicken Sie auf **Ausnahmeereignisse überwachen**, um im Business Steward-Modul ein Protokoll zu verwalten, in dem festgehalten wird, wann Ausnahmedatensätze erstellt, gelesen, aktualisiert oder gelöscht werden.
2. Klicken Sie auf **Fortschritt nachverfolgen**, um nachzuverfolgen, wann Ausnahmedatensätze im Business Steward Portal genehmigt werden. Wenn Sie diese Option ausschalten, werden die Fortschrittsdiagramme auf dem Dashboard im Business Steward Portal nicht angezeigt.

Data Normalization-Modul

Data Normalization-Modul

Das Data Normalization-Modul prüft Begriffe in einem Datensatz und ermittelt, ob es sich bei dem Begriff um die bevorzugte Form handelt.

Komponenten

Das Data Normalization-Modul setzt sich aus folgenden Komponenten zusammen:

- **Advanced Transformer:** In diesem Schritt werden Datenzeichenfolgen durchsucht und in mehrere Felder unterteilt. Dabei werden die extrahierten und nicht extrahierten Daten in einem vorhandenen Feld oder einem neuen Feld platziert.
- **Open Parser:** In diesem Schritt werden Ihre Eingabedaten aus vielen Kulturen der Welt mit einer einfachen, aber leistungsstarken Parsing-Grammatik geparkt. Mithilfe dieser Grammatik können Sie eine Sequenz von Ausdrücken definieren, die Domänenmuster zum Parsen Ihrer Eingabedaten darstellen. Open Parser sammelt auch statistische Daten und bewertet die Parsing-Vergleiche, um Ihnen bei der Bestimmung der Effektivität Ihrer Parsing-Grammatiken zu helfen.
- **Table Lookup:** In diesem Schritt wird ein Begriff ausgewertet und mit einem zuvor überprüften Format dieses Begriffs verglichen. Wenn der Begriff nicht das richtige Format aufweist, wird er durch die Standardversion ersetzt. Table Lookup umfasst das Ändern vollständiger Wörter in Abkürzungen, das Ändern von Abkürzungen in vollständige Wörter, das Ändern von Spitznamen in vollständige Namen oder von Rechtschreibfehlern in die korrigierte Schreibweise.
- **Transliterator:** Transliterator konvertiert eine Zeichenfolge zwischen lateinischen und anderen Skripts.

Advanced Transformer

Der Auftrag „Advanced Transformer“ scannt und teilt die Datenzeichenfolgen in mehrere Felder auf und verwendet dazu Tabellen oder reguläre Ausdrücke. Er extrahiert einen bestimmten Begriff oder eine angegebene Anzahl von Wörtern rechts oder links von einem Begriff. Extrahierte und nicht extrahierte Daten können in ein vorhandenes Feld oder ein neues Feld eingefügt werden.

Zum Beispiel sollen die Suite-Informationen aus diesem Adressfeld extrahiert und in ein separates Feld eingefügt werden.

2300 BIRCH RD STE 100

Dazu könnten Sie einen „Advanced Transformer“ erstellen, der den Begriff STE und alle Wörter rechts vom Begriff STE extrahiert. Das verbleibende Feld sieht dann so aus:

2300 BIRCH RD

Eingabe

Der Advanced Transformer verwendet jedes definierte Eingabefeld im Datenfluss.

Optionen

Advanced Transformer-Optionen können in der Schrittebene durch beliebige der Spectrum™ Technology Platform-Clients oder zur Laufzeit mit Datenflussoptionen konfiguriert werden.

Konfigurieren von Optionen

Um die Optionen für den Advanced Transformer anzugeben, erstellen Sie eine Regel. Sie können mehrere Regeln erstellen und dann die Reihenfolge angeben, in der die Regeln angewendet werden sollen. So erstellen Sie eine Regel:

1. Doppelklicken Sie auf die Instanz von Advanced Transformer auf der Arbeitsfläche. Das Dialogfeld Advanced Transformer-Optionen wird angezeigt.
2. Wählen Sie die Anzahl der Laufzeitinstanzen aus und klicken Sie auf **OK**. Verwenden Sie die Option „Laufzeitinstanzen“, um einen Datenfluss zu konfigurieren, um mehrere, parallele Instanzen eines Schrittes auszuführen und so potenziell die Leistung zu erhöhen.
3. Klicken Sie auf die Schaltfläche **Hinzufügen**. Das Dialogfeld Regeloptionen für Advanced Transformer wird angezeigt.

Anmerkung: Wenn Sie mehrere Transformer-Regeln hinzufügen, können Sie die Schaltflächen **Nach oben** und **Nach unten** verwenden, um die Reihenfolge zu ändern, in der die Regeln angewendet werden.

4. Wählen Sie den Typ der Transformationsaktion aus, die Sie ausführen möchten, und klicken Sie auf **OK**. Die Optionen sind in der nachfolgenden Tabelle aufgelistet.

Tabelle 23: Advanced Transformer-Optionen

| Option | Beschreibung |
|--------|---|
| Quelle | Gibt das Quelleingabefeld an, das zum Suchen und Aufteilen ausgewertet werden soll. |

| Option | Beschreibung |
|------------------------------|--|
| Extrahieren mit | <p>Wählen Sie Tabellendaten oder Reguläre Ausdrücke aus.</p> <p>Wählen Sie Tabellendaten, wenn Sie mit den XML-Tabellen unter <code><Drive>:\Program Files\Pitney Boves\Spectrum\server\modules\advancedtransformer\data</code> scannen und teilen möchten. Weitere Informationen zu jeder Option finden Sie in den Tabellendateninformationen.</p> <p>Wählen Sie Reguläre Ausdrücke aus, wenn Sie zum Scannen und Aufteilen reguläre Ausdrücke verwenden möchten. Reguläre Ausdrücke bieten viele zusätzliche Optionen zum Aufteilen von Daten. Sie können einen der vordefinierten regulären Ausdrücke verwenden, indem Sie ihn aus der Liste auswählen, oder Sie können mit der RegEx-Syntax einen eigenen erstellen.</p> <p>Beispiel: Sie könnten die Daten aufteilen, wenn der erste numerische Wert gefunden wird, so wie in „John Smith 123 Main St.“, wo „John Smith“ in ein Feld und „123 Main St.“ in ein anderes Feld übernommen würde. Weitere Informationen zu den einzelnen Optionen finden Sie unten unter „Optionen für reguläre Ausdrücke“.</p> |
| Tabellendatenoptionen | |
| Nicht extrahierte Daten | <p>Gibt das Ausgabefeld an, das die transformierten Daten enthalten soll. Wenn der ursprüngliche Wert ersetzt werden soll, geben Sie im Feld „Ziel“ dasselbe Feld an wie zuvor im Dropdown-Feld „Quelle“.</p> <p>Sie können in das Feld „Ziel“ auch einen neuen Feldnamen eingeben. Wenn Sie einen neuen Feldnamen eingeben, ist dieser Feldname in den Schritten in Ihrem Datenfluss verfügbar, die dem Advanced Transformer nachgeschaltet sind.</p> |
| Extrahierte Daten | <p>Gibt das Ausgabefeld an, in das die extrahierten Daten eingefügt werden sollen.</p> <p>Sie können in das Feld „Extrahierte Daten“ auch einen neuen Feldnamen eingeben. Wenn Sie einen neuen Feldnamen eingeben, ist dieser Feldname in den Schritten in Ihrem Datenfluss verfügbar, die dem Advanced Transformer nachgeschaltet sind.</p> |
| Tokenisierungszeichen | <p>Gibt beliebige Sonderzeichen an, die Sie tokenisieren möchten. Bei der Tokenisierung werden Begriffe getrennt. Beispiel: Bei einem Feld mit den Daten „Smith, John“ würden Sie das Komma tokenisieren. Das ergäbe die Begriffe:</p> <ul style="list-style-type: none"> • Smith • , • John <p>Nun, wo die Begriffe getrennt sind, können die Daten durch Scannen auf und Extrahieren des Kommas aufgeteilt werden. „Smith“ und „John“ werden so sauber als zu standardisierende Daten identifiziert.</p> |

| Option | Beschreibung | | | | | | |
|--|--|----------------------------|---|--|---|---|---|
| Tabelle | <p>Gibt die Tabelle an, die die Begriffe enthält, anhand denen das Feld aufgeteilt werden soll. Eine Liste der Tabellen finden Sie unter Advanced Transformer-Tabellen auf Seite 164. Weitere Informationen zum Erstellen oder Ändern von Tabellen finden Sie unter Einführung in Suchtabellen auf Seite 164.</p> | | | | | | |
| Aus mehreren Wörtern bestehende Begriffe suchen | <p>Aktivieren Sie dieses Kästchen, um mehrere Wortsuchläufe innerhalb einer gegebenen Zeichenfolge zu erlauben. Beispiel:</p> <p>Eingabezeichenfolge = „Cedar Rapids 52401“, Geschäftsregel = „Cedar Rapids“ in Zeichenfolge auf Basis einer Tabelle mit dem Eintrag identifizieren; Cedar Rapids = US Output = Identifiziert Vorkommen von „Cedar Rapids“ und übernimmt die Begriffe in ein neues Feld, z. B. Ort.</p> <p>Bei der Mehrwortsuche stoppt die Suche beim ersten Vorkommen einer Übereinstimmung.</p> <p>Anmerkung: Die Auswahl dieser Option kann die Leistung beeinträchtigen.</p> | | | | | | |
| Extrahieren | <p>Gibt den Typ der durchzuführenden Extrahierung an. Wählen Sie aus Folgenden aus:</p> <table border="0"> <tr> <td data-bbox="552 1060 779 1102">Begriff extrahieren</td> <td data-bbox="828 1060 1429 1134">Extrahiert den durch die ausgewählte Tabelle identifizierten Begriff.</td> </tr> <tr> <td data-bbox="552 1144 779 1249">N Wörter auf der rechten Seite des Begriffs</td> <td data-bbox="828 1144 1429 1312">Extrahiert Wörter auf der rechten Seite des Begriffs. Sie geben die Anzahl der zu extrahierenden Wörter an. Wenn Sie beispielsweise die zwei Wörter rechts vom identifizierten Begriff extrahieren möchten, geben Sie 2 an.</td> </tr> <tr> <td data-bbox="552 1323 779 1386">N Wörter auf der linken Seite des Begriffs</td> <td data-bbox="828 1323 1429 1480">Extrahiert Wörter auf der linken Seite des Begriffs. Sie geben die Anzahl der zu extrahierenden Wörter an. Wenn Sie beispielsweise die zwei Wörter links vom identifizierten Begriff extrahieren möchten, geben Sie 2 an.</td> </tr> </table> <p>Wenn Sie Wörter rechts oder links des Begriffs extrahieren möchten, können Sie angeben, ob der Begriff selbst in die Zieldaten oder die extrahierten Daten eingefügt werden soll. Sie haben beispielsweise dieses Feld:</p> <p>2300 BIRCH RD STE 100</p> <p>Sie möchten „STE 100“ extrahieren und in das in „Extrahierte Daten“ angegebene Feld übernehmen. Sie würden den Begriff in das Feld mit den extrahierten Daten einfügen, somit also die Abkürzung „STE“ und das Wort „100“.</p> <p>Wenn Sie weder „Ziel“ noch „Extrahierte Daten“ auswählen, wird der Begriff nicht eingefügt und verworfen.</p> | Begriff extrahieren | Extrahiert den durch die ausgewählte Tabelle identifizierten Begriff. | N Wörter auf der rechten Seite des Begriffs | Extrahiert Wörter auf der rechten Seite des Begriffs. Sie geben die Anzahl der zu extrahierenden Wörter an. Wenn Sie beispielsweise die zwei Wörter rechts vom identifizierten Begriff extrahieren möchten, geben Sie 2 an. | N Wörter auf der linken Seite des Begriffs | Extrahiert Wörter auf der linken Seite des Begriffs. Sie geben die Anzahl der zu extrahierenden Wörter an. Wenn Sie beispielsweise die zwei Wörter links vom identifizierten Begriff extrahieren möchten, geben Sie 2 an. |
| Begriff extrahieren | Extrahiert den durch die ausgewählte Tabelle identifizierten Begriff. | | | | | | |
| N Wörter auf der rechten Seite des Begriffs | Extrahiert Wörter auf der rechten Seite des Begriffs. Sie geben die Anzahl der zu extrahierenden Wörter an. Wenn Sie beispielsweise die zwei Wörter rechts vom identifizierten Begriff extrahieren möchten, geben Sie 2 an. | | | | | | |
| N Wörter auf der linken Seite des Begriffs | Extrahiert Wörter auf der linken Seite des Begriffs. Sie geben die Anzahl der zu extrahierenden Wörter an. Wenn Sie beispielsweise die zwei Wörter links vom identifizierten Begriff extrahieren möchten, geben Sie 2 an. | | | | | | |

| Option | Beschreibung |
|--|--|
| Optionen für reguläre Ausdrücke | |
| Reguläre Ausdrücke | <p>Wählen Sie einen der vordefinierten regulären Ausdrücke aus der Liste aus oder erstellen Sie im Textfeld einen eigenen. Der Advanced Transformer unterstützt die RegEx-Standardsyntax.</p> <p>Java 2 Plattform enthält ein Paket mit Namen java.util.regex, das die Verwendung regulärer Ausdrücke ermöglicht. Weitere Informationen finden Sie unter: java.sun.com/docs/books/tutorial/essential/regex/index.html.</p> |
| Auslassungsschaltfläche | Klicken Sie auf diese Schaltfläche, um einen neuen regulären Ausdruck hinzuzufügen oder zu entfernen. |
| Gruppe auffüllen | Nachdem Sie einen vordefinierten RegEx-Ausdruck ausgewählt oder einen neuen eingegeben haben, klicken Sie auf Gruppe auffüllen , um beliebige RegEx-Gruppen zu extrahieren und den vollständigen Ausdruck sowie beliebige gefundene RegEx-Gruppen in die Liste „Gruppen“ zu übernehmen. |
| Gruppen | <p>Diese Spalte zeigt die regulären Ausdrücke für die ausgewählte Gruppe „Reguläre Ausdrücke“ an.</p> <p>Wenn Sie beispielsweise den RegEx-Ausdruck „Date“ auswählen, wird der folgende Ausdruck im Textfeld angezeigt: <code>(1[012]{1,2} 0?[1-9])[-./]([12][0-9] 3[01]{1,2} 0?[1-9])[-./](((0-9){4}))</code>. Dieser RegEx-Ausdruck besteht aus drei Teilen, und der vollständige Ausdruck und jeder der drei Teile kann an ein anderes Ausgabefeld gesendet werden. Der vollständige Ausdruck wird im Quellenfeld gesucht. Wird eine Übereinstimmung im Quellenfeld gefunden, werden die zugehörigen Teile in das zugewiesene Ausgabefeld verschoben. Beispiel: Das Quellenfeld ist „On 12/14/2006“, Sie wenden den Ausdruck „Date“ darauf an und legen fest, dass das vollständige Datum (d. h. „12/14/2006“) in das Feld DATE, „12“ in das Feld MONTH, „14“ in das Feld DAY und „2006“ in das Feld YEAR übernommen werden sollen. In diesem Fall wird nach dem Datum gesucht. Wird es gefunden, werden die entsprechenden Informationen in das entsprechende Ausgabefeld übernommen.</p> <p>Quellenfeld: „On 12/14/2006“ DATE: „12/14/2006“ MONTH: „12“ DAY: „14“ YEAR: „2006“</p> |
| Ausgabefeld | Dropdown-Menü zum Auswählen eines Ausgabefeldes. |

Konfigurieren von Optionen zur Laufzeit

„Advanced Transformer“-Regeln können konfiguriert und zur Laufzeit übergeben werden, wenn sie als Datenflussoptionen verfügbar gemacht werden. Das ermöglicht Ihnen, die vorhandene Konfiguration durch Zeichenfolgen im JSON-Format zu überschreiben. Sie können auch Schrittoptionen festlegen, wenn Sie den Auftrag über einen Prozessfluss oder über das Befehlszeilentool „Job Executor“ aufrufen.

Sie finden Schemas für „AdvancedTransformerRules“ im folgenden Ordner:

```
<Spectrum Location>\server\modules\jsonSchemas\advancedTransformer
```

So definieren Sie „Advanced Transformer“-Regeln zur Laufzeit:

1. Öffnen Sie im Enterprise Designer einen Datenfluss, der den „Advanced Transformer“-Schritt verwendet.
2. Speichern Sie diesen Datenfluss und machen Sie ihn verfügbar.
3. Öffnen Sie **Bearbeiten > Datenflussoptionen**.
4. Erweitern Sie in der Tabelle **Datenflussoptionen zu Schritten zuordnen** den Eintrag „Advanced Transformer“. Aktivieren Sie das Kästchen für „AdvancedTransformerRules“.
5. Optional: Ändern Sie den Namen der Optionen im Feld **Optionsbeschriftung**.
6. Klicken Sie zweimal auf **OK**.

Ausgabe

Advanced Transformer erstellt keine neuen Ausgabefelder. Nur die Felder, die Sie definieren, werden in die Ausgabe geschrieben.

Open Parser

Open Parser analysiert Ihre Eingabedaten aus vielen Kulturen der Welt mit einer einfachen, aber leistungsstarken Parsing-Grammatik. Mithilfe dieser Grammatik können Sie eine Sequenz von Ausdrücken definieren, die Domänenmuster zum Parsen Ihrer Eingabedaten darstellen. Open Parser sammelt auch statistische Daten und bewertet die Parsing-Vergleiche, um Ihnen bei der Bestimmung der Effektivität Ihrer Parsing-Grammatiken zu helfen.

Verwenden Sie Open Name Parser, um:

- Eingabedaten mit domänenspezifischen und kulturspezifischen Parsing-Grammatiken zu parsen, die Sie im Domain Editor definieren.
- Eingabedaten mithilfe von domänenunabhängigen Parsing-Grammatiken zu parsen, die Sie in Open Parser mit derselben einfachen, aber leistungsstarken Parsing-Grammatik im Domain Editor definieren.
- Eingabedaten mithilfe von domänenunabhängigen Parsing-Grammatiken zur Laufzeit zu parsen, die Sie in Datenflussoptionen definieren.

- Eine Vorschau von Parsing-Grammatiken anzuzeigen und zu testen, wie Beispielingabedaten geparkt werden, bevor der Auftrag mit der Zieldatendatei ausgeführt wird.
- Parsing-Grammatikergenergebnisse zu verfolgen und zu sehen, inwieweit Tokens mit den von Ihnen definierten Ausdrücken übereinstimmen oder nicht übereinstimmen, und um den entsprechenden Prozess besser verstehen.

Eingabe

Open Parser akzeptiert die Eingabefelder, die Sie in Ihrer Parser-Grammatik definieren. Weitere Informationen finden Sie unter [Befehle des Kopfzeilenbereichs](#) auf Seite 30.

Wenn Sie kulturspezifisches Parsing durchführen, können Sie optional ein CultureCode-Feld in die Eingabedaten einfügen, um eine Grammatik einer bestimmten Kultur für einen Datensatz zu verwenden. Wenn Sie das CultureCode-Feld weglassen oder wenn es leer ist, wird jede im „Open Parser“-Schritt aufgeführte Kultur in der angegebenen Reihenfolge angewendet. Es wird das Ergebnis aus der Kultur mit der höchsten Parser-Punktzahl oder die erste Kultur mit einer Punktzahl von 100 zurückgegeben. Weitere Informationen zum CultureCode-Feld finden Sie unter [Zuweisen einer Parsing-Kultur zu einem Datensatz](#) auf Seite 12.

Optionen

In den folgenden Tabellen sind die Optionen für den „Open Parser“-Schritt aufgelistet.

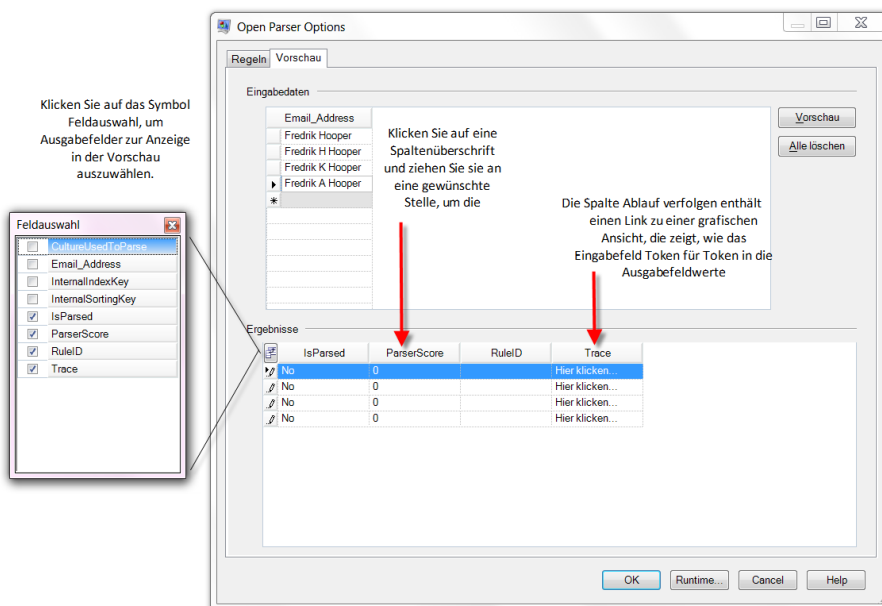
Registerkarte „Regeln“

| Option | Beschreibung |
|---|--|
| Grammatik von kulturspezifischer Domäne verwenden | <p>Gibt an, dass eine sprachen- und domänenspezifische Parsing-Grammatik verwendet wird, die bereits im Open Parser Domänen Editor im Enterprise Designer definiert wurde. Weitere Informationen zum Definieren von Domänen finden Sie unter Definieren einer kulturspezifischen Parsing-Grammatik auf Seite 10.</p> <p>Wenn Sie diese Option auswählen, werden auch folgende Optionen angezeigt:</p> <p>Domäne Gibt die zu verwendende Parsing-Grammatik an.</p> <p>Kulturen Gibt die Sprache oder Kultur der Daten an, die Sie parsen möchten. Klicken Sie auf die Schaltfläche Hinzufügen, um eine Kultur hinzuzufügen. Sie können die Reihenfolge ändern, in der Open Parser versucht, die Daten mit jeder einzelnen Kultur zu parsen, indem Sie die Schaltflächen Nach oben und Nach unten verwenden. Weitere Informationen über Kulturen finden Sie unter Definieren einer kulturspezifischen Parsing-Grammatik auf Seite 10.</p> <p>Mehrere geparte Datensätze zurückgeben Aktivieren Sie diese Option, damit Open Parser Datensätze für jede Kultur zurückgibt, die erfolgreich die Eingabe parst. Wenn Sie dieses Kästchen nicht aktivieren, wird Open Parser die Ergebnisse für den ersten Datensatz zurückgeben, der unabhängig von der Kultur eine Parser-Punktzahl von 100 erreicht. Wenn alle Kulturen ohne einen Datensatz mit einer Parser-Punktzahl von 100 ausgeführt werden, gibt Open Parser den Datensatz mit der Punktzahl zurück, die am nächsten an 100 liegt. Wenn mehrere Kulturen Datensätze mit der gleichen Punktzahl unter 100 zurückgeben, bestimmt die in Schritt 4 festgelegte Reihenfolge, welcher Kulturdatensatz zurückgegeben wird.</p> |
| Domänenunabhängige Grammatik definieren | <p>Aktivieren diese Option, wenn Sie eine Parsing-Grammatik definieren möchten, die ohne Berücksichtigung der Sprache oder Domäne der Eingabedaten angewendet werden soll. Wenn Sie diese Option aktivieren, wird der Grammatik-Editor angezeigt und Sie können die Parsing-Grammatik direkt im „Open Parser“-Schritt definieren, anstatt den Open Parser Domänen Editor im Enterprise Designer zu verwenden.</p> <p>Anmerkung: Sie können auch zur Laufzeit domänenunabhängige Grammatiken definieren. Klicken Sie hier für weitere Informationen.</p> |

Registerkarte „Vorschau“

Das Erstellen einer funktionierenden Parsing-Grammatik ist ein iterativer Prozess. Eine Vorschau ist beim Testen von Variationen in Ihrer Eingabe hilfreich, damit die Parsing-Grammatik die erwarteten Ergebnisse erzeugt.

Geben Sie Testwerte in das Eingabefeld ein, und klicken Sie dann auf **Vorschau**.



Die geparsen Ausgabefelder werden im Raster **Ergebnisse** angezeigt. Weitere Informationen über Ausgabefelder finden Sie unter [Ausgabe](#) auf Seite 310. Weitere Informationen zur Ablaufverfolgung finden Sie unter [Ablaufverfolgung von endgültigen Parsing-Ergebnissen](#) auf Seite 37. Wenn die Ergebnisse nicht Ihren Erwartungen entsprechen, klicken Sie auf die Registerkarte **Regeln** und fahren Sie mit der Bearbeitung der Parsing-Grammatik und dem Testen der Eingabedaten fort, bis die erwarteten Ergebnisse vorliegen.

Ausgabe

Tabelle 24: Open Parser-Ausgabe

| Feldname | Beschreibung/gültige Werte |
|--------------------|--|
| <Input Field> | Das ursprüngliche Eingabefeld, das in der Parsing-Grammatik definiert ist. |
| <Output Fields...> | Die in der Parsing-Grammatik definierten Ausgabefelder. |
| CultureCode | Die Kulturcodes, die in den Eingabedaten enthalten sind. Eine vollständige Liste der unterstützten Kulturcodes finden Sie unter Zuweisen einer Parsing-Kultur zu einem Datensatz auf Seite 12. |

| Feldname | Beschreibung/gültige Werte |
|---|---|
| CultureUsedtoParseSelect a match results in the Match Results List and then click Remove . | Der Kulturcodewert, der zur Analyse jedes Ausgabedatensatzes verwendet wird. Dieser Wert basiert auf Vergleichen mit einer kulturspezifischen Parsing-Grammatik. |
| IsParsed | Gibt an, ob ein Ausgabedatensatz geparkt wurde. Möglich sind die Werte „Yes“ oder „No“. |
| ParserScoreSelect a match results in the Match Results List and then click Remove . | Gibt die durchschnittliche Punktzahl an. Der Wert „ParserScore“ liegt zwischen 0 und 100 wie in der Parsing-Grammatik definiert. 0 wird zurückgegeben, wenn keine Übereinstimmungen zurückgegeben werden. Weitere Informationen finden Sie unter Bewertung . |
| Trace | Klicken Sie auf dieses Steuerelement, um eine grafische Ansicht zu sehen, wie jeder Token in der Parsing-Grammatik zu einem Ausgabefeld für die ausgewählte Zeile im Raster „Ergebnisse“ geparkt wurde. |

Table Lookup

Der „Table Lookup“-Schritt standardisiert Begriffe mit Bezug auf eine zuvor überprüfte Form dieses Begriffs und wendet die Standardversion an. Diese Auswertung erfolgt durch das Durchsuchen einer Tabelle nach dem zu standardisierenden Begriff.

Beispiel:

| | Vorname | Nachname |
|--------------------------|---------|----------|
| Quelleneingabe: | Bill | Smith |
| Standardisierte Ausgabe: | William | Smith |

Es gibt drei Typen von Aktionen, die Sie ausführen können: Standardisieren, Identifizieren und Kategorisieren.

Wenn der Begriff beim Durchführen der Standardisierungsaktion gefunden wird, werden beim Table Lookup entweder das gesamte Feld oder einzelne Begriffe innerhalb des Feldes mit dem standardisierten Begriff ersetzt, auch wenn das Feld mehrere Wörter enthält. Das Table Lookup kann das Ändern vollständiger Worte in Abkürzungen, das Ändern von Abkürzungen in vollständige Worte, das Ändern von Spitznamen in vollständige Namen oder das Ändern von Rechtschreibfehlern in die korrigierte Schreibweise umfassen.

Wenn der Begriff beim Durchführen der Identifizierungsaktion gefunden wird, wird beim Table Lookup der Datensatz markiert, der einen standardisierbaren Begriff enthält, es wird aber keine Aktion ausgeführt.

Wenn der Begriff beim Durchführen der Kategorisierungsaktion gefunden wird, wird beim Table Lookup der Quellwert als Schlüssel verwendet und der entsprechende Wert aus dem Tabelleneintrag in das ausgewählte Feld kopiert. Wenn keiner der Quellbegriffe übereinstimmt, wird bei **Kategorisieren** der angegebene Standardwert verwendet.

Eingabe

Tabelle 25: „Table Lookup“-Eingabefelder

| Feldname | Beschreibung/gültige Werte |
|----------------------|---|
| Source | Gibt das Quelleingabefeld an, das zum Suchen und Aufteilen ausgewertet werden soll. |
| StandardizationTable | Eine der in „ Table Lookup “-Tabellen auf Seite 167 aufgeführten Tabellen. |

Optionen

„TableLookup“-Optionen können auf Schrittebene über einen der Spectrum™ Technology Platform-Clients oder zur Laufzeit mit Datenflussoptionen konfiguriert werden.

Konfigurieren von Optionen

Um die Optionen für das Table Lookup anzugeben, müssen Sie eine Regel erstellen. Sie können mehrere Regeln erstellen und dann die Reihenfolge angeben, in der die Regeln angewendet werden sollen. Um eine Regel zu erstellen, öffnen Sie den „Table Lookup“-Schritt und klicken Sie auf **Hinzufügen**. Füllen Sie anschließend die folgenden Felder aus.

Anmerkung: Wenn Sie mehrere „Table Lookup“-Regeln hinzufügen, können Sie die Schaltflächen **Nach oben** und **Nach unten** verwenden, um die Reihenfolge zu ändern, in der die Regeln angewendet werden.

| Option | Beschreibung |
|--------|---|
| Aktion | <p data-bbox="535 367 1429 409">Gibt den Typ der Aktion an, die das Quellfeld übernehmen soll. Zur Auswahl stehen:</p> <ul data-bbox="535 420 1429 957" style="list-style-type: none"><li data-bbox="535 420 1429 619">Standardisieren Ändert die Daten in einem Feld entsprechend dem standardisierten Begriff, der in der Nachschlagetabelle gefunden wurde. Wenn das Feld mehrere Begriffe enthält, werden nur die Begriffe, die in der Nachschlagetabelle gefunden wurden, durch den standardisierten Begriff ersetzt. Die anderen Daten im Feld werden nicht geändert.<li data-bbox="535 630 1429 829">Identifizieren Markiert den Datensatz mit einem Begriff, der standardisiert werden kann, führt aber keine Aktion für die Daten im Feld aus. Das Ausgabefeld StandardizedTermIdentified wird zu dem Datensatz mit dem Wert „Yes“ hinzugefügt, wenn das Feld standardisiert werden kann, und zu dem Datensatz mit dem Wert „No“, wenn dies nicht möglich ist.<li data-bbox="535 840 1429 957">Kategorisieren Verwendet den Quellwert als Schlüssel und kopiert den entsprechenden Wert aus der Tabelle in das in der Zielliste ausgewählte Feld. Dadurch wird ein neues Feld in Ihren Daten erstellt, mit dem Datensätze kategorisiert werden können. |

| Option | Beschreibung |
|--------|---|
| Am | <p data-bbox="535 367 1425 472">Gibt an, ob das gesamte Feld als Suchbegriff verwendet werden soll oder ob die Nachschlagetabelle für jeden Begriff im Feld durchsucht werden soll. Zur Auswahl stehen:</p> <p data-bbox="535 483 1425 546">Vollständiges Feld Behandelt das gesamte Feld als einen Begriff, was zu Folgendem führt:</p> <ul data-bbox="714 556 1425 955" style="list-style-type: none"> • Wenn Sie die Aktion Standardisieren auswählen, wird das gesamte Feld beim Table Lookup als eine Zeichenfolge und versucht, das Feld mit der Zeichenfolge als Ganzes zu standardisieren. „International Business Machines“ würde beispielsweise in „IBM“ geändert. • Wenn Sie die Aktion Identifizieren auswählen, wird das gesamte Feld beim Table Lookup als eine Zeichenfolge behandelt und der Datensatz markiert, wenn die Zeichenfolge als Ganzes standardisiert werden kann. • Wenn Sie die Aktion Kategorisieren auswählen, wird das gesamte Feld beim Table Lookup als eine Zeichenfolge behandelt und der Datensatz markiert, wenn die Zeichenfolge als Ganzes kategorisiert werden kann. <p data-bbox="535 976 1425 1039">Einzelne Begriffe im Feld Behandelt jedes Wort im Feld als eigenen zugehörigen Begriff, was zu Folgendem führt:</p> <ul data-bbox="714 1060 1425 1596" style="list-style-type: none"> • Wenn Sie die Aktion Standardisieren auswählen, wird das Feld beim Table Lookup geparkt und versucht, die einzelnen Begriffe innerhalb des Feldes zu standardisieren. Zum Beispiel würde „Bill Mike Smith“ in „William Michael Smith“ geändert. • Wenn Sie die Aktion Identifizieren auswählen, wird das Feld beim Table Lookup geparkt und der Datensatz markiert, wenn ein einzelner Begriff innerhalb des Feldes standardisiert werden kann. • Wenn Sie die Aktion Kategorisieren auswählen, Im Gegensatz zu Standardisieren wird bei Kategorisieren nicht der Quellbegriff kopiert, wenn keine Tabellenübereinstimmung vorhanden ist. Wenn keiner der Quellbegriffe übereinstimmt, wird bei Kategorisieren der angegebene Standardwert verwendet. Im Gegensatz zu Standardisieren wird bei Kategorisieren nur dieser Tabellenwert zurückgegeben und nichts von Quelle. Wenn keiner der Quellbegriffe übereinstimmt, wird bei Kategorisieren der angegebene Standardwert verwendet. |
| Quelle | Gibt das Feld an, das den Begriff enthalten soll, den Sie suchen möchten. |

| Option | Beschreibung | | | | |
|---|---|------------------|---|---------------|--|
| Ziel | <p>Gibt das Feld an, in das die beim Table Lookup zurückgegebenen Begriffe geschrieben werden sollen.</p> <p>Wenn Sie den Wert ersetzen möchten, geben Sie im Feld Ziel das gleiche Feld wie im Feld Quelle an. Sie können auch ein neues Feld erstellen, indem Sie den Namen des Feldes eingeben, das Sie erstellen möchten.</p> <p>Das Feld Ziel ist nicht verfügbar, wenn Sie die Aktion Identifizieren auswählen.</p> | | | | |
| Tabelle | <p>Gibt die Tabelle an, die Sie für die Suche nach Begriffen verwenden möchten, die mit den Daten in Ihrem Datenfluss übereinstimmen.</p> <p>Eine Liste der Tabellen, die Sie bearbeiten können, finden Sie unter „Table Lookup“-Tabellen auf Seite 167. Weitere Informationen zum Erstellen oder Ändern von Tabellen finden Sie unter Einführung in Suchtabellen auf Seite 164.</p> | | | | |
| Aus mehreren Wörtern bestehende Begriffe suchen | <p>Ermöglicht Mehrwortsuchvorgänge innerhalb einer angegebenen Zeichenfolge. Beispiel:</p> <p>Eingabezeichenfolge: „Generalmajor John Smith“ Geschäftsregel: Identifizieren Sie „Generalmajor“ in einer Zeichenfolge basierend auf einer Tabelle, die den Eintrag enthält Ausgabe: Ersetzen Sie „Generalmajor“ durch „Gen. Maj.“</p> <p>Bei der Mehrwortsuche stoppt die Suche beim ersten Vorkommen einer Übereinstimmung.</p> <p>Diese Option ist deaktiviert, wenn Ein auf Vollständiges Feld festgelegt ist.</p> <p>Anmerkung: Die Auswahl dieser Option kann die Leistung beeinträchtigen.</p> | | | | |
| Wenn Tabellenwert nicht gefunden wird, Zielwert festlegen auf | <p>Gibt den Wert an, der in das Zielfeld eingefügt werden soll, wenn ein übereinstimmender Begriff in der Nachschlagetabelle nicht gefunden werden kann. Zur Auswahl stehen:</p> <table border="0"> <tr> <td>Quellwert</td> <td>Geben Sie den Wert aus dem Quellfeld in das Zielfeld ein.</td> </tr> <tr> <td>Andere</td> <td>Geben Sie einen bestimmten Wert in das Zielfeld ein.</td> </tr> </table> | Quellwert | Geben Sie den Wert aus dem Quellfeld in das Zielfeld ein. | Andere | Geben Sie einen bestimmten Wert in das Zielfeld ein. |
| Quellwert | Geben Sie den Wert aus dem Quellfeld in das Zielfeld ein. | | | | |
| Andere | Geben Sie einen bestimmten Wert in das Zielfeld ein. | | | | |

Konfigurieren von Optionen zur Laufzeit

„Table Lookup“-Optionen können zur Laufzeit konfiguriert und weitergeleitet werden, wenn sie als Datenflussoptionen verfügbar gemacht werden. Das ermöglicht Ihnen, die vorhandene Konfiguration durch Zeichenfolgen im JSON-Format zu überschreiben. Sie können auch Schrittoptionen festlegen, wenn Sie den Auftrag über einen Prozessfluss oder über das Befehlszeilentool „Job Executor“ aufrufen.

Sie finden im folgenden Ordner ein Schema für „LookupRule“:

```
<Spectrum Location>\server\modules\jsonSchemas\tableLookup
```

So definieren Sie „Table Lookup“-Regeln zur Laufzeit:

1. Öffnen Sie im Enterprise Designer einen Datenfluss, der den „Table Lookup“-Schritt verwendet.
2. Speichern Sie diesen Datenfluss und machen Sie ihn verfügbar.
3. Öffnen Sie `Edit > Dataflow Options`.
4. Erweitern Sie in der Tabelle **Datenflussoptionen zu Schritten zuordnen** den Eintrag „Table Lookup“. Aktivieren Sie das Kästchen für „LookupRule“.
5. Optional: Ändern Sie den Namen der Optionen im Feld **Optionsbeschriftung**.
6. Klicken Sie zweimal auf **OK**.

Ausgabe

Tabelle 26: Table Lookup-Ausgaben

| Feldname | Beschreibung/gültige Werte |
|----------------------------|---|
| StandardizedTermIdentified | Gibt an, ob das Feld einen Begriff enthält, der standardisiert werden kann. Die Ausgabe erfolgt nur, wenn Sie die Option Vollständiges Feld oder Einzelne Begriffe im Feld auswählen. Yes Der Datensatz enthält einen Begriff, der standardisiert werden kann. No Der Datensatz enthält keinen Begriff, der standardisiert werden kann. |

Transliteratör

Transliteratör konvertiert eine Zeichenfolge zwischen lateinischen und anderen Schriften. Beispiel:

| Quelle | Transkription |
|-----------------------|---------------|
| キャンパス | kyanpasu |
| Αλφαβητικός Κατάλογος | |

| Quelle | Transkription |
|--------|---------------|
|--------|---------------|

| | |
|---------------|-----------------|
| биологическом | biologichyeskom |
|---------------|-----------------|

Es ist wichtig, zu verstehen, dass die Transkription keine Übersetzung ist. Die Transkription ist die Konvertierung von Buchstaben einer Schrift in die einer anderen Schrift, ohne die zugrundeliegenden Wörter zu übersetzen.

Anmerkung: Standardmäßige Transkriptionsmethoden befolgen oft nicht die Ausspracheregeln einer bestimmten Sprache in der Zielschrift.

Der Transliteratorschritt unterstützt folgende Schriften. Im Allgemeinen folgt der Transliteratorschritt den Richtlinien der UNGEGN-Arbeitsgruppe für Umschriftsysteme (Working Group on Romanization Systems). Weitere Informationen finden Sie unter www.eki.ee/wgrs.

| | |
|-------------------|--|
| Arabisch | Diese Schrift wird von mehreren asiatischen und afrikanischen Sprachen verwendet einschließlich Arabisch, Persisch und Urdu. |
| Kyrillisch | Diese Schrift wird von osteuropäischen und asiatischen Sprachen verwendet einschließlich slawischen Sprachen wie Russisch. Der Transliteratorschritt folgt im Allgemeinen der ISO 9 für den kyrillischen Basissatz. |
| Devanagari | Die Schrift, die von verschiedenen indischen Sprachen verwendet wird, einschließlich Hindi und Sanskrit. Diese Schrift ist ein Abkömmling der Brahmi-Schrift, die eines der ältesten Schreibsysteme im alten Indien und in Süd- und Zentralasien ist. |
| Griechisch | Diese Schrift wird von der griechischen Sprache verwendet. Diese Schrift gehört zum hellenischen Zweig der indoeuropäischen Sprachfamilie. |
| Gujarati | Das vom Bundesstaat Gujarat im Westen Indiens verwendete Skript. Es ist eine der modernen Schriften Indiens, die von der Devanagari-Schrift übernommen wurde. |
| Gurmukhi | Die Schrift der indischen Sprache Punjabi. Diese Schrift weist einen beträchtlichen Einfluss durch die Nagari-Schrift auf, die eine frühere Form der Devanagari-Schrift ist. |
| Hangul | Diese Schrift wird von der koreanischen Sprache verwendet. Der Transliteratorschritt folgt den Transkriptionsvorschriften des koreanischen Ministeriums für Kultur und Tourismus. Weitere Informationen finden Sie auf der Webseite des Nationalen Instituts für die koreanische Sprache . |
| Han | Diese Schrift wird von der chinesischen Sprache verwendet. Sie ist ein Zweig der tibetisch-burmesischen Sprachfamilie und wurde mit auf Thailändisch und Chinesisch basierenden Schriften geschrieben. |

| | |
|--|--|
| Traditionelles/vereinfachtes Chinesisch | Der Transliterator-Schritt unterstützt sowohl traditionelles als auch vereinfachtes Chinesisch. Hier ein Beispiel für traditionelles Chinesisch: 人人生而自由. Und hier für vereinfachtes Chinesisch: 人人人生而自由 |
| Kannada | Die Schrift, die von verschiedenen südindischen Sprachen wie Konkani verwendet wird. Diese Schrift ist ein Abkömmling der Brahmi-Schrift des alten Indiens. |
| Katakana und Hiragana | Dies ist eine der verschiedenen Schriften, in denen Japanisch geschrieben werden kann. Der Transliterator-Schritt verwendet ein leicht abgewandeltes Hepburn-System. Beim Hepburn-System werden sowohl „ZI“ (ジ) als auch „DI“ (ヂ) durch „ji“ und sowohl „ZU“ (ズ) als auch „DU“ (ヅ) durch „zu“ dargestellt. Dies wurde leicht abgeändert zu „dji“ für „DI“ und „dzu“ für „DU“, um Reversibilität zu erreichen. Die Katakana-Transkription ist umkehrbar. Die Hiragana-Katakana-Transkription ist nicht vollständig umkehrbar, da es mehrere Katakana-Buchstaben ohne entsprechende Hiragana-Äquivalente gibt. Außerdem wird die Längenmarkierung bei Hiragana nicht verwendet. Die Hiragana-Lateinisch-Transkription ist ebenfalls nicht umkehrbar, da sie intern eine Kombination aus Katakana-Hiragana und Hiragana-Lateinisch ist. |
| Halbe Breite/volle Breite | Der Transliterator-Schritt kann zwischen schmalen Halbbreitenschriften und breiteren Vollbreitenschriften konvertieren. Hier ein Beispiel für halbe Breite: アルアノリウ. Hier das gleiche Beispiel in voller Breite: アルアノリウ. |
| Latin | Diese Schrift wird von den meisten europäischen Sprachen wie Englisch verwendet. Sie wurde ursprünglich von den alten Römern verwendet, um die lateinische Sprache zu schreiben. |
| Malayalam | Die Schrift der Malayalam-Sprache, der Amtssprache des indischen Bundesstaates Kerala. Diese Schrift wurde zuerst mit dem Vattelutu-Alphabet geschrieben. „Vattelutu“ bedeutet „rundes Schreiben“. Diese Schrift ist aus der Brahmi-Schrift des alten Indiens hervorgegangen. |
| Oriya | Die Schrift der Oriya-Sprache, der Amtssprache des indischen Bundesstaates Odisha. Die Oriya-Schrift ist aus der Kalinga-Schrift hervorgegangen, einem der vielen Abkömmlinge der Brahmi-Schrift des alten Indiens. |
| Tamil | Die von der tamilischen Sprache in mehreren Bundesstaaten Indiens, Sri Lankas und Malaysias verwendete Schrift. Diese Schrift wurde ursprünglich mit einer Version der Brahmi-Schrift geschrieben, die als tamilisches Brahmi bekannt ist. |

| | |
|---------------|--|
| Telugu | Die Schrift, die von mehreren Sprachen Südindiens verwendet wird. Diese Schrift ist ein Abkömmling der Brahmi-Schrift des alten Indiens. |
| Thai | Diese Schrift wird von der thailändischen Sprache verwendet. Diese Schrift ist von der Brahmi-Schrift des alten Indien und den Khmer-Alphabeten beeinflusst. |

Transliteratordaten sind Bestandteil des Data Normalization-Moduls. Eine Auflistung anderer Schritte finden Sie unter [Data Normalization-Modul](#) auf Seite 302.

Transkriptionskonzepte

Es gibt eine Reihe allgemeiner erwünschter Eigenschaften bei der Transkription von Schriften. Eine gute Transkription sollte über folgende Eigenschaften verfügen:

- Abgeschlossen
- vorhersagbar
- aussprechbar
- eindeutig

Diese Eigenschaften werden selten gleichzeitig erfüllt. Der Transliteratordaten-Schritt versucht daher, eine gute Balance zu finden.

Abgeschlossen

Jede wohlgeformte Zeichenfolge der Quellschrift sollte in eine Zeichenfolge der Zielschrift transkribiert werden.

vorhersagbar

Die Buchstaben selbst sollten (ohne Kenntnis der Sprachen, die in dieser Schrift geschrieben werden), basierend auf einer relativ kleinen Anzahl von Regeln, für die Transkription ausreichend sein. So kann die Transkription mechanisch durchgeführt werden.

aussprechbar

Die Transkription ist weniger nützlich, wenn der Prozess die Zeichen einfach zuordnet, ohne ihre Aussprache zu beachten. Das einfache Mapping von „αβγδεζηθ...“ mit „abcdefgh...“ würde Zeichenfolgen ergeben, die vielleicht vollständig und eindeutig sind, aber nicht ausgesprochen werden können.

Standardmäßige Transkriptionsmethoden befolgen oft nicht die Ausspracheregeln einer bestimmten Sprache in der Zielschrift. Das japanische Hepburn-System verwendet beispielsweise ein „j“ mit dem englischen Lautwert (im Gegensatz zum deutschen, französischen oder spanischen Lautwert), jedoch auch Vokale, die nicht den englischen Standardklang aufweisen. Eine Transkriptionsmethode könnte auch besondere Kenntnisse erfordern, um die korrekte Aussprache zu erreichen. Beispielsweise wird im japanischen Kunrei-System „tu“ als „tsu“ ausgesprochen. Ähnlich verhält es sich in Situationen, in denen unterschiedliche Sprachen dieselbe Schrift verwenden. Ein sachkundiger Leser wird beispielsweise beim Wort „Vase“ als deutsches Wort das „V“ wie ein „W“ aussprechen.

In einigen Fällen kann die Transkription stark von Traditionen beeinflusst sein. Beispielsweise klingt der Buchstabe „beta“ (β) des modernen Griechisch wie ein „w“, doch könnte eine Transformation weiterhin ein „b“ wie in „Biologie“ verwenden. In diesem Fall müsste der Benutzer wissen, dass ein „b“ im transkribierten Wort „beta“ (β) entspricht und in modernem Griechisch als „w“ ausgesprochen werden muss. Buchstaben können auch je nach Kontext unterschiedlich transkribiert werden, um die Aussprache vorhersagbarer zu gestalten. Da zum Beispiel die griechische Folge „GAMMA GAMMA“ ($\gamma\gamma$) als „ng“ ausgesprochen wird, kann das erste „GAMMA“ als „n“ transkribiert werden.

Anmerkung: Im Allgemeinen erzeugt englischer Text keine phonetischen Ergebnisse, um bei der Transkription lateinischer Schrift in andere Schriften vorhersagbare Ergebnisse zu liefern. Der Grund liegt darin, dass die englische Aussprache nicht einfach aus den Buchstaben eines Worts vorhergesagt werden kann. Beispielsweise enden die Wörter „grove“, „move“ und „love“ alle mit „ove“, werden aber unterschiedlich ausgesprochen.

eindeutig

Es sollte immer möglich sein, den Text in der Quellschrift aus der Transkription in die Zielschrift wiederherzustellen. Beispielsweise sollte es möglich sein, von „Elláda“ wieder zurück zum ursprünglichen Wort Ελλάδα zu gelangen. Bei der Transkription können jedoch mehrere Zeichen Zweideutigkeiten erzeugen. Der griechische Buchstabe „PSI“ (ψ) wird beispielsweise „ps“ zugeordnet, doch könnte „ps“ auch aus der Folge „PI SIGMA“ ($\pi\sigma$) stammen, da „PI“ (π) zu „p“ und „SIGMA“ (σ) zu „s“ zugeordnet ist.

Um dem Problem der Zweideutigkeit Rechnung zu tragen, verwendet Transliteratör ein Apostroph, um Zeichenfolgen eindeutig zu machen. Der griechischen Zeichenfolge „PI SIGMA“ ($\pi\sigma$) wird bei diesem Verfahren „p's“ zugeordnet. Wenn im Japanischen eine zweideutige Folge in der Zielschrift nicht aus einem einzelnen Buchstaben resultiert, verwendet die Transformation ein Apostroph, um Eindeutigkeit zu schaffen. Durch dieses Verfahren wird zum Beispiel zwischen „man'ichi“ und „manichi“ unterschieden.

Anmerkung: Einige Zeichen einer Zielschrift sind normalerweise nicht außerhalb eines gewissen Kontexts zu finden. Das kleine japanische Zeichen „ya“ wie in „kya“ (キヤ) steht normalerweise nicht alleine. Zur Verarbeitung von solchen Zeichen verwendet Transliteratör eine Tilde. Beispielsweise würde die Eingabe „~ya“ ein isoliertes kleines „ya“ erzeugen. Bei der Transkription in Griechisch würde die Eingabe „a~s“ ein nicht am Ende stehendes griechisches „sigma“ ($\alpha\sigma$) am Ende eines Worts erzeugen. Ähnlich würde die Eingabe „~sa“ ein am Ende stehendes „sigma“ an einer mittleren Position ($\zeta\alpha$) erzeugen.

Bei der allgemeinen Schrifttransformation besteht eine gebräuchliche Technik zur Reversibilität darin, zusätzliche Akzente zu verwenden, um Buchstaben zu unterscheiden, die möglicherweise nicht anders unterschieden werden können. Im Folgenden finden Sie beispielsweise einen griechischen Text der voll reversibler lateinischer Schrift zugeordnet wurde:


Eingabe

| Feldname | Beschreibung | | | | | | | | | | |
|----------------------------------|---|---------------------|-------------------------|--------------------|---------------------------|---------------------|-----------------------|-----------------------|-------------------------|----------------------------|-------------------------------------|
| Ein beliebiges Zeichenfolgenfeld | Der Transliterator-Schritten kann alle Zeichenfolgenfelder transkribieren. Sie können in den Optionen des Transliterator-Schrittes angeben, welche Felder transkribiert werden sollen. | | | | | | | | | | |
| TransliteratorID | <p>Überschreibt die standardmäßige Transkription, wie sie in den Optionen des Transliterator-Schrittes angegeben sind. Verwenden Sie dieses Feld, wenn Sie für jeden Datensatz eine andere Transkription angeben möchten.</p> <p>Beispiel:</p> <table border="0"> <tr> <td>Arabic-Latin</td> <td>Von Arabisch zu Latein.</td> </tr> <tr> <td>Greek-Latin</td> <td>Von Griechisch zu Latein.</td> </tr> <tr> <td>Latin-Hangul</td> <td>Von Latein zu Hangul.</td> </tr> <tr> <td>Latin-Katakana</td> <td>Von Latein zu Katakana.</td> </tr> <tr> <td>Fullwidth-Halfwidth</td> <td>Von voller Breite zu halber Breite.</td> </tr> </table> | Arabic-Latin | Von Arabisch zu Latein. | Greek-Latin | Von Griechisch zu Latein. | Latin-Hangul | Von Latein zu Hangul. | Latin-Katakana | Von Latein zu Katakana. | Fullwidth-Halfwidth | Von voller Breite zu halber Breite. |
| Arabic-Latin | Von Arabisch zu Latein. | | | | | | | | | | |
| Greek-Latin | Von Griechisch zu Latein. | | | | | | | | | | |
| Latin-Hangul | Von Latein zu Hangul. | | | | | | | | | | |
| Latin-Katakana | Von Latein zu Katakana. | | | | | | | | | | |
| Fullwidth-Halfwidth | Von voller Breite zu halber Breite. | | | | | | | | | | |

Optionen

Tabelle 27: Optionen für Transliterator

| Option | Beschreibung/Gültige Werte |
|--------|--|
| Von | <p>Die Schrift, die in den Feldern verwendet wird, die Sie transkribieren möchten. Eine Beschreibung der unterstützten Schriften finden Sie unter Transliterator auf Seite 316.</p> <p>Anmerkung: Der Transliterator-Schritt unterstützt nicht die Transkription zwischen allen Schriften. Die Felder Von und Zu enthalten automatisch gültige Werte basierend auf Ihrer Auswahl.</p> |

| Option | Beschreibung/Gültige Werte |
|--|--|
| Zu | <p>Die Schrift, in die das Feld konvertiert werden soll. Eine Beschreibung der unterstützten Schriften finden Sie unter Transliterator auf Seite 316.</p> <p>Anmerkung: Der Transliterator-Schritt unterstützt nicht die Transkription zwischen allen Schriften. Die Felder Von und Zu enthalten automatisch gültige Werte basierend auf Ihrer Auswahl.</p> |
| Schaltfläche „Tauschen“  | <p>Klicken Sie auf die Schaltfläche „Tauschen“, um die Sprachen in den Feldern Von und Zu zu vertauschen.</p> |
| Zu transkribierende Felder | Gibt die Felder an, die transkribiert werden sollen. |
| Akzentzeichen entfernen | Aktivieren Sie das Kästchen, um Akzentzeichen aus dem Feld zu entfernen. Diese Option ist standardmäßig deaktiviert. |

Ausgabe

Der Transliterator-Schritt transkribiert die von Ihnen angegebenen Felder. Er produziert keine andere Ausgabe.

Universal Name-Modul

Universal Name-Modul

Für eine möglichst genaue Standardisierung müssen Sie gegebenenfalls Datenzeichenfolgen in mehrere Felder aufteilen. Spectrum™ Technology Platform bietet erweiterte Parsing-Features, mit denen Sie Personennamen, Unternehmensnamen und viele andere Begriffe und Abkürzungen parsen können. Zusätzlich können Sie eine eigene Liste mit benutzerdefinierten Begriffen erstellen, die als Grundlage für die Scan- und Extraktionsvorgänge verwendet werden sollen.

Komponenten

Das Universal Name-Modul setzt sich aus folgenden Komponenten zusammen:

- **Name Parser (veraltet):** Diese Komponente schlüsselt Personen- und Unternehmensnamen sowie andere Begriffe im Namensdatenfeld in ihre Komponententeile auf. Diese geparsten Namenselemente sind anschließend für andere automatisierte Vorgänge verfügbar, z. B. Namensvergleiche, die Namensstandardisierung oder die Namenskonsolidierung über mehrere Datensätze.

Achtung: Der „Name Parser“-Schritt ist veraltet und wird in zukünftigen Releases möglicherweise nicht unterstützt. Mit „Open Name Parser“ können Sie Namen parsen.

- **Name Variant Finder:** Diese Komponente funktioniert in einem Modus für Vornamen oder für Nachnamen zur Abfrage der Namensdatenbank und zur Rückgabe von Namensvarianten. „Name Variant Finder“ kann die Anzahl der Ergebnisse nach Geschlecht und Kultur begrenzen. Spectrum™ Technology Platform enthält eine Datei mit Basisnamen. Dateien mit Erweiterungsamen aus anderen Kulturen sind verfügbar und können durch Kopieren der entsprechenden JAR-Datei im Ordner `modules/tables/ext` bereitgestellt werden. Arabische Namen sind als Erweiterung verfügbar. Die Daten in dieser Datei stammen von Nomino, Inc.
- **Open Name Parser:** Der „Open Name Parser“ schlüsselt Personen- und Unternehmensnamen sowie andere Begriffe im Namensdatenfeld in ihre Komponententeile auf. Diese geparsten Namenselemente sind anschließend für andere automatisierte Vorgänge verfügbar, z. B. Namensvergleiche, die Namensstandardisierung oder die Namenskonsolidierung über mehrere Datensätze.

Name Parser (VERALTET)

Achtung: Der „Name Parser“-Schritt ist veraltet und wird in zukünftigen Releases möglicherweise nicht unterstützt. Mit „Open Name Parser“ können Sie Namen parsen.

„Name Parser“ schlüsselt Personen- und Unternehmensnamen sowie andere Begriffe im Namensdatenfeld in ihre Komponententeile auf. Der Parsing-Prozess umfasst eine Erläuterung von Funktion, Form und der syntaktischen Beziehung der einzelnen Komponenten zum Ganzen. Diese geparsten Namenselemente sind anschließend für andere automatisierte Vorgänge verfügbar, z. B. Namensvergleiche, die Namensstandardisierung oder die Namenskonsolidierung über mehrere Datensätze.

Beim Namensparsing werden folgende Aktionen ausgeführt:

- Es bestimmt den Entitätstyp, um die Funktion dieses Namens zu beschreiben. Namensentitätstypen lassen sich in zwei größere Gruppen aufteilen: Personennamen und Unternehmensnamen mit Untergruppen innerhalb dieser größeren Gruppen.
- Bestimmt die Form eines Namens, um herauszufinden, welche Syntax der Parser beim Parsen verwendet sollte. Personennamen liegen üblicherweise in normaler Reihenfolge (Unterschrift) oder in umgekehrter Reihenfolge vor. Unternehmensnamen sind normalerweise hierarchisch sortiert.
- Bestimmt die Komponenten eines Namens und beschriftet sie, um die syntaktische Beziehung jedes Namensteils zum gesamten Namen so zu identifizieren. Die Syntax von Personennamen umfasst neben anderen Komponenten von Personennamen Präfixe, Vornamen, zweiten Vornamen,

Nachnamen, Suffixe und Begriffe zur Kontobeschreibung. Die Syntax von Unternehmensnamen umfasst neben anderen Komponenten von Unternehmensnamen den primären Text, nicht signifikante Begriffe, Präpositionen und Suffixe.

- Bestimmt das Geschlecht des Namens. Das Geschlecht wird basierend auf von Ihnen angegebenen kulturellen Voraussetzungen bestimmt. Beispiel: Jean ist in Frankreich ein männlicher Name, aber in den Vereinigten Staaten ein weiblicher Name. Wenn Sie wissen, dass die Namen, die Sie gerade verarbeiten, aus Frankreich stammen, könnten Sie Französisch als Kultur für die Geschlechtsbestimmung angeben. Der „Name Parser“ verwendet für die Geschlechtsbestimmung die Tabellen „Vorname“ und „Zusammengesetzte Vornamen“. Wenn ein Name in einer der Tabellen nicht gefunden werden kann und der Name einen Titel enthält, prüft der Parser die Tabelle „Titel“ zur Geschlechtsbestimmung. Andernfalls wird das Geschlecht als unbekannt gekennzeichnet.

Anmerkung: Wenn ein Feld Ihres Eingabedatensatzes bereits eine der unterstützten Kulturen enthält, können Sie das Feld „GenderDeterminationSource“ in Ihrer Eingabe vordefinieren, um die Quelle für die Geschlechtsbestimmung in der grafischen Benutzeroberfläche außer Kraft zu setzen.

- Weist eine Parsing-Punktzahl zu, die angibt, dass der Zuverlässigkeitsgrad des Parsers korrekt ist.

Eingabe

Achtung: Der „Name Parser“-Schritt ist veraltet und wird in zukünftigen Releases möglicherweise nicht unterstützt. Mit „Open Name Parser“ können Sie Namen parsen.

Tabelle 28: Eingabe für „Name Parser“

| Feldname | Beschreibung/gültige Werte | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---------------------------|---|---------------|---------------------------|-----------------|-----------|----------------|--|---------------|-------------|---------------------|---|-----------------|---|--------------|---------------|------------------|---------|----------------|----------|-------------------|-----------|----------------|-----------|-----------------|----------|---------------|-----------|
| GenderDeterminationSource | <p>Die Kultur der Namensdaten für die Geschlechtsbestimmung. Standardmäßig werden kulturübergreifende Regeln verwendet. Beispiel: Jean ist häufig ein weiblicher Name und standardmäßig wird er auch als weiblicher Name angegeben. Wenn Sie Französisch auswählen, wird er jedoch als männlicher Name angegeben. Die Optionen werden nachfolgend zusammen mit Beispielländern für jede Kultur aufgeführt. Beachten Sie, dass die Liste der Länder unter den einzelnen Kulturen nicht vollständig ist.</p> <table border="0"> <tr> <td>SLAVIC</td> <td>Bosnien, Polen, Albanien.</td> </tr> <tr> <td>ARMENIAN</td> <td>Armenien.</td> </tr> <tr> <td>DEFAULT</td> <td>Bulgarien, Caymaninseln, Irland, Großbritannien, Vereinigte Staaten.</td> </tr> <tr> <td>FRENCH</td> <td>Frankreich.</td> </tr> <tr> <td>SCANDINAVIAN</td> <td>Dänemark, Finnland, Island, Norwegen, Schweden.</td> </tr> <tr> <td>GERMANIC</td> <td>Österreich, Deutschland, Luxemburg, Schweiz, die Niederlande.</td> </tr> <tr> <td>GREEK</td> <td>Griechenland.</td> </tr> <tr> <td>HUNGARIAN</td> <td>Ungarn.</td> </tr> <tr> <td>ITALIAN</td> <td>Italien.</td> </tr> <tr> <td>PORTUGUESE</td> <td>Portugal.</td> </tr> <tr> <td>ROMANIA</td> <td>Rumänien.</td> </tr> <tr> <td>HISPANIC</td> <td>Spanien.</td> </tr> <tr> <td>ARABIC</td> <td>Tunesien.</td> </tr> </table> <p>„Name Variant Finder“ verwendet das Feld „GenderDeterminationSource“ auch zur Begrenzung zurückgegebener Namensvarianten basierend auf der Kultur. Weitere Informationen finden Sie unter Name Variant Finder auf Seite 345.</p> | SLAVIC | Bosnien, Polen, Albanien. | ARMENIAN | Armenien. | DEFAULT | Bulgarien, Caymaninseln, Irland, Großbritannien, Vereinigte Staaten. | FRENCH | Frankreich. | SCANDINAVIAN | Dänemark, Finnland, Island, Norwegen, Schweden. | GERMANIC | Österreich, Deutschland, Luxemburg, Schweiz, die Niederlande. | GREEK | Griechenland. | HUNGARIAN | Ungarn. | ITALIAN | Italien. | PORTUGUESE | Portugal. | ROMANIA | Rumänien. | HISPANIC | Spanien. | ARABIC | Tunesien. |
| SLAVIC | Bosnien, Polen, Albanien. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ARMENIAN | Armenien. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| DEFAULT | Bulgarien, Caymaninseln, Irland, Großbritannien, Vereinigte Staaten. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| FRENCH | Frankreich. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| SCANDINAVIAN | Dänemark, Finnland, Island, Norwegen, Schweden. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| GERMANIC | Österreich, Deutschland, Luxemburg, Schweiz, die Niederlande. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| GREEK | Griechenland. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| HUNGARIAN | Ungarn. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ITALIAN | Italien. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| PORTUGUESE | Portugal. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ROMANIA | Rumänien. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| HISPANIC | Spanien. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ARABIC | Tunesien. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Name | Der zu parsende Name. Dieses Feld ist ein Pflichtfeld. | | | | | | | | | | | | | | | | | | | | | | | | | | |

Optionen

Achtung: Der „Name Parser“-Schritt ist veraltet und wird in zukünftigen Releases möglicherweise nicht unterstützt. Mit „Open Name Parser“ können Sie Namen parsen.

Doppelklicken Sie für die Angabe der „Name Parser“-Optionen auf die „Name Parser“-Instanz auf der Arbeitsfläche. Das Dialogfeld Name Parser-Optionen wird angezeigt.

Tabelle 29: „Name Parser“-Optionen

| Option | Beschreibung |
|--|---|
| Personennamen parsen | Aktivieren Sie dieses Kästchen zum Parsen von Personennamen. |
| Trennen Sie verbundene Namen in mehrere Datensätze. Wählen Sie aus der Liste „Übereinstimmungsergebnisse“ ein Übereinstimmungsergebnis aus und klicken Sie anschließend auf Entfernen . | <p>Klicken Sie in dieses Feld, um Namen von mehreren Personen in mehrere Datensätze zu trennen. Beispiel: Bill & Sally Smith.</p> <p>Wenn ein verbundener Datensatz zu zwei separaten Namensdatensätzen führt, wird das Ausgabefeld „Parser-Datensatz-ID“ generiert. Jedes Paar separater Namensdatensätze wird mit der gleichen Parser-Datensatz-ID identifiziert.</p> |
| Quelle für Geschlechtsbestimmung: Wählen Sie aus der Liste der Übereinstimmungsergebnisse ein Übereinstimmungsergebnis aus und klicken Sie anschließend auf Entfernen . | <p>Bestimmt, wie der „Name Parser“ dem Namen ein Geschlecht zuweist. In den meisten Fällen ist „Standard“ die beste Einstellung, da mit ihr ein breites Spektrum von Namen abgedeckt wird. Wenn Sie Namen aus einer bestimmten Kultur verarbeiten, wählen Sie diese Kultur aus. Die Auswahl einer bestimmten Kultur hilft bei der Sicherstellung, dass den Namen das richtige Geschlecht zugewiesen wird. Beispiel: Wenn Sie „Standard“ aktiviert lassen, wird der Name Jean als weiblicher Name identifiziert. Wenn Sie Französisch auswählen, wird der Name als männlicher Name identifiziert.</p> <p>Anmerkung: Wenn Sie eine Kultur auswählen, der Name in dieser Kultur jedoch nicht gefunden werden kann, wird das Geschlecht über „Standardkultur“ bestimmt, die Daten aus einer Vielzahl von Kulturen enthält.</p> |
| Sortierung | <p>Gibt an, wie die Namensfelder in Ihren Eingabedatensätzen sortiert sind. Zur Auswahl stehen:</p> <p>Normal Die Namensfelder sind nach Titel, Vorname, zweiter Vorname, Nachname und Suffix sortiert.</p> <p>Umgekehrt Die Namensfelder sind mit Nachnamen zuerst sortiert.</p> <p>Groß- und Kleinbuchstaben Die Namensfelder sind nach einer Kombination aus normaler und umgekehrter Reihenfolge sortiert.</p> |
| Punkte beibehalten | Behält die Interpunktion im Feld mit geparsten Personennamen bei. |
| Unternehmensnamen parsen | Aktivieren Sie dieses Kästchen, um Unternehmensnamen zu Parsen. |
| Punkte beibehalten | Aktivieren Sie dieses Kästchen, um die Interpunktion in das Feld mit geparsten Personennamen zurückzugeben. |

| Option | Beschreibung |
|----------------------------|---|
| Benutzerdefinierte Tabelle | Klicken Sie auf eine beliebige der benutzerdefinierten Tabellen, um in den verschiedenen Parser-Tabellen Werte zu den vorhandenen Werten hinzuzufügen. Diese Funktion ermöglicht es Ihnen, Tabellen für Ihre individuelle Geschäftsumgebung anzupassen. Klicken Sie auf „Konfigurieren“, um eine XML-Datei mit den von Ihnen hinzugefügten Werte auszuwählen. Weitere Informationen zu benutzerdefinierten Tabellen finden Sie unter Ändern benutzerdefinierter Tabellen von „Name Parser“ auf Seite 327. |

[Ändern benutzerdefinierter Tabellen von „Name Parser“](#)

Achtung: Der „Name Parser“-Schritt ist veraltet und wird in zukünftigen Releases möglicherweise nicht unterstützt. Mit „Open Name Parser“ können Sie Namen parsen.

Sie können Werte in den Tabellen von „Name Parser“ hinzufügen, ändern und löschen, um sie an Ihre individuelle Geschäftsumgebung anzupassen.

Bei den benutzerdefinierten Tabellen von „Name Parser“ handelt es sich um XML-Dateien, die sich standardmäßig im Ordner <Drive>:\Program Files\Pitney Bowes\Spectrum\server\modules\parser\data befinden. Spectrum™ Technology Platform enthält die folgenden benutzerdefinierten Tabellen:

[UserAccountDescriptions.xml](#)

Tabelle 30: Spalten in UserAccountDescriptions.xml

| Spaltenname | Beschreibung/gültige Werte |
|-------------|--|
| LookupValue | Ein Suchbegriff, der häufig in Kontobeschreibungen vorkommt. Beliebiges Einzelwort. Groß-/Kleinschreibung wird nicht beachtet. |

Beispieleintrag:

```
<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

        LookupValue
        ART
        AND
```

```

    </deleted-entry-group>
</deleted-entries>
<added-entries delimiter-character="|">

    LookupValue
    A/C
    ACCOUNT
    EXP

</added-entries>
</table-data>

```

UserCompanyPrepositions.xml

Tabelle 31: Spalten in UserCompanyPrepositions.xml

| Spaltenname | Beschreibung/gültige Werte |
|-------------|--|
| LookupValue | Beliebige Präposition (beispielsweise „of“ oder „on“), die häufig in Unternehmensnamen vorkommt. Beliebiges Einzelwort. Groß-/Kleinschreibung wird nicht beachtet. |

Beispielintrag:

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      LookupValue
      AROUND
      NEAR

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

    LookupValue
    ABOUT
    AFTER
    ACROSS

  </added-entries>
</table-data>

```


[UserCompanySuffixes.xml](#)**Tabelle 32: Spalten in UserCompanySuffixes.xml**

| Spaltenname | Beschreibung/gültige Werte |
|-------------|--|
| LookupValue | Beliebiges Suffix, das häufig in Unternehmensnamen vorkommt. Beispiele hierfür sind „Inc.“ und „Co.“ Beliebiges Einzelwort. Groß-/Kleinschreibung wird nicht beachtet. |

Beispieleintrag:

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      LookupValue
      SANDY
      CLUE

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

    LookupValue
    LTD
    LLC
    CO
    INC

  </added-entries>
</table-data>

```

[UserCompanyTerms.xml](#)**Tabelle 33: Spalten in UserCompanyTerms.xml**

| Spaltenname | Beschreibung/gültige Werte |
|-------------|---|
| LookupValue | Beliebiger Begriff, der häufig in Unternehmensnamen vorkommt. Beliebiges Einzelwort. Groß-/Kleinschreibung wird nicht beachtet. |

Beispielintrag:

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      LookupValue
      MARY
      BLUE

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

    LookupValue
    ARC
    ARCADE
    ASSEMBLY
    ARIZONA

  </added-entries>
</table-data>

```

UserCompoundFirstNames.xml

Diese Tabelle enthält benutzerdefinierte zusammengesetzte Vornamen. Zusammengesetzte Namen sind Namen, die aus zwei Wörtern bestehen.

Tabelle 34: Spalten in UserCompoundFirstNames.xml

| Spaltenname | Beschreibung/gültige Werte |
|-------------|---|
| FirstName | Der zusammengesetzte Vorname. Maximal zwei Wörter. Groß-/Kleinschreibung wird nicht beachtet. |
| Kultur | Die Kultur, in der diese Kombination aus Vorname/Geschlecht gültig ist. Sie können jeden Wert verwenden, der im Eingabefeld „GenderDeterminationSource“ gültig ist. Weitere Informationen finden Sie unter Eingabe auf Seite 324. |

| Spaltenname | Beschreibung/gültige Werte |
|-------------|----------------------------|
|-------------|----------------------------|

| | |
|------------|---|
| Geschlecht | <p>Das Geschlecht, das am häufigsten mit dieser Kombination aus Vorname/Kultur verbunden ist. Zur Auswahl stehen:</p> <p>M Bei dem Namen handelt es sich um einen männlichen Namen.</p> <p>F Bei dem Namen handelt es sich um einen weiblichen Namen.</p> <p>A Mehrdeutig. Der Name kann entweder männlich oder weiblich sein.</p> <p>U Unbekannt. Das Geschlecht dieses Namens ist nicht bekannt. Wenn Sie dieses Feld leer lassen, wird „Unbekannt“ angenommen.</p> |
|------------|---|

| | |
|------------|--|
| Häufigkeit | Wird in dieser Version nicht verwendet. Sie können diese Spalte leer lassen. |
|------------|--|

Beispieleintrag:

```
<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      FirstName
      ANN MARIE
      BILLY JOE

    </deleted-entry-group>
  </deleted-entry-group>

  FirstName|Frequency
  KAREN SUE|0.126
  BILLY JOE|0.421

</deleted-entry-group>
</deleted-entry-group>

  FirstName|Gender|Culture
  JEAN ANN|M|DEFAULT
  JEAN CLUADE|F|FRENCH

</deleted-entry-group>
</deleted-entries>
<added-entries delimiter-character="|">

  FirstName|Gender|Culture
  JOHN Henry|M|DEFAULT
  A'SHA A'MAR|F|ARABIC
  BILLY JO|A|DEFAULT
```

```
</added-entries>
</table-data>
```

UserConjunctions.xml

Diese Tabelle enthält eine Liste von benutzerdefinierten Konjunktionen wie „und“, „oder“ oder „&“.

Tabelle 35: Spalten in UserConjunctions.xml

| Spaltenname | Beschreibung/gültige Werte |
|-------------|---|
| LookupValue | Beliebige Konjunktion. Muss ein einzelnes Wort sein. Groß-/Kleinschreibung wird nicht beachtet. |

Beispielinträge:

```
<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      LookupValue
      FIND
      CARE
      %

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

    LookupValue
    &
    AND
    OR

  </added-entries>
</table-data>
```

*UserFirstNames.xml***Tabelle 36: Spalten in UserFirstNames.xml**

| Spaltenname | Beschreibung/gültige Werte |
|-------------|---|
| FirstName | Der Vorname, der durch diese Tabellenzeile beschrieben wird. Groß-/Kleinschreibung wird nicht beachtet. |
| Geschlecht | <p>Das Geschlecht, das am häufigsten mit dieser Kombination aus Vorname/Kultur verbunden ist. Zur Auswahl stehen:</p> <p>M Bei dem Namen handelt es sich um einen männlichen Namen.</p> <p>F Bei dem Namen handelt es sich um einen weiblichen Namen.</p> <p>A Mehrdeutig. Der Name kann entweder männlich oder weiblich sein.</p> <p>U Unbekannt. Das Geschlecht dieses Namens ist nicht bekannt. Wenn Sie dieses Feld leer lassen, wird „Unbekannt“ angenommen.</p> |
| Kultur | Die Kultur, in der diese Kombination aus Vorname/Geschlecht gültig ist. Sie können jeden Wert verwenden, der im Eingabefeld „GenderDeterminationSource“ gültig ist. Weitere Informationen finden Sie unter Eingabe auf Seite 324. |

Beispieleintrag:

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      FirstName
      AADEL
      AADIL

    </deleted-entry-group>
    <deleted-entry-group>

      FirstName
      A'SACE
      A'BOCKETT

    </deleted-entry-group>
  </deleted-entry-group>

  FirstName | Gender | Culture

```

```

                ALII|M|DEFAULT
                AISHA|F|ARABIC

        </deleted-entry-group>
</deleted-entry-group>

                FirstName|Gender
                JOHE|M

        </deleted-entry-group>
</deleted-entries>
<added-entries delimiter-character="|">

                FirstName|Gender|Culture
                JOHE|M|DEFAULT
                A'SHAN|F|ARABIC

        </added-entries>
</table-data>

```

UserGeneralSuffixes.xml

Diese Tabelle enthält eine Liste von benutzerdefinierten Suffixen, die in Personennamen verwendet werden, wie „Dr.“, „MD“ oder „PhD“. Generations-Suffixe sind ausgeschlossen.

Tabelle 37: Spalten in UserGeneralSuffixes.xml

| Spaltenname | Beschreibung/gültige Werte |
|-------------|--|
| LookupValue | Beliebiges Suffix, das oft in Personennamen vorkommt und kein Generations-Suffix ist. Muss ein einzelnes Wort sein. Groß-/Kleinschreibung wird nicht beachtet. |

Beispielintrag:

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

                LookupValue
                AND
                WILL
                TUNA

    </deleted-entry-group>
  </deleted-entries>

```

```

<added-entries delimiter-character="|">
    LookupValue
    ACCOUNTANT
    ATTORNEY
    ANALYST
    ASSISTANT
</added-entries>
</table-data>

```

UserLastNamePrefixes.xml

Diese Tabelle enthält eine Liste von benutzerdefinierten Präfixen, die in Nachnamen von Personen vorkommen, wie „Van“, „De“ oder „La“.

Tabelle 38: Spalten in UserLastNamePrefixes.xml

| Spaltenname | Beschreibung/gültige Werte |
|-------------|---|
| LookupValue | Beliebiges Präfix, das als Teil eines Nachnamens einer Person vorkommt. Beliebiges Einzelwort. Groß-/Kleinschreibung wird nicht beachtet. |

Beispieleintrag:

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>
      LookupValue
      DO
      RUN
      ANIMAL
    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">
    LookupValue
    D'
    DA
    DEN
    DEL
  </added-entries>

```

```
</table-data>
```

UserLastNames.xml

Tabelle 39: Spalten in UserLastNames.xml

| Spaltenname | Beschreibung/gültige Werte |
|-------------|---|
| LastName | Der Nachname, der durch diese Tabellenzeile beschrieben wird. Groß-/Kleinschreibung wird nicht beachtet. |
| Geschlecht | Das Geschlecht, das am häufigsten mit dieser Kombination aus Vorname/Kultur verbunden ist. Zur Auswahl stehen: <ul style="list-style-type: none"> M Bei dem Namen handelt es sich um einen männlichen Namen. F Bei dem Namen handelt es sich um einen weiblichen Namen. A Mehrdeutig. Der Name kann entweder männlich oder weiblich sein. U Unbekannt. Das Geschlecht dieses Namens ist nicht bekannt. Wenn Sie dieses Feld leer lassen, wird „Unbekannt“ angenommen. |
| Kultur | Die Kultur, in der diese Kombination aus Vorname/Geschlecht gültig ist. Sie können jeden Wert verwenden, der im Eingabefeld „GenderDeterminationSource“ gültig ist. Weitere Informationen finden Sie unter Eingabe auf Seite 324. |

Beispieleintrag:

```
<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      LastName
      Rusod
      AADIL

    </deleted-entry-group>
  </deleted-entry-group>

  LastName
  KAASEEY
  JOIEN
```



```

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

    LastName|Culture|Gender
    SMITH|ENGLISH|A
    WILSON|ENGLISH|A
    JONES|ENGLISH|A

  </added-entries>
</table-data>

```

UserMaturitySuffixes.xml

Diese Tabelle enthält benutzerdefinierte Generations-Suffixe, die in Personennamen vorkommen, wie „Jr.“ oder „Sr.“

Tabelle 40: Spalten in UserMaturitySuffixes.xml

| Spaltenname | Beschreibung/gültige Werte |
|-------------|--|
| LookupValue | Ein Generations-Suffix, das in Personennamen vorkommt. Beliebiges Einzelwort. Groß-/Kleinschreibung wird nicht beachtet. |

Beispieleintrag:

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      LookupValue
      I
      V
      18
      VI

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

    LookupValue
    I
    II
    III

  </added-entries>
</table-data>

```

```
</added-entries>
</table-data>
```

UserTitles.xml

Diese Tabelle enthält benutzerdefinierte Titel die in Personennamen vorkommen, wie „Herr“, „Frau“, „Mr.“ oder „Ms.“

Tabelle 41: Spalten in UserTitles.xml

| Spaltenname | Beschreibung/gültige Werte |
|-------------|--|
| LookupValue | Ein Titel, der in Personennamen vorkommt. Beliebiges Einzelwort. Groß-/Kleinschreibung wird nicht beachtet. |
| Geschlecht | Das Geschlecht, das am häufigsten mit diesem Titel verbunden ist. Zur Auswahl stehen: <ul style="list-style-type: none"> M Bei dem Namen handelt es sich um einen männlichen Namen. F Bei dem Namen handelt es sich um einen weiblichen Namen. A Mehrdeutig. Der Name kann entweder männlich oder weiblich sein. U Unbekannt. Das Geschlecht dieses Namens ist nicht bekannt. Wenn Sie dieses Feld leer lassen, wird „Unbekannt“ angenommen. |

Beispieleintrag:

```
<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>
      LookupValue
      Belt
      Friend
      Thursday
      Red
    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">
    LookupValue|Gender
    Mrs|F
```

```

        Mr | M
        Most | F

    </added-entries>
</table-data>

```

Beispiel einer benutzerdefinierten Tabellen

Im Folgenden finden Sie ein Beispiel für eine UserFirstNames.xml-Tabelle und die beim Ändern von benutzerdefinierten Tabellen zu verwendende Syntax.

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      FirstName
      AADEL
      AADIL

    </deleted-entry-group>
  </deleted-entry-group>

    <deleted-entry-group>

      FirstName | Frequency
      A | SACE | 0.126
      A | BECKETT | 0.421

    </deleted-entry-group>
  </deleted-entry-group>

    <deleted-entry-group>

      FirstName | Gender | Culture | VariantGroup
      ALI | M | DEFAULT | GROUP88
      AISHA | F | ARABIC | GROUP43

    </deleted-entry-group>
  </deleted-entry-group>

    <deleted-entry-group>

      FirstName | Gender
      JOHN | M

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

    FirstName | Gender | Culture
    JOHN | M | DEFAULT
    A | SHA | F | ARABIC
    JAMES | M | DEFAULT

  </added-entries>

```

```
</table-data>
```

Ausgabe

Achtung: Der „Name Parser“-Schritt ist veraltet und wird in zukünftigen Releases möglicherweise nicht unterstützt. Mit „Open Name Parser“ können Sie Namen parsen.

Tabelle 42: Ausgabe für „Name Parser“

| Feldname | Format | Beschreibung/gültige Werte |
|--------------------|--------------|--|
| AccountDescription | Zeichenfolge | Eine Kontobeschreibung, die Teil des Namens ist. Im Beispiel „Mary Jones Account # 12345“ lautet die Kontobeschreibung „Account#12345“. |
| EntityType | Zeichenfolge | Gibt den Typ des Namens an. Zur Auswahl stehen: <ul style="list-style-type: none"> Firm Bei dem Namen handelt es sich um einen Unternehmensnamen. Personal Bei dem Namen handelt es sich um einen Personennamen. |

Mit Unternehmensnamen verwandte Felder

| | | |
|----------------------------|--------------|--|
| FirmModifier.1.Object | Zeichenfolge | Das erste Objekt einer Präposition, das im Unternehmensnamen auftritt. Beispiel: Bei dem Unternehmensnamen „Pratt & Whitney Division of United Technologies“ ist „United Technologies“ das erste Objekt einer Präposition. |
| FirmModifier.1.Preposition | Zeichenfolge | Die erste Präposition, die im Unternehmensnamen auftritt. Beispiel: Bei dem Unternehmensnamen „Pratt & Whitney Division of United Technologies“ wäre „of“ die erste Präposition. |
| FirmModifier.2.Object | Zeichenfolge | Das zweite Objekt einer Präposition, das im Unternehmensnamen auftritt. Beispiel: Bei dem Firmennamen „Church of Our Lady of Lourdes“ ist „Lourdes“ das zweite Objekt einer Präposition. |

| Feldname | Format | Beschreibung/gültige Werte |
|--|--------------|---|
| FirmModifier.2.Preposition | Zeichenfolge | Die zweite Präposition, die im Unternehmensnamen auftritt. Beispiel: Bei dem Firmennamen „Church of Our Lady of Lourdes“ ist „of“ die zweite Präposition. |
| FirmName | Zeichenfolge | Der Name eines Unternehmens. Beispiel: „Pitney Bowes, Inc.“ |
| FirmPrimary | Zeichenfolge | Der Basisteil eines Unternehmensnamens. Zum Beispiel: Pitney Bowes. |
| FirmSuffix | Zeichenfolge | Das Unternehmenssuffix. Zum Beispiel: „Co.“ und „GmbH“ |
| Mit den Namen von Einzelpersonen verwandte Felder | | |
| FirstName | Zeichenfolge | Der Vorname einer Person. |
| FirstNameVariantGroup | Zeichenfolge | Eine numerische ID, die die Gruppe der Namen angibt, die dem Vornamen ähneln. Beispiel: Muhammad, Mohammed und Mehmet gehören alle in dieselbe „Name Variant Group“. Die tatsächliche Gruppen-ID wird beim Laden der Erweiterungsdaten zugewiesen. Dieses Feld wird nur ausgefüllt, wenn Sie das Feature „Name Variant Group“ erworben haben. |
| GenderCode | Zeichenfolge | Das durch die Analyse des Vornamens bestimmte Geschlecht einer Person. Zur Auswahl stehen: <ul style="list-style-type: none"> A Mehrdeutig. Bei dem Namen handelt es sich sowohl um einen männlichen als auch um einen weiblichen Namen. Beispiel: Pat. F Female. Bei dem Namen handelt es sich um einen weiblichen Namen. M Male. Bei dem Namen handelt es sich um einen männlichen Namen. U Unbekannt. Der Name konnte in der Geschlechtstabelle nicht gefunden werden. |

| Feldname | Format | Beschreibung/gültige Werte |
|---|--------------|--|
| GenderDeterminationSource | Zeichenfolge | Die zur Geschlechtsbestimmung eines Namens verwendete Kultur. Dieses Feld ist leer, wenn der Name in der Geschlechtstabelle nicht gefunden werden konnte. |
| GeneralSuffix | Zeichenfolge | Das allgemeine/berufliche Suffix einer Person. Zum Beispiel Dr. oder PhD. |
| LastName | Zeichenfolge | Der Nachname einer Person. |
| MaturitySuffix | Zeichenfolge | Das Generationssuffix einer Person. Zum Beispiel Jr. oder Sr. |
| MiddleName | Zeichenfolge | Der zweite Vorname einer Person. |
| NameScore | Zeichenfolge | Wert, der die Qualität des Parsing-Vorgangs darstellt, von 0 bis 100. 0 steht für schlechte Qualität und 100 für hohe Qualität. |
| ParserRecordID | Zeichenfolge | Eine den einzelnen Eingabedatensätzen zugewiesene eindeutige ID. |
| TitleOfRespect | Zeichenfolge | Der Titel einer Person, z. B. „Mr.“, „Mrs.“, „Dr.“ oder „Rev.“. |
| Mit verbundenen Namen verwandte Felder | | |
| PersonalName.2.FirstName | Zeichenfolge | Der Vorname der zweiten Person in einem verbundenen Namen. Ein Beispiel für einen verbundenen Namen ist „John und Jane Smith“. |
| PersonalName.2.FirstNameVariantGroup | Zeichenfolge | Eine numerische ID, die die Gruppe der Namen angibt, die dem Vornamen der zweiten Person in einem verbundenen Namen ähneln. Beispiel: Muhammad, Mohammed und Mehmet gehören alle in dieselbe „Name Variant Group“. Die tatsächliche Gruppen-ID wird beim Laden der Erweiterungsdaten zugewiesen. Dieses Feld wird nur ausgefüllt, wenn Sie das Feature „Name Variant Group“ erworben haben. |

| Feldname | Format | Beschreibung/gültige Werte |
|--|--------------|--|
| PersonalName.2.GenderCode | Zeichenfolge | Das Geschlecht der zweiten Person in einem verbundenen Namen, das durch die Analyse des Vornamens mit dem „Name Parser“ bestimmt wird. Ein Beispiel für einen verbundenen Namen ist „John und Jane Smith“. Einer der folgenden Werte: <ul style="list-style-type: none"> A Mehrdeutig. Bei dem Namen handelt es sich sowohl um einen männlichen als auch um einen weiblichen Namen. Beispiel: Pat. F Female. Bei dem Namen handelt es sich um einen weiblichen Namen. M Male. Bei dem Namen handelt es sich um einen männlichen Namen. U Unbekannt. Der Name konnte in der Geschlechtstabelle nicht gefunden werden. |
| PersonalName.2.GenderDeterminationSource | Zeichenfolge | Die zur Geschlechtsbestimmung der zweiten Person in einem verbundenen Namen verwendete Kultur. Ein Beispiel für einen verbundenen Namen ist „John und Jane Smith“. |
| PersonalName.2.GeneralSuffix | Zeichenfolge | Das allgemeine/berufliche Suffix der zweiten Person in einem verbundenen Namen. Ein Beispiel für einen verbundenen Namen ist „John und Jane Smith“. Beispiele für allgemeine Suffixe sind „MD“ und „PhD“. |
| PersonalName.2.LastName | Zeichenfolge | Der Nachname der zweiten Person in einem verbundenen Namen. Ein Beispiel für einen verbundenen Namen ist „John und Jane Smith“. |
| PersonalName.2.MaturitySuffix | Zeichenfolge | Das Generationssuffix der zweiten Person in einem verbundenen Namen. Ein Beispiel für einen verbundenen Namen ist „John und Jane Smith“. Beispiele für Generationssuffixe sind „Jr.“ und „Sr.“. |
| PersonalName.2.MiddleName | Zeichenfolge | Der zweite Vorname der zweiten Person in einem verbundenen Namen. Ein Beispiel für einen verbundenen Namen ist „John und Jane Smith“. |
| PersonalName.2.TitleOfRespect | Zeichenfolge | Der Titel des zweiten Namens in einem verbundenen Namen. Beispiel: „Mr. und Mrs. Smith“ ist ein verbundener Name. Beispiele für Titel sind „Mr.“, „Mrs.“ und „Dr.“. |

| Feldname | Format | Beschreibung/gültige Werte |
|--|--------------|--|
| PersonalName.3.FirstName | Zeichenfolge | Der Vorname der dritten Person in einem verbundenen Namen. Beispiel: „Mr. & Mrs. John Smith & Dr. Mary Jones“ ist ein verbundener Name. |
| PersonalName.3.FirstNameVariantGroup | Zeichenfolge | Eine numerische ID, die die Gruppe der Namen angibt, die dem Vornamen der zweiten Person in einem verbundenen Namen ähneln. Beispiel: Muhammad, Mohammed und Mehmet gehören alle in dieselbe „Name Variant Group“. Die tatsächliche Gruppen-ID wird beim Laden der Erweiterungsdaten zugewiesen. Dieses Feld wird nur ausgefüllt, wenn Sie das Feature „Name Variant Group“ erworben haben. |
| PersonalName.3.GenderCode | Zeichenfolge | Das Geschlecht der dritten Person in einem verbundenen Namen, das durch die Analyse des Vornamens mit dem Name Parser bestimmt wird. Ein Beispiel für einen verbundenen Namen ist „Mr. & Mrs. John Smith & Adam Jones“. Zur Auswahl stehen: <ul style="list-style-type: none"> A Mehrdeutig. Bei dem Namen handelt es sich sowohl um einen männlichen als auch um einen weiblichen Namen. Beispiel: Pat. F Female. Bei dem Namen handelt es sich um einen weiblichen Namen. M Male. Bei dem Namen handelt es sich um einen männlichen Namen. U Unbekannt. Der Name konnte in der Geschlechtstabelle nicht gefunden werden. |
| PersonalName.3.GenderDeterminationSource | Zeichenfolge | Die zur Geschlechtsbestimmung der dritten Person in einem verbundenen Namen verwendete Kultur. „Mr. & Mrs. John Smith & Adam Jones“. |
| PersonalName.3.GeneralSuffix | Zeichenfolge | Das allgemeine/berufliche Suffix der dritten Person in einem verbundenen Namen. Ein Beispiel für einen verbundenen Namen ist „Mr. & Mrs. John Smith & Adam Jones PhD“. Beispiele für allgemeine Suffixe sind „MD“ und „PhD“. |
| PersonalName.3.LastName | Zeichenfolge | Der Nachname der dritten Person in einem verbundenen Namen. Beispiel: „Mr. & Mrs. John Smith & Dr. Mary Jones“ ist ein verbundener Name. |

| Feldname | Format | Beschreibung/gültige Werte |
|-------------------------------|--------------|--|
| PersonalName.3.MaturitySuffix | Zeichenfolge | Das Generationssuffix der dritten Person in einem verbundenen Namen. Ein Beispiel für einen verbundenen Namen ist „Mr. & Mrs. John Smith & Adam Jones Sr.“. Beispiele für Generationssuffixe sind „Jr.“ und „Sr.“. |
| PersonalName.3.MiddleName | Zeichenfolge | Der zweite Vorname der dritten Person in einem verbundenen Namen. Beispiel: „Mr. & Mrs. John Smith & Dr. Mary Jones“ ist ein verbundener Name. |
| PersonalName.3.TitleOfRespect | Zeichenfolge | Der Titel des dritten Namens in einem verbundenen Namen. Beispiel: „Mr. & Mrs. John Smith & Dr. Mary Jones“ ist ein verbundener Name. Beispiele für Titel sind „Mr.“, „Mrs.“ und „Dr.“. |

Name Variant Finder

„Name Variant Finder“ funktioniert in einem Modus für Vornamen oder Nachnamen, um eine Datenbank für die Rückgabe alternativer Versionen eines Namens abzufragen. Beispiel: „John“ und „Jon“ sind Varianten des Namens „Johnathan“. Für „Name Variant Finder“ sind Erweiterungsbibliotheken erforderlich, die mit der Database Load Utility des Universal Name-Moduls, des Data Normalization-Moduls und des Advanced Matching-Moduls installiert werden können. Wenden Sie sich an Ihren Vertriebsbeauftragten, um weitere Informationen zur Vorgehensweise beim Abrufen dieser optionalen kulturspezifischen Wörterbücher zu erhalten.

Eingabe

Tabelle 43: Eingabefelder von Name Variant Finder

| Feldname | Beschreibung/gültige Werte |
|-----------|---|
| FirstName | Der Name, für den Sie Varianten finden möchten, wenn der Name ein Vorname ist. |
| LastName | Der Name, für den Sie Varianten finden möchten, wenn der Name ein Nachname ist. |

| Feldname | Beschreibung/gültige Werte |
|------------|--|
| GenderCode | <p>Das Geschlecht des Namens im Feld „FirstName“. Zur Auswahl stehen:</p> <p>Anmerkung: Geschlechtscodes sind nur auf Vornamen, nicht auf Nachnamen anwendbar.</p> <p>M Bei dem Namen handelt es sich um einen männlichen Namen.</p> <p>F Bei dem Namen handelt es sich um einen weiblichen Namen.</p> <p>A Mehrdeutig. Der Name kann entweder männlich oder weiblich sein.</p> <p>U Unbekannt. Das Geschlecht dieses Namens ist nicht bekannt.</p> |
| Ethnicity | <p>Die Kultur, die am häufigsten mit dem Namen im Feld „FirstName“ oder „LastName“ verknüpft ist. Sie können den „Name Parser“- oder den „Open Parser“-Schritt verwenden, um dieses Feld auszufüllen, wenn Sie die Ethnizität eines Namens nicht kennen.</p> <p>Anmerkung: Dieses Feld hieß vorher „GenderDeterminationSource“.</p> |

Optionen

Tabelle 44: Optionen für Name Variant Finder

| Option | Beschreibung |
|-----------------|--|
| Vorname | Findet Namensvarianten basierend auf dem Vornamen. |
| Nachname | Findet Namensvarianten basierend auf dem Nachnamen. |
| Geschlechtscode | Gibt nur Namensvarianten für das im Feld „GenderCode“ des Datensatzes spezifizierte Geschlecht zurück. Weitere Informationen zum Feld „GenderCode“ finden Sie unter Eingabe auf Seite 345. |
| Ethnizität | Gibt nur Namensvarianten für die im Feld „Ethnizität“ des Datensatzes spezifizierte Kultur zurück. Weitere Informationen zum Feld „Ethnizität“ finden Sie unter Eingabe auf Seite 345. |

| Option | Beschreibung |
|---------------|--|
| Romanisiert | Gibt die englisch romanisierte Version des Namens zurück. Ein romanisierter Name ist ein Name, der von einer nicht lateinischen Schrift in die lateinische Schrift konvertiert wurde. Beispielsweise ist „Achin“ die romanisierte Version des koreanischen Namens 아진. |
| Muttersprache | Gibt den Namen in der muttersprachlichen Schrift der Kultur des Namens zurück. Ein koreanischer Name würde beispielsweise in Hangul zurückgegeben. |
| Kana | Wenn Sie Muttersprache auswählen, können Sie japanische Namen durch Auswählen dieser Option in Kana zurückgeben lassen. Kana ist eine Kombination aus den Schriften Hiragana und Katakana. Anmerkung: Sie müssen die Asian Plus Pack-Datenbank lizenziert haben, um nach japanischen Namensvarianten zu suchen. Wenden Sie sich für weitere Informationen an Ihren Kundenbetreuer. |
| Kanji | Wenn Sie Muttersprache auswählen, können Sie japanische Namen durch Auswählen dieser Option in Kanji zurückgeben lassen. Kanji ist eine der Schriften, die in der japanischen Sprache verwendet werden. Anmerkung: Sie müssen die Asian Plus Pack-Datenbank lizenziert haben, um nach japanischen Namensvarianten zu suchen. Wenden Sie sich für weitere Informationen an Ihren Kundenbetreuer. |

Ausgabe

Tabelle 45: Ausgaben für Name Variant Finder

| Feldname | Format | Beschreibung/gültige Werte |
|----------------|--------------|--|
| CandidateGroup | Zeichenfolge | Identifiziert eine Gruppierung eines Eingabensamens und seiner Namensvarianten. Jedem Eingabensamens wird eine CandidateGroup-Nummer zugewiesen. Den Varianten dieses Eingabensamens wird dieselbe CandidateGroup-Nummer zugewiesen. |

| Feldname | Format | Beschreibung/gültige Werte |
|-------------------------|--------------|--|
| Ethnicity | Zeichenfolge | Die Kultur eines Namens, die durch die Core Name-Bibliothek und Erweiterungsbibliotheken festgelegt wird. Anmerkung: Dieses Feld hieß vorher „GenderDeterminationSource“. |
| FirstName | Zeichenfolge | Der Vorname einer Person. |
| GenderCode | Zeichenfolge | Das Geschlecht eines Namens, das durch die Core Name-Bibliothek und Erweiterungsbibliotheken festgelegt wird. Zur Auswahl stehen: M Bei dem Namen handelt es sich um einen männlichen Namen. F Bei dem Namen handelt es sich um einen weiblichen Namen. A Mehrdeutig. Der Name kann entweder männlich oder weiblich sein. U Unbekannt. Das Geschlecht dieses Namens ist nicht bekannt. |
| LastName | Zeichenfolge | Der Nachname einer Person. |
| TransactionalRecordType | Zeichenfolge | Gibt an, wie der Namen im Vergleichsprozess verwendet wurde. Zur Auswahl stehen: Suspect Ein Kopfdubletten-Datensatz wird als Eingabe für eine Abfrage verwendet. Candidate Ein Kandidatendatensatz ist ein Ergebnis, das aus einer Abfrage zurückgegeben wird. |

OpenNameParser

OpenNameParser schlüsselt Personen- und Unternehmensnamen sowie andere Begriffe im Namensdatenfeld in ihre Komponententeile auf. Diese geparsten Namenselemente sind anschließend für andere automatisierte Vorgänge verfügbar, z. B. Namensvergleiche, die Namensstandardisierung oder die Namenskonsolidierung über mehrere Datensätze.

OpenNameParser führt die folgenden Aktionen aus:

- Es bestimmt den Typ eines Namens, um die Funktion dieses Namens zu beschreiben. Namensentitätstypen lassen sich in zwei größere Gruppen aufteilen: Personennamen und Unternehmensnamen. Innerhalb dieser größeren Gruppen gibt es Untergruppen.
- Bestimmt die Form eines Namens, um herauszufinden, welche Syntax der Parser beim Parsen verwendet sollte. Personennamen liegen üblicherweise in normaler Reihenfolge (Unterschrift) oder in umgekehrter Reihenfolge vor. Unternehmensnamen sind normalerweise hierarchisch sortiert.
- Bestimmt die Komponenten eines Namens und beschriftet sie, um die syntaktische Beziehung jedes Namensteils zum gesamten Namen so zu identifizieren. Die Syntax von Personennamen umfasst neben anderen Teilen von Personennamen Präfixe, Vorname, zweiter Vorname, Nachname, Suffixe und Begriffe zur Kontobeschreibung. Die Syntax für Unternehmensnamen umfasst den Firmennamen und Suffixe.
- Parst verbundene Personen- und Unternehmensnamen, und behält sie entweder als einen Datensatz bei oder teilt sie in mehrere Datensätze auf. Beispiele für verbundene Namen sind „Mr. und Mrs. John Smith“ und „Baltimore Gas & Electric dba Constellation Energy“.
- Parst die Ausgabe als Datensätze oder als Liste.
- Ermöglicht Ihnen, den Open Parser Domain Editor zu verwenden, um neue Domänen zu erstellen, die Sie in den erweiterten Optionen von Open Name Parser verwenden können.
- Weist eine Parsing-Punktzahl zu, die den Zuverlässigkeitsgrad des Parsing widerspiegelt.

Eingabe

Tabelle 46: Eingabe für Open Name Parser

| Feldname | Beschreibung | | | | | | | | |
|---------------------|--|---------------------|---------------------------|-----------|---------|-----------|----------|-----------|-----------|
| CultureCode | <p>Die Kultur der Eingabenamensdaten. Die Optionen sind unten aufgelistet.</p> <table border="0"> <tr> <td>Null (empty)</td> <td>Globale Kultur (Standard)</td> </tr> <tr> <td>de</td> <td>Deutsch</td> </tr> <tr> <td>es</td> <td>Spanisch</td> </tr> <tr> <td>ja</td> <td>Japanisch</td> </tr> </table> <p>Anmerkung: Wenn Sie über den Open Parser Domain Editor Ihre eigene Domäne hinzugefügt haben, sind auch die Kulturen und Kulturcodes für diese Domäne gültig.</p> | Null (empty) | Globale Kultur (Standard) | de | Deutsch | es | Spanisch | ja | Japanisch |
| Null (empty) | Globale Kultur (Standard) | | | | | | | | |
| de | Deutsch | | | | | | | | |
| es | Spanisch | | | | | | | | |
| ja | Japanisch | | | | | | | | |
| Name | Der zu parsende Name. Dieses Feld ist ein Pflichtfeld. | | | | | | | | |

Optionen

Open Name Parser-Optionen können auf Schrittebene über einen der Spectrum™ Technology Platform-Clients oder zur Laufzeit mit Datenflussoptionen konfiguriert werden.

Parsing-Optionen

Die folgende Tabelle führt die Optionen auf, die das Parsen von Namen steuern.

Tabelle 47: Parsing-Optionen für Open Name Parser

| Name der Option | Beschreibung |
|---|---|
| Personennamen parsen | <p>Gibt an, ob Personennamen geparkt werden sollen.</p> <p>Normal Die Namensfelder sind nach Titel, Vorname, zweiter Vorname, Nachname und Suffix sortiert.</p> <p>Umgekehrt Die Namensfelder sind mit Nachnamen zuerst sortiert.</p> <p>Beide Die Namensfelder sind nach einer Kombination aus normaler und umgekehrter Reihenfolge sortiert.</p> |
| Verbundene Namen | Gibt an, ob verbundene Namen geparkt werden sollen. |
| Verbundene Namen auf mehrere Datensätze aufteilen | <p>Gibt an, ob Namen, die mehr als eine Einzelperson enthalten, wie <i>Bill & Sally Smith</i>, in mehrere Datensätze aufgeteilt werden sollen.</p> <p>Verwenden Sie einen „Unique ID Generator“-Schritt, um für jeden der aufgeteilten Datensätze eine ID zu erstellen.</p> |
| Unternehmensnamen parsen | Gibt an, ob Unternehmensnamen geparkt werden sollen. |
| Ergebnisse als Liste ausgeben | Gibt an, ob geparkte Namenselemente in Listenform zurückgegeben werden sollen. |

| Name der Option | Beschreibung |
|-------------------------|---|
| Abkürzungsschwellenwert | Gibt an, wie Leistung gegenüber Qualität ausbalanciert werden soll. Eine höhere Leistung resultiert in einer niedrigeren Qualität bei der Ausgabe. Anders herum resultiert eine höhere Qualität in einer niedrigeren Leistung. Wenn dieser Schwellenwert erreicht ist, wird bei dem Datensatz keine weitere Verarbeitung mehr durchgeführt. Der Standardwert ist 100. |

Kulturoptionen

Die folgende Tabelle führt die Optionen auf, die Namenskulturen steuern.

Tabelle 48: Kulturoptionen für Open Name Parser

| Name der Option | Beschreibung |
|-----------------|--|
| Kulturen | <p>Gibt an, welche Kultur(en) Sie in der Parsing-Grammatik einschließen möchten. Die Standardauswahl ist „Globale Kultur“.</p> <p>Anmerkung: Wenn Sie über den Open Parser Domain Editor Ihre eigene Domäne hinzugefügt haben, erscheinen hier auch die Kulturen und Kulturcodes für diese Domäne.</p> <p>Klicken Sie auf die Schaltflächen Aufwärts und Abwärts, um die Reihenfolge festzulegen, in der die Kulturen ausgeführt werden sollen.</p> |

Erweiterte Optionen

Die folgende Tabelle führt die erweiterten Optionen für das Parsen von Namen auf.

Tabelle 49: Erweiterte Optionen für Open Name Parser

| Option | Beschreibung |
|---------------------|--|
| Erweiterte Optionen | <p>Verwenden Sie die Dropdown-Liste „Domänen“, um für jeden Namen die entsprechende Domäne auszuwählen.</p> <p>Klicken Sie auf die Schaltflächen Aufwärts und Abwärts, um die Reihenfolge festzulegen, in der die Parser ausgeführt werden sollen. Ergebnisse werden für die erste Domäne zurückgegeben, die eine höhere Punktzahl erzielt, als im Feld Abkürzungsschwellenwert angegeben ist. Wenn keine Domäne diesen Schwellenwert erreicht, werden Ergebnisse für die Domäne mit der höchsten Punktzahl zurückgegeben. Wenn mehrere Domänen den Schwellenwert gleichzeitig erreichen, erhält die Domäne Priorität, die zuerst ausgeführt wurde (wie durch die Reihenfolge hier festgelegt). Ergebnisse werden dann für diese Domäne zurückgegeben.</p> <p>Anmerkung: Wenn Sie über den Open Parser Domain Editor Ihre eigene Domäne hinzugefügt haben, erscheint auch diese Domäne hier.</p> |

Konfigurieren von Optionen zur Laufzeit

Optionen von Open Name Parser können zur Laufzeit konfiguriert und weitergeleitet werden, wenn sie als Datenflussoptionen verfügbar gemacht werden. Das ermöglicht Ihnen, die vorhandene Konfiguration durch Zeichenfolgen für Namensparsing im JSON-Format zu überschreiben. Sie können auch Schrittoptionen festlegen, wenn Sie den Auftrag über einen Prozessfluss oder über das Befehlszeilentool „Job Executor“ aufrufen.

So definieren Sie die Optionen von Open Name Parser zur Laufzeit:

1. Öffnen Sie im Enterprise Designer einen Datenfluss, der den „Open Name Parser“-Schritt verwendet.
2. Speichern Sie diesen Datenfluss und machen Sie ihn verfügbar.
3. Öffnen Sie „Bearbeiten > Datenflussoptionen“.
4. Erweitern Sie Open Name Parser in der Tabelle **Datenflussoptionen zu Schritten zuordnen**, und bearbeiten Sie die Optionen nach Bedarf. Aktivieren Sie das Kästchen für die zu bearbeitende Option, und ändern Sie anschließend den Wert in der Dropdown-Liste **Standardwert**.
5. Optional: Ändern Sie den Namen der Optionen im Feld **Optionsbeschriftung**.
6. Klicken Sie zweimal auf **OK**.

Ausgabe

Tabelle 50: Ausgabe für Open Name Parser

| Feldname | Format | Beschreibung |
|--|--------------|---|
| AccountDescription | Zeichenfolge | Eine Kontobeschreibung, die Teil des Namens ist. Im Beispiel „Mary Jones Account # 12345“ lautet die Kontobeschreibung „Account#12345“. |
| Names | Zeichenfolge | Ein hierarchisches Feld, das eine Liste von geparsten Elementen enthält. Dieses Feld wird zurückgegeben, wenn Sie in den Parsing-Optionen das Kästchen Ergebnisse als Liste ausgeben aktivieren. |
| Mit Unternehmensnamen verwandte Felder | | |
| FirmConjunction | Zeichenfolge | Gibt an, dass der Name eines Unternehmens eine Konjunktion wie „d/b/a“ (doing business as – geschäftlich tätig als), „o/a“ (operating as – tätig als) oder „t/a“ (trading as – handelnd als) enthält. |
| FirmName | Zeichenfolge | Der Name eines Unternehmens. Zum Beispiel: Pitney Bowes. |
| FirmSuffix | Zeichenfolge | Das Unternehmenssuffix. Zum Beispiel: „Co.“ und „GmbH“ |
| IsFirm | Zeichenfolge | Gibt an, ob der Name eher ein Unternehmen als eine Einzelperson ist. |
| Mit den Namen von Einzelpersonen verwandte Felder | | |
| Conjunction | Zeichenfolge | Gibt an, dass der Name eine Konjunktion wie „und“, „oder“ oder „&“ enthält. |
| CultureCode | Zeichenfolge | Die Kulturcodes, die in den Eingabedaten enthalten sind. |

| Feldname | Format | Beschreibung |
|------------------------|--------------|---|
| CultureCodeUsedToParse | Zeichenfolge | <p>Identifiziert die kulturspezifische Grammatik, die zum Parsen der Daten verwendet wurde.</p> <p>Null (empty) Globale Kultur (Standard)</p> <p>de Deutsch</p> <p>es Spanisch</p> <p>ja Japanisch</p> <p>Anmerkung: Wenn Sie über den Open Parser Domain Editor Ihre eigene Domäne hinzugefügt haben, erscheinen in diesem Feld auch die Kulturen und Kulturcodes für diese Domäne.</p> |
| FirstName | Zeichenfolge | Der Vorname einer Person. |
| GeneralSuffix | Zeichenfolge | Das allgemeine/berufliche Suffix einer Person. Zum Beispiel Dr. oder PhD. |
| IsParsed | Zeichenfolge | Gibt an, ob ein Ausgabedatensatz geparkt wurde. Die Werte sind „wahr“ oder „falsch“. |
| IsPersonal | Zeichenfolge | Gibt an, ob der Name eher eine Einzelperson als ein Unternehmen ist. Die Werte sind „wahr“ oder „falsch“. |
| IsReverseOrder | Zeichenfolge | Gibt an, ob der Eingabename in umgekehrter Reihenfolge ist. Die Werte sind „wahr“ oder „falsch“. |
| LastName | Zeichenfolge | Der Nachname einer Person. Beinhaltet den elterlichen Nachnamen. |
| LeadingData | Zeichenfolge | Information ohne Namen, die vor einem Namen erscheint. |
| MaturitySuffix | Zeichenfolge | Das Generationssuffix einer Person. Zum Beispiel Jr. oder Sr. |
| MiddleName | Zeichenfolge | Der zweite Vorname einer Person. |

| Feldname | Format | Beschreibung |
|-------------------|--------------|--|
| Name. | Zeichenfolge | Der persönliche Name oder Firmenname, der in der Eingabe bereitgestellt wurde. |
| NameScore | Zeichenfolge | Gibt die durchschnittliche Punktzahl von bekannten und unbekannt Tokens für jeden Namen an. Der Wert von „NameScore“ liegt zwischen 0 und 100 wie in der Parsing-Grammatik definiert. 0 wird zurückgegeben, wenn keine Übereinstimmungen zurückgegeben werden. |
| SecondaryLastName | Zeichenfolge | In der spanischen Parsing-Grammatik der Nachname der Mutter einer Person. |
| TitleOfRespect | Zeichenfolge | Eine Information, die vor einem Namen erscheint, z. B. Fr., Hr. oder Dr. |
| TrailingData | Zeichenfolge | Information ohne Namen, die nach einem Namen erscheint. |

Mit verbundenen Namen verwandte Felder

| | | |
|--------------|--------------|---|
| Conjunction2 | Zeichenfolge | Gibt an, dass ein zweiter verbundener Name eine Konjunktion wie „und“, „oder“ oder „&“ enthält. |
| Conjunction3 | Zeichenfolge | Gibt an, dass ein dritter verbundener Name eine Konjunktion wie „und“, „oder“ oder „&“ enthält. |
| FirmName2 | Zeichenfolge | Der Name eines zweiten verbundenen Unternehmens. Zum Beispiel: Baltimore Gas & Electric dba Constellation Energy. |
| FirmSuffix2 | Zeichenfolge | Das Suffix eines zweiten verbundenen Unternehmens. |
| FirstName2 | Zeichenfolge | Der Vorname eines zweiten verbundenen Namens. |
| FirstName3 | Zeichenfolge | Der Vorname eines dritten verbundenen Namens. |

| Feldname | Format | Beschreibung |
|-----------------|--------------|--|
| GeneralSuffix2 | Zeichenfolge | Das allgemeine/berufliche Suffix für einen zweiten verbundenen Namen. Zum Beispiel Dr. oder PhD. |
| GeneralSuffix3 | Zeichenfolge | Das allgemeine/berufliche Suffix für einen dritten verbundenen Namen. Zum Beispiel Dr. oder PhD. |
| IsConjoined | Zeichenfolge | Gibt an, ob der Eingabename verbunden ist. Ein Beispiel für einen verbundenen Namen ist „John und Jane Smith“. |
| LastName2 | Zeichenfolge | Der Nachname eines zweiten verbundenen Namens. |
| LastName3 | Zeichenfolge | Der Nachname eines dritten verbundenen Namens. |
| MaturitySuffix2 | Zeichenfolge | Das Generationssuffix für einen zweiten verbundenen Namen. Zum Beispiel Jr. oder Sr. |
| MaturitySuffix3 | Zeichenfolge | Das Generationssuffix für einen dritten verbundenen Namen. Zum Beispiel Jr. oder Sr. |
| MiddleName2 | Zeichenfolge | Der zweite Vorname eines zweiten verbundenen Namens. |
| MiddleName3 | Zeichenfolge | Der zweite Vorname eines dritten verbundenen Namens. |
| TitleOfRespect2 | Zeichenfolge | Eine Information, die vor einem zweiten verbundenen Namen erscheint, z. B. Fr., Hr. oder Dr. |
| TitleOfRespect3 | Zeichenfolge | Eine Information, die vor einem dritten verbundenen Namen erscheint, z. B. Fr., Hr. oder Dr. |

Zusammenfassungsbericht für Open Name Parser

Der Zusammenfassungsbericht für Open Name Parser liefert zusammenfassende Statistiken zum Auftrag, z. B. die Summe aller Eingabedatensätze und die Summe der Datensätze ohne

Namensdaten, sowie mehrere Parsingstatistiken. Anleitungen zur Verwendung von Berichten finden Sie im *Datenfluss-Designer-Handbuch von Spectrum™ Technology Platform*.

Allgemeine Ergebnisse

- **Gesamtzahl der Eingabedatensätze:** Die Anzahl der Datensätze in der Eingabedatei.
- **Gesamtzahl der Datensätze, die keine Namensdaten enthalten haben:** Die Anzahl der Datensätze in der Eingabedatei, die keine zu parsenden Namensdaten enthalten haben.
- **Gesamtzahl der geparsten Namen:** Die Anzahl der Namen in der Eingabedatei, die geparkt wurden.
- **Gesamtzahl der Datensätze:** Die Gesamtanzahl der verarbeiteten Datensätze.
- **Niedrigste Namensparsing-Punktzahl:** Die niedrigste Parsing-Punktzahl, die ein Name in der Eingabedatei erhalten hat.
- **Höchste Namensparsing-Punktzahl:** Die höchste Parsing-Punktzahl, die ein Name in der Eingabedatei erhalten hat.
- **Durchschnittliche Namensparsing-Punktzahl:** Die durchschnittliche Parsing-Punktzahl aller geparsten Namen in der Eingabedatei.

Parsing-Ergebnisse – Personennamen

- **Anzahl der geschriebenen Datensätze mit Personennamen:** Die Anzahl der Personennamen in der Eingabedatei.
- **Anzahl geparster Namen aus verbundenen Namen:** Die Anzahl geparster Namen aus Datensätzen mit verbundenen Namen. Wenn Ihre Eingabedatei beispielsweise fünf Datensätze mit zwei verbundenen Namen und sieben Datensätze mit drei verbundenen Namen enthält, ist der Wert für dieses Feld gemäß der Gleichung $(5 \times 2) + (7 \times 3)$ gleich 31.
- **Datensätze mit zwei verbundenen Namen:** Die Anzahl der Eingabedatensätze mit zwei verbundenen Namen.
- **Datensätze mit drei verbundenen Namen:** Die Anzahl der Eingabedatensätze mit drei verbundenen Namen.
- **Anzahl von Namen mit vorhandenem Titel:** Die Anzahl der geparsten Namen mit Titel.
- **Anzahl von Namen mit vorhandenem Generations-Suffix:** Die Anzahl der geparsten Namen mit Generations-Suffix.
- **Anzahl von Namen mit vorhandenem allgemeinen Suffix:** Die Anzahl der geparsten Namen mit allgemeinem Suffix.
- **Anzahl von Namen mit Kontobeschreibungen:** Die Anzahl der geparsten Namen mit einer Kontobeschreibung.
- **Gesamtzahl an Namen in umgekehrter Reihenfolge:** Die Anzahl der geparsten Namen in umgekehrter Reihenfolge, sodass das Ausgabefeld „isReversed“ als „Wahr“ angezeigt wird.

Parsing-Ergebnisse – Unternehmensnamen

- **Anzahl der geschriebenen Datensätze mit Unternehmensnamen:** Die Anzahl der Unternehmensnamen in der Eingabedatei.

- **Anzahl von Namen mit vorhandenem Unternehmenssuffix:** Die Anzahl der geparsten Namen mit einem Unternehmenssuffix.
- **Anzahl von Namen mit Kontobeschreibungen:** Die Anzahl der Eingabedatensätze mit einer Kontobeschreibung.
- **Gesamtzahl der DBA-Datensätze:** Die Anzahl der Eingabedatensätze mit Konjunktionen für Firmennamen, sodass die beiden Ausgabefelder „isPersonal“ und „isFirm“ als „Wahr“ angezeigt werden.

9 - ISO-Ländercodes und Modulunterstützung

[In this section](#)

ISO-Ländercodes und Modulunterstützung

360

ISO-Ländercodes und Modulunterstützung

Diese Tabelle enthält die ISO-Codes für jedes Land sowie die Module, die die Adressierung, das Geocoding und die Routenführung für jedes Land unterstützen.

Beachten Sie, dass das Enterprise Geocoding-Modul Datenbanken für Afrika (30 Länder), den Mittleren Osten (8 Länder) und Lateinamerika (20 Länder) enthält. Diese Datenbanken decken die kleineren Länder in diesen Regionen ab, die nicht über ihre eigenen landesspezifischen Geocoding-Datenbanken verfügen. Die Spalte „Unterstützte Module“ gibt an, welche Länder von diesen Datenbanken für Afrika, den Mittleren Osten und Lateinamerika abgedeckt werden.

Außerdem stellt die Geocode Address World-Datenbank geografisches und begrenztes Post-Geocoding (keine Geocoding auf Straßenebene) für alle Länder bereit.

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|--------------------|-------------------------|-----------------------|--|
| Afghanistan | AF | AFG | Universal Addressing-Modul |
| Ålandinseln | AX | ALA | Universal Addressing-Modul |
| Albanien | AL oder SQ (Routing) | ALB | Universal Addressing-Modul Enterprise Geocoding-Modul Enterprise Routing-Modul |
| Algerien | DZ | DZA | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Amerikanisch-Samoa | AS | ASM | Universal Addressing-Modul |
| Andorra | AD | AND | Enterprise Geocoding-Modul. (Andorra wird durch den spanischen Geocoder abgedeckt.) Universal Addressing-Modul GeoComplete-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|---------------------|-----------------------|-----------------------|---|
| Angola | AO | AGO | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Anguilla | AI | AIA | Universal Addressing-Modul |
| Antarktika | AQ | ATA | Universal Addressing-Modul |
| Antigua und Barbuda | AG | ATG | Universal Addressing-Modul |
| Argentinien | AR | ARG | Enterprise Geocoding-Modul Universal Addressing-Modul Enterprise Routing-Modul |
| Armenien | AM | ARM | Universal Addressing-Modul |
| Aruba | AW | ABW | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Australien | AU | AUS | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul |
| Österreich | AT | AUT | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Aserbaidshon | AZ | AZE | Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|----------------|-----------------------|-----------------------|---|
| Bahamas | BS | BHS | Enterprise Geocoding-Modul Universal Addressing-Modul Enterprise Routing-Modul |
| Bahrain | BH | BHR | Enterprise Geocoding-Modul (Mittlerer Osten) Universal Addressing-Modul |
| Bangladesch | BD | BGD | Universal Addressing-Modul |
| Barbados | BB | BRB | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Weißrussland | BY | BLR | Universal Addressing-Modul Enterprise Routing-Modul |
| Belgien | BE | BEL | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Belize | BZ | BLZ | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Benin | BJ | BEN | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Bermuda | BM | BMU | Universal Addressing-Modul Enterprise Routing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|---|-----------------------|-----------------------|--|
| Bhutan | BT | BTN | Universal Addressing-Modul |
| Bolivien | BO | BOL | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Bonaire, Sint Eustatius und Saba | BQ | BES | Universal Addressing-Modul |
| Bosnien und Herzegowina | BA | BIH | Enterprise Geocoding-Modul Universal Addressing-Modul Enterprise Routing-Modul Enterprise Geocoding-Modul |
| Botswana | BW | BWA | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Bouvet-Insel | BV | BVT | Universal Addressing-Modul |
| Brasilien | BR | BRA | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Britisches Territorium im Indischen Ozean | IO | IOT | Universal Addressing-Modul |
| Brunei Darussalam | BN | BRN | Enterprise Geocoding-Modul Universal Addressing-Modul |
| Bulgarien | BG | BGR | Enterprise Geocoding-Modul Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|------------------------------|-----------------------|-----------------------|---|
| Burkina Faso | BF | BFA | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Burundi | BI | BDI | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Kambodscha | KH | KHM | Universal Addressing-Modul |
| Kamerun | CM | CMR | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Kanada | CA | CAN | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Kap Verde | CV | CPV | Universal Addressing-Modul |
| Caymaninseln | KY | CYM | Universal Addressing-Modul |
| Zentralafrikanische Republik | CF | CAF | Universal Addressing-Modul |
| Tschad | TD | TCD | Universal Addressing-Modul |
| Chile | CL | CHL | Enterprise Geocoding-Modul Universal Addressing-Modul Enterprise Routing-Modul GeoComplete-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|------------------------------|----------------------------|-----------------------|---|
| China | CN oder zh_CN (Routing) | CHN | Enterprise Geocoding-Modul Universal Addressing-Modul Enterprise Routing-Modul |
| Weihnachtsinsel | CX | CXR | Universal Addressing-Modul |
| Cocosinseln | CC | CCK | Universal Addressing-Modul |
| Kolumbien | CO | COL | Enterprise Geocoding-Modul Universal Addressing-Modul |
| Komoren | KM | COM | Universal Addressing-Modul |
| Kongo, Republik | CG | COG | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Demokratische Republik Kongo | CD | COD | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul Enterprise Routing-Modul |
| Cookinseln | CK | COK | Universal Addressing-Modul |
| Costa Rica | CR | CRI | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Elfenbeinküste | CI | CIV | Universal Addressing-Modul |
| Kroatien | HR | HRV | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|-------------------------|-------------------------|-----------------------|---|
| Kuba | CU | CUB | Enterprise Geocoding-Modul (Lateinamerika) Enterprise Routing-Modul Universal Addressing-Modul |
| Curaçao | CW | CUW | Universal Addressing-Modul |
| Zypern | CY | CYP | Enterprise Geocoding-Modul Universal Addressing-Modul |
| Tschechien | CZ oder CS (Routing) | CZE | Enterprise Geocoding-Modul Universal Addressing-Modul Enterprise Routing-Modul GeoComplete-Modul |
| Dänemark | DK | DNK | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Dschibuti | DJ | DJI | Universal Addressing-Modul |
| Dominica | DM | DMA | Universal Addressing-Modul |
| Dominikanische Republik | DO | DOM | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Ecuador | EC | ECU | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Ägypten | EG | EGY | Enterprise Geocoding-Modul (Mittlerer Osten) Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|---------------------------|-----------------------|-----------------------|---|
| El Salvador | SV | SLV | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Äquatorialguinea | GQ | GNQ | Universal Addressing-Modul |
| Eritrea | ER | ERI | Universal Addressing-Modul |
| Estland | EE | EST | Enterprise-Geocoding-Modul Enterprise-Routing-Modul Universal Addressing-Modul |
| Äthiopien | ET | ETH | Universal Addressing-Modul |
| Falklandinseln (Malwinen) | FK | FLK | Universal Addressing-Modul |
| Färöer-Inseln | FO | FRO | Universal Addressing-Modul |
| Fidschi | FJ | FJI | Universal Addressing-Modul |
| Finnland | FI | FIN | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Frankreich | FR | FRA | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|---|-----------------------|-----------------------|---|
| Französisch-Guayana | GF | GUF | Enterprise Geocoding-Modul (<i>Französisch-Guayana wird durch den französischen Geocoder abgedeckt.</i>) Universal Addressing-Modul |
| Französisch-Polynesien | PF | PYF | Universal Addressing-Modul |
| Französische Gebiete im südlichen Indischen Ozean | TF | ATF | Universal Addressing-Modul |
| Gabun | GA | GAB | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Gambia | GM | GMB | Universal Addressing-Modul |
| Georgien | GE | GEO | Universal Addressing-Modul |
| Deutschland | DE | DEU | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Ghana | GH | GHA | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul Enterprise Routing-Modul |
| Gibraltar | GI | GIB | Enterprise Geocoding-Modul (<i>Gibraltar wird durch den spanischen Geocoder abgedeckt.</i>) Universal Addressing-Modul |
| Griechenland | GR | GRC | Enterprise Geocoding-Modul Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|--------------------------|-----------------------|-----------------------|---|
| Grönland | GL | GRL | Universal Addressing-Modul |
| Grenada | GD | GRD | Universal Addressing-Modul |
| Guadeloupe | GP | GLP | Enterprise Geocoding-Modul (<i>Guadeloupe wird durch den französischen Geocoder abgedeckt.</i>) Universal Addressing-Modul |
| Guam | GU | GUM | Universal Addressing-Modul |
| Guatemala | GT | GTM | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Guernsey | GG | GGY | Universal Addressing-Modul |
| Guinea | GN | GIN | Universal Addressing-Modul |
| Guinea-Bissau | GW | GNB | Universal Addressing-Modul |
| Guyana | GY | GUY | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Haiti | HT | HTI | Universal Addressing-Modul |
| Heard und McDonaldinseln | HM | HMD | Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|------------------------------------|-----------------------|-----------------------|--|
| Apostolischer Stuhl (Vatikanstaat) | VA | MwSt | Enterprise Geocoding-Modul (<i>Der Vatikan wird durch den italienischen Geocoder abgedeckt.</i>) Universal Addressing-Modul |
| Honduras | HN | HND | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Hongkong | HK | HKG | Enterprise Geocoding-Modul Universal Addressing-Modul |
| Ungarn | HU | HUN | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Island | IS | ISL | Enterprise Geocoding-Modul Universal Addressing-Modul |
| Indien | IN | IND | Enterprise Geocoding-Modul Universal Addressing-Modul |
| Indonesien | ID | IDN | Enterprise Geocoding-Modul Universal Addressing-Modul |
| Islamische Republik Iran | IR | IRN | Universal Addressing-Modul |
| Irak | IQ | IRQ | Enterprise Geocoding-Modul (Mittlerer Osten) Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|----------------|-----------------------|-----------------------|---|
| Irland | IE | IRL | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Isle of Man | IM | IMN | Universal Addressing-Modul |
| Israel | IL | ISR | Universal Addressing-Modul Enterprise Routing-Modul |
| Italien | IT | ITA | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Jamaika | JM | JAM | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Japan | JP | JPN | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Jersey | JE | JEY | Universal Addressing-Modul |
| Jordanien | JO | JOR | Universal Addressing-Modul Enterprise Geocoding-Modul (Mittlerer Osten) Enterprise Routing-Modul |
| Kasachstan | KZ | KAZ | Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|-----------------------------------|-----------------------|-----------------------|---|
| Kenia | KE | KEN | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul Enterprise Routing-Modul |
| Kiribati | KI | KIR | Universal Addressing-Modul |
| Demokratische Volksrepublik Korea | KP | PRK | Universal Addressing-Modul |
| Republik Korea | KR | KOR | Enterprise Geocoding-Modul Universal Addressing-Modul |
| Kosovo | KS | KOS | Enterprise Geocoding-Modul Universal Addressing-Modul GeoComplete-Modul |
| Kuwait | KW | KWT | Enterprise Geocoding-Modul (Mittlerer Osten) Universal Addressing-Modul |
| Kirgisistan | KG | KGZ | Universal Addressing-Modul |
| Volksdemokratische Republik Laos | LA | LAO | Universal Addressing-Modul |
| Lettland | LV | LVA | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul |
| Libanon | LB | LBN | Enterprise Geocoding-Modul (Mittlerer Osten) Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|--|-----------------------|-----------------------|---|
| Lesotho | LS | LSO | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul Enterprise Routing-Modul |
| Liberia | LR | LBR | Universal Addressing-Modul |
| Libysch-Arabische Dschamahirija | LY | LBY | Universal Addressing-Modul |
| Liechtenstein | LI | LIE | Enterprise Geocoding-Modul (<i>Liechtenstein wird durch den Schweizer Geocoder abgedeckt.</i>) Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Litauen | LT | LTU | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul |
| Luxemburg | LU | LUX | Enterprise Geocoding-Modul (<i>Luxemburg wird durch den belgischen Geocoder abgedeckt.</i>) Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Macao | MO | MAC | Enterprise Geocoding-Modul Universal Addressing-Modul |
| Ehemalige jugoslawische Republik Mazedonien | MK | MKD | Enterprise Geocoding-Modul Universal Addressing-Modul |
| Madagaskar | MG | MDG | Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|----------------|-----------------------|-----------------------|---|
| Malawi | MW | MWI | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Malaysia | MY | MYS | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul |
| Malediven | MV | MDV | Universal Addressing-Modul |
| Mali | MT | MLI | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Malta | MT | MLT | Enterprise Geocoding-Modul – Universal Addressing-Modul |
| Marshallinseln | MH | MHL | Universal Addressing-Modul |
| Martinique | MQ | MTQ | Enterprise Geocoding-Modul (<i>Martinique wird durch den französischen Geocoder abgedeckt.</i>) Universal Addressing-Modul |
| Mauretanien | MR | MRT | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Mauritius | MU | MUS | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Mayotte | YT | MYT | Enterprise Geocoding-Modul (<i>Mayotte wird durch den französischen Geocoder abgedeckt.</i>) Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|------------------------------------|-----------------------|-----------------------|---|
| Mexiko | MX | MEX | Enterprise Geocoding-Modul Universal Addressing-Modul |
| Föderierte Staaten von Mikronesien | FM | FSM | Universal Addressing-Modul |
| Republik Moldau | MD | MDA | Universal Addressing-Modul Enterprise Routing-Modul |
| Monaco | MC | MCO | Enterprise Geocoding-Modul (<i>Monaco wird durch den französischen Geocoder abgedeckt.</i>) Universal Addressing-Modul |
| Mongolei | MN | MNG | Universal Addressing-Modul |
| Montenegro | ME | MNE | Enterprise Geocoding-Modul Universal Addressing-Modul |
| Montserrat | MS | MSR | Universal Addressing-Modul |
| Marokko | MA | MAR | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Mosambik | MZ | MOZ | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul Enterprise Routing-Modul |
| Myanmar | MM | MMR | Universal Addressing-Modul |
| Namibia | NA | NAM | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|--------------------|-----------------------|-----------------------|---|
| Nauru | NR | NRU | Universal Addressing-Modul |
| Nepal | NP | NPL | Universal Addressing-Modul |
| Niederlande | NL | NLD | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Neukaledonien | NC | NCL | Universal Addressing-Modul |
| Neuseeland | NZ | NZL | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul |
| Nicaragua | NI | NIC | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Niger | NE | NER | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Nigeria | NG | NGA | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Niue | NU | NIU | Universal Addressing-Modul |
| Norfolkinsel | NF | NFK | Universal Addressing-Modul |
| Nördliche Marianen | MP | MNP | Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|------------------------------------|-----------------------|-----------------------|---|
| Norwegen | NO | NOR | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Oman | OM | OMN | Enterprise Geocoding-Modul (Mittlerer Osten) Universal Addressing-Modul |
| Pakistan | PK | PAK | Universal Addressing-Modul |
| Palau | PW | PLW | Universal Addressing-Modul |
| Besetztes palästinensisches Gebiet | PS | PSE | Universal Addressing-Modul |
| Panama | PA | PAN | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Papua-Neuguinea | PG | PNG | Universal Addressing-Modul |
| Paraguay | PY | PRY | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Peru | PE | PER | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Philippinen | PH | PHL | Enterprise Geocoding-Modul Universal Addressing-Modul Enterprise Routing-Modul |
| Pitcairn | PN | PCN | Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|----------------------|-----------------------|-----------------------|--|
| Polen | PL | POL | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Portugal | PT | PRT | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Puerto Rico | PR | PRI | Universal Addressing-Modul |
| Katar | QA | QAT | Enterprise Geocoding-Modul (Mittlerer Osten) Universal Addressing-Modul |
| Réunion | RE | REU | Enterprise Geocoding-Modul (<i>Réunion wird durch den französischen Geocoder abgedeckt.</i>) Universal Addressing-Modul |
| Rumänien | RO | ROU | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul |
| Russische Föderation | RU | RUS | Enterprise Geocoding-Modul Universal Addressing-Modul GeoComplete-Modul |
| Ruanda | RW | RWA | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Saint Barthelemy | BL | BLM | Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|--|-----------------------|-----------------------|---|
| St. Helena, Ascension und Tristan da Cunha | SH | SHE | Universal Addressing-Modul |
| Saint Kitts und Nevis | KN | KNA | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| St. Lucia | LC | LCA | Universal Addressing-Modul |
| Saint-Martin (französischer Teil) | MF | MAF | Universal Addressing-Modul |
| Saint Pierre und Miquelon | PM | SPM | Universal Addressing-Modul |
| St. Vincent / Grenadinen | VC | VCT | Universal Addressing-Modul |
| Samoa | WS | WSM | Universal Addressing-Modul |
| San Marino | SM | SMR | Enterprise Geocoding-Modul (<i>San Marino wird durch den italienischen Geocoder abgedeckt.</i>) Universal Addressing-Modul |
| Sao Tome und Principe | ST | STP | Universal Addressing-Modul |
| Saudi-Arabien | SA | SAU | Enterprise Geocoding-Modul (Mittlerer Osten) Universal Addressing-Modul |
| Senegal | SN | SEN | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|--------------------------------------|-----------------------|-----------------------|---|
| Serbien | RS | SRB | Enterprise Geocoding-Modul Universal Addressing-Modul |
| Seychellen | SC | SYC | Universal Addressing-Modul |
| Sierra Leone | SL | SLE | Universal Addressing-Modul |
| Singapur | SG | SGP | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Sint Maarten (niederländischer Teil) | SX | SXM | Universal Addressing-Modul |
| Slowakei | SK | SVK | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Slowenien | SI | SVN | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Salomoninseln | SB | SLB | Universal Addressing-Modul |
| Somalia | SO | SOM | Universal Addressing-Modul |
| Südafrika | ZA | ZAF | Enterprise Geocoding-Modul Universal Addressing-Modul GeoComplete-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|--|-----------------------|-----------------------|---|
| Südgeorgien und Südliche Sandwichinseln | GS | SGS | Enterprise Geocoding-Modul – Universal Addressing-Modul |
| Südsudan | SS | SSD | Universal Addressing-Modul |
| Spanien | ES | ESP | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Sri Lanka | LK | LKA | Universal Addressing-Modul |
| Sudan | SD | SDN | Universal Addressing-Modul |
| Surinam | SR | SUR | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Svalbard und Jan Mayen | SJ | SJM | Universal Addressing-Modul |
| Swasiland | SZ | SWZ | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Schweden | SE | SWE | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Schweiz | CH | CHE | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|------------------------------|----------------------------|-----------------------|---|
| Syrische Arabische Republik | SY | SYR | Universal Addressing-Modul |
| Taiwan, Provinz Chinas | TW oder zh_TW (Routing) | TWN | Universal Addressing-Modul Enterprise Routing-Modul |
| Tadschikistan | TJ | TJK | Universal Addressing-Modul |
| Vereinigte Republik Tansania | TZ | TZA | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul Enterprise Routing-Modul |
| Thailand | TH | THA | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul |
| Timor-Leste | TL | TLS | Universal Addressing-Modul |
| Togo | TG | TGO | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Tokelau | TK | TKL | Universal Addressing-Modul |
| Tonga | TO | TON | Universal Addressing-Modul |
| Trinidad und Tobago | TT | TTO | Enterprise Geocoding-Modul (Lateinamerika) Universal Addressing-Modul |
| Tunesien | TN | TUN | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|--------------------------------------|-----------------------|-----------------------|---|
| Türkei | TR | TUR | Enterprise Geocoding-Modul Universal Addressing-Modul GeoComplete-Modul |
| Turkmenistan | TM | TKM | Universal Addressing-Modul |
| Turks- und Caicosinseln | TC | TCA | Universal Addressing-Modul |
| Tuvalu | TV | TUV | Universal Addressing-Modul |
| Uganda | UG | UGA | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Ukraine | UA | UKR | Enterprise Geocoding-Modul Universal Addressing-Modul |
| Vereinigte Arabische Emirate | AE | ARE | Enterprise Geocoding-Modul (Mittlerer Osten) Universal Addressing-Modul |
| Großbritannien | GB | GBR | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| Vereinigte Staaten von Amerika | US | USA | Enterprise Geocoding-Modul Enterprise Routing-Modul Universal Addressing-Modul GeoComplete-Modul |
| United States Minor Outlying Islands | UM | UMI | Universal Addressing-Modul |

| ISO-Ländername | ISO 3116-1 Alpha-2 | ISO 3116-1 Alpha-3 | Unterstützte Module |
|---------------------------------|-----------------------|-----------------------|--|
| Uruguay | UY | URY | Enterprise Geocoding-Modul – Universal Addressing-Modul |
| Usbekistan | UZ | UZB | Universal Addressing-Modul |
| Vanuatu | VU | VUT | Universal Addressing-Modul |
| Bolivarische Republik Venezuela | VE | VEN | Enterprise Geocoding-Modul Universal Addressing-Modul |
| Vietnam | VN | VNM | Enterprise Geocoding-Modul Universal Addressing-Modul |
| Britische Jungferninseln | VG | VGB | Universal Addressing-Modul |
| Amerikanische Jungferninseln. | VI | VIR | Universal Addressing-Modul |
| Wallis und Futuna | WF | WLF | Universal Addressing-Modul |
| Westsahara | EH | ESH | Universal Addressing-Modul |
| Jemen | YE | YEM | Enterprise Geocoding-Modul (Mittlerer Osten) Universal Addressing-Modul |
| Sambia | ZM | ZMB | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |
| Zimbabwe | ZW | ZWE | Enterprise Geocoding-Modul (Afrika) Universal Addressing-Modul |

Notices

© 2018 Pitney Bowes Software Inc. Alle Rechte vorbehalten. MapInfo und Group 1 Software sind Marken von Pitney Bowes Software Inc. Alle anderen Marken und Markenzeichen sind Eigentum ihrer jeweiligen Besitzer.

USPS® Urheberrechtshinweise

Pitney Bowes Inc. wurde eine nicht-ausschließliche Lizenz erteilt, die die Veröffentlichung und den Verkauf von ZIP + 4® Postleitzahl-Datenbanken auf optischen und magnetischen Medien genehmigt. Folgende Marken sind Markenzeichen des United States Postal Service: CASS, CASS Certified, DPV, eLOT, FASTforward, First-Class Mail, Intelligent Mail, LACS^{Link}, NCOA^{Link}, PAVE, PLANET Code, Postal Service, POSTNET, Post Office, RDI, Suite^{Link}, United States Postal Service, Standard Mail, United States Post Office, USPS, ZIP Code, und ZIP + 4. Hierbei handelt es sich jedoch nicht um eine vollständige Liste der Marken, die zum United States Postal Service gehören.

Pitney Bowes Inc. ist nicht-exklusiver Lizenznehmer von USPS® für die Verarbeitungsprozesse von NCOA^{Link}®.

Die Preisgestaltung jeglicher Pitney Bowes Softwareprodukte, -optionen und -dienstleistungen erfolgt nicht durch USPS® oder die Regierung der Vereinigten Staaten. Es wird auch keine Regulierung oder Genehmigung der Preise durch USPS® oder die US-Regierung durchgeführt. Bei der Verwendung von RDI™-Daten zur Berechnung von Paketversandkosten wird die Entscheidung, welcher Paketlieferdienst genutzt wird, nicht von USPS® oder der Regierung der Vereinigten Staaten getroffen.

Datenbereitstellung und Hinweise

Hier verwendete Datenprodukte und Datenprodukte, die in Software-Anwendungen von Pitney Bowes verwendet werden, sind durch verschiedene Markenzeichen und mindestens eines der folgenden Urheberrechte geschützt:

© Copyright United States Postal Service. Alle Rechte vorbehalten.

© 2014 TomTom. Alle Rechte vorbehalten. TomTom und das TomTom Logo sind eingetragene Marken von TomTom N.V.

© 2016 HERE

Fuente: INEGI (Instituto Nacional de Estadística y Geografía)

Basierend auf elektronischen Daten © National Land Survey Sweden.

© Copyright United States Census Bureau

© Copyright Nova Marketing Group, Inc.

Teile dieses Programms sind urheberrechtlich geschützt durch © Copyright 1993-2007 Nova Marketing Group Inc. Alle Rechte vorbehalten.

© Copyright Second Decimal, LLC

© Copyright Canada Post Corporation

Diese CD-ROM enthält Daten einer urheberrechtlich geschützten Datenerfassung der Canada Post Corporation.

© 2007 Claritas, Inc.

Das Geocode Address World Dataset enthält lizenzierte Daten des GeoNames-Projekts (www.geonames.org), die unter den Bedingungen der Creative Commons Attribution License ("Attribution License") bereitgestellt werden. Die Attribution License können Sie unter <http://creativecommons.org/licenses/by/3.0/legalcode> einsehen. Ihre Nutzung der GeoNames-Daten (wie im Spectrum™ Technology Platform Nutzerhandbuch beschrieben) unterliegt den Bedingungen der Attribution License. Bei Konflikten zwischen Ihrer Vereinbarung mit Pitney Bowes Software, Inc. und der Attribution License hat die Attribution License lediglich bezüglich der Nutzung von GeoNames-Daten Vorrang.



3001 Summer Street
Stamford CT 06926-0700
USA

www.pitneybowes.com