pitney bowes

Spectrum™ Technology Platform

Version 12.0 SP1

Datenfluss-Designer-Handbuch

Inhalt

1 - Erste Schritte

Installieren der Clienttools 4
Starten des Enterprise Designer 5
Erster Blick auf den Enterprise Designer 5
Mein erster Datenfluss (Auftrag) 9
Mein erster Datenfluss (Dienst) 12
Datenflussvorlagen 15
Importieren und Exportieren von Datenflüssen 16

2 - Entwerfen eines Flusses

Flusstypen 19
Flusseingabe 21
Felder 27
Steuerungsschritte 44
Flussausgabe 111
Eingebettete Datenflüsse 116
Berichte 122
Überlegungen zur Leistung 126
Datenflussversionen 141

3 - Überprüfen und Testen

Prüfen eines Flusses auf Fehler 146
Überprüfen eines Datenflusses 147
Testen eines Dienstes mit Management Console 152

4 - Ausführen eines Flusses

Ausführen eines Auftrags- oder Prozessflusses 154
Verfügbarmachen eines Dienstes 178
Laufzeitoptionen 181
Konfigurieren einer E-Mail-Benachrichtigung für einen Datenfluss 184

5 - Kombinieren von Flüssen zu einem Prozessfluss

Einführung in Prozessflüsse 187
Entwerfen von Prozessflüssen 187

6 - Erstellen von wiederverwendbaren Flusskomponenten

Einführung in die Unterflüsse 226
Verwenden eines Unterflusses als Quelle 226
Verwenden eines Unterflusses in der Mitte eines Datenflusses 228
Verwenden eines Unterflusses als ein Laden von Daten 229
Ändern eines Unterflusses 230
Löschen eines Unterflusses 231
Verfügbarmachen eines Unterflusses und Aufhebung der Verfügbarkeit 231
Konvertieren eines Schrittes in einen Unterfluss 231

7 - Info zu Spectrum™ Technology Platform

Was ist Spectrum™ Technology Platform? 234
Architektur der Unternehmensdatenverwaltung 235
Architektur von Spectrum™ Technology Platform 239
Module und Komponenten 244

1 - Erste Schritte

In this section

Installieren der Clienttools 4
Starten des Enterprise Designer 5
Erster Blick auf den Enterprise Designer 5
Mein erster Datenfluss (Auftrag) 9
Mein erster Datenfluss (Dienst) 12
Datenflussvorlagen 15
Importieren und Exportieren von Datenflüssen 16

Installieren der Clienttools

Bei den Spectrum™ Technology Platform-Clienttools handelt es sich um Anwendungen, mit denen Sie Ihren Server verwalten und Datenflüsse sowie Prozessflüsse entwerfen und ausführen können. Sie müssen Ihren Spectrum™ Technology Platform-Server installieren, bevor Sie die Clienttools installieren können.

Stellen Sie sicher, dass Sie vor der Installation die Versionshinweise lesen. Die Versionshinweise enthalten wichtige Informationen zur Kompatibilität und versionsspezifische Installationshinweise.

In dieser Prozedur wird beschrieben, wie die folgenden Clienttools installiert werden:

- **Enterprise Designer**: Mit Enterprise Designer können Sie Datenflüsse erstellen, ändern und ausführen.
- **Job Executor**: Job Executor ist ein Befehlszeilentool, mit dem Sie einen Auftrag über eine Befehlszeile oder ein Skript ausführen können. Der Auftrag muss vorher erstellt und mit Enterprise Designer in Spectrum™ Technology Platform gespeichert worden sein.
- **Process Flow Executor**: Process Flow Executor ist ein Befehlszeilentool, mit dem ein Prozessfluss über eine Befehlszeile oder ein Skript ausgeführt werden kann. Der Prozessfluss muss vorher erstellt und mit Enterprise Designer in Spectrum™ Technology Platform gespeichert worden sein.
- **Administrationsumgebung**: Die Administrationsumgebung bietet Befehlszeilenzugriff auf mehrere administrative Funktionen. Sie können diese in einem Skript verwenden und so bestimmte administrative Aufgaben automatisieren. Sie können sie auch interaktiv verwenden.

  Anmerkung: Ab Spectrum-Version 11.0 stellt die Management Console anstatt eines installierbaren Clients, wie dies in Vorgängerversionen der Fall war, ein webbasiertes Tool dar.

So installieren Sie die Clienttools:

1. Öffnen Sie einen Webbrowser und navigieren Sie zur Spectrum™ Technology Platform-Begrüßungsseite unter:

   http://<servername>:<port>

   Wenn Sie beispielsweise Spectrum™ Technology Platform auf einem Computer mit dem Namen „myspectrumplatform“ installiert haben und dieser den HTTP-Standardport 8080 verwendet, navigieren Sie zu:

   http://myspectrumplatform:8080

2. Klicken Sie auf **Platform-Clienttools**.
3. Laden Sie das Clienttool herunter, das Sie installieren möchten.

Starten des Enterprise Designer

Der Enterprise Designer ist eine Windows-Anwendung zum Erstellen von Datenflüssen. So starten Sie den Enterprise Designer:

1. Wählen Sie **Start** > **Programme** > **Pitney Bowes** > **Spectrum™ Technology Platform** > **Clienttools** > **Enterprise Designer** aus.
2. Geben Sie den Servernamen oder die IP-Adresse ein oder wählen Sie diese Angaben im Listenfeld aus. Wenn Sie Spectrum™ Technology Platform in einem Cluster einsetzen, geben Sie den Namen oder die IP-Adresse des Load Balancer des Clusters ein.
3. Geben Sie Ihren Benutzernamen und das dazugehörige Kennwort ein.
4. Geben Sie in das Feld „Port“ den Netzwerkport ein, für den der Server zur Spectrum™ Technology Platform-Kommunikation konfiguriert wurde. Die Standardportnummer ist 8080.
5. Klicken Sie auf **Sichere Verbindung**, wenn die Kommunikation zwischen dem Client und dem Server über eine HTTPS-Verbindung erfolgen soll.

   Anmerkung: Eine sichere Verbindung ist nur verfügbar, wenn HTTPS-Kommunikation auf dem Server konfiguriert wurde. Wenn Sie den Enterprise Designer unter Windows 7 ausführen, funktioniert die Verwendung der IP-Adresse im Feld **Servername** möglicherweise nicht. Dies hängt vom Typ des zum Schützen der Kommunikation zwischen dem Enterprise Designer und dem Server verwendeten Zertifikats ab. Wenn die IP-Adresse nicht funktioniert, verwenden Sie stattdessen den Hostnamen.

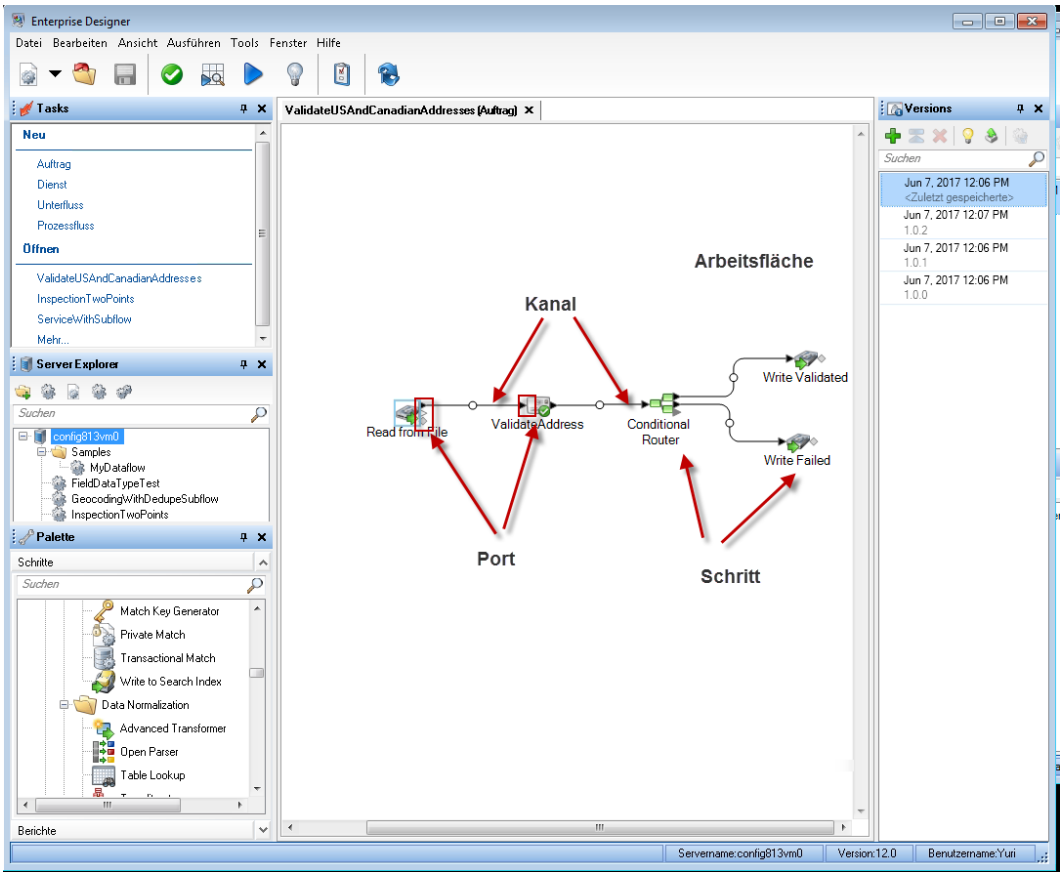
6. Klicken Sie auf **Anmeldung**.

Erster Blick auf den Enterprise Designer

Der Enterprise Designer ist ein visuelles Tool zum Erstellen von Datenflüssen. Mit diesem Client können Sie:

- Aufträge, Dienste, Unterflüsse und Prozessflüsse erstellen und ändern
- Datenflüsse auf Probleme testen
- Dienste verfügbar machen und ausblenden
- Berichte generieren

Das Fenster des Enterprise Designer sieht wie folgt aus:



Um mit Datenflüssen zu arbeiten, müssen Sie einige wichtige Begriffe verstehen:

Arbeitsfläche Die Arbeitsfläche ist der Hauptarbeitsbereich. Das Bild oben zeigt die geöffnete Arbeitsfläche mit einem Datenfluss mit Namen „ValidateUSAndCanadianAddresses“. Dies ist ein Auftragsdatenfluss, was bedeutet, dass er eine Batch-Verarbeitung ausführt, indem er Daten aus einer Datei liest und die Ausgabe in eine Datei schreibt. In diesem Fall schreibt der Datenfluss die Ausgabe in zwei Dateien.

Schritt Schritte, die durch Symbole auf der Arbeitsfläche dargestellt werden, führen eine bestimmte Art von Aktivität aus, wie das Sortieren von Datensätzen, das Überprüfen von Adressen, das Vergleichen ähnlicher Datensätze usw. Um einen Schritt hinzuzufügen, ziehen Sie den Schritt aus der Palette (auf der linken Seite des Fensters) auf die Arbeitsfläche.

Wenn ein Schritt Ihre Aufmerksamkeit erfordert, wird ein blauer Kreis auf dem Symbol angezeigt:

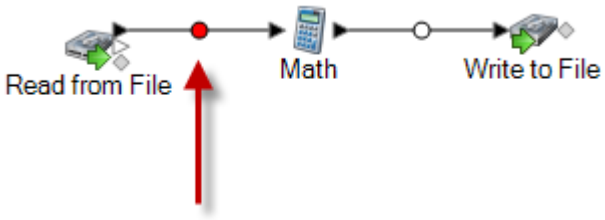
Ein Datenfluss kann nicht erfolgreich ausgeführt werden, wenn er Schritte enthält, die Ihre Aufmerksamkeit erfordern. Doppelklicken Sie also auf den Schritt, um die erforderlichen Einstellungen zu konfigurieren. Sobald Sie alle erforderlichen Einstellungen konfiguriert haben, wird der blaue Kreis nicht mehr angezeigt:

Kanal

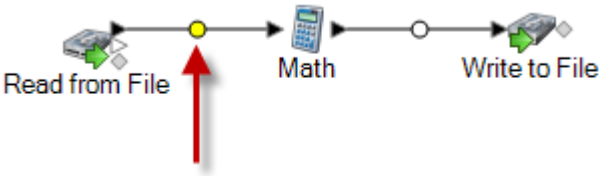
Ein Kanal ist eine Verbindung zwischen zwei oder mehr Schritten, über die Datensätze von einem Schritt an einen anderen übergeben werden. Im obigen Beispiel können Sie sehen, dass der „Read from File“-Schritt über einen Kanal mit dem „ValidateAddress“-Schritt verbunden ist. Datensätze werden in den Datenfluss in „Read from File“ eingelesen und dann über diesen Kanal an „ValidateAddress“ gesendet. „ValidateAddress“ wird dann über einen Kanal mit „Conditional Router“ verbunden. „Conditional Router“ analysiert Datensätze und sendet sie über verschiedene Pfade in einem Datenfluss, je nach den vom Datenfluss-Designer definierten Bedingungen. Aus „Conditional Router“ gehen zwei Kanäle heraus, einer an einen „Write Validated“-Schritt und einer an einen „Write Failed“-Schritt.

Der Punkt in der Mitte eines Kanals kann seine Farbe ändern, um verschiedene Bedingungen anzugeben:

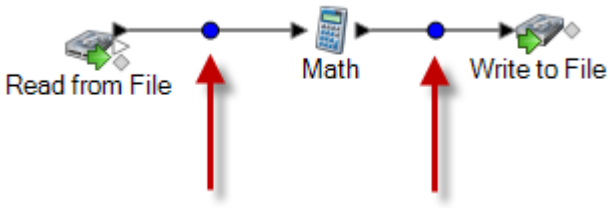
Rot Gibt einen Fehler wie einen Typkonvertierungsfehler an, der ein Feld für den Downstream-Schritt unbrauchbar macht.



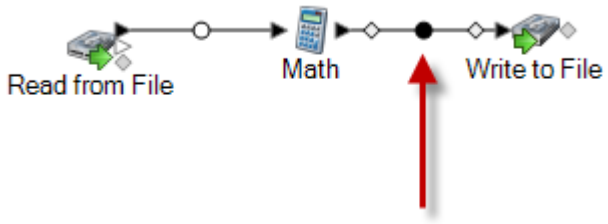
Gelb Sie haben ein Feld entfernt, das für einen Downstream-Schritt erforderlich ist.



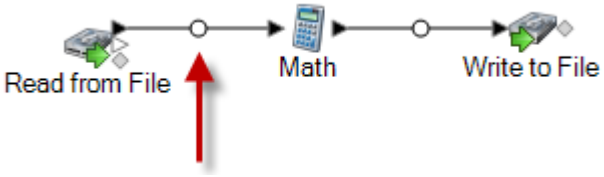
Blau Die automatische Typkonvertierung hat ein Feld erfolgreich in den für den Downstream-Schritt erforderlichen Datentyp konvertiert.



Schwarz Ein Feld wird im Kanal umbenannt.



Weiß Es wird keine Aktion für die Felder ausgeführt.



Port Wenn Sie sich die Schrittsymbole genauer ansehen, sehen Sie die kleinen dreieckigen oder rautenförmigen Ports an der Seite jedes Schritttes. Ein Port ist der Mechanismus, mit dem ein Schritt Daten in einen Kanal sendet oder aus einem Kanal liest. Schritte, die Daten in den Datenfluss einlesen (als „Quellen“ bezeichnet), haben nur Ausgabeports, da sie immer am Anfang eines Datenflusses sind. Schritte, die Daten aus dem Datenfluss senden (als „Datenladungen“ bezeichnet), haben Eingabeports, da sie immer am Ende eines Datenflusses sind. Alle anderen Schritte haben sowohl Eingabe- als auch Ausgabeports. Außerdem haben manche Schritte Fehlerports, die zum Ausgeben von Datensätzen genutzt werden, die während der Verarbeitung des Schrittes Fehler verursachen. Manche Schritte haben auch Berichtsports, die dazu dienen, Berichte über die Ausgabe des Schrittes zu generieren.

Außerdem hat das Fenster des Enterprise Designer die folgenden Features:

Feature	Beschreibung
Aufgaben	Bietet eine schnelle Möglichkeit, einen neuen Auftrag, Dienst, Unterfluss oder Prozessfluss zu erstellen. Erlaubt Ihnen auch, Datenflüsse zu öffnen, die zuletzt geöffnet waren.
Server-Explorer	Zeigt alle Flüsse an, die auf dem Spectrum™ Technology Platform-Server gespeichert sind. Falls der Server-Explorer nicht sichtbar ist, wählen Sie **Ansicht** > **Server-Explorer** aus. Sie können Flüsse in Ordnern verwalten. Um einen Ordner zu erstellen, klicken Sie mit der rechten Maustaste auf den Servernamen und wählen Sie **Neuer Ordner** aus. Flussnamen müssen über alle Ordner hinweg eindeutig sein. Es dürfen keine zwei Flüsse mit demselben Namen vorhanden sein, auch wenn diese sich in unterschiedlichen Ordnern befinden.

Feature	Beschreibung
Palette	Enthält alle Schritte und Berichte, dass Sie zu Ihrem Datenfluss hinzufügen können. Die in der Palette verfügbaren Schritte hängen von den Modulen ab, die Sie lizenziert haben.
Arbeitsfläche	Der Arbeitsbereich, auf den Sie Schritte ziehen und sie mit Kanälen verbinden, um Datenflüsse zu erzeugen. Sie können mehrere Datenfluss-Arbeitsfläche gleichzeitig geöffnet haben.
Versionen	Das Feature „Version“ im Enterprise Designer ermöglicht Ihnen, einen Revisionsverlauf Ihrer Datenflüsse beizubehalten. Sie können vorherige Versionen eines Datenflusses anzeigen, ältere Versionen zur Ausführung verfügbar machen und einen Verlauf Ihrer Änderungen beibehalten – für den Fall, dass Sie jemals zu einer vorherigen Version eines Datenflusses zurückkehren müssen.
Navigator	Listet die Schritte und Berichte im Fluss auf. Sie können mit der rechten Maustaste auf ein Element im Bereich **Navigator** klicken, um die zugehörigen Optionen zu bearbeiten.

Mein erster Datenfluss (Auftrag)

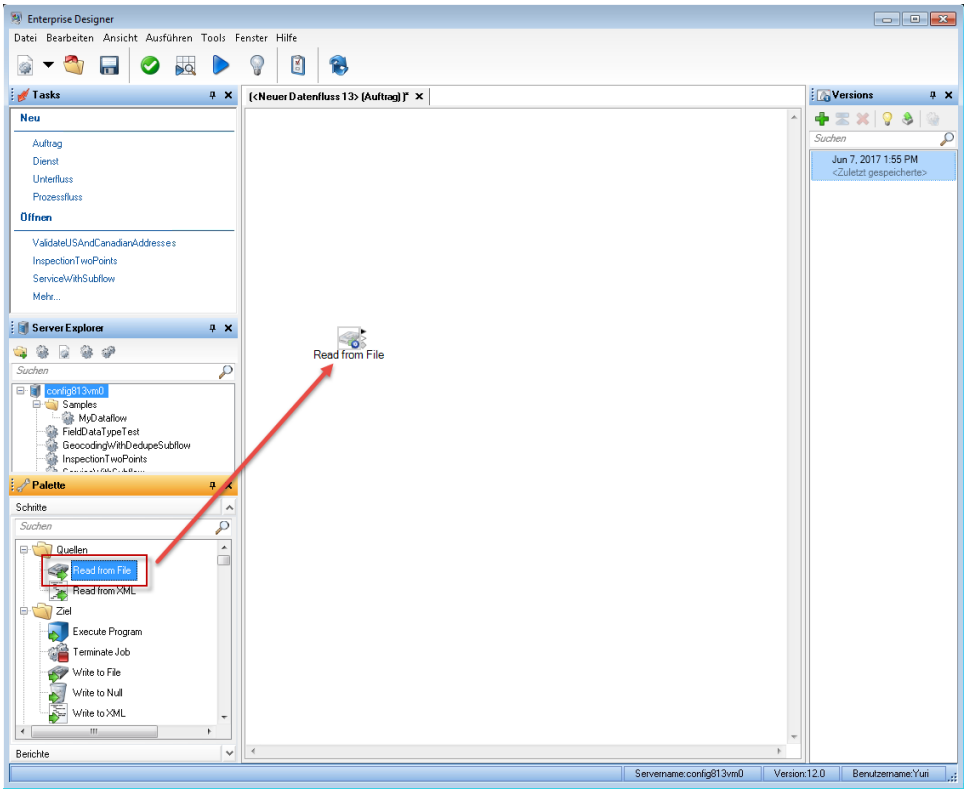
In diesem Abschnitt erstellen Sie einen einfachen Datenfluss, der Daten aus einer Datei liest, sie sortiert und dann in eine Datei schreibt. Da dieser Datenfluss Daten aus einer Datei liest und seine Ausgabe in eine Datei schreibt, handelt es sich um einen „Auftrag“. Dies ist ein Datenfluss, der eine Stapelverarbeitung durchführt. (Der andere primäre Datenflusstyp, ein „Dienst“, führt die interaktive Verarbeitung über einen API- oder Webservice-Aufruf zum Server durch.)

1. Der erste Schritt besteht darin, einige Beispieldaten zu erstellen, die als Eingabe für Ihren Datenfluss verwendet werden sollen. Erstellen Sie mithilfe eines Texteditors eine Datei, die wie folgt aussieht:

```
FirstName,LastName,Region,Amount
Alan,Smith,East,18.23
Jeannie,Wagner,North,45.43
Joe,Simmons,East,10.87
Pam,Hiznay,Central,98.78
```

2. Speichern Sie die Datei an einem geeigneten Ort.
3. Wählen Sie **Start** > **Programme** > **Pitney Bowes** > **Spectrum™ Technology Platform** > **Client-Tools** > **Enterprise Designer** aus.

4. Wählen Sie **Datei** > **Neu** > **Datenfluss** > **Auftrag** aus.
5. Sie können nun mit der Erstellung Ihres Datenflusses beginnen. Der erste Schritt besteht darin, die Eingabe des Datenflusses zu definieren. Gehen Sie dazu wie folgt vor:
   a) Ziehen Sie einen „Read from File“-Schritt auf die Arbeitsfläche:



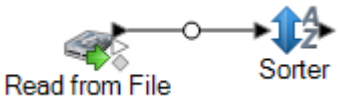
   b) Doppelklicken Sie auf den „Read from File“-Schritt auf der Arbeitsfläche.
   c) Geben Sie in das Feld **Dateiname** die in Schritt 1 auf Seite 9 erstellte Datei an.
   d) Wählen Sie im Feld **Datensatztyp** die Option **Mit Begrenzungszeichen** aus.
   e) Wählen Sie im Feld **Feldtrennzeichen** den Eintrag **Komma (,)** aus.
   f) Aktivieren Sie das Kästchen **Erste Zeile ist Kopfzeilen-Datensatz**.
   g) Klicken Sie auf die Registerkarte **Felder**.
   h) Klicken Sie auf **Erneut generieren** und anschließend auf **Ja**.

      Der Schritt wird automatisch für die Felder in Ihrer Eingabedatei konfiguriert.

   i) Klicken Sie auf **Typ erkennen**. Dadurch wird die Eingabedatei gescannt und für jedes Feld der entsprechende Datentyp ermittelt. Beachten Sie, dass der Typ für das Feld **Betrag** von „string“ zu „double“ wechselt.
   j) Sie haben die Konfiguration von „Read from File“ abgeschlossen. Klicken Sie auf **OK**.
6. Fügen Sie als nächstes einen Schritt hinzu, der die Datensätze nach Region sortiert. Gehen Sie dazu wie folgt vor:
   a) Ziehen Sie den Sorter-Schritt auf die Arbeitsfläche.

b) Klicken Sie auf das schwarze Dreieck auf der rechten Seite des „Read from File“-Schrittes (dem Ausgabeport) und ziehen Sie es auf die linke Seite des Sorter-Schrittes auf der Arbeitsfläche, um einen Kanal zu erstellen, der „Read from File“ und Sorter verbindet.

Ihr Datenfluss sollte wie folgt aussehen:



c) Doppelklicken Sie auf der Arbeitsfläche auf den Sorter-Schritt.
d) Klicken Sie auf **Hinzufügen**.
e) Wählen Sie im Feld **Feldname** den Eintrag **Region** aus.
f) Sie haben die Konfiguration von Sorter abgeschlossen. Klicken Sie auf **OK**.

7. Definieren Sie schließlich die Ausgabedatei, in der der Datenfluss seine Ausgabe schreibt. Gehen Sie dazu wie folgt vor:
   a) Ziehen Sie einen „Write to File“-Schritt auf die Arbeitsfläche.
   b) Klicken Sie auf das schwarze Dreieck auf der rechten Seite des Sorter-Schrittes und ziehen Sie es auf die linke Seite des „Write to File“-Schrittes auf der Arbeitsfläche.

   Ihr Datenfluss sollte wie folgt aussehen:

   c) Doppelklicken Sie auf den „Write to File“-Schritt.
   d) Geben Sie im Feld **Dateiname** eine Ausgabedatei an. Dies kann jede beliebige Datei sein.
   e) Wählen Sie im Feld **Feldtrennzeichen** den Eintrag **Komma (,)** aus.
   f) Aktivieren Sie das Kästchen **Erste Zeile ist Kopfzeilen-Datensatz**.
   g) Klicken Sie auf die Registerkarte **Felder**.
   h) Klicken Sie auf **Schnell hinzufügen**.
   i) Klicken Sie auf **Alle auswählen** und anschließend auf **OK**.
   j) Verwenden Sie die Schaltflächen **Nach oben** und **Nach unten**, um die Felder in der folgenden Reihenfolge zu sortieren:

      FirstName
      LastName
      Region
      Amount

      Dadurch enthalten die Datensätze in Ihrer Ausgabedatei die Felder in der gleichen Reihenfolge wie in Ihrer Eingabedatei.

   k) Sie haben die Konfiguration von „Write to File“ abgeschlossen. Klicken Sie auf **OK**.

8. Wählen Sie im Enterprise Designer die Optionen **Datei** > **Speichern** aus.
9. Geben Sie Ihrem Datenfluss einen Namen, und klicken Sie auf **OK**.

10. Ihr Datenfluss ist nun einsatzbereit. Wählen Sie **Ausführen** > **Aktuellen Fluss ausführen** aus.
11. Das Fenster **Ausführungsdetails** wird mit dem Status des Auftrags angezeigt. Klicken Sie auf **Aktualisieren**. Sobald der Status **Erfolgreich** angezeigt wird, klicken Sie auf **Schließen**.

Öffnen Sie die Ausgabedatei, die Sie im „Write to File“-Schritt angegeben haben. Sie werden sehen, dass die Datensätze nach Region sortiert wurden, wie Sie es im Sorter-Schritt angegeben haben.

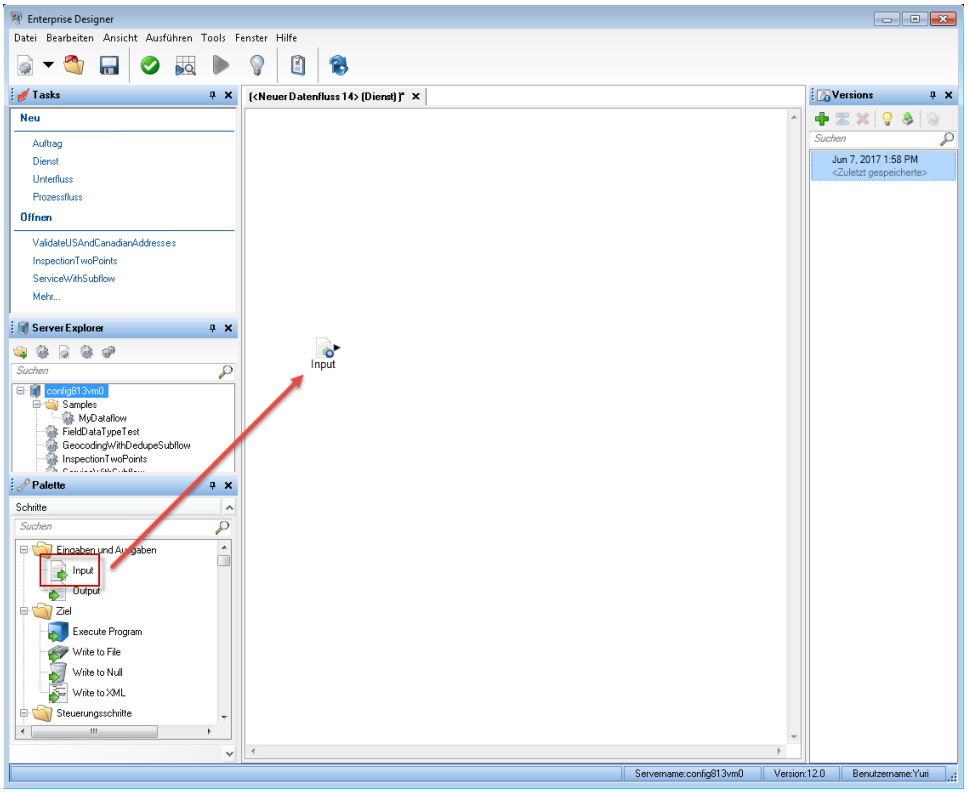
```
FirstName,LastName,Region,Amount
Pam,Hiznay,Central,98.78
Alan,Smith,East,18.23
Joe,Simmons,East,10.87
Jeannie,Wagner,North,45.43
```

Herzlichen Glückwunsch! Sie haben Ihren ersten Auftragsdatenfluss entworfen und ausgeführt.

Mein erster Datenfluss (Dienst)

In diesem Abschnitt erstellen Sie einen einfachen Datenfluss, der Daten aus einem API- oder Webservice-Aufruf akzeptiert, die Daten verarbeitet und eine Antwort über die API oder den Webservice zurückgibt. Da dieser Datenfluss als Dienst auf dem Spectrum™ Technology Platform-Server bereitgestellt werden soll, handelt es sich um einen Datenfluss vom Typ „Dienst“. (Der andere primäre Datenflusstyp, ein „Auftrag“, führt eine Batch-Verarbeitung aus, liest Daten aus einer Datei oder Datenbank, verarbeitet die Daten und schreibt dann die Ausgabe in eine Datei oder Datenbank.)

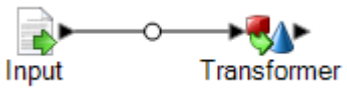
1. Wählen Sie **Start** > **Programme** > **Pitney Bowes** > **Spectrum™ Technology Platform** > **Client-Tools** > **Enterprise Designer** aus.
2. Wählen Sie **Datei** > **Neu** > **Datenfluss** > **Dienst** aus.
3. Sie können nun mit der Erstellung Ihres Datenflusses beginnen. Der erste Schritt besteht darin, die Eingabe des Datenflusses zu definieren. Ihr Datenfluss nutzt zwei Felder als Eingabe: Vorname und Nachname.
   a) Ziehen Sie einen Input-Schritt aus der Palette auf die Arbeitsfläche.



b) Doppelklicken Sie auf den Input-Schritt auf der Arbeitsfläche.
c) Klicken Sie auf **Hinzufügen** und anschließend erneut auf **Hinzufügen**.
d) Geben Sie in das Feld **Feldname** den Eintrag `Vorname` ein.
e) Klicken Sie auf **OK** und anschließend erneut auf **OK**.
f) Klicken Sie auf **Hinzufügen** und anschließend erneut auf **Hinzufügen**.
g) Geben Sie in das Feld **Feldname** den Eintrag `Nachnahme` ein.
h) Klicken Sie auf **OK** und anschließend erneut auf **OK**.
i) Sie haben die Definition der Datenflusseingabe abgeschlossen. Klicken Sie auf **OK**.

4. Fügen Sie als nächstes einen Schritt hinzu, um die Schreibweise der Daten in den Feldern „Vorname“ und „Nachname“ auf Großbuchstaben zu ändern.
   a) Ziehen Sie einen Transformer-Schritt aus der Palette auf die Arbeitsfläche.
   b) Klicken Sie auf das schwarze Dreieck auf der rechten Seite des Input-Schrittes (der Ausgabeport) und ziehen Sie es auf die linke Seite des Transformer-Schrittes auf der Arbeitsfläche, um einen Kanal zu erstellen, der Input und Transformator verbindet.

      Ihr Datenfluss sollte wie folgt aussehen:

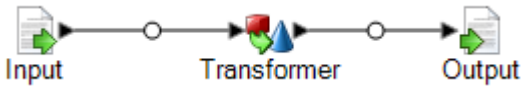


   c) Doppelklicken Sie auf den „Transformer“-Schritt.
   d) Klicken Sie auf **Hinzufügen**.

e) Klicken Sie im Baum auf der linken Seite unter **Formatierung** auf **Groß-/Kleinbuchstaben**.
f) Wählen Sie im Feld **Feld** den Eintrag **Vorname** aus. Lassen Sie **Großbuchstaben** ausgewählt.
g) Klicken Sie auf **Hinzufügen**.
h) Wählen Sie im Feld **Feld** den Eintrag **Nachname** aus. Lassen Sie **Großbuchstaben** ausgewählt.
i) Klicken Sie auf **Hinzufügen**.
j) Klicken Sie auf **Schließen**.
k) Sie haben die Konfiguration von Transformer abgeschlossen, um den Wert in den Feldern „Vorname“ und „Nachname“ in Großbuchstaben zu ändern. Klicken Sie auf **OK**.

5. Letztlich definieren Sie die Ausgabe für den Datenfluss. Ihr Datenfluss gibt die Felder „Vorname“ und „Nachname“ als Ausgabe zurück.
   a) Ziehen Sie einen Output-Schritt auf die Arbeitsfläche.
   b) Klicken Sie auf das schwarze Dreieck auf der rechten Seite des Transformer-Schrittes und ziehen Sie es auf die linke Seite des Output-Schrittes auf der Arbeitsfläche.

      Ihr Datenfluss sollte wie folgt aussehen:

   c) Doppelklicken Sie auf den Output-Schritt auf der Arbeitsfläche.
   d) Aktivieren Sie das Kästchen **Verfügbar machen**. Die Kästchen neben „Vorname“ und „Nachname“ sollten nun aktiviert sein.
   e) Klicken Sie auf **OK**.
6. Wählen Sie im Enterprise Designer die Optionen **Datei** > **Speichern** aus.
7. Geben Sie Ihrem Datenfluss den Namen `MyFirstDataflow-Service`, und klicken Sie auf **OK**.
8. Wählen Sie **Datei** > **Verfügbar machen/Verfügbarkeit aufheben und Speichern** aus. Dadurch wird Ihr Datenfluss verfügbar gemacht und als Dienst auf dem Server bereitgestellt.
9. So testen Sie Ihren Dienst:
   a) Öffnen Sie die Management Console, indem Sie die folgende URL in einem Webbrowser laden:

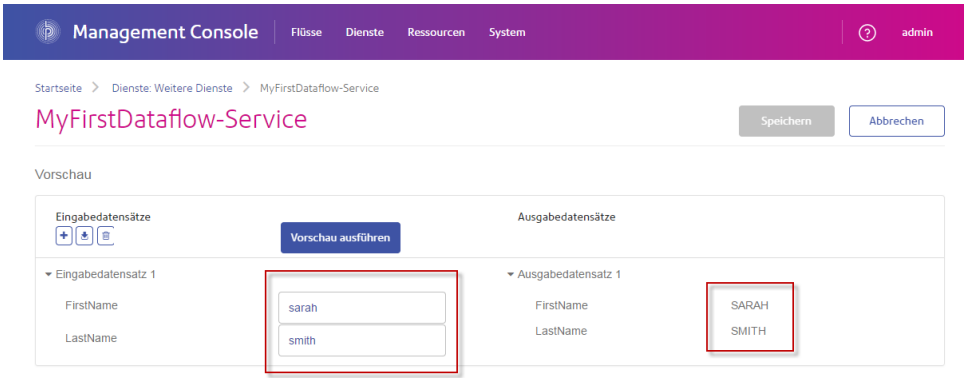
      http://*server*:*port*/managementconsole

      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.

   b) Öffnen Sie **Dienste** > **Weitere Dienste**.
   c) Aktivieren Sie in der Liste der Dienste das Kästchen neben **MyFirstDataflow-Service**, und klicken Sie dann auf die Schaltfläche „Bearbeiten“ .
   d) Geben Sie in das Feld „Vorname“ einen Namen in Kleinbuchstaben ein.

e) Geben Sie in das Feld „Nachname“ einen Namen in Kleinbuchstaben ein.
f) Klicken Sie auf **Vorschau ausführen**.

Sie können sehen, dass der Dienst die Namensfelder als Großbuchstaben schreibt, wie Sie es im Transformer-Schritt Ihres Datenflusses angegeben haben.



Herzlichen Glückwunsch! Sie haben Ihren ersten Dienst-Datenfluss entworfen und ausgeführt. Der Dienst ist nun auf dem Server verfügbar und kann über einen API- oder Webservice-Aufruf zugänglich gemacht werden. Die Ressourcen-URL für den SOAP-Endpunkt dieses Dienstes lautet:

```
http://<ServerName>:<Port>/soap/MyFirstDataflow-Service
```

Die Ressourcen-URL für den REST-Endpunkt dieses Dienstes lautet:

```
http://<ServerName>:<Port>/rest/MyFirstDataflow-Service
```

Datenflussvorlagen

Datenflussvorlagen veranschaulichen, wie Sie Spectrum™ Technology Platform und seine Module verwenden können, um Ihre Geschäftsanforderungen zu realisieren. Sie stellen dar, wie bestimmte Module unterschiedliche Anforderungen lösen, z. B. Parsing, Standardisierung, Überprüfung von Namen und Adressen, Geocoding von Adressen usw.

Mit jedem Modul, das Sie lizenzieren, werden Datenflussvorlagen mitgeliefert. Wenn Sie beispielsweise für das „Data Normalization“-Modul lizenziert sind, erhalten Sie die Datenflussvorlage „Personennamen standardisieren“. Wenn Sie für das Universal Addressing-Modul lizenziert sind, erhalten Sie die Datenflussvorlagen „US-amerikanische Adressen überprüfen“ und „Kanadische Adressen überprüfen“.

Je nach Zweck jeder Vorlage kann es sich um einen Auftrag mit Beispieldaten oder einen Dienst ohne Beispieldaten handeln. Sie können Datenflüsse in ihrem Originalstatus verwenden und diejenigen ausführen, die als Aufträge eingehen, um ihre Funktionsweise nachzuvollziehen. Sie können die Datenflüsse alternativ auch manipulieren, indem Sie die Eingabe- und Ausgabedateien

ändern oder Dienste für Ihre eigenen Aufträge verwenden und Eingabe- und Ausgabedateien hinzufügen.

> **Anmerkung:** Diese Beispiele dienen zur Veranschaulichung unterschiedlicher Spectrum™ Technology Platform-Features. Es handelt sich dabei nicht um vollständige Lösungen für Ihre individuelle Geschäftsumgebung.

Erstellen eines Datenflusses mithilfe einer Vorlage

Mit jedem Modul, das Sie lizenzieren, werden Datenflussvorlagen mitgeliefert. Um einen Datenfluss mithilfe einer Vorlage zu erstellen,

- navigieren Sie im Enterprise Designer zu **Datei** > **Neu** > **Datenfluss** > **Aus Vorlage**
- oder klicken Sie alternativ auf das Symbol **Neu** und wählen Sie „Neuer Datenfluss aus Vorlage“ aus.

Eine Liste von für die installierten Module verfügbaren Vorlagen wird angezeigt.

Importieren und Exportieren von Datenflüssen

Mithilfe der Import- und Export-Features können Sie Datenflüsse mit anderen Enterprise Designer-Benutzern austauschen.

> **Anmerkung:** Datenflüsse können nur zwischen identischen Versionen von Spectrum™ Technology Platform ausgetauscht werden.

- Wählen Sie zum Exportieren eines Datenflusses die Optionen **Datei** > **Exportieren** aus. Wenn Sie das Feature „Versionen“ zum Speichern von Versionen des Datenflusses verwendet haben, ist die aktuell ausgewählte Version die Version, die exportiert wird.

  > **Anmerkung:** Verwenden Sie keine Sonderzeichen in den Namen der Dienste und Aufträge, die Sie definieren. Dies könnte zu einem Fehler während des Exports führen.

- Wählen Sie zum Importieren eines Prozessflusses die Optionen **Datei** > **Prozessfluss** > **importieren** aus.
- Wählen Sie zum Importieren eines Datenflusses die Optionen **Datei** > **Datenfluss** > **importieren** aus. Die Schritte im Datenfluss müssen in Ihrem System verfügbar sein, bevor Sie den Datenfluss importieren. Wenn der zu importierende Datenfluss nicht verfügbare Schritte enthält, wird ein Fehler angezeigt.
- Wenn Sie Server-Explorer zur Organisation Ihrer Datenflüsse verwenden, können Sie einen Datenfluss auch exportieren, indem Sie mit der rechten Maustaste darauf klicken und die Option **Exportieren** auswählen. Klicken Sie zum Importieren eines Datenflusses mithilfe von

Server-Explorer mit der rechten Maustaste an die Stelle im Server-Explorer, wo Sie den Datenfluss importieren möchten, und klicken Sie auf **Importieren**.

2 - Entwerfen eines Flusses

In this section

Flusstypen	19
Flusseingabe	21
Felder	27
Steuerungsschritte	44
Flussausgabe	111
Eingebettete Datenflüsse	116
Berichte	122
Überlegungen zur Leistung	126
Datenflussversionen	141

Flusstypen

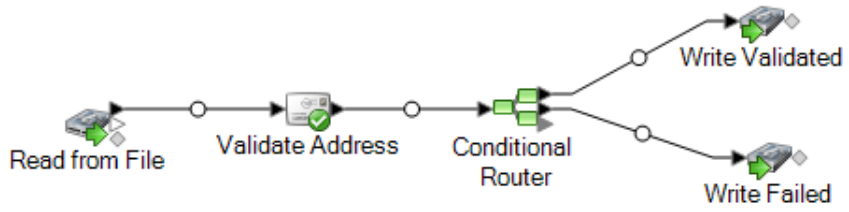
Ein Datenfluss ist eine Reihe von Operationen, die Daten aus einer beliebigen Quelle abrufen, diese Daten verarbeiten und dann die Ausgabe in ein beliebiges Ziel schreiben. Bei der Verarbeitung der Daten kann es um einfaches Sortieren bis hin zu komplexeren Aktionen für Datenqualität und -ergänzung gehen. Das Konzept eines Datenflusses ist einfach, Sie können aber äußerst komplexe Datenflüsse mit Verzweigungspfaden, mehreren Eingabequellen und mehreren Ausgabezielen entwerfen.

Es gibt vier Typen von Datenflüssen: Aufträge, Dienste, Unterflüsse und Prozessflüsse.

Auftrag

Ein Auftrag ist ein Datenfluss, der eine Batch-Verarbeitung ausführt. Ein Auftrag liest Daten aus einer oder mehr Dateien oder Datenbanken, verarbeitet diese Daten und schreibt die Ausgabe in eine oder mehr Dateien oder Datenbanken. Aufträge können manuell im Enterprise Designer oder in einer Befehlszeile mit Job Executor ausgeführt werden.

Der folgende Datenfluss ist ein Auftrag. Beachten Sie, dass er den „Read from File“-Schritt für die Eingabe und zwei „Write to File“-Schritte als Ausgabe nutzt.

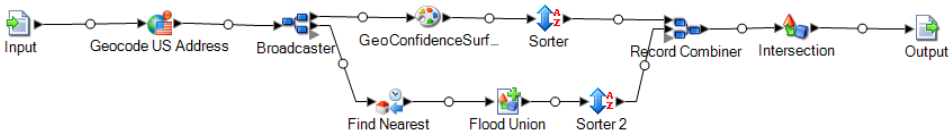


Dienst

Ein Dienst ist ein Datenfluss, auf den Sie in Form von Webservices oder über die Spectrum™ Technology Platform-API zugreifen können. Sie übergeben dem Dienst einen Datensatz und geben optional die bei der Verarbeitung des Datensatzes zu verwendenden Optionen an. Der Dienst verarbeitet die Daten und gibt die Daten zurück.

Manche Dienste werden verfügbar, wenn Sie ein Modul installieren. Wenn Sie z. B. das Universal Addressing-Modul installieren, wird der Dienst „ValidateAddress“ auf Ihrem System verfügbar. In anderen Fällen müssen Sie einen Dienst im Enterprise Designer erstellen und diesen dann auf Ihrem System als benutzerdefinierten Dienst verfügbar machen. Zum Beispiel sind die Schritte des Location Intelligence-Moduls nicht als Dienste verfügbar, sofern Sie nicht zuerst einen Dienst erstellen, der die Schritte des Moduls nutzt.

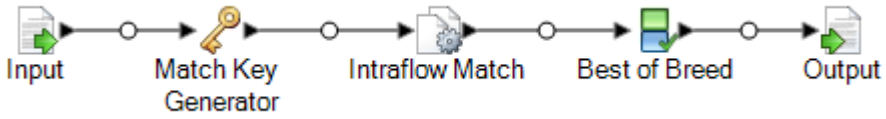
Sie können auch eigene benutzerdefinierte Dienste im Enterprise Designer entwerfen. Zum Beispiel ermittelt der folgende Datenfluss, ob für eine Adresse ein Flooding-Risiko besteht:



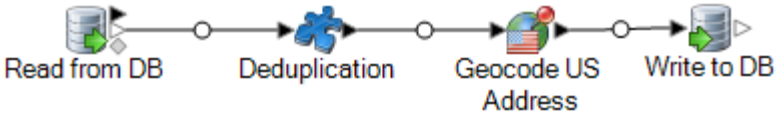
Anmerkung: Da der Dienstname, Optionsname und Feldname letztendlich zu XML-Elementen werden, dürfen sie keine Zeichen enthalten, die in XML-Elementnamen ungültig sind (zum Beispiel sind Leerzeichen nicht erlaubt). Dienste, die dieser Anforderung nicht entsprechen, funktionieren weiterhin, werden aber nicht als Webservices verfügbar gemacht.

Unterfluss

Ein Unterfluss ist ein Datenfluss, der innerhalb anderer Datenflüsse wiederverwendet werden kann. Unterflüsse sind nützlich, wenn Sie einen wiederverwendbaren Prozess erstellen möchten, der ohne großen Aufwand in Datenflüsse eingebunden werden kann. Sie könnten beispielsweise einen Unterfluss erstellen, der eine Deduplizierung ausführt und in jedem Schritt bestimmte Einstellungen nutzt, sodass Sie denselben Deduplizierungsprozess in mehreren Datenflüssen verwenden können. Zu diesem Zweck könnten Sie einen Unterfluss wie diesen erstellen:



Danach könnten Sie diesen Unterfluss in einem Datenfluss verwenden. Sie könnten beispielsweise den Deduplizierungsunterfluss innerhalb eines Datenflusses verwenden, der Geocoding ausführt, sodass die Daten vor der Geocoding-Operation dedupliziert werden:



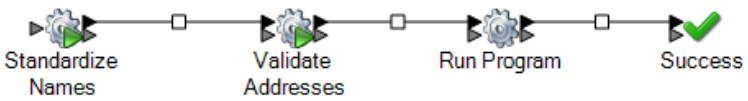
In diesem Beispiel würden Daten aus einer Datenbank eingelesen und dann dem Deduplizierungsunterfluss übergeben, wo sie durch Match Key Generator, dann Intraflow Match, dann Best of Breed verarbeitet und schließlich aus dem Unterfluss heraus in den nächsten Schritt im übergeordneten Datenfluss, in diesem Fall „Geocode US Address“, gesendet würden. Unterflüsse werden durch ein Puzzleteil-Symbol im Datenfluss dargestellt (siehe oben).

Unterflüsse, die gespeichert und verfügbar gemacht werden, werden im Ordner **Benutzerdefinierte Schritte** im Enterprise Designer angezeigt.

Prozessfluss

In einem Prozessfluss werden eine Reihe von Aktivitäten ausgeführt, wie z. B. Aufträge und externe Anwendungen. Jede Aktivität im Prozessfluss wird nach Beendigung der vorherigen Aktivität ausgeführt. Prozessflüsse sind nützlich, wenn Sie mehrere Datenflüsse nacheinander ausführen möchten oder wenn Sie ein externes Programm ausführen möchten. In einem Prozessfluss könnte beispielsweise ein Auftrag zum Standardisieren von Namen und Überprüfen von Adressen ausgeführt werden und anschließend eine externe Anwendung aufgerufen werden, um die Datensätze in die

richtige Reihenfolge zu bringen, damit Portoermäßigungen geltend gemacht werden können. Ein solcher Prozessfluss würde wie folgt aussehen:



In diesem Beispiel wurden die Aufträge „Namen standardisieren“ und „Adressen überprüfen“ auf dem Spectrum™ Technology Platform-Server verfügbar gemacht. Über die Aktivität „Programm ausführen“ wird eine externe Anwendung aufgerufen und die Aktivität „Erfolg“ gibt das Ende des Prozessflusses an.

Flusseingabe

Verwenden Sie zum Definieren der Eingabe für einen Datenfluss den Schritt „Quelle“. Bei einer Quelle handelt es sich um den ersten Schritt in einem Datenfluss. Hier werden die zu verarbeitenden Eingabedaten definiert.

Eingabe für einen Auftrag

Die Eingabedaten für einen Auftrag können aus einer Datei oder Datenbank stammen. Spectrum™ Technology Platform kann Daten aus vielen Dateiformaten und Datenbanktypen lesen. Die Typen von Datenquellen, aus denen Sie lesen können, hängen von den lizenzierten Modulen ab. Das Enterprise Data Integration-Modul bietet Zugriff auf die meisten Datenquellen eines beliebigen Moduls.

> **Anmerkung:** Wenn Sie einen Auftrag erstellen, sollten Sie stets die Möglichkeit falsch formatierter Eingabedatensätze in Betracht ziehen. Ein falsch formatierter Datensatz ist ein Datensatz, der nicht mit einer der Parser-Klassen von Spectrum™ Technology Platform geparst werden kann. Informationen zur Handhabung falsch formatierter Eingabedatensätze finden Sie unter **Verwalten von falsch formatierten Eingabedatensätzen** auf Seite 22.

Eingabe für einen Dienst

Eingabedaten für einen Dienst werden in einem Input-Schritt definiert. In diesem Schritt werden die Felder definiert, die der Dienst von einer Webservice-Anforderung oder einem API-Aufruf zulässt.

Definieren der Auftragseingabe

Die Eingabedaten für einen Auftrag können abhängig von den von Ihnen lizenzierten Modulen aus einer Datei, einer Datenbank oder einem Cloud-Dienst stammen. Jedes Modul unterstützt Eingaben

aus unterschiedlichen Quellen, und die Prozeduren zum Konfigurieren der einzelnen Typen von Quellen unterscheiden sich deutlich. Weitere Informationen finden Sie in den Lösungshandbüchern für Ihre Module unter **support.pb.com/spectrum**.

Verwalten von falsch formatierten Eingabedatensätzen

Ein falsch formatierter Datensatz ist ein Datensatz, den Spectrum™ Technology Platform nicht parsen kann. Wenn Spectrum™ Technology Platform auf einen falsch formatierten Datensatz trifft, kann es eine oder mehrere der folgenden Aktionen ausführen:

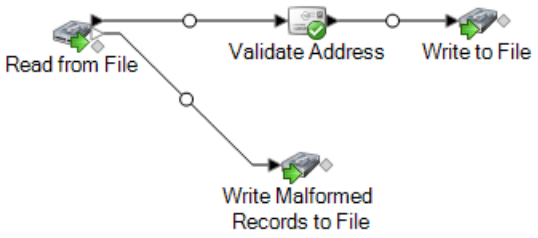
- Auftrag beenden
- Verarbeitung fortsetzen
- Verarbeitung fortsetzen, bis eine bestimmte Anzahl defekter Datensätzen vorgefunden wurde
- Verarbeitung fortsetzen, aber defekte Datensätze in eine Protokolldatei schreiben (über einen optionalen Datenladeschritt)

Anmerkung: Die Funktionalität falsch formatierter Datensätze ist auf Quellen beschränkt, die zum Lesen aus lokalen Serverdateien konfiguriert sind und für die keine Sortierung konfiguriert ist. Wenn eine Quelle entweder mit einer Remote-Datei oder mit Sortierfeldern konfiguriert ist und die Quelle auf einen falsch formatierten Datensatz trifft, wird der Auftrag unabhängig von der Konfiguration für falsch formatierte Datensätze beendet.

So verwalten Sie falsch formatierte Datensätze:

1. Öffnen Sie den Auftrag im Enterprise Designer.
2. Fügen Sie eine Datenladung falsch formatierter Datensätze in Ihrem Datenfluss hinzu.
   a) Erstellen Sie Ihren Auftrag, indem Sie Ihre Eingabedatei und Quellschritte definieren und Dienste und Unterflüsse zu Ihrem Datenfluss hinzufügen.
   b) Führen Sie eine der folgenden Aktionen aus:
      - Verbinden Sie einen Datenladeschritt mit dem optionalen Ausgabeport des Quellschrittes in Ihrem Datenfluss. Der optionale Port ist der helle Ausgabeport direkt unter dem schwarzen Ausgabeport Ihres Quellschrittes. Wenn Sie den Zeiger auf diesen Port bewegen, sehen Sie die QuickInfo „error_port“. Falsch formatierte Datensätze werden in diese Datenladung geschrieben.
      - Verbinden Sie nichts mit dem optionalen Ausgabeport des Quellschrittes in Ihrem Datenfluss, was dazu führt, dass Spectrum™ Technology Platform falsch formatierte Datensätze ignoriert.

   Der abgeschlossene Datenfluss sollte in etwa wie folgt aussehen:



Wenn Sie Ihren Auftrag ausführen, enthält der Ausführungsverlauf eine Spalte, die die Anzahl falsch formatierter Datensätze zeigt, die während des Auftrags gefunden wurden.

3. Standardmäßig beendet Spectrum™ Technology Platform einen Auftrag, wenn es auf einen falsch formatierten Datensatz trifft. Dieses Standardverhalten kann in der Management Console geändert werden. Unabhängig vom Standardverhalten Ihres Systems können Sie das Standardverhalten für einen Auftrag mit diesen Schritten überschreiben:
   a) Öffnen Sie den Auftrag im Enterprise Designer.
   b) Rufen Sie innerhalb eines offenen Auftrags **Bearbeiten** > **Auftragsoptionen** auf.
   c) Wählen Sie entweder **Auftrag bei falsch formatiertem Datensatz nicht beenden** oder **Auftrag nach Auftreten dieser Anzahl falsch formatierter Datensätze beenden** und geben Sie die Anzahl falsch formatierter Datensätze ein, auf die ein Auftrag treffen darf, bevor er beendet wird.

Definieren der Diensteingabe

Der Input-Schritt definiert die Eingabefelder für einen Dienst oder Unterfluss. Sie definiert auch Testdaten, die während der Datenüberprüfung verwendet werden sollen.

Registerkarte „Eingabefelder“

Diese Registerkarte listet die Felder auf, die der Datenfluss als Eingabe akzeptiert. Wenn der Input-Schritt mit einem anderen Schritt im Datenfluss verbunden ist, wird eine Liste von Feldern angezeigt, die von den Schritten im Datenfluss verwendet werden. Weitere Informationen finden Sie unter **Definieren von Eingabefeldern für einen Dienst oder Unterfluss** auf Seite 23.

Registerkarte „Überprüfungseingabe“

Diese Registerkarte erlaubt Ihnen, Testeingabedatensätze anzugeben, die mit dem Datenüberprüfungstool verwendet werden sollen. Weitere Informationen zur Datenüberprüfung finden Sie unter **Überprüfen eines Datenflusses** auf Seite 147.

Definieren von Eingabefeldern für einen Dienst oder Unterfluss

Um die Eingabefelder für einen Dienst oder Unterfluss zu definieren, verwenden Sie den Input-Schritt.

Anmerkung: Wenn Sie hierarchische Daten in den Eingabefeldern definieren, können Sie keine Daten importieren oder vertikal anzeigen.

1. Ziehen Sie einen Input-Schritt auf die Arbeitsfläche.
2. Verbinden Sie diesen Schritt mit dem folgenden Schritt im Datenfluss.
3. Doppelklicken Sie auf den Input-Schritt.
4. Wählen Sie die Felder aus, die Sie zur Eingabe verwenden möchten. Welche Felder in der Liste angezeigt werden, hängt von dem Schritt ab, mit dem der Input-Schritt verbunden ist.
5. Möchten Sie ein neues Feld zur Feldliste hinzufügen, klicken Sie auf **Hinzufügen**. Das Fenster **Benutzerdefiniertes Feld hinzufügen** wird angezeigt.
6. Klicken Sie erneut auf **Hinzufügen**.
7. Geben Sie im Feld **Feldname** den Namen ein, den Sie für dieses Feld verwenden möchten.
8. Wählen Sie den Datentyp aus.

   Die folgenden Datentypen werden unterstützt:

   bigdecimal Ein numerischer Datentyp, der eine Genauigkeit von 38 Dezimalstellen unterstützt. Verwenden Sie diesen Datentyp für Daten in mathematischen Berechnungen, die einen hohen Grad von Genauigkeit erfordern, speziell solche mit Finanzdaten. Der Datentyp „bigdecimal“ unterstützt genauere Berechnungen als der Datentyp „double“.

   boolean Ein logischer Typ mit zwei Werten: wahr und falsch.

   bytearray Ein Array (Liste) von Bytes.

   Anmerkung: Bytearray wird nicht als Eingabe für einen REST-Dienst unterstützt.

   date Ein Datentyp, der einen Monat, einen Tag und ein Jahr enthält. Beispiel: 2012-01-30 oder 30. Januar 2012. Sie können ein standardmäßiges Datumsformat in der Management Console angeben.

   datetime Ein Datentyp, der Tag, Monat, Jahr und Stunden, Minuten und Sekunden enthält. Beispiel: 30.01.2012 18:15:00.

   double Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit doppelter Genauigkeit zwischen 2^{-1074} und $(2-2^{-52})\times2^{1023}$ enthält. In der E-Notation liegt der Wertebereich bei -1,79769313486232E+308 bis 1,79769313486232E+308.

   float Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit einzelner Genauigkeit zwischen 2^{-149} und $(2-2^{23})\times2^{127}$ enthält. In der E-Notation liegt der Wertebereich bei -3,402823E+38 bis 3,402823E+38.

   integer Ein numerischer Datentyp, der sowohl positive als auch negative ganze Zahlen zwischen -2^{31} (-2,147,483,648) und $2^{31}-1$ (2,147,483,647) enthält.

list Streng genommen ist eine Liste kein Datentyp. Wenn jedoch ein Feld hierarchische Daten enthält, wird es als „list“-Feld behandelt. In Spectrum™ Technology Platform ist eine Liste eine Sammlung von Daten und besteht aus mehreren Werten. So kann etwa ein Feld „Name“ eine Liste von Namenswerten enthalten. Das kann in einer XML-Struktur wie folgt aussehen:

```
<Names>
    <Name>John Smith</Name>
    <Name>Ann Fowler</Name>
</Names>
```

Beachten Sie, dass sich der Spectrum™ Technology Platform-Datentyp „list“ vom Datentyp „list“ des XML-Schemas unterscheidet. Der XML-Datentyp „list“ ist ein einfacher Datentyp, der aus mehreren Werten besteht, während der Spectrum™ Technology Platform-Datentyp „list“ einem komplexen XML-Datentyp ähnelt.

long Ein numerischer Datentyp, der sowohl negative als auch positive ganze Zahlen zwischen -2^{63} (-9.223.372.036.854.775.808) und $2^{63}-1$ (9.223.372.036.854.775.807) enthält.

string Eine Folge von Zeichen.

time Ein Datentyp, der die Tageszeit enthält. Beispiel: 21:15:59 oder 9:15:59 PM.

Sie können bei Bedarf auch einen neuen, benutzerdefinierten Datentyp hinzufügen, und dieser neue Typ kann eine Liste eines beliebigen definierten Datentyps sein. Sie könnten beispielsweise eine Liste von Namen (Zeichenfolge) oder einen neuen Datentyp von Adressen mit „AddressLine1“ (Zeichenfolge), „City“ (Zeichenfolge), „StateProvince“ (Zeichenfolge) und „PostalCode“ (Zeichenfolge) definieren. Nachdem Sie das Feld erstellt haben, können Sie den Datentyp anzeigen, indem Sie auf das Dialogfeld „Eingabeoptionen“ zugreifen und auf die Schaltfläche in der Spalte „Datentyp“ klicken. Das Dialogfeld **Datentypdetails** wird angezeigt. Es enthält die Struktur des Feldes.

9. Klicken Sie erneut auf **OK**.
10. Aktivieren Sie in der Spalte **Verfügbar machen** das Kästchen, um das Feld für Schrittvorgänge verfügbar zu machen. Wenn Sie das Kästchen deaktivieren und auf **OK** klicken, wird das Feld aus der Feldliste gelöscht.
11. Im Feld **Datentypname** wird der Standardelementname angezeigt, der für Eingabedatensätze in SOAP- und REST-Webdienstanforderungen an diesen Dienst verwendet werden soll. Der Standardwert ist `Row`. Wenn Sie einen anderen Elementnamen für Eingabedatensätze verwenden möchten, geben Sie ihn hier ein.

    Beispiel: Mit dem Standardwert `Row` würde eine JSON-Webdienstanforderung `Row` als Elementname für den Eingabedatensatz verwenden (siehe Abbildung):

```
{
  "Input":
```

```
  {
    "Row": [
      {
        "AddressLine1": "1825 Kramer Ln",
        "City": "Austin",
        "StateProvince": "TX"
      }
    ]

  }
}
```

Wenn Sie den Wert im Feld **Datentypname** in `Address` ändern möchten, müsste die JSON-Anforderung `Address` statt `Row` als Elementnamen für den Datensatz verwenden (siehe Abbildung):

```
{
  "Input":
  {
    "Address": [
      {
        "AddressLine1": "1825 Kramer Ln",
        "City": "Austin",
        "StateProvince": "TX"
      }
    ]

  }
}
```

DeDefinieren eines Webservice-Datentyps

Im Feld **Datentypname** können Sie die Schnittstellen von WSDL (SOAP) und WADL (REST) für den Dienst, den Sie erstellen, steuern. Der Name des Elements „Spalten“ wird durch den Namen festgelegt, den Sie diesem Schritt im Dienst zuweisen. Der Name des Elements „Spalte“ wird durch den Text festgelegt, den Sie hier eingeben.

> **Anmerkung:** Bei WSDL betrifft dies sowohl Anforderungen als auch Antworten, bei WADL sind jedoch nur Antworten betroffen.

Bevor Sie diesen Schritt benennen und Text in dieses Feld eingeben, könnte Ihr Code folgendermaßen aussehen:

```
<Rows>
     <Row>
          <FirstName>John</FirstName>
          <LastName>Doe</LastName>
     </Row>
     <Row>
```

```
          <FirstName>Jane</FirstName>
    <LastName>Doe></LastName>
     </Row>
</Rows>
```

Nachdem Sie diesen Schritt benannt und Text in dieses Feld eingeben haben, könnte Ihr Code folgendermaßen aussehen:

```
<Names>
     <Name>
          <FirstName>John</FirstName>
          <LastName>Doe</LastName>
     </Name>
     <Name>
          <FirstName>Jane</FirstName>
          <LastName>Doe></LastName>
     </Name>
</Names>
```

Felder

Flache und hierarchische Daten

Spectrum™ Technology Platform unterstützt flache und hierarchische Daten. Im Allgemeinen können Sie entweder flache oder hierarchische Daten als Eingabe und Ausgabe für einen Datenfluss verwenden. Einige Schritte im Enterprise Routing-Modul erfordern Daten in einem hierarchischen Format.

Flache Daten

Flache Daten bestehen aus Datensätzen, die in Zeilen angeordnet sind. Diese Datensätze enthalten Felder. Felder werden durch ein bestimmtes Zeichen begrenzt oder an einer definierten Position in der Zeile platziert. Dies sind z. B. flache Daten mit kommabegrenzten Feldern:

```
Sam,43,United States
Jeff,32,Canada
Mary,61,Ireland
```

Zum Lesen von flachen Daten in einem Datenfluss können Sie die Schritte „Read From File“, „Read From DB“ oder „Input“ verwenden. Zum Schreiben von flachen Daten aus einem Datenfluss können Sie die Schritte „Write to File“, „Write to DB“ oder „Output“ verwenden.

Hierarchische Daten

Hierarchische Daten sind mit Datenelementen in einer Baumstruktur angeordnet, die übergeordnete und untergeordnete Beziehungen enthält. Spectrum™ Technology Platform kann hierarchische Daten im XML- und Variable Format File-Format lesen und schreiben. Dies sind z. B. hierarchische Daten im XML-Format:

```
<customers>
 <customer>
  <name>Sam</name>
  <age>43</age>
  <country>United States</country>
 </customer>
 <customer>
  <name>Jeff</name>
  <age>32</age>
  <country>Canada</country>
 </customer>
 <customer>
  <name>Mary</name>
  <age>61</age>
  <country>Ireland</country>
 </customer>
</customers>
```

Dieses Beispiel zeigt eine Struktur, in der `<customer>` für einen Datensatz steht. Jeder Datensatz besteht aus einfachen XML-Elementen (`<name>`,`<age>`, und `<country>`).

Konvertieren von Daten

Es gibt viele Situationen, in denen Sie eventuell flache Daten in hierarchische Daten (oder umgekehrt) konvertieren müssen. So können Sie beispielsweise über eine Datenflusseingabe im hierarchischen Format verfügen, es ist jedoch notwendig, dass der Datenfluss flache Daten ausgibt. Sie müssen vielleicht außerdem für bestimmte Schritte flache Eingabedaten in hierarchische Daten konvertieren (besonders bei Schritten im Location Intelligence-Modul). Anschließend müssen die Daten eventuell für die Ausgabe zurück in das flache Format konvertiert werden.

Gehen Sie wie folgt vor, um die Daten vom flachen in das hierarchische Format zu konvertieren:

- Das Tool „Prozessliste“
- Der Aggregator-Schritt in einem Datenfluss

Verwenden Sie den Splitter-Schritt, um hierarchische Daten in flache Daten zu konvertieren.

Konvertieren von flachen Daten in eine Liste

„Liste verarbeiten“ ist ein Tool, das Sie innerhalb eines Dienstes oder Unterflusses verwenden können, um flache Daten in eine Liste umzuwandeln. Dies ist nützlich, wenn Ihre Datenflüsse Schritte enthalten, die eine Listeneingabe erfordern, so wie die im Location Intelligence-Modul.

1. Klicken Sie bei einem vorhandenen Datenfluss mit der rechten Maustaste auf den Schritt, dessen Ausgabe Sie in eine Liste konvertieren möchten. Dies könnte jeder Schritt außer „Input“ oder „Output“ sein.
2. Wählen Sie **Liste verarbeiten** aus. Sie sehen die Schritte innerhalb eines blauen quadratischen Hintergrunds.
3. Um einen Schritt in die bzw. aus der Prozessliste zu verschieben, drücken Sie die **Umschalttaste**, während Sie den zusätzlichen Schritt ziehen.

   Anmerkung: Wenn Sie die Daten mehrerer Schritte mit „Liste verarbeiten“ verarbeiten möchten, können Sie einen Unterfluss erstellen, diesen in Ihren Datenfluss aufnehmen und dann das Feature „Liste verarbeiten“ auf den Unterfluss als Ganzes anwenden.

4. Die Eingabe- und Ausgabefelder einer Prozessliste werden als „ListField“ bezeichnet. Ordnen Sie mithilfe der Funktion „Felder umbenennen“ Ihr Eingabeschrittfeld „ListField“ im Eingabekanal zu und ordnen Sie „ListField“ Ihrem Ausgabeschrittfeld zu. Weitere Informationen finden Sie unter **Ändern eines Feldnamen** auf Seite 43.
5. Sollen in der Liste die Daten in derselben Reihenfolge beibehalten werden, in der sie eingegeben wurden, klicken Sie mit der rechten Maustaste auf das Kästchen „Liste verarbeiten“ und wählen Sie „Optionen“ aus. Aktivieren Sie dann das Kästchen „Sortierreihenfolge beibehalten“.
6. Um sich zu vergewissern, dass die in den nächsten Schritt eingegebenen Daten als Liste formatiert werden, überprüfen Sie den Datenfluss. Weitere Informationen zum Überprüfen von Daten finden Sie unter **Überprüfen eines Datenflusses** auf Seite 147.

Datentypen

Spectrum™ Technology Platform unterstützt eine Vielzahl von numerischen und komplexen Datentyp sowie Zeichenfolgedatentypen. Je nach Art der auszuführenden Verarbeitung können Sie einen oder mehrere dieser Datentypen verwenden. Für einen Adressenüberprüfungsdatenfluss müssen Sie eventuell nur Zeichenfolgendaten verwenden. Für Datenflüsse, die mathematische Berechnungen beinhalten, können Sie numerische und boolesche Datentypen verwenden. Für Datenflüsse, die räumliche Verarbeitungen ausführen, können Sie einen komplexen Datentyp verwenden. Für Datenflüsse, die diese Verarbeitungen kombinieren, können Sie eine Vielzahl von Datentypen verwenden.

Spectrum™ Technology Platform unterstützt die folgenden Datentypen.

bigdecimal Ein numerischer Datentyp, der eine Genauigkeit von 38 Dezimalstellen unterstützt. Verwenden Sie diesen Datentyp für Daten in mathematischen Berechnungen, die einen hohen Grad von Genauigkeit erfordern, speziell solche mit Finanzdaten. Der Datentyp „bigdecimal“ unterstützt genauere Berechnungen als der Datentyp „double“.

boolean Ein logischer Typ mit zwei Werten: wahr und falsch.

bytearray Ein Array (Liste) von Bytes.

> **Anmerkung:** Bytearray wird nicht als Eingabe für einen REST-Dienst unterstützt.

date Ein Datentyp, der einen Monat, einen Tag und ein Jahr enthält. Beispiel: 2012-01-30 oder 30. Januar 2012. Sie können ein standardmäßiges Datumsformat in der Management Console angeben.

datetime Ein Datentyp, der Tag, Monat, Jahr und Stunden, Minuten und Sekunden enthält. Beispiel: 30.01.2012 18:15:00.

double Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit doppelter Genauigkeit zwischen 2^{-1074} und $(2-2^{-52})\times2^{1023}$ enthält. In der E-Notation liegt der Wertebereich bei -1,79769313486232E+308 bis 1,79769313486232E+308.

float Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit einzelner Genauigkeit zwischen 2^{-149} und $(2-2^{23})\times2^{127}$ enthält. In der E-Notation liegt der Wertebereich bei -3,402823E+38 bis 3,402823E+38.

integer Ein numerischer Datentyp, der sowohl positive als auch negative ganze Zahlen zwischen -2^{31} (-2,147,483,648) und $2^{31}-1$ (2,147,483,647) enthält.

list Streng genommen ist eine Liste kein Datentyp. Wenn jedoch ein Feld hierarchische Daten enthält, wird es als „list"-Feld behandelt. In Spectrum™ Technology Platform ist eine Liste eine Sammlung von Daten und besteht aus mehreren Werten. So kann etwa ein Feld „Name" eine Liste von Namenswerten enthalten. Das kann in einer XML-Struktur wie folgt aussehen:

```
<Names>
    <Name>John Smith</Name>
    <Name>Ann Fowler</Name>
</Names>
```

Beachten Sie, dass sich der Spectrum™ Technology Platform-Datentyp „list" vom Datentyp „list" des XML-Schemas unterscheidet. Der XML-Datentyp „list" ist ein einfacher Datentyp, der aus mehreren Werten besteht, während der Spectrum™ Technology Platform-Datentyp „list" einem komplexen XML-Datentyp ähnelt.

long Ein numerischer Datentyp, der sowohl negative als auch positive ganze Zahlen zwischen -2^{63} (-9.223.372.036.854.775.808) und $2^{63}-1$ (9.223.372.036.854.775.807) enthält.

string Eine Folge von Zeichen.

time Ein Datentyp, der die Tageszeit enthält. Beispiel: 21:15:59 oder 9:15:59 PM.

Festlegen des Datentyps eines Feldes

Sie können den Datentyp für ein Feld in den folgenden Situationen festlegen:

- **Quellschritte:** Durch das Festlegen der Datentypen können Sie den Datentyp am Anfang eines Datenflusses einstellen, und dadurch die Notwendigkeit von Datentypkonvertierungen später im

Datenfluss eliminieren. Beachten Sie, dass für den „Read from DB“-Schritt der Datentyp automatisch ausgewählt wird und nicht geändert werden kann.

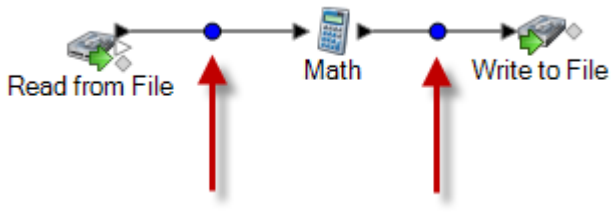
- **Datenladeschritte:** Durch Festlegen von Datentypen können Sie das Datenformat steuern, das durch den Datenfluss zurückgegeben wird. Beachten Sie, dass für den „Write to DB“-Schritt der Datentyp automatisch ausgewählt wird und nicht geändert werden kann.
- **Transformer-Schritt:** Sie können die Datentypen in diesem Schritt festlegen, wenn Sie ein benutzerdefiniertes Skript verwenden.
- **Math-Schritt und „Group Statistics“-Schritt:** Da diese Schritte mathematische Berechnungen ausführen, kann die Auswahl eines bestimmten numerischen Datentyps die Ergebnisse der Berechnungen beeinflussen, wie z. B. die Genauigkeit eines Teilungsvorgangs. Wenn Sie einen Datentyp für ein Feld festlegen, das sich vom Datentyp des in den Schritt eintretenden Feldes unterscheidet, konvertiert der Downstream-Kanal das Feld in den von Ihnen festgelegten Datentyp, wie unter **Automatische Datentypkonvertierung** auf Seite 31 beschrieben.

  Anmerkung: Jeder Schritt unterstützt unterschiedliche Datentypen. Eine Beschreibung der unterstützten Datentypen für jeden Schritt finden Sie in der Dokumentation des jeweiligen Schrittes.

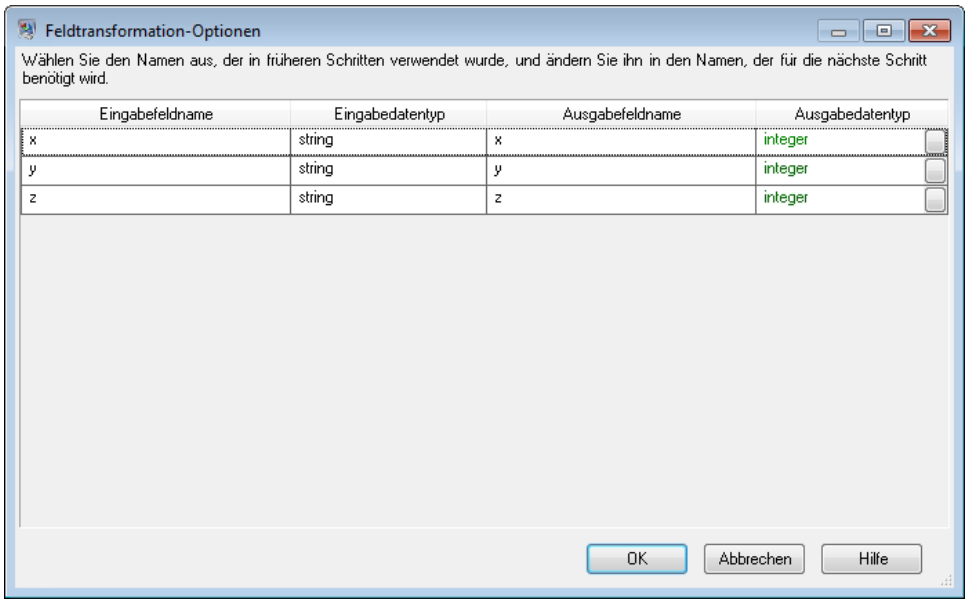
Automatische Datentypkonvertierung

Wenn die in einem Schritt angezeigten Daten einen falschen Typ aufweisen, kann Spectrum™ Technology Platform in manchen Fällen automatisch die Daten in den benötigten Typ konvertieren. So akzeptiert „Validate Address“ z. B. nur Zeichenfolgendaten als Eingabe. Wenn das Eingabefeld „PostalCode“ eine ganze Zahl erfordert, kann Spectrum™ Technology Platform automatisch das Feld in eine Zeichenfolge konvertieren und das Feld „PostalCode“ erfolgreich verarbeiten. Gleichmaßen benötigt der Math-Schritt Daten in numerischer Form. Wenn die eingehenden Daten eine Zeichenfolge sind, kann Spectrum™ Technology Platform die Daten in den Datentyp konvertieren, der auf der Registerkarte „Felder“ des Math-Schrittes angegeben ist.

Automatische Datentypkonvertierungen erfolgen in den Kanälen eines Datenflusses. Wenn ein Kanal erfolgreich einen Datentyp konvertiert, wird ein blauer Punkt in der Mitte des Kanals angezeigt:

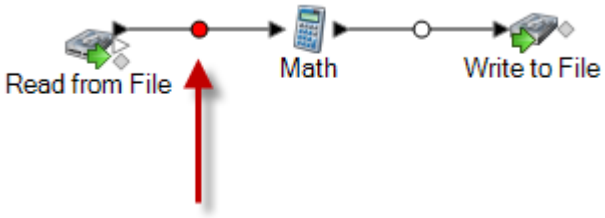


Wenn Sie auf den Kanal doppelklicken, wird die stattfindende Datentypkonvertierung angezeigt. In diesem Fall werden die Zeichenfolgendaten in ganzzahlige Daten konvertiert:



Beachten Sie, dass Sie den Datentyp im Dialogfeld für automatische Datentypkonvertierungen nicht ändern können. Der Ausgabedatentyp wird durch die Einstellungen im Downstream-Schritt bestimmt.

Wenn Felder keine gültigen Werte enthalten oder nicht konvertiert werden können, wird ein roter Kreis im Kanal angezeigt.



Mithilfe der Typkonvertierungsoptionen können Sie angeben, was der Datenfluss tun soll, wenn die Konvertierung fehlschlägt.

Einstellen von Optionen zur Datentypkonvertierung für einen Datenfluss

Die Datentypkonvertierung erfolgt, wenn ein Datenfluss automatisch ein Feld in den für einen Schritt erforderlichen Datentyp konvertiert. Die Datentypkonvertierung erfolgt ebenfalls innerhalb einiger Schritte. Im „Read from DB“-Schritt können Sie z. B. auswählen, dass ein Feld den Zeichenfolgendatentyp verwendet, obwohl die Quelldaten einen numerischen Datentyp ausweisen. Die Daten werden in das Zeichenfolgenformat konvertiert, wenn sie in den Datenfluss eingelesen werden.

Es gibt zwei Einstellungen, mit denen Sie die Datentypkonvertierungen steuern können. Zunächst gibt es Einstellungen, die bestimmen, wie die in eine Zeichenfolge konvertierten numerischen, Datums- und Zeitdaten formatiert werden. So können Sie z. B. Datumsangaben, die in eine Zeichenfolge konvertiert wurden, im Format mm/tt/jjjj anstatt tt/mm/jjjj darstellen. Die andere Einstellung steuert, was passieren soll, wenn das System ein Feld nicht von einem Datentyp in den anderen konvertieren kann.

Die standardmäßigen Datentypkonvertierungseinstellungen für Ihr System werden in der Management Console festgelegt. Sie können die Standardformate für einzelne Datenflüsse im Enterprise Designer überschreiben.

Diese Prozedur beschreibt, wie Sie standardmäßige Datentypkonvertierungsoptionen für einen Datenfluss überschreiben.

> **Anmerkung:** Unterflüsse erben die Typkonvertierungseinstellungen von dem Datenfluss, in dem sie sich befinden. Sie können die Typkonvertierungseinstellungen für Unterflüsse nicht festlegen.

1. Öffnen Sie den Datenfluss im Enterprise Designer.
2. Wählen Sie **Bearbeiten** > **Typkonvertierungsoptionen** aus.
3. Aktivieren Sie das Kästchen **System-Standardoptionen mit den folgenden Werten überschreiben**.
4. Geben Sie im Feld **Fehlerbehandlung** ein, was zu tun ist, wenn ein Feldwert nicht automatisch in den von einem Schritt erforderlichen Datentyp konvertiert werden kann.

Datenfluss als ungültig erklären	Wenn ein Feld nicht konvertiert werden kann, schlägt der Datenfluss fehl.
Datensatz als ungültig erklären	Wenn ein Feld nicht konvertiert werden kann, schlägt der Datensatz fehl, aber der Datenfluss wird weiterhin ausgeführt.
Das Feld mit Standardwerten initialisieren	Wenn ein Feld nicht konvertiert werden kann, wird der Feldwert durch einen Wert ersetzt, den Sie hier angeben. Diese Option ist nützlich, wenn Sie wissen, dass manche Datensätze fehlerhafte Daten enthalten und Sie die fehlerhaften Daten durch einen Standardwert ersetzen möchten. Geben Sie einen Wert für jeden Datentyp an.

5. Geben Sie die Formate an, die Sie für Datums- und Zeitdaten verwenden möchten, die in eine Zeichenfolge konvertiert werden. Wenn das Datum oder die Zeit in eine Zeichenfolge konvertiert wird, hat die Zeichenfolge das von Ihnen hier festgelegte Format.
   a) Wählen Sie im Feld **Gebietsschema** das Land aus, dessen Format Sie für Datumsangaben verwenden möchten, die in eine Zeichenfolge konvertiert werden. Ihre Auswahl bestimmt die Standardwerte in den Feldern **Datum**, **Zeit**, und **DateTime**. Ihre Auswahl bestimmt außerdem die verwendete Sprache, wenn ein Monat ausgeschrieben wird. Wenn Sie beispielsweise „Englisch“ auswählen, wäre der erste Monat des Jahres „January“, wenn Sie aber „Französisch“ angeben, wäre es „Janvier“.
   b) Wählen Sie im Feld **Datum** das Format für Datumsangaben aus, wenn diese in eine Zeichenfolge konvertiert werden. Es wird eine Liste der am häufigsten verwendeten Formate für das ausgewählte Gebietsschema bereitgestellt.

      Wenn Sie das Format **M/D/YY** auswählen, und ein Datumsfeld den Wert „2012-3-2“ enthält, würden diese Datumsangaben in die folgende Zeichenfolge konvertiert werden: `3/2/12`.

c) Wählen Sie im Feld **Zeit** das Format für Zeitangaben aus, wenn diese in eine Zeichenfolge konvertiert werden. Es wird eine Liste der am häufigsten verwendeten Formate für das ausgewählte Gebietsschema bereitgestellt.

   Wenn Sie das Format **h:mm a** auswählen, und ein Datumsfeld den Wert „23:00“ enthält, würden diese Zeitangaben in die folgende Zeichenfolge konvertiert werden: `11:00 PM`.

d) Wählen Sie im Feld **DateTime** das Format für die Felder mit DateTime-Datentyp aus, wenn diese in eine Zeichenfolge konvertiert werden. Es wird eine Liste der am häufigsten verwendeten Formate für das ausgewählte Gebietsschema bereitgestellt.

   Wenn Sie das Format **M/d/yy h:mm a** auswählen, und ein DateTime-Feld den Wert „2012-3-2 23:00“ enthält, würden diese DateiTime-Angaben in die folgende Zeichenfolge konvertiert werden: `3/2/12 11:00 PM`.

e) Wählen Sie im Feld **Ganze Zahlen** die Formatierung aus, die Sie für ganze Zahlen (float- und double.Datentypen) verwenden möchten.

   Wenn Sie beispielsweise das Format **#.###** auswählen, wird die Zahl 4324 wie folgt formatiert:`4.324`.

   > **Anmerkung:** Wenn Sie dieses Feld leer lassen, werden die Zahlen so wie in Spectrum™ Technology Platform 8.0 oder früher formatiert. Es wird insbesondere kein Tausendertrennzeichen verwendet. Der Punkt („.“) wird als Dezimalzeichen verwendet. Zahlen, die kleiner als 10^{-3} oder größer gleich 10^{7} sind, werden in wissenschaftlicher Schreibweise angezeigt. Negative Zahlen tragen ein vorangestellte Minus („-“). Beachten Sie auch, dass wenn Sie das Feld leer lassen, Zahlen mit dem bigdecimal-Datentyp immer das Format #,###.000 aufweisen.

f) Wählen Sie im Feld **Dezimalzahlen** die Formatierung für Zahlen aus, die einen Dezimalwert enthalten (integer- und long-Datentypen).

   Wenn Sie beispielsweise das Format **#.##0,0#** auswählen, wird die Zahl 4324,25 wie folgt formatiert:`4.324,25`.

   > **Anmerkung:** Wenn Sie dieses Feld leer lassen, werden die Zahlen so wie in Spectrum™ Technology Platform 8.0 oder früher formatiert. Es wird insbesondere kein Tausendertrennzeichen verwendet. Der Punkt („.“) wird als Dezimalzeichen verwendet. Zahlen, die kleiner als 10^{-3} oder größer gleich 10^{7} sind, werden in wissenschaftlicher Schreibweise angezeigt. Negative Zahlen tragen ein vorangestellte Minus („-“). Beachten Sie auch, dass wenn Sie das Feld leer lassen, Zahlen mit dem bigdecimal-Datentyp immer das Format #,###.000 aufweisen.

Sie können außerdem Ihre eigenen Datums-, Zeit- und Zahlenformate festlegen, wenn die verfügbare Auswahl nicht Ihren Anforderungen entspricht. Geben Sie zur Festlegung eines eigenen Datums- oder Zeitformats das Format mithilfe der unter **Datums- und Zeitmuster** auf Seite 35 erläuterten Schreibweise in das Feld ein. Geben Sie zur Festlegung eines eigenen Zahlenformats das Format mithilfe der unter **Zahlenmuster** auf Seite 38 erläuterten Schreibweise in das Feld ein.

6. Wählen Sie unter **NULL-Behandlung** aus, was zu tun ist, wenn ein Feld, das eine Typkonvertierung erfordert, einen Nullwert enthält. Wenn Sie eine der folgenden Optionen auswählen, schlägt entweder der Datenfluss oder der Datensatz mit dem Nullwert basierend auf der Auswahl von **Datenfluss als ungültig erklären** oder **Datensatz als ungültig erklären** unter **Typkonvertierungsfehler** fehl.

NULL-Zeichenfolge als ungültig erklären	„Datenfluss/Datensatz als ungültig erklären“ gilt, wenn die Typkonvertierung für ein Zeichenfolgenfeld benötigt wird, das einen Nullwert enthält.
Booleschen NULL-Wert als ungültig erklären	„Datenfluss/Datensatz als ungültig erklären“ gilt, wenn die Typkonvertierung für ein boolesches Feld benötigt wird, das einen Nullwert enthält.
Numerischen NULL-Wert als ungültig erklären	„Datenfluss/Datensatz als ungültig erklären“ gilt, wenn die Typkonvertierung für ein numerisches Feld benötigt wird, das einen Nullwert enthält. Zu den numerischen Feldern gehören double-, float-, long-, integer- und bigdecimal-Felder.
NULL-Datum als ungültig erklären	Datenfluss/Datensatz als ungültig erklären gilt, wenn die Typkonvertierung für ein Datumsfeld benötigt wird, das einen Nullwert enthält. Dazu gehören Datums-, Zeit- und DateTime-Felder.

Datums- und Zeitmuster

Beim Definieren von Datentypoptionen für Datums- und Zeitdaten können Sie Ihr eigenes benutzerdefiniertes Datums- und Zeitmuster erstellen, falls die vordefinierten Muster nicht Ihren Anforderungen entsprechen. Verwenden Sie zum Erstellen eines Datums- oder Zeitmusters das in der folgenden Tabelle beschriebene Format. Das Beispielmuster:

dd MMMM yyyy

Würde ein Datum in der folgenden Schreibweise zurückgeben:

14 Dezember 2012

Buchstabe	Beschreibung	Beispiel
G	Epochenkennung	AD
yy	Zweistelliges Jahr	96
yyyy	Vierstelliges Jahr	1996
M	Numerischer Monat des Jahres.	7

Buchstabe	Beschreibung	Beispiel
MM	Numerischer Monat des Jahres. Wenn die Zahl kleiner als 10 ist, wird eine Null angefügt, um eine zweistellige Zahl zu erzeugen.	07
MMM	Kurzname des Monats	Jul
MMMM	Langer Name des Monats	Juli
w	Woche des Jahres	27
ww	Zweistellige Woche des Jahres. Wenn die Woche kleiner als 10 ist, wird eine Null angefügt.	06
W	Woche des Monats	2
D	Tag des Jahres	189
DDD	Dreistelliger Tag des Jahres. Wenn die Zahl weniger als drei Stellen enthält, werden Nullen angeführt.	006
d	Tag des Monats	10
dd	Zweistelliger Tag des Monats. Zahlen kleiner als 10 enthalten eine vorausgehende Null.	09
F	Tage der Woche im Monat	2
E	Kurzname des Tages der Woche	Di
EEEE	Langer Name des Tages der Woche	Dienstag
a	Kennzeichnung für Vormittag/Nachmittag	PM
H	Stunde des Tages, wobei die erste Stunde 0 und die letzte Stunde 23 ist.	0
HH	Zweistellige Stunde des Tages, wobei die erste Stunde 0 und die letzte Stunde 23 ist. Zahlen kleiner als 10 enthalten eine vorausgehende Null.	08

Buchstabe	Beschreibung	Beispiel
k	Stunde des Tages, wobei die erste Stunde 1 und die letzte Stunde 24 ist.	24
kk	Zweistellige Stunde des Tages, wobei die erste Stunde 1 und die letzte Stunde 24 ist. Zahlen kleiner als 10 enthalten eine vorausgehende Null.	02
K	Stunde am Vormittag (AM) oder Nachmittag (PM), wobei 0 für die erste Stunde und 11 für die letzte Stunde steht.	0
KK	Zweistellige Stunde des Tages, wobei die erste Stunde 1 und die letzte Stunde 24 ist. Zahlen kleiner als 10 enthalten eine vorausgehende Null.	02
h	Stunde am Vormittag (AM) oder Nachmittag (PM), wobei 1 für die erste Stunde und 12 für die letzte Stunde steht.	12
hh	Zweistellige Stunde am Vormittag (AM) oder Nachmittag (PM), wobei 1 für die erste Stunde und 12 für die letzte Stunde steht. Zahlen kleiner als 10 enthalten eine vorausgehende Null.	09
m	Minute der Stunde	30
mm	Zweistellige Minuten der Stunde. Zahlen kleiner als 10 enthalten eine vorausgehende Null.	05
s	Sekunde der Minute	55
ss	Zweistellige Sekunden der Minute. Zahlen kleiner als 10 enthalten eine vorausgehende Null.	02
S	Millisekunde der Sekunde	978
SSS	Dreistellige Millisekunde der Sekunde. Zahlen, die weniger als drei Stellen enthalten, tragen eine oder zwei Nullen, um eine dreistellige Zahl zu erzeugen.	978 078 008
z	Zeitabkürzung des Namen der Zeitzone. Wenn die Zeitzone keinen Namen hat, wird die GMT-Abweichung angezeigt.	PST GMT-08:00

Buchstabe	Beschreibung	Beispiel
zzzz	Der vollständige Name der Zeitzone. Wenn die Zeitzone keinen Namen hat, wird die GMT-Abweichung angezeigt.	Pacific Standard Time GMT-08:00
Z	Die Zeitzone RFC 822.	-0800
X	The Zeitzone ISO 8601.	-08Z
XX	Die Zeitzone gemäß ISO 8601 mit Minuten.	-0800Z
XXX	Die Zeitzone gemäß ISO 8601 mit Minuten und Doppelpunkt als Trennzeichen zwischen Stunden und Minuten.	-08:00Z

Zahlenmuster

Beim Definieren von Datentypoptionen für numerische Daten können Sie Ihr eigenes benutzerdefiniertes numerisches Muster erstellen, falls die vordefinierten Muster nicht Ihren Anforderungen entsprechen. Ein grundlegendes Zahlenmuster besteht aus den folgenden Elementen:

- Ein Präfix wie ein Währungssymbol (optional)
- Ein Muster aus Zahlen mit einem optionalen Gruppierungszeichen (z. B. ein Punkt als Tausendertrennzeichen)
- Ein Suffix (optional)

Das Beispielmuster:

###.###,00 €

würde eine Zahl erzeugen, die wie folgt formatiert ist (beachten Sie die Verwendung des Tausendertrennzeichens nach den ersten drei Stellen):

232.998,60 €

Muster für negative Zahlen

Negative Zahlen werden standardmäßig so formatiert wie positive Zahlen, nur das ihnen ein Negativzeichen als Präfix hinzugefügt wird. Das als Zahlenzeichen verwendete Zeichen basiert auf dem Gebietsschema. In den meisten Gebietsschemata ist das Negativzeichen ein Minus („-“). Wenn Sie z. B. das folgende Zahlenmuster festlegen:

```
0.00
```

wird die Zahl „Minus Zehn“ in den meisten Gebietsschemata wie folgt formatiert:

```
-10.00
```

Wenn Sie jedoch ein anderes Präfix oder Suffix für negative Zahlen definieren möchten, geben Sie ein zweites Muster an, das durch ein Semikolon („;“) vom ersten Muster abgetrennt ist. Beispiel:

```
0.00;(0.00)
```

In diesem Muster stehen negative Zahlen in Klammern:

```
(10.00)
```

Wissenschaftliche Schreibweise

Wenn Sie eine Zahl in wissenschaftlicher Schreibweise formatieren möchten, verwenden Sie das Zeichen `E` gefolgt von der minimalen Anzahl von Stellen, die Sie in den Exponent einbeziehen möchten. Zum Beispiel dieses Muster:

```
0.###E0
```

Die Zahl `1234` würde wie folgt formatiert werden:

```
1.234E3
```

Mit anderen Worten: 1.234 x 10^3.

Beachten Sie Folgendes:

- Die Anzahl der Stellen nach dem Exponentenzeichen gibt die minimale Anzahl der Exponentenstellen an. Es gibt kein Maximum.
- Negative Exponenten werden mithilfe des lokalisierten Minuszeichens formatiert. Es wird kein Präfix oder Suffix aus dem Muster verwendet.
- Muster mit wissenschaftlicher Schreibweise dürfen keine Gruppierungstrennzeichen (z. B. ein Tausendertrennzeichen) enthalten.

Sonderzeichen für Zahlenmuster

Die folgenden Zeichen werden zur Erzeugung anderer Zeichen verwenden, d. h. sie werden nicht buchstaben- oder zahlengetreu in der resultierenden Zahl dargestellt. Wenn Sie eines dieser Sonderzeichen buchstaben- oder zahlengetreu im Präfix oder Suffix Ihres Zahlenmusters verwenden möchten, schließen Sie es in gerade Anführungszeichen ein.

Symbol	Beschreibung

`0`	Steht für eine Stelle im Muster, bei dem Nullen zur Füllung des Musters verwendet werden. Durch Anwendung des folgenden Musters auf die Zahl 27:
	`0000`
	würde die Zahl wie folgt aussehen:
	`0027`

Symbol	Beschreibung

`#`	Steht für Stelle, für die Nullen ausgelassen werden. Durch Anwendung des folgenden Musters auf die Zahl 27:
	`####`
	würde die Zahl wie folgt aussehen:
	`27`

`.`	Das Dezimaltrennzeichen oder das monetäre Dezimaltrennzeichen, das im ausgewählten Gebietsschema verwendet wird. In den USA wird z. B. der Punkt (.) als Dezimaltrennzeichen verwendet. In Deutschland ist jedoch das Komma (,) das Dezimaltrennzeichen.
`-`	Das Negativzeichen, das im ausgewählten Gebietsschema verwendet wird. Für die meisten Gebietsschemata ist dies das Minus (-).

`,`	Das Gruppierungszeichen, das im ausgewählten Gebietsschema verwendet wird. Es wird das entsprechende Zeichen für das ausgewählte Gebietsschema verwendet. In den USA wird z. B. das Komma (,) als Trennzeichen verwendet.
	Das Gruppierungstrennzeichen wird häufig für Tausender verwendet. In einigen Ländern trennt es jedoch die Zehntausend ab. Die Gruppierungsgröße ist eine konstante Anzahl von Stellen zwischen den Gruppierungszeichen, wie z. B. 3 für 100.000.000 oder 4 für 1.0000.0000. Wenn Sie ein Muster mit mehreren Gruppierungszeichen anlegen, wird das Intervall zwischen dem letzten Zeichen und dem Ende der ganzen Zahl verwenden. Alle folgenden Muster erzeugen beispielsweise das gleiche Ergebnis:
	`#,##,###,####`
	`######,####`
	`##,####,####`

`E`	Trennt Mantisse und Exponent in wissenschaftlicher Schreibweise. Sie müssen das E in Ihrem Muster nicht in Anführungszeichen einschließen. Siehe **Wissenschaftliche Schreibweise** auf Seite 39.
`;`	Trennt positive und negative Untermuster. Siehe **Muster für negative Zahlen** auf Seite 38.

Symbol	Beschreibung

%	Multipliziert die Zahl mit 100 und zeigt die Zahl als Perzentil an. Zum Beispiel würde die Zahl .35, wenn sie auf dieses Muster angewendet wird:
	`##%`
	Folgendes Ergebnis aufweisen:
	`35%`

¤	Das Währungssymbol für das ausgewählte Gebietsschema. Bei einer Dopplung wird das internationale Währungssymbol verwendet. In einem Muster wird das monetäre Dezimaltrennzeichen anstelle des normalen Dezimaltrennzeichens verwendet.

'	Dient der Angabe von Sonderzeichen in einem Präfix oder Suffix. Beispiel:
	`"'#'#"`
	Formatiert 123 in:
	`"#123"`
	Verwenden Sie zwei aufeinander folgende Anführungsstriche, um einen einzelnen zu erzeugen:
	`"# o''clock"`

Ändern des Datentyps eines Feldes

Spectrum™ Technology Platform ändert automatisch die Felddatentypen mithilfe der Typkonvertierungseinstellungen in der Management Console oder gemäß den Typkonvertierungsoptionen des Datenflusses im Enterprise Designer. In den meisten Fällen müssen Sie die Felddatentypen nicht manuell ändern, da alle notwendigen Datentypkonvertierungen automatisch erfolgen. In Situationen, in denen ein Schritt die eingehenden Daten nicht in den notwendigen Datentyp konvertieren kann, müssen Sie den Datentyp manuell im Upstream-Kanal ändern.

Es gibt nur einige wenige Typkonvertierungen, die manuell durchgeführt werden können. Diese sind:

- Polygon- und MultiPolygon-Typen können in und aus einem Geometrietyp konvertiert werden.
- Die Datentypen „Datum“, „Zeit“ und „DateTime“ können in und aus einem Zeichenfolgentyp konvertiert werden.

Gehen Sie wie folgt vor, um einen Felddatentyp manuell zu ändern:

1. Doppelklicken Sie im Enterprise Designer auf den Kanal, in dem Sie den Felddatentyp ändern möchten. Ein Kanal ist die Linie, die zwei Schritte auf der Arbeitsfläche verbindet.

2. Klicken Sie auf die kleine quadratische Schaltfläche neben dem Datentyp, den Sie ändern möchten.

   Anmerkung: Wenn die kleine quadratische Schaltfläche nicht neben dem Datentyp angezeigt wird, ist in diesem Fall keine manuelle Datentypkonvertierung möglich.

3. Gehen Sie für die Datentypen „Datum“, „Zeit“ und „DateTime“ wie folgt vor:

   Anmerkung: Es werden nur die passenden Optionen je nach ausgewähltem Datentyp angezeigt.

   a) Wählen Sie im Feld **Gebietsschema** das Land aus, dessen Format Sie für Datumsangaben verwenden möchten, die in eine Zeichenfolge konvertiert werden. Ihre Auswahl bestimmt die Standardwerte in den Feldern **Datum**, **Zeit**, und **DateTime**. Ihre Auswahl bestimmt außerdem die verwendete Sprache, wenn ein Monat ausgeschrieben wird. Wenn Sie beispielsweise „Englisch“ auswählen, wäre der erste Monat des Jahres „January“, wenn Sie aber „Französisch“ angeben, wäre es „Janvier“.

   b) Wählen Sie im Feld **Datum** das Format für Datumsangaben aus, wenn diese in eine Zeichenfolge konvertiert werden. Es wird eine Liste der am häufigsten verwendeten Formate für das ausgewählte Gebietsschema bereitgestellt.

      Wenn Sie das Format **M/D/YY** auswählen, und ein Datumsfeld den Wert „2012-3-2“ enthält, würden diese Datumsangaben in die folgende Zeichenfolge konvertiert werden: `3/2/12`.

   c) Wählen Sie im Feld **Zeit** das Format für Zeitangaben aus, wenn diese in eine Zeichenfolge konvertiert werden. Es wird eine Liste der am häufigsten verwendeten Formate für das ausgewählte Gebietsschema bereitgestellt.

      Wenn Sie das Format **h:mm a** auswählen, und ein Datumsfeld den Wert „23:00“ enthält, würden diese Zeitangaben in die folgende Zeichenfolge konvertiert werden: `11:00 PM`.

   d) Wählen Sie im Feld **DateTime** das Format für die Felder mit DateTime-Datentyp aus, wenn diese in eine Zeichenfolge konvertiert werden. Es wird eine Liste der am häufigsten verwendeten Formate für das ausgewählte Gebietsschema bereitgestellt.

      Wenn Sie das Format **M/d/yy h:mm a** auswählen, und ein DateTime-Feld den Wert „2012-3-2 23:00“ enthält, würden diese DateiTime-Angaben in die folgende Zeichenfolge konvertiert werden: `3/2/12 11:00 PM`.

   e) Wählen Sie im Feld **Ganze Zahlen** die Formatierung aus, die Sie für ganze Zahlen (float- und double.Datentypen) verwenden möchten.

      Wenn Sie beispielsweise das Format **#.###** auswählen, wird die Zahl 4324 wie folgt formatiert:`4.324`.

      Anmerkung: Wenn Sie dieses Feld leer lassen, werden die Zahlen so wie in Spectrum™ Technology Platform 8.0 oder früher formatiert. Es wird insbesondere kein Tausendertrennzeichen verwendet. Der Punkt („.“) wird als Dezimalzeichen verwendet. Zahlen, die kleiner als 10^{-3} oder größer gleich 10^{7} sind, werden in wissenschaftlicher Schreibweise angezeigt. Negative Zahlen tragen ein vorangestellte

Minus („-“). Beachten Sie auch, dass wenn Sie das Feld leer lassen, Zahlen mit dem bigdecimal-Datentyp immer das Format #,###.000 aufweisen.

f) Wählen Sie im Feld **Dezimalzahlen** die Formatierung für Zahlen aus, die einen Dezimalwert enthalten (integer- und long-Datentypen).

Wenn Sie beispielsweise das Format **#.##0,0#** auswählen, wird die Zahl 4324,25 wie folgt formatiert: `4.324,25`.

Anmerkung: Wenn Sie dieses Feld leer lassen, werden die Zahlen so wie in Spectrum™ Technology Platform 8.0 oder früher formatiert. Es wird insbesondere kein Tausendertrennzeichen verwendet. Der Punkt („.“) wird als Dezimalzeichen verwendet. Zahlen, die kleiner als 10^{-3} oder größer gleich 10^{7} sind, werden in wissenschaftlicher Schreibweise angezeigt. Negative Zahlen tragen ein vorangestellte Minus („-“). Beachten Sie auch, dass wenn Sie das Feld leer lassen, Zahlen mit dem bigdecimal-Datentyp immer das Format #,###.000 aufweisen.

4. Klicken Sie auf **OK**.

   Die Farbe der Datentypnamen ändert sich in grün.

5. Klicken Sie erneut auf **OK**, um die Änderung zu speichern.

Ändern eines Feldnamen

Es gibt eine Vielzahl von Situationen, in denen Sie möglicherweise ein Feld in einem Datenfluss umbenennen müssen. Beispiel:

- Die Eingabe eines Schrittes erfordert bestimmte Feldnamen, aber die Ausgabe des vorherigen Schrittes verwendet andere Feldnamen.
- Es gibt Daten in einem Feld, die Sie beibehalten möchten, wenn ein Downstream-Schritt Daten in ein Feld mit demselben Namen schreibt.

Anmerkung: Nachdem ein Feld umbenannt wurde, ist es in nachfolgenden Schritten nicht mehr unter dem alten Namen verfügbar.

1. Doppelklicken Sie in einem Datenfluss auf den Kanal zwischen zwei Schritten. Das Dialogfeld **Feldtransformationsoptionen** wird angezeigt.
2. Ändern Sie je nach Bedarf den bzw. die Feldnamen.

   So könnte beispielsweise der letzte Schritt „AddressLine3“ erfordern, jedoch verwendet der vorherige Schritt stattdessen „FirmName“. In diesem Fall würden Sie auf den Listenfeldpfeil bei „Eingabefeldname“ klicken, der „AddressLine3“ als „Ausgabefeldname“ entspricht, und dann „FirmName“ auswählen.

   Die Farbe des Ausgabefeldnamens ändert sich in Grün.

3. Klicken Sie auf **OK**.

Reservierte Feldnamen

Die folgenden Feldnamen werden vom System verwendet, weshalb Sie sie nicht in einem Datenfluss verwenden sollten.

Status
Status.Code
Status.Description

Steuerungsschritte

Verwenden Sie Steuerungsschritte, um Daten über unterschiedliche Pfade in einem Fluss zu verschieben, Datensätze aufzuteilen oder zu gruppieren sowie einfache Datenumwandlungen und mathematische Operationen durchzuführen.

Aggregator

Aggregator wandelt flache Daten in hierarchische Daten um. Er verwendet Eingabedaten aus einer einzelnen Quelle, erstellt ein Schema (eine strukturierte Hierarchie von Daten), indem er die Daten basierend auf den von Ihnen angegebenen Feldern gruppiert, und erstellt anschließend die Gruppen im Schema.

> **Anmerkung:** Wenn in Ihren Daten ein Feld enthalten ist, nach dem Sie alle Ihre Daten gruppieren werden, z. B. ein ID-Feld, müssen Sie Ihre Daten sortieren, bevor Sie sie durch einen Aggregator laufen lassen. Hierfür können Sie die Daten sortieren, bevor Sie sie in den Datenfluss aufnehmen. Sie können auch die Eingabedatei innerhalb des Enterprise Designer (für Aufträge oder Unterflüsse, jedoch nicht für Services) sortieren oder ihrem Datenfluss einen Sorter-Schritt hinzufügen (für Aufträge, Services oder Unterflüsse).

Gruppieren nach

Wählen Sie das Feld aus, das Sie als Grundlage zur Aggregation in die Hierarchie verwenden möchten, indem Sie **Gruppieren nach** im Baum auswählen und anschließend auf **Hinzufügen** klicken. Die Daten von Datensätzen mit demselben Wert in dem von Ihnen ausgewählten Feld werden in einer einzelnen Hierarchie aggregiert. Wenn Sie mehrere Felder auswählen, müssen die Daten aus allen Feldern übereinstimmen, damit die Datensätze in einer Hierarchie gruppiert werden.

Wenn Sie Daten beispielsweise nach Kontonummer gruppieren möchten, müssten Sie das Kontonummernfeld auswählen. Die Daten aller eingehenden Datensätze mit demselben Wert im Kontonummernfeld würden in einem einzelnen Hierarchiedatensatz gruppiert werden.

> **Anmerkung:** Sie müssen einen Schritt mit dem Eingangsport des Aggregators verbinden, damit eine Liste von Feldern zur Auswahl verfügbar wird.

Ausgabelisten

Die Felder, die Sie unter **Ausgabelisten** auswählen, bestimmen, welche Felder in jeden vom Aggregator erstellten Datensatz aufgenommen werden. Wählen Sie zum Hinzufügen eines Feldes **Ausgabelisten** aus, klicken Sie anschließend auf **Hinzufügen**, und wählen Sie eine der folgenden Optionen aus:

Vorhandenes Feld	Wählen Sie diese Option aus, wenn Sie ein Feld aus dem Datenfluss zur Hierarchie hinzufügen möchten.
Neuer Datentyp	Wählen Sie diese Option aus, wenn Sie ein übergeordnetes Feld erstellen möchten, zu dem Sie anschließend untergeordnete Felder hinzufügen können.
Vorlage	Mithilfe dieser Option können Sie ein Feld basierend auf den Daten im Schritt hinzufügen, der mit dem Ausgangsport des Aggregators verbunden ist.

Wenn Sie untergeordnete Felder benötigen, müssen Sie das Kästchen **Liste** aktivieren.

Geben Sie den Namen des Feldes in das Textfeld **Name** ein oder lassen Sie ihn unverändert, wenn er automatisch ausgefüllt wurde und Sie mit dem Namen zufrieden sind. Beachten Sie, dass der Aggregator-Schritt keine ungültigen XML-Zeichen in Feldnamen zulässt. Er lässt jedoch alphanumerische Zeichen, Punkte (.), Unterstriche (_) und Bindestriche (-) zu.

Klicken Sie auf **Hinzufügen**, um das Feld hinzuzufügen. Sie können ein weiteres Feld festlegen, um es zur selben Stufe in der Hierarchie hinzuzufügen, oder Sie können auf **Schließen** klicken.

Um einem vorhandenen Feld untergeordnete Felder hinzuzufügen, müssen Sie das übergeordnete Feld auswählen und anschließend auf **Hinzufügen** klicken.

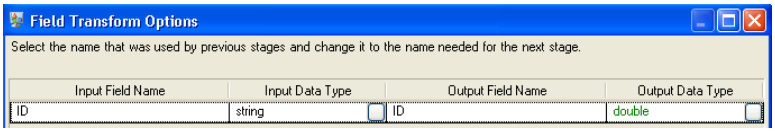
> **Anmerkung:** Sie können Änderungen an der Feldgruppe vornehmen, indem Sie eine Zeile markieren und auf **Ändern** klicken. Sie können außerdem eine Feldgruppe entfernen, indem Sie eine Zeile markieren und auf **Entfernen** klicken. Sie können außerdem die Reihenfolge der Felder ändern, indem Sie auf ein Feld und anschließend auf **Nach oben** oder **Nach unten** klicken.

Beispiel für Aggregator

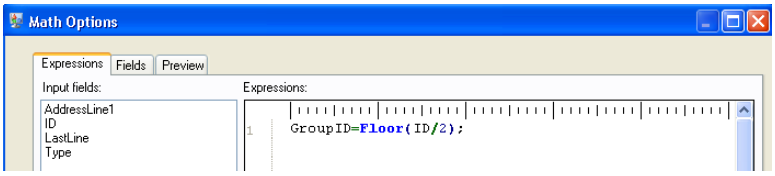
Ein Beispiel für das Aggregator-Feature ist es, eine Gruppe von Adressen auszuwählen und sie in Anfahrtsbeschreibungen umzuwandeln. Dies kann mit zweierlei Punkten erfolgen, z. B. einem Startpunkt und einem Endpunkt, oder mit mehreren Punkten entlang einer Route. Der Datenfluss für diese Art von Feature könnte folgendermaßen aussehen:

Der Datenfluss führt das Feature folgendermaßen aus:

1. Der **Read from File**-Schritt enthält Adressen in einer Flatfile. Die Felder in dieser Datei enthalten Folgendes:
   - eine **ID**, die eine bestimmte Adresse in der Datei identifiziert
   - einen **Typ**, der angibt, ob es sich bei der Adresse um eine „Startadresse“ oder eine „Zieladresse“ handelt
   - das Feld **AddressLine1**, das die Adresse bezeichnet
   - das Feld **LastLine**, in dem Informationen wie Ort, Bundesland/Kanton und/oder Postleitzahl bereitgestellt werden
2. Die **Feldtransformation** zwischen dem „Read from File“-Schritt und dem Math-Schritt ändert das Format des ID-Feldes von „string“ in „double“, da der Math-Schritt keine string-Daten zulässt.



3. Beim **Math**-Schritt wird ein Ausdruck erstellt, der ein Feld für die Gruppen-ID einrichtet, das nachgeschaltet im Datenfluss verwendet werden soll. In diesem Beispiel berechnet sie die Gruppen-ID als die Untergrenze des Wertes des ID-Feldes geteilt durch 2 oder rundet die ID ab. Wenn die ID also 3 ist, ist der Ausdruck 3/2, was 1,5 ergibt. Wenn Sie 1,5 abrunden, erhalten Sie 1. Wenn die ID 2 ist, ist der Ausdruck 2/2, was 1 ergibt, sodass Sie nicht abrunden müssen. Die IDs 2 und 3 haben dementsprechend dieselbe Gruppen-ID 1.



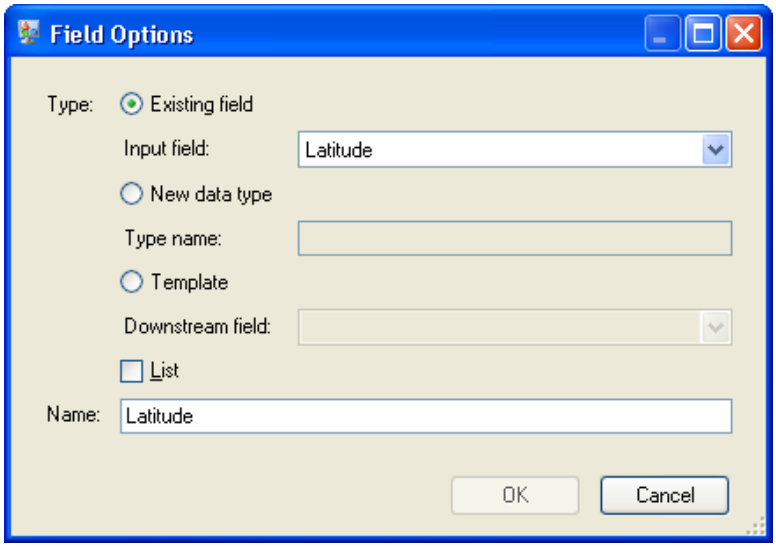
4. **Geocode US Address** ruft Breitengrade und Längengrade für jede Adresse ab.
5. Im **Aggregator**-Schritt wird festgelegt, dass die Daten nach dem Feld für die Gruppen-ID gruppiert werden sollen und dass die Ausgabelisten Routenpunkte aus Breitengraden und Längengraden enthalten sollen.

   In den Anleitungen unten wird erläutert, wie Sie der Aggregator-Schritt **manuell** für diesen Datenfluss konfigurieren:

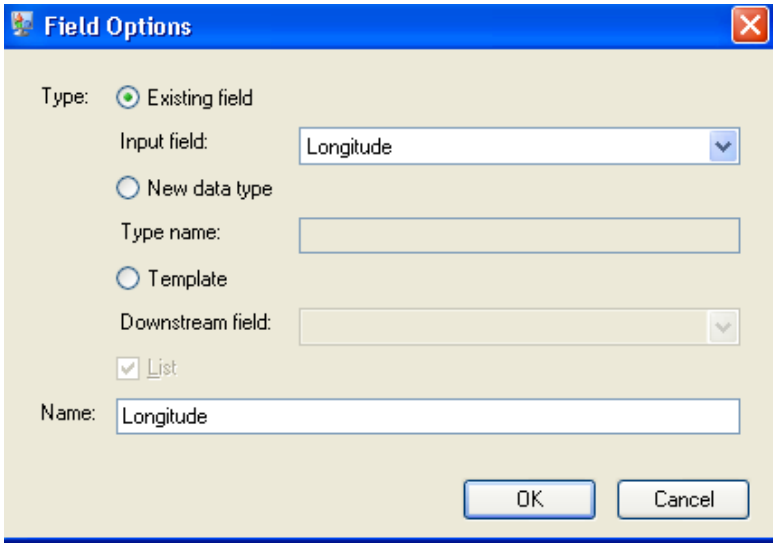
   - Klicken Sie doppelt auf den Aggregator-Schritt, und klicken Sie anschließend doppelt auf **Gruppieren nach**.
   - Wählen Sie das Feld **GroupID** aus, und klicken Sie auf **OK**. Mithilfe dieses Feldes können wir Routenpunkte für den nächsten Schritt im Datenfluss

aufnehmen. Routenpunkte sind wichtige Elemente für einen Datenfluss, der Wegbeschreibungen erzeugt.

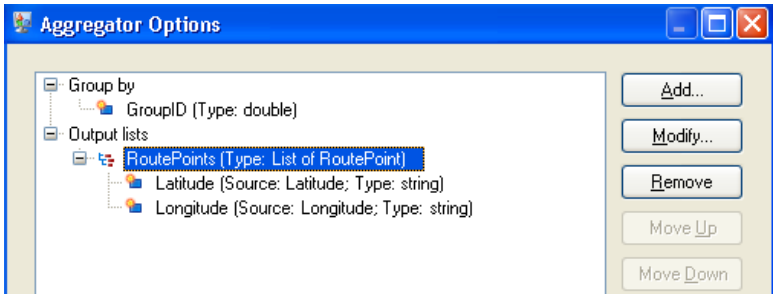
- Klicken Sie doppelt auf **Ausgabelisten**. Das Dialogfeld **Feldoptionen** wird angezeigt.
- Wählen Sie **Neuer Datentyp** aus. Geben Sie in das Feld **Typname** `Routenpunkt` ein. Geben Sie in das Feld **Name** `Routenpunkte` ein. Standardmäßig ist dies eine Liste, die nicht geändert werden kann, sodass das Kästchen ausgegraut ist.
- Klicken Sie auf **OK**.
- Klicken Sie auf **Routenpunkte**, und klicken Sie auf **Hinzufügen**. Das Dialogfeld **Feldoptionen** wird erneut angezeigt.
- Routenpunkte setzen sich aus Breitengraden und Längengraden zusammen, sodass wir zunächst ein **Vorhandenes Feld** aus dem vorhandenen Eingabefeld `Breitengrad` hinzufügen müssen. Das Feld **Name** wird automatisch ausgefüllt.



Wiederholen Sie diesen Schritt für die Längengrade.



Der abgeschlossene Aggregator-Schritt wird folgendermaßen dargestellt:



6. **Get Travel Directions** liefert Fahranweisungen von den Punkte-IDs 0, 2, und 4 zu jeweils den Punkte-IDs 1, 3 und 5.
7. Im **Splitter**-Schritt wird festgelegt, dass die Daten am Feld „Wegbeschreibung“ aufgeteilt werden sollen und dass die Ausgabelisten alle möglichen Felder aus dem „Get Travel Directions“-Schritt enthalten sollen.
8. Der **Write to File**-Schritt schreibt Wegbeschreibungen direkt in eine Ausgabedatei.

Broadcaster

Ein Broadcaster wählt eine Reihe von Datensätzen und teilt diese in mehrere Datenströme auf. Dadurch können Sie Datensätze zur gleichzeitigen Verarbeitung an mehrere Schritte senden.

Der Broadcaster verfügt über keine zu ändernden Einstellungen.

Conditional Router

Der **Conditional Router**-Schritt sendet Datensätze abhängig von den von Ihnen angegebenen Kriterien an unterschiedliche Pfade im Datenfluss. Der Schritt kann einen oder mehrere Ausgangsports haben, je nachdem, welche Kriterien definiert wurden. Ausgangsports sind fortlaufend nummeriert und beginnen mit 1 (was als „Port“ angezeigt wird).

Die Ausgangsports sind mit unterschiedlichen Schritten verbunden, an die die Daten gesendet werden sollen, je nachdem, welche Bedingungen definiert sind. Sie können beispielsweise eine Datensatzgruppe bei einer erfolgreichen Übereinstimmung an Port 1 senden, während eine andere Datensatzgruppe bei einer fehlenden Übereinstimmung an Port 2 gesendet wird.

Ein Eingabedatensatz wird nur in den Ausgangsport des „Conditional Router“ geschrieben, wenn der gesamte Ausdruck als „wahr“ ausgewertet wird.

Konfigurieren eines „Conditional Router“

1. Klicken Sie unter **Steuerungsschritte** auf **Conditional Router**, und ziehen Sie den Schritt auf die Arbeitsfläche. Platzieren Sie sie im gewünschten Speicherort im Datenfluss.
2. Verbinden Sie den Router mit anderen Schritten auf der Arbeitsfläche.

   Anmerkung: Dies ist ein obligatorischer Schritt vor der Definition der Porteinstellungen. Andernfalls können die Ports nicht bearbeitet werden.

3. Klicken Sie doppelt auf den Schritt **Conditional Router** auf der Arbeitsfläche. Das Fenster **„Conditional Router“-Optionen** wird angezeigt.
4. Klicken Sie auf das Quadrat in der Spalte **Bedingung/Ausdruck** gegenüber der Zeile **Port**. Das Fenster **Ausdrucks-Editor** wird angezeigt.
5. Wählen Sie im Abschnitt **Ausdruckstyp auswählen** eine der folgenden Optionen aus:

   - **Mit Ausdrucks-Generator erzeugter Ausdruck**: Wählen Sie diese Option aus, um einen grundlegenden Ausdruck zu erstellen, in den Sie Gruppen und Ausdrücke hinzufügen können, die anhand unterschiedlicher logischer Operatoren kombiniert werden können. Weitere Informationen finden Sie unter **Verwenden des Ausdrucks-Generators** auf Seite 50.
   - **Benutzerdefinierter Ausdruck**: Wählen Sie diese Option aus, um mithilfe der Groovy-Skriptsprache einen Ausdruck zu schreiben. Weitere Informationen finden Sie unter **Schreiben eines benutzerdefinierten Ausdrucks** auf Seite 53.
   - **Standardausdruck**: Wählen Sie diese Option aus, um Datensätze standardmäßig an diesen Port zu leiten. Datensätze, die mit keinem der Ausdrücke der anderen Ports übereinstimmen, werden an diesen Port geleitet. Sie sollten stets einen Ausgangsport mit „default“ als Ausdruck haben, um sicherzustellen, dass im Fall einer fehlenden Portübereinstimmung keine Zeilen übersehen werden und dass alle Zeilen aus dem Router geschrieben werden.

6. Klicken Sie auf **OK**. Das Fenster **Ausdrucks-Editor** wird geschlossen.
7. Klicken Sie im Fenster **„Conditional Router“-Optionen** auf **OK**.

Verwenden des Ausdrucks-Generators

Mit dem **Ausdrucks-Generator** des **Conditional Router**-Schritts können Sie einen Ausdruck erstellen, der für einen Eingabedatensatz positiv erfüllt sein muss, damit dieser an den Ausgabeport des Schritts weitergeleitet wird.

1. Jede übergeordnete Gruppe besteht aus einer erwünschten bedingten Kombination aus untergeordneten Ausdrücken und untergeordneten Gruppen.
2. Jeder Ausdruck besteht aus einem linken Operanden, einem rechten Operanden und einem logischen Operator.
3. Für jede Gruppe muss angegeben sein, ob *alle* oder *beliebige* der einzelnen Bedingungen der gesamten Gruppe erfüllt sein müssen, um in der Auswertung wahr zu sein.

So generieren Sie mithilfe des Ausdrucks-Generators einen Ausdruck:

1. Wählen Sie im **Ausdrucks-Editor** die Option **Mit Ausdrucks-Generator erzeugter Ausdruck** aus.

   Standardmäßig ist die Option „Ausdrucks-Generator“ ausgewählt, und die übergeordnete Gruppe wird in der Struktur für die Ausdruckshierarchie links vom **Ausdrucks-Generator** angezeigt.

2. Klicken Sie auf **Gruppe hinzufügen**, um in der ausgewählten Gruppe eine untergeordnete Gruppe hinzuzufügen.

   Diese neu hinzugefügte Gruppe wird der übergeordneten Gruppe als untergeordnete Gruppe hinzugefügt und standardmäßig in der Struktur ausgewählt. Sie können in jeder Gruppe untergeordnete Ausdrücke und untergeordnete Gruppen hinzufügen.

3. Wählen Sie für jede Gruppe unter der Überschrift **Ausdrucksmethode kombinieren** entweder **Alle Übereinstimmungen** oder **Beliebige Übereinstimmung** aus.

   - `Alle Übereinstimmungen`: Die Gruppe wird nur mit „Wahr“ ausgewertet, wenn alle untergeordneten Kriterien der Gruppe erfüllt sind.
   - `Beliebige Übereinstimmung`: Die Gruppe wird mit „Wahr“ ausgewertet, wenn auch nur eines der untergeordneten Kriterien der Gruppe erfüllt ist.

4. Klicken Sie auf **Ausdruck hinzufügen**, um in der ausgewählten Gruppe einen untergeordneten Ausdruck hinzuzufügen.

   Der neu hinzugefügte Ausdruck wird der übergeordneten Gruppe als untergeordneter Ausdruck hinzugefügt und standardmäßig in der Struktur ausgewählt.

   So definieren Sie diesen untergeordneten Ausdruck:

   a) Geben Sie den linken Operanden des ausgewählten Ausdrucks an, und verwenden Sie die Dropdown-Liste **Feld**, um beliebige Spalten in der Eingabedatei auszuwählen.
   b) Geben Sie den logischen Operator an, der die beiden Komponenten des ausgewählten Ausdrucks verbindet, indem Sie den entsprechenden Operator wie unten erläutert über das Feld **Operator** auswählen:

Tabelle 1: Operatoren des Ausdrucks-Generators

Operator	Beschreibung
Ist gleich	Prüft, ob der Wert im Feld mit dem angegebenen Wert oder Feld übereinstimmt.
Ist ungleich	Prüft, ob der Wert im Feld nicht mit dem angegebenen Wert oder Feld übereinstimmt.
Ist NULL	Prüft, ob es sich bei dem Feld um einen Nullwert handelt.
Ist nicht NULL	Prüft, ob es sich bei dem Feld nicht um einen Nullwert handelt.
Ist leer	Prüft, ob das Feld NULL oder eine Zeichenfolge mit einer Länge von 0 ist. **Anmerkung:** Dieser Vorgang ist nur für Felder mit dem Datentyp „string“ verfügbar.
Ist nicht leer	Prüft, ob das Feld weder NULL noch eine Zeichenfolge mit einer Länge von 0 ist. **Anmerkung:** Dieser Vorgang ist nur für Felder mit dem Datentyp „string“ verfügbar.
Ist kleiner als	Prüft, ob das Feld einen numerischen Wert enthält, der kleiner als der angegebene Wert ist. Dieser Operator funktioniert bei numerischen Datentypen sowie Zeichenfolgenfeldern mit Zahlen. **Anmerkung:** Dieser Vorgang ist nicht für Felder mit dem Datentyp „boolean“ verfügbar.
Ist kleiner oder gleich	Prüft, ob das Feld einen numerischen Wert enthält, der kleiner als oder gleich dem angegebenen Wert ist. Dieser Operator funktioniert bei numerischen Datentypen sowie Zeichenfolgenfeldern mit Zahlen. **Anmerkung:** Dieser Vorgang ist nicht für Felder mit dem Datentyp „boolean“ verfügbar.

Operator	Beschreibung
Ist größer als	Prüft, ob das Feld einen numerischen Wert enthält, der größer als der angegebene Wert ist. Dieser Operator funktioniert bei numerischen Datentypen sowie Zeichenfolgenfeldern mit Zahlen. **Anmerkung:** Dieser Vorgang ist nicht für Felder mit dem Datentyp „boolean“ verfügbar.
Ist größer oder gleich	Prüft, ob das Feld einen numerischen Wert enthält, der größer als oder gleich dem angegebenen Wert ist. Dieser Operator funktioniert bei numerischen Datentypen sowie Zeichenfolgenfeldern mit Zahlen. **Anmerkung:** Dieser Vorgang ist nicht für Felder mit dem Datentyp „boolean“ verfügbar.
Beginnt mit	Prüft, ob das Feld mit den angegebenen Zeichen beginnt. **Anmerkung:** Dieser Vorgang ist nur für Felder mit dem Datentyp „string“ verfügbar.
Beginnt nicht mit	Prüft, ob das Feld nicht mit den angegebenen Zeichen beginnt. **Anmerkung:** Dieser Vorgang ist nur für Felder mit dem Datentyp „string“ verfügbar.
Enthält	Prüft, ob das Feld die angegebene Zeichenfolge enthält. **Anmerkung:** Dieser Vorgang ist nur für Felder mit dem Datentyp „string“ verfügbar.
Enthält nicht	Prüft, ob das Feld nicht die angegebene Zeichenfolge enthält. **Anmerkung:** Dieser Vorgang ist nur für Felder mit dem Datentyp „string“ verfügbar.
Endet mit	Prüft, ob das Feld mit den angegebenen Zeichen endet. **Anmerkung:** Dieser Vorgang ist nur für Felder mit dem Datentyp „string“ verfügbar.

Operator	Beschreibung
Endet nicht mit	Prüft, ob das Feld mit den angegebenen Zeichen endet. **Anmerkung:** Dieser Vorgang ist nur für Felder mit dem Datentyp „string“ verfügbar.
Stimmt überein mit regulärem Ausdruck	Gleicht das Feld mit einem regulären Ausdruck ab, um Textzeichenfolgen von Interesse zu ermitteln, z. B. bestimmte Zeichen, Wörter oder Zeichenmuster. Das Wertefeld sollte ein gültiges reguläres Ausdrucksmuster enthalten. **Anmerkung:** Dieser Vorgang ist nur für Felder mit dem Datentyp „string“ verfügbar.

c) Geben Sie den rechten Operanden des ausgewählten Ausdrucks an, indem Sie entweder einen `Wert` oder ein `Feld` auswählen.

- `Wert`: Der linke Operand des ausgewählten Ausdrucks wird mit diesem Wert verglichen.
- `Feld`: Der linke Operand des ausgewählten Ausdrucks wird mit dieser Spalte derselben Eingabedatei verglichen. Wählen Sie die Spalte für den rechten Operanden aus der Dropdown-Liste aus.

5. Wählen Sie eine beliebige Entität in der Struktur aus und klicken Sie auf **Ausdruck hinzufügen** oder **Gruppe hinzufügen**, um einen gleichgeordneten Ausdruck bzw. eine gleichgeordnete Gruppe zu dieser Entität hinzuzufügen.
6. Um einen untergeordneten Ausdruck oder eine untergeordnete Gruppe von einer übergeordneten Gruppe zu einer anderen übergeordneten Gruppe zu verschieben, ziehen Sie den Ausdruck bzw. die Gruppe zu der Überschrift der gewünschten übergeordneten Gruppe in der Kriterienstruktur auf der linken Seite.
7. Wiederholen Sie die obigen Schritten, um so viele untergeordnete Ausdrücke und untergeordnete Gruppen wie erforderlich hinzuzufügen, um die gewünschten finalen Ausdruckskriterien zu erstellen.
8. Klicken Sie auf **OK**.

In der Spalte **Bedingung/Ausdruck** im Fenster **„Conditional Router“-Optionen** werden die definierten Ausdruckskriterien angezeigt, der als „wahr“ ausgewertet werden müssen, damit ein Datensatz in den zugehörigen Ausgabeport des Schritts geschrieben wird.

Schreiben eines benutzerdefinierten Ausdrucks

Sie können Ihre eigenen benutzerdefinierten Ausdrücke schreiben, um zu steuern, wie der „Conditional Router“ Datensätze weiterleitet. Verwenden Sie dazu die Skriptsprache Groovy, um einen Ausdruck zu erstellen.

Verwenden von Groovy-Scripting

Informationen zu Groovy finden Sie unter **groovy-lang.org**.

Die im „Conditional Router“-Schritt verwendeten Groovy-Ausdrücke müssen als boolescher Wert (wahr oder falsch) ausgewertet werden. Dadurch wird angegeben, ob der Datensatz in den Port geschrieben werden soll. Der Datensatz wird in den ersten Ausgangsport geleitet, dessen Ausdruck als „Wahr“ ausgewertet wird.

Wenn Sie beispielsweise Datensätze mit einem Konfidenzniveau für die Überprüfung >= 85 an einen Schritt und Datensätze mit einem Konfidenzniveau für die Überprüfung < 85 an einen anderen Schritt leiten müssen, würde Ihr Skript folgendermaßen aussehen:

```
data['Confidence']>=85
```

Das Skript für den anderen Port würde folgendermaßen aussehen:

```
data['Confidence']<85
```

Der Router würde den Wert des Zuverlässigkeitsfeldes mit Ihren Kriterien abgleichen, um den Zielausgangsport zu bestimmen.

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;
```

Scripting-Richtlinien

1. Spaltennamen müssen entweder in einzelne oder doppelte Anführungszeichen eingeschlossen werden.

   Diese Syntax ist beispielsweise falsch, da der Spaltenname `PostalCode` nicht in einzelne oder doppelte Anführungszeichen eingeschlossen ist.

   ```
   return data[PostalCode];
   ```

2. Ein Spaltenname muss angegeben werden.

   Diese Syntax ist beispielsweise falsch, da keine Spalte angegeben ist.

   ```
   return data[];
   ```

3. Eine `return`-Anweisung muss einen `boolean`-Wert zurückgeben.

   Dieses Skript ist beispielsweise falsch, da `row.set('PostalCode', '88989')` keinen `boolean`-Wert zurückgibt. Es legt einfach den Wert des Felds `PostalCode` auf `88989` fest.

   ```
   return row.set('PostalCode', '88989');
   ```

4. Verwenden Sie ein einfaches Gleichheitszeichen (=), um den Wert eines Feldes festzulegen, und ein doppeltes Gleichheitszeichen (==), um den Wert eines Feldes zu überprüfen.

Group Statistics

Der „Group Statistics“-Schritt ermöglicht Ihnen, statistische Vorgänge für mehrere in Gruppen unterteilte Datenzeilen auszuführen, die Sie analysieren möchten. Wenn keine Gruppen definiert sind, werden alle Zeilen als zu einer Gruppe gehörig behandelt.

Gruppen werden durch ein oder mehr Felder definiert, die in mehreren Datenzeilen denselben Wert haben.

So könnten beispielsweise die Daten in dieser Tabelle nach Region, State oder beidem gruppiert werden.

Region	State
Osten	MD
Osten	MD
Osten	CT
Westen	CA
Westen	CA

Eine Gruppe nach Region würde Osten und Westen ergeben. Eine Gruppe nach State würde Kalifornien, Connecticut und Maryland ergeben. Eine Gruppe nach Region und State würde Osten/Maryland, Osten/Connecticut und Westen/Kalifornien ergeben.

Eingabe

Der „Group Statistics“-Schritt akzeptiert jedes Feld als Eingabe. Die Gruppierung kann für numerische Daten oder Zeichenfolgendaten ausgeführt werden.

Optionen

Tabelle 2: Registerkarte „Vorgänge“

Option	Beschreibung
Eingabefelder	Listet die Felder im Datenfluss auf, die Sie zum Gruppieren von Datensätzen und Ausführen von Berechnungen verwenden können.
Row	Gibt das Feld bzw. die Felder an, die Sie als Kategorien für die Berechnungen verwenden möchten. Beispiel: Wenn Sie bei Daten mit einem Feld „Region“ die Gesamtbevölkerung nach Region berechnen wollten, würden Sie nach dem Feld „Region“ gruppieren. Wählen Sie zum Hinzufügen eines Feldes das Feld in der Liste **Eingabefelder** aus, und klicken Sie dann auf **>>**.
Spalte	Optional: Gibt zum Erstellen einer Pivot-Tabelle das Feld bzw. die Felder an, deren Werte Sie zum Zweck der Kreuztabellierung in Spalten pivotisieren möchten. Wählen Sie zum Hinzufügen eines Feldes das Feld in der Liste **Eingabefelder** aus, und klicken Sie dann auf **>>**. Beispiel: Wenn Sie bei Daten mit Regionen und Versanddaten die Anzahl der Lieferungen pro Tag für jedes Bundesland nachzählen möchten, müssen Sie das Feld „State“ als Zeile und das Feld „Versanddatum“ als Spalte angeben.

Option	Beschreibung
Zeilen und Spalten werden in der konfigurierten Reihenfolge vorsortiert	Gibt an, ob die Eingabedaten bereits sortiert sind. Wenn dieses Kästchen aktiviert ist, werden die Daten nicht im Schritt sortiert und der angegebene Vorgang wird direkt mit den Eingabedaten ausgeführt.
Bedienung	Gibt die Berechnung an, die für jede Gruppe ausgeführt werden soll. Wählen Sie zum Hinzufügen eines Vorgangs in der Liste **Eingabefelder** das Feld aus, das Sie für den Vorgang verwenden möchten, und klicken Sie dann auf **>>**. Weitere Informationen über die unterstützten „Group Statistics"-Vorgänge finden Sie unter **Vorgänge** auf Seite 58.
Typ	Gibt für die Eingabe- und Ausgabefelder den Datentyp an. **Ganze Zahl** Ein numerischer Datentyp, der sowohl negative als auch positive ganze Zahlen zwischen -2^{31} (-2.147.483.648) und $2^{31}-1$ (2.147.483.647) enthält. **Lang** Ein numerischer Datentyp, der sowohl negative als auch positive ganze Zahlen zwischen -2^{63} (-9.223.372.036.854.775.808) und $2^{63}-1$ (9.223.372.036.854.775.807) enthält. **Float** Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit einfacher Genauigkeit zwischen 2^{-149} (1,4E-45) und $(2-2^{23})\times2^{127}$ (3,4028235E38) enthält. **Double** Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit doppelter Genauigkeit zwischen 2^{-1074} (4,9E-324) und $(2-2-52)\times2^{1023}$ (1,7976931348623157E308) enthält. **Anmerkung:** Bei Verwendung der Typen „Ganze Zahl" und „Lang" können Daten verlorengehen, wenn die Eingabezahl oder berechnete Zahl aus einem Vorgang dezimale Daten enthält.
Anzahl der Datensätze aufrufen, die berechnet werden	Gibt die tatsächliche Anzahl der Datensätze in einer Gruppe zurück, mit denen der ausgewählte Vorgang ausgeführt wird. In der Spalte *Computational Count* sind die Eingabedatensätze ausgeschlossen, bei denen die Spalte, mit der der Vorgang ausgeführt wird, den Wert `null` enthält.

Registerkarte „Felder"

Die Registerkarte „Felder" wird beim Erstellen einer Pivot-Tabelle verwendet. Weitere Informationen finden Sie unter **Erstellen einer Pivot-Tabelle** auf Seite 65.

Registerkarte „Ausgabe“

Option	Beschreibung
Eine Zeile pro Gruppe zurückgeben	Gibt für jede Gruppe von Zeilen eine einzelne Zeile zurück, die die aggregierten Daten für alle Zeilen in der Gruppe enthält. Einzelnen Zeilen werden verworfen. Wenn diese Option nicht ausgewählt wird, werden alle Zeilen zurückgegeben. In diesem Fall werden keine Daten verworfen. Diese Option ist nicht verfügbar, wenn Sie die Vorgänge Prozentrang oder Z-Score verwenden.
Eine Anzahl von Zeilen in jeder Gruppe zurückgeben	Gibt die Anzahl der Zeilen in jeder Gruppe zurück. Der Standard-Ausgabefeldname, der die Anzahl enthält, ist GroupCount.
Eine eindeutige ID für jede Gruppe zurückgeben	Gibt eine eindeutige ID für jede Gruppe von Zeilen zurück. Die ID beginnt bei 1 und wird für jede zusätzliche gefundene Gruppe um 1 inkrementiert. Die Standardfeldname ist GroupID.

Vorgänge

Die verfügbaren Berechnungen sind:

Durchschnitt Berechnet für jede Gruppe den Mittelwert eines gegebenen Feldes. Beispiel: Bei einer Gruppe von Datensätzen mit den Werten 10, 12, 1 und 600 in einem gegebenen Feld wäre der Mittelwert dieses Feldes für diese Gruppe 155,75, berechnet mit (10+12+1+600)÷4.

Maximum Gibt für jede Gruppe den größten Wert in einem gegebenen Feld zurück. Beispiel: Bei einer Gruppe von Datensätzen mit den Werten 10, 12, 1 und 600 in einem gegebenen Feld wäre der Maximalwert dieses Feldes für diese Gruppe 600.

Minimum Gibt für jede Gruppe den kleinsten Wert in einem gegebenen Feld zurück. Beispiel: Bei einer Gruppe von Datensätzen mit den Werten 10, 12, 1 und 600 in einem gegebenen Feld wäre der Minimalwert dieses Feldes für diese Gruppe 1.

Prozentrang Berechnet für jeden Datensatz innerhalb einer Gruppe den Perzentilrang eines Wertes in einem gegebenen Feld relativ zu anderen Datensätzen in der Gruppe. Der Perzentilrang stellt den Prozentsatz von Datensätzen in der Gruppe mit niedrigeren Werten im Feld dar.

Perzentil Berechnet für jede Gruppe den Wert, der das Perzentil darstellen würde, das Sie für ein gegebenes Feld angeben (0–100). Ein Perzentil stellt den Prozentsatz von Datensätzen dar, die einen niedrigeren Wert haben.

	Beispiel: Wenn Sie bei einer Gruppe von Datensätzen mit den Werten 22, 26 und 74 eine Perzentilberechnung ausführen, die das 60. Perzentil angibt, würde der Vorgang 35,6 zurückgeben. Das bedeutet, dass ein Datensatz mit dem Wert 35,6 im gegebenen Feld im 60. Perzentil der Datensätze in der Gruppe wäre.
Standardabweichung	Berechnet für jede Gruppe die Standardabweichung für ein gegebenes Feld. Die Standardabweichung misst den Betrag der Streuung innerhalb der Gruppe. Je niedriger die Standardabweichung, desto stärker sind die Werte um den Mittelwert zentriert und dementsprechend weniger stark gestreut. Je höher der Wert, desto weiter sind die Werte gestreut. Die Standardabweichung wird in denselben Einheiten wie die Daten ausgedrückt. Die Standardabweichung ist die Quadratwurzel der Abweichung.
Summe	Berechnet für jede Gruppe die Summe der Werte für ein gegebenes Feld.
Streuung	Berechnet für jede Gruppe die Streuung für ein gegebenes Feld. Die Streuung misst den Betrag der Streuung innerhalb der Gruppe. Sie ist das Quadrat der Standardabweichung.
Z-Score	Gibt für jeden Datensatz in einer Gruppe den Z-Score zurück. Der Z-Score gibt an, wie viele Standardabweichungen ein Wert über oder unter dem Mittelwert der Gruppe liegt.
Alphabetisch erste	Gibt für jede Gruppe den ersten Wörterbuchwert zurück. Bei mehr als einem Feldwert mit derselben Länge oder Wörterbuchposition wird das erste Vorkommen dieses Wertes zurückgegeben. Beispiel: Bei einer Gruppe mit einem Datensatz mit den Werten Joel und Joey in einem Feld ist der alphabetisch erste Wert für eine Gruppe Joel, da „l“ im Alphabet vor „y“ kommt.
Alphabetisch letzte	Gibt für jede Gruppe den letzten Wörterbuchwert zurück. Bei mehr als einem Feldwert mit derselben Länge oder Wörterbuchposition wird das letzte Vorkommen dieses Wertes zurückgegeben. Beispiel: Bei einer Gruppe mit einem Datensatz mit den Werten Joel und Joey in einem Feld ist der alphabetisch letzte Wert für eine Gruppe Joey, da „y“ im Alphabet nach „l“ kommt.
Längste	Gibt für jede Gruppe den längsten Wert zurück. Beispiel: Bei einer Gruppe mit einem Datensatz mit den Werten Joel und Jacob in einem Feld ist der Wert mit der längsten Länge für eine Gruppe Jacob, da Jacob 5 Buchstaben umfasst, Joel hingegen nur 4.
Kürzeste	Gibt für jede Gruppe den kürzesten Wert zurück. Beispiel: Bei einer Gruppe mit einem Datensatz mit den Werten Joel und Jacob in einem Feld ist der Wert mit der kürzesten Länge für eine Gruppe Joel, da Joel 4 Buchstaben umfasst, Jacob hingegen 5.
Späteste	Gibt für jede Gruppe den spätesten Wert für Datum/Datum und Uhrzeit zurück. Beispiel: Wenn eine Gruppe einen Datensatz mit den Werten

	15-12-2014 und 24-12-2014 in einem Feld enthält, dann ist der späteste Wert für die Gruppe der 24-12-2014.
Früheste	Gibt für jede Gruppe den frühesten Wert für Datum/Datum und Uhrzeit zurück. Beispiel: Wenn eine Gruppe einen Datensatz mit den Werten 15-12-2014 und 24-12-2014 in einem Feld enthält, dann ist der früheste Wert für die Gruppe der 15-12-2014.

Ausgabespalten

Feldname	Beschreibung/gültige Werte
<Vorgang>Of<Eingabefeldname>	Enthält das Ergebnis einer Berechnung. Group Statistics erstellt ein Ausgabefeld pro Vorgang und benennt das Feld basierend auf dem Vorgang und Feld. So wäre etwa der Standardfeldname für den Vorgang `Sum` für ein Feld namens `Population` `SumOfPopulation`.
<Wert>_<Vorgang>	Enthält das Ergebnis eines Pivots, wobei `<Value>` einer der Werte in einer Pivot-Spalte und `<Operation>` der für die Spalte ausgeführte Vorgang ist. Weitere Informationen finden Sie unter **Erstellen einer Pivot-Tabelle** auf Seite 65.
GroupCount	Gibt die Anzahl der Datensätze in der Gruppe an.
GroupID	Eine jeder Gruppe sequenziell zugewiesene eindeutige Zahl. Die erste Gruppe hat den GroupID-Wert 1, die zweite den Wert 2 usw.
ComputationalCount<Vorgang> Of<Eingabefeldname>	Zeigt die tatsächliche Anzahl der Datensätze in einer Gruppe an, mit denen der Vorgang ausgeführt wird. Für den Vorgang `Average`, der mit der Spalte `Salary` ausgeführt wird, wird die Spalte `ComputationalCountAverageOfSalary` generiert.
Status	Meldet den Erfolg oder das Fehlschlagen der „Group Statistics“-Berechnungen. **null** Erfolg **F** Fehler

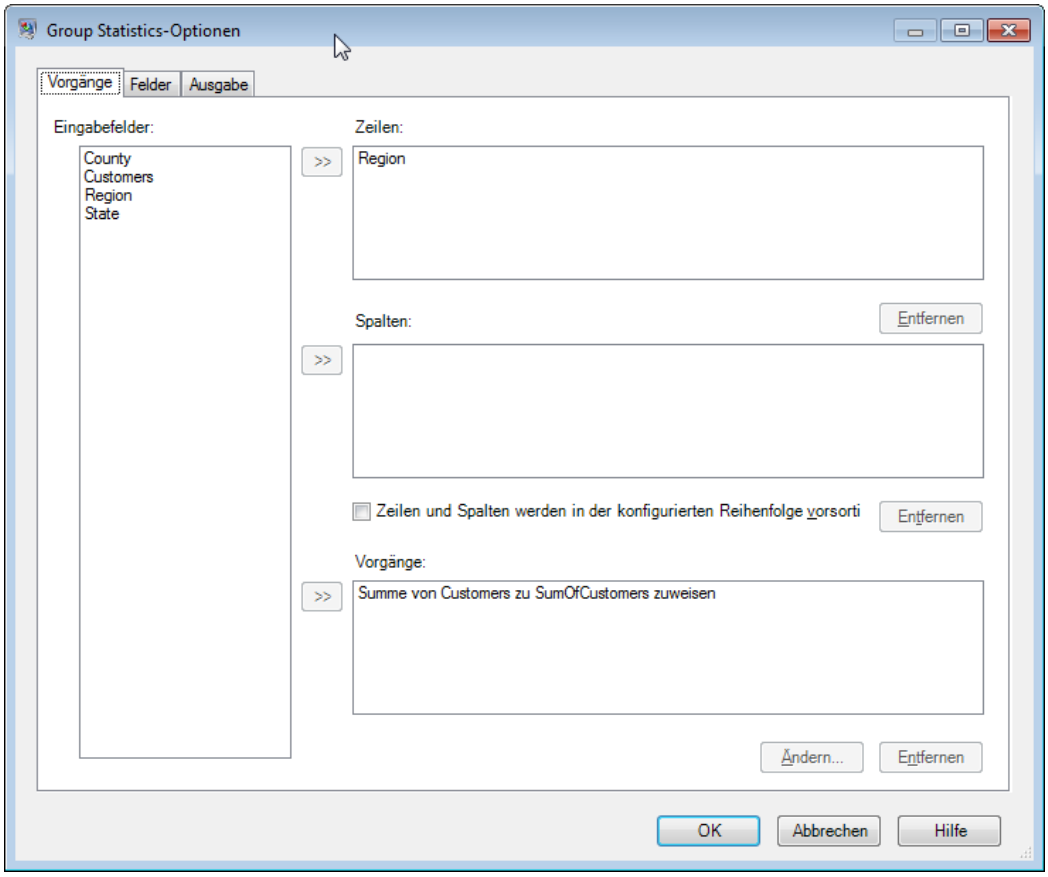
Feldname	Beschreibung/gültige Werte
Status.Code	Ursache für den Fehler, falls vorhanden. Die verfügbaren Statuscodes sind: **UnableToDoGroupStatistics** – Der „Group Statistics“-Schritt konnte ihre Berechnungen nicht ausführen. **Fehler beim Berechnen des Perzentilwertes** – Der Perzentilwert konnte anhand der bereitgestellten Eingabedaten nicht berechnet werden.
Status.Description	Eine ausführliche Beschreibung des Fehlers. **Der Eingabefeldwert konnte nicht in den Feldtyp konvertiert werden. Möglicherweise ein Überlauf!** – Eine Zahl in einem Eingabefeld ist größer, als der Datentyp zulässt. Versuchen Sie, sie in einen Datentyp zu konvertieren, der größere Zahlen unterstützt (wie „Double“).

Group Statistics-Beispiel

Diese Eingabedaten zeigen die Anzahl der Kunden, die Sie in bestimmten Countys haben. Die Daten zeigen auch den Staat der USA, in denen sich das County befindet (MD, VA, CA und NV), sowie die Region (Osten oder Westen). Die erste Zeile ist ein Kopfzeilen-Datensatz.

```
Region|State|County|Customers
East|MD|Calvert|25
East|MD|Calvert|30
East|MD|Prince Georges|30
East|MD|Montgomery|20
East|MD|Baltimore|25
East|VA|Fairfax|45
East|VA|Clarke|35
West|CA|Alameda|74
West|CA|Los Angeles|26
West|NV|Washoe|22
```

Wenn Sie die Gesamtzahl der Kunden für jede Region berechnen wollten, würden Sie auf der Registerkarte **Vorgänge** das Feld „Region“ als Zeile definieren. Für den Vorgang würden Sie einen „Sum“-Vorgang für das Feld „Kunden“ ausführen.



Das Ergebnis:

```
Region|SumOfCustomers
    East|210.0
    West|122.0
```

Anmerkung: Dieses Beispiel zeigt einen einfachen Gruppenstatistikvorgang nur mit Zeilen zum Aggregieren der Daten. Sie können auch eine Pivot-Tabelle erstellen, die sowohl Zeilen als auch Spalten aggregiert, indem Sie auf der Registerkarte **Vorgänge** eine Spalte angeben, nach der gruppiert werden soll.

Weitere Informationen zum Erstellen einer Pivot-Tabelle finden Sie unter **Erstellen einer Pivot-Tabelle** auf Seite 65.

Pivot-Tabellen

Eine Pivot-Tabelle aggregiert und transponiert Spaltenwerte im Datenfluss, um es einfacher zu machen, Daten visuell zu analysieren. Mit dem Pivot-Feature können Sie Eingabespalten in einem Kreuztabellierungsformat anordnen, das Zeilen, Spalten und zusammengefasste Werte erzeugt.

Sie können auch Felder als Eingabe verwenden und sie nicht anzeigen. Sie können Pivot zur Pivotisierung in zwei Dimensionen oder zum Gruppieren von Aggregatdaten in einer Dimension verwenden.

Dieses Beispiel zeigt Umsatzdaten für Hemden.

Tabelle 3: Eingabedaten

Region	Geschlecht	Stil	Versanddatum	Maßeinheiten	Preis	Kosten
Osten	Herren	Tee	31.01.2010	12	11,04	10,42
Osten	Herren	Golf	31.06.2010	12	13,00	10,60
Osten	Herren	Fancy	25.02.2010	12	11,96	11,74
Osten	Damen	Tee	31.01.2010	10	11,27	10,56
Osten	Damen	Golf	31.06.2010	10	12,12	11,95
Osten	Damen	Fancy	31.01.2010	10	13,74	13,33
Westen	Herren	Tee	31.01.2010	11	11,44	10,94
Westen	Herren	Golf	25.02.2010	11	12,63	11,73
Westen	Herren	Fancy	25.02.2010	11	12,06	10,51
Westen	Damen	Tee	25.02.2010	15	13,42	13,29
Westen	Damen	Golf	31.06.2010	15	11,48	10,67
Norden	Herren	Tee	25.02.2010	17	16,04	10,42
Norden	Herren	Fancy	25.02.2010	12	11,56	12,42
Norden	Damen	Tee	25.02.2010	16	12,32	18,42
Norden	Herren	Golf	31.01.2010	18	11,78	13,23
Norden	Damen	Tee	25.02.2010	12	18,45	11,64

Region	Geschlecht	Stil	Versanddatum	Maßeinheiten	Preis	Kosten
Norden	Damen	Golf	25.02.2010	14	11,23	19,85
Norden	Herren	Fancy	31.01.2010	16	12,54	13,42
Norden	Damen	Tee	25.02.2010	17	181,73	15,83
Süden	Herren	Fancy	31.01.2010	19	14,15	13,42
Süden	Damen	Tee	25.02.2010	11	11,85	12,92
Süden	Damen	Fancy	31.01.2010	13	11,54	14,35
Süden	Herren	Tee	25.02.2010	15	14,14	14,73
Süden	Herren	Golf	25.02.2010	16	17,83	17,83
Süden	Damen	Fancy	31.06.2010	11	18,24	12,35
Süden	Damen	Tee	31.01.2010	20	19,94	12,95
Süden	Herren	Golf	25.02.2010	12	21,25	19,56

Wir möchten ermitteln können, wie viele Einheiten wir in jeder Region zu jedem Versanddatum verkauft haben. Dazu generieren wir mit dem Pivot-Feature diese Tabelle:

Tabelle 4: Pivot-Tabelle

Region	31.01.2010_Versanddatum	25.02.2010_Versanddatum	31.06.2010_Versanddatum
Osten	32	12	22
Norden	34	88	
Süden	52	54	11
Westen	11	37	15

In diesem Fall ist die Spalte „Versanddatum“, die Zeile „Region“ und die Daten, die wir sehen möchten, „Einheiten“. Die Gesamtzahl der versandten Einheiten wird hier mithilfe einer Summenaggregation angezeigt.

Erstellen einer Pivot-Tabelle

In einer Pivot-Tabelle werden Daten zur einfacheren Analyse zusammengefasst. Dazu werden Kategorien von Tabellenzeilen und -spalten auf Basis der Eingabedaten erstellt. Weitere Informationen finden Sie unter **Pivot-Tabellen** auf Seite 62.

In den Optionen des **Group Statistics**-Schritts:

1. Wählen Sie auf der Registerkarte **Vorgänge** ein Feld aus den **Eingabefeldern** aus, das die als Zeilenbeschriftungen in der Pivot-Tabelle zu verwendenden Daten enthält. Klicken Sie dann auf die Schaltfläche **>>** neben dem Feld **Zeilen**.
2. Wählen Sie ein Feld aus, das die als Spalten in der Pivot-Tabelle zu verwendenden Daten enthält. Klicken Sie dann auf die Schaltfläche **>>** neben dem Feld **Spalten**.

   Tipp: Führen Sie an diesem Punkt eine Überprüfung aus, um die Ergebnisse Ihrer Auswahl zu sehen. Dies hilft Ihnen, die Ergebnisse der Kreuztabellierungen auf Basis der ausgewählten Spalten und Zeilen zu visualisieren.

3. Aktivieren Sie **Zeilen und Spalten werden in der konfigurierten Reihenfolge vorsortiert**, um die Sortierung der Eingabedatensätze zu überspringen.

   Wenn dieses Feld aktiviert ist, verarbeitet der Schritt die Eingabedatensätze, ohne sie zu sortieren.

   Anmerkung: Aktivieren Sie diese Option, wenn die Datensätze bereits sortiert sind.

4. Klicken Sie auf die Schaltfläche **>>** neben dem Feld **Vorgänge**, um die auszuführenden Vorgänge zu definieren.

   Im Fenster **Vorgang hinzufügen**:

   a) Wählen Sie den **Vorgang** aus, der ausgeführt werden soll.
   b) Wählen Sie im Bereich **Eingabefeld** den **Namen** und den **Typ** des Eingabefeld aus, für das der Vorgang durchgeführt werden muss.
   c) Geben Sie im Bereich **Ausgabefeld** den **Namen** des Ausgabefelds ein, der beim Ausführen des Vorgangs generiert werden soll, und wählen Sie seinen **Typ** aus.
   d) Aktivieren Sie **Anzahl der Datensätze aufrufen, die berechnet werden**, um die tatsächliche Anzahl von Eingabedatensätzen, mit denen der Vorgang ausgeführt wird, als separate Ausgabespalte abzurufen.

      Datensätze mit Werten von `null` werden bei der Zählung `ComputationalCount<Operation>Of<InputFieldName>` nicht berücksichtigt.

      Funktionen mit Unterstützung für die berechnete Anzahl:

      • Durchschnitt

- Streuung
- Z-Score
- Standardabweichung
- Perzentil
- Prozentrang
- Summe

Bei allen anderen Vorgängen bleibt dieses Kästchen deaktiviert.

5. Klicken Sie auf die Registerkarte **Felder**, um die Ausgabefelder für jede Spalte in der Pivot-Tabelle zu definieren.

   Tipp: Um die Felder genau zu definieren, führen Sie vor diesem Schritt einmal einen Überprüfungsfluss aus, um die durch Ihre Daten generierten Spaltennamen anzuzeigen.

   a) Klicken Sie auf **Add** (Hinzufügen).
   Das Fenster **Feld hinzufügen** wird angezeigt.

   b) Die Spalten des Rasters im Fenster **Feld hinzufügen** basieren auf den Feldern **Spalten**, die Sie auf der Registerkarte **Vorgänge** ausgewählt haben. Geben Sie in diese Rasterspalten die Werte ein, die Sie beim Ausführen des Überprüfungsflusses als Spaltenüberschriften angezeigt bekommen.

   Die Datensätze in der Spalte „Daten“ können über das Feature „Import“ in einem Vorgang ausgefüllt werden. Importieren von Daten aus einer CSV- oder TXT-Datei:

   1. Klicken Sie auf **Importieren**.
   2. Durchsuchen Sie die Quelldatei anhand des Felds **Dateiname**.
   3. Geben Sie die Werte für **Feld- und Datensatz-Trennzeichen** ein.
   4. Klicken Sie auf OK.

   Die Tabelle „Spaltendaten“ aller Datensätze in der Datei wird ausgefüllt.

   Anmerkung: Die Quelldatei sollte keine Kopfzeile aufweisen.

   Wenn Sie beispielsweise auf der Registerkarte **Vorgänge** unter **Spalten** ein Eingabefeld namens `ShipDat e` ausgewählt haben, ist im Raster im Fenster **Feld hinzufügen** eine Spalte mit der Beschriftung `ShipDate` vorhanden. Geben Sie in dieser Rasterspalte die genauen Werte für `ShipDate` ein, die in den Eingabedaten Ihres Datenflusses vorhanden sind, zum Beispiel `2/25/2010`, `1/31/2010`.

   c) Wählen Sie im Feld **Vorgänge** einen oder mehrere Vorgänge aus, für die Ausgabespalten für jeden eingegebenen Spaltenfeldwert generiert werden. Beachten Sie, dass der ausgewählte Vorgang nur den Feldnamen betrifft und nicht die tatsächliche Berechnung steuert.

   Um die im Feld **Vorgang** der Registerkarte **Felder** aufgelisteten Vorgänge zu ändern, ändern Sie die Werte im Feld **Vorgänge** auf der Registerkarte **Vorgänge**.

   Achtung: Die Option `ComputationalCount<Operation>Of<InputFieldName>` des Vorgangs „Computational Count“ wird nur aufgeführt, wenn das Kästchen **Anzahl der**

Datensätze aufrufen, die berechnet werden beim Definieren des **Vorgangs** auf der Registerkarte **Vorgänge** aktiviert wurde.

d) Klicken Sie auf **Add** (Hinzufügen).

6. Klicken Sie auf **OK**.

Für jeden Eingabewert, den Sie im obigen Raster eingegeben haben, werden automatisch Ausgabespalten erzeugt, indem diese jedem einzelnen der ausgewählten Werte **Vorgang** zugeordnet werden. Zur automatischen Erzeugung der finalen Ausgabespalten wird ein *kartesisches Produkt* der im Raster eingegebenen Eingabespaltenwerte und der ausgewählten Vorgänge verwendet.

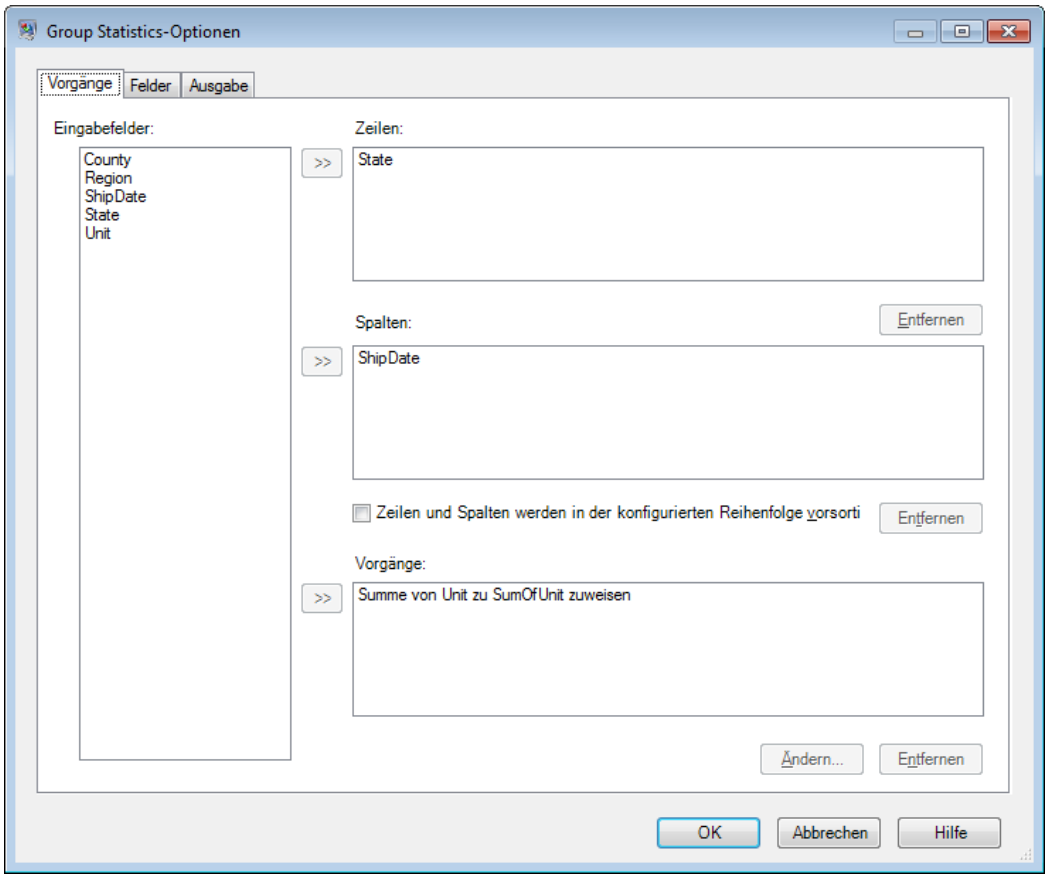
Die Namen dieser Ausgabespalten folgen der Benennungskonvention `<Data>_<Operation>Of<InputFieldName>`, wobei `<Data>` der Wert ist, den Sie im ersten Feld eingegeben haben, `<Operation>` der Vorgang ist, den Sie im Feld **Vorgang** ausgewählt haben, und `<InputFieldName>` die Eingabespalte ist, mit der `operation` ausgeführt wird.

Beispiel für Pivot-Tabelle

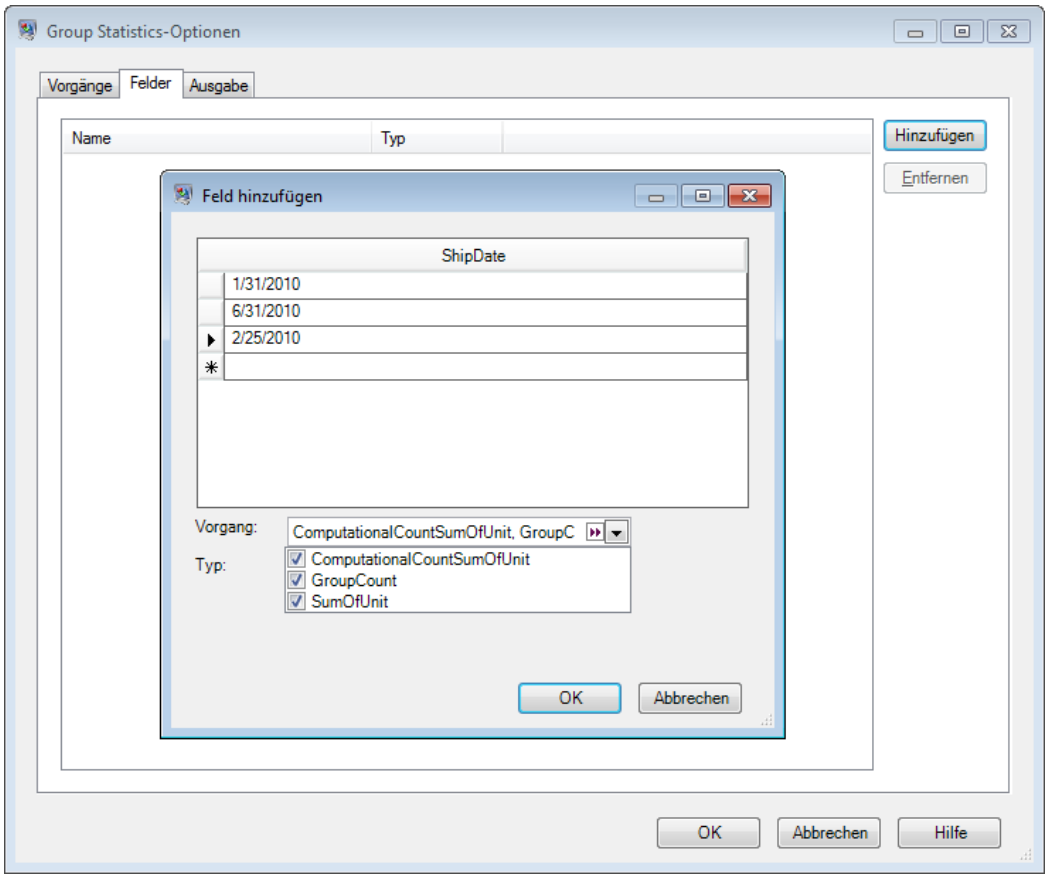
Die Eingabedaten, die Versandinformationen der Fulfillment-Abteilung zeigen:

```
Region,State,County,ShipDate,Unit
East,MD,Calvert,1/31/2010,
East,MD,Calvert,6/31/2010,212
East,MD,Calvert,1/31/2010,633
East,MD,Calvert,6/31/2010,234
East,MD,Prince Georges,2/25/2010,112
East,MD,Montgomery,1/31/2010,120
East,MD,Baltimore,6/31/2010,210
East,VA,Fairfax,1/31/2010,710
West,CA,SanJose,1/31/2010,191
West,CA,Alameda,2/25/2010,411
West,CA,Los Angeles,2/25/2010,
West,CA,Los Angeles,2/25/2010,215
West,CA,Los Angeles,6/31/2010,615
West,CA,Los Angeles,6/31/2010,727
```

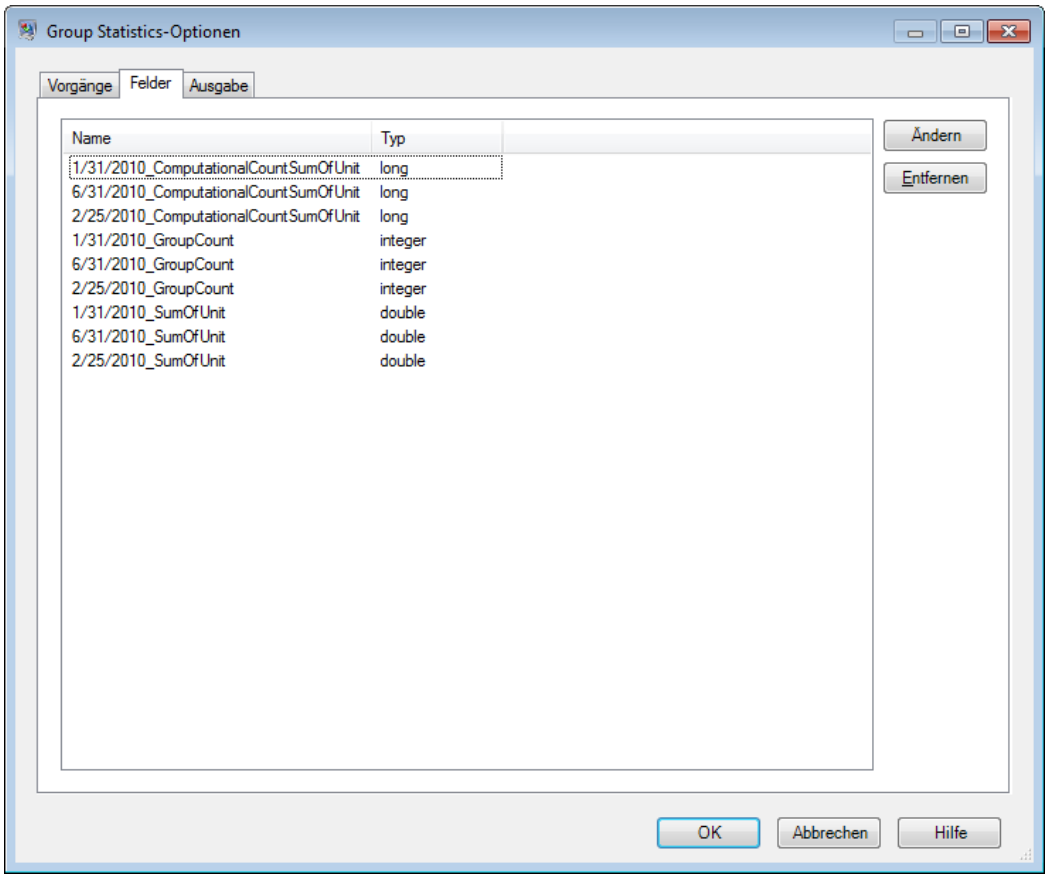
Um die Anzahl von Sendungen zu ermitteln, die an einzelnen Versandtagen in die einzelnen Bundesländer/Kantone versendet wurden, erstellen Sie eine Pivot-Tabelle, indem Sie den „Group Statistics“-Schritt wie dargestellt konfigurieren:



Fügen Sie auf der Registerkarte **Felder** der Schrittoptionen die genauen Daten in das Raster ein, das im Feld `ShipDate` der Eingabedaten des Datenflusses erscheint, und wählen Sie die Werte **Vorgang** aus, die für jeden dieser Spaltenwerte angezeigt werden sollen.



Sobald Sie im Fenster **Feld hinzufügen** auf **OK** klicken, werden die zu erstellenden Ausgabespalten automatisch auf der Registerkarte **Felder** aufgelistet. Diese Ausgabespalten sind ein *kartesisches Produkt* der genauen Eingabewerte und der Vorgänge, die Sie im Fenster **Feld hinzufügen** ausgewählt haben.



Ausgabe

```
State,1/31/2010_GroupCount,1/31/2010_ComputationalCountSumOfUnit,
1/31/2010_SumOfUnit,2/25/2010_GroupCount,2/25/2010_ComputationalCountSumOfUnit,
2/25/2010_SumOfUnit,6/31/2010_GroupCount,6/31/2010_ComputationalCountSumOfUnit,
6/31/2010_SumOfUnit
VA,1,1,710,,,,,,
CA,1,1,191,3,2,626,2,2,1342
MD,3,2,753,1,1,112,3,3,656
```

Math

Im „Math“-Schritt werden mathematische Berechnungen in einer einzelnen Datenzeile verarbeitet. Zudem können Sie in diesem Schritt unter Verwendung von mindestens einem Ausdruck eine Vielzahl von mathematischen Funktionen durchführen. Daten werden als Zeichenfolgen eingegeben. Basierend auf dem auf die Daten angewendeten Vorgangstyp muss es sich bei den Werten jedoch um numerische oder boolesche Werte handeln.

1. Klicken Sie unter „Steuerungsschritte“ auf den „Math“-Schritte und ziehen Sie diese auf die Arbeitsfläche. Platzieren Sie ihn in dem von Ihnen ausgewählten Datenfluss.
2. Verbinden Sie den Schritt mit anderen Schritten auf der Arbeitsfläche.
3. Doppelklicken Sie auf den „Math“-Schritt. Das Dialogfeld „**Math-Optionen**“ wird angezeigt, in dem die Registerkarte „Ausdrücke“ geöffnet ist. In dieser Ansicht werden die Eingabefelder, der Rechner und die Arbeitsfläche „Ausdrücke“ angezeigt. Alternativ können Sie auf die Registerkarte „Funktionen“ klicken, um anstelle des Rechners Funktionen zu verwenden.

Das Steuerelement „Eingabefelder“ listet die gültigen Felder auf, die am Eingangsport gefunden wurden. Die Feldnamenssyntax ist sehr flexibel, weist jedoch basierend auf Groovy-Scripting-Regeln einige Einschränkungen auf. Wenn Sie mit Groovy-Scripting nicht vertraut sind, finden Sie auf der folgenden Website vollständige Informationen zu Groovy: **groovy-lang.org**.

Verwenden des Rechners

Das Steuerelement „Rechner“ enthält Schaltflächen für die Eingabe numerischer Konstanten und Operatoren in einen Ausdruck. Durch Doppelklicken auf Felder, Konstanten, Operatoren und Funktionen werden diese in einen Ausdruck eingefügt.

Tabelle 5: Rechneroperatoren

Operator	Beschreibung
Rücktaste	Wird verwendet, um in einen Ausdruck ein Leerzeichen zurückzugehen
pi	Pi, eine mathematische Konstante, die das Verhältnis des Umfangs eines Kreises zu seinem Durchmesser darstellt
e	Eulersche Zahl, eine mathematische Konstante, die die Basis des natürlichen Logarithmus ist
/	Division
*	Multiplikation
+	Addition
-	Subtraktion
x^y	Potenz von (x^2 ist z. B. x potenziert mit 2)

Operator	Beschreibung
Mod	Modulo, der Restwert eines Vorgangs
;	Semikolon, wird am Ende von Ausdrücken verwendet
=	Zuweisungsoperator
()	Klammern, für die Darstellung einer Hierarchie in einem Ausdruck
.	Dezimalkomma
if\else	Bedingungsanweisung zum Ergreifen von Maßnahmen, wenn eine Bedingung „Wahr“ ergibt; andernfalls müssen andere Maßnahmen ergriffen werden
if\else if\else	Mehrere Bedingungsanweisungen zum Ergreifen von Maßnahmen, wenn eine Bedingung „Wahr“ ergibt; andernfalls müssen andere Maßnahmen ergriffen werden
==	Gleich, in einer mathematischen Funktion
!=	Nicht gleich
&&	Logisch und
||	Logisch oder
>	Größer als
>=	Größer oder gleich
<	Kleiner als
<=	Kleiner oder gleich

Verwenden von Funktionen und Konstanten

Im „Math“-Schritt werden mehrere Funktionen zur Verfügung gestellt, die in einem Ausdruck verwendet werden können. Funktionen weisen die allgemeine Form Funktion(Parameter); Funktion(Parameter,Parameter); Funktion(Parameter,...) auf, wobei „Parameter“ für eine numerische Konstante, eine Variable oder einen mathematischen Ausdruck steht. Funktionen können mit anderen mathematischen Ausdrücken verwendet werden (z. B. x = Sin(y) * Cos(z)).

Konstanten, Konvertierung, Math und Trigonometrie. Alle unterstützten Funktionen werden nachfolgend in der zugehörigen Kategorie aufgeführt.

Tabelle 6: Unterstützte Funktionen

Funktion	Beschreibung
Konstanten	
e	Eine mathematische Konstante, die die Basis des natürlichen Algorithmus ist.
false	Eine boolesche Konstante, die den Wert „Falsch“ darstellt.
Infinity	Eine mathematische Konstante, die Unendlichkeit darstellt.
NaN	Eine mathematische Konstante, die einen Wert darstellt, der keine Nummer ist.
Pi	Eine mathematische Konstante, die das Verhältnis des Umfangs eines Kreises zu seinem Durchmesser darstellt.
true	Eine boolesche Konstante, die den wahren Wert darstellt.
Konvertierung	
Abs (Wert)	Akzeptiert einen Parameter. Gibt den absoluten Wert des angegebenen Wertes zurück.
Ceil (Wert)	Akzeptiert einen Parameter. Gibt einen aufgerundeten Wert zurück (z. B. Ceil(5,5) gibt 6 zurück).
DegToRad (Wert)	Akzeptiert einen Parameter. Konvertiert einen angegebenen Wert von Grad in Radiant.
Floor (Wert)	Akzeptiert einen Parameter. Gibt einen abgerundeten Wert zurück (z. B. Floor(5,5) gibt 5 zurück).
RadToDeg (Wert)	Akzeptiert einen Parameter. Konvertiert einen angegebenen Wert von Radiant in Grad.

Funktion	Beschreibung
Round (Wert)	Akzeptiert einen Parameter. Gibt einen gerundeten Wert zurück.
Math	
Avg (Wert, Wert,...)	Akzeptiert einen oder mehrere Parameter. Gibt den Mittelwert aller angegebenen Werte zurück.
Exp (Wert)	Akzeptiert einen Parameter. Gibt die Eulersche Zahl potenziert mit dem Wert zurück.
Fac (Wert)	Akzeptiert einen Parameter. Gibt die Fakultät eines angegebenen Wertes zurück (z. B. Fac(6) wird mit 6*5*4*3*2*1 berechnet und gibt 720 zurück).
Ln (Wert)	Akzeptiert einen Parameter. Gibt den natürlichen Logarithmus (Basis e) eines angegebenen Wertes zurück.
Log (Wert)	Akzeptiert einen Parameter. Gibt den natürlichen Logarithmus (Basis 10) eines angegebenen Wertes zurück.
Max (Wert, Wert,...)	Akzeptiert einen oder mehrere Parameter. Gibt den maximalen übergebenen Wert zurück.
Min (Wert, Wert,...)	Akzeptiert einen oder mehrere Parameter. Gibt den minimalen übergebenen Wert zurück.
Sqrt (Wert)	Akzeptiert einen oder mehrere Parameter. Gibt die Quadratwurzel des übergebenen Wertes zurück.
Sum (Wert)	Akzeptiert einen Parameter. Gibt die Summe der angegebenen Werte zurück.
Trigonometrie	

Funktion	Beschreibung
ArcCos (Wert)	Akzeptiert einen Parameter. Gibt den Arkuskosinus eines Wertes zurück.
ArcSin (Wert)	Akzeptiert einen Parameter. Gibt den Arkussinus eines Wertes zurück.
ArcTan (Wert)	Akzeptiert einen Parameter. Gibt den Arkustangens eines Wertes zurück.
Cos (Wert)	Akzeptiert einen Parameter. Gibt den Kosinus eines Wertes zurück.
Ln (Wert)	Akzeptiert einen Parameter. Gibt den natürlichen Logarithmus (Basis e) eines angegebenen Wertes zurück.
Sin (Wert)	Akzeptiert einen Parameter. Gibt den Sinus eines Wertes zurück.
Tan (Wert)	Akzeptiert einen Parameter. Gibt den Tangens eines Wertes zurück.

Verwenden von Bedingungsanweisungen

Abhängig davon, ob verschiedene Bedingungen „Wahr“ oder „Falsch“ ergeben, können zum Ergreifen von Maßnahmen Bedingungsanweisungen verwendet werden. Eine Gruppierung mithilfe von Klammern, (und), kann bei komplexeren Bedingungen verwendet werden.

Tabelle 7: Bedingungen

Bedingung	Beschreibung
Gleich	Ausdruck = = Ausdruck
Ungleich	Ausdruck != Ausdruck

Bedingung	Beschreibung
Größer als	Ausdruck > Ausdruck
Größer oder gleich	Ausdruck >= Ausdruck
Kleiner als	Ausdruck < Ausdruck
Kleiner oder gleich	Ausdruck <= Ausdruck
Not-Bedingung	!Bedingung
Und	Bedingung && Bedingung
Oder	Bedingung || Bedingung

If-Anweisung

```
if(condition)
{
     actions to take if condition is true
}
```

Klammern sind nur dann erforderlich, wenn nach dem „if“ mehrere Anweisungen ausgeführt werden.

If-Else-If-Anweisungen

```
if(condition)
{
     actions to take if condition is true
}
else if(condition)
{
     actions to take if condition is true
}
else if...
```

```
if(SideLength != NaN)
 {
 AreaOfPolygon=
 ((SideLength^2)*NumberOfSides)/
 (4*Tan(pi/NumberOfSides));
 }
 else if(Radius != NaN)
 {
 AreaOfPolygon=
```

```
 (Radius^2)*NumberOfSides*Sin((2*pi)/NumberOfSides)/2;
 }
```

Es kann mindestens eine Else-if-Anweisung angegeben werden. Klammern sind nur dann erforderlich, wenn nach dem „if-else-if-else“ mehrere Anweisungen ausgeführt werden.

Else-If-Anweisung

```
if(condition)
 {
  actions to take if condition is true
 }
 else if(condition)
 {
  actions to take if condition is true
 }
 else if...
 else
 {
  actions to take if no conditions are met
 }
```

Verwenden des Steuerelements „Ausdrücke“

Mit dem Steuerelement „Ausdrücke“ können Sie mathematische Ausdrücke eingeben, die im „Math“-Schritt ausgewertet werden sollen. Die Steuerelemente „Eingabe“, „Rechner“ und „Funktionen“ werden zum Einfügen von Werten in dieses Steuerelement verwendet. Sie können Ausdrücke auch manuell in das Steuerelement eingeben. Ausdrücke weisen die Form einer Konstanten, Variablen oder eines mathematischen Vorgangs auf und bestehen aus numerischen Konstanten und Variablen. Numerische Konstanten sind ganze Zahlen oder Dezimalzahlen, die signiert werden können. Variablen stellen Daten aus der Eingangszeile dar. Wenn in der Eingabe beispielsweise die Felder x, y und z definiert sind, können x, y, und z in einem Ausdruck verwendet werden. Variablen werden zur Laufzeit durch Feldwerte ersetzt.

Im „Math“-Schritt sind zudem gruppierte Ausdrücke zugelassen. Dies beinhaltet die Verwendung von Klammern zum Gruppieren der Ausdrücke und zum Überschreiben der Operatorrangfolge. Beispiel: 2 * 5^2 ist gleich 50, während (2 * 5)^2 gleich 100 ist.

> **Anmerkung:** Jeder von Ihnen eingegebene Ausdruck muss mit einem Semikolon enden.

Zusätzlich können im Steuerelement „Ausdrücke“ – abhängig davon, ob verschiedene Bedingungen „Wahr“ oder „Falsch“ ergeben – Bedingungsanweisungen zum Ergreifen von Maßnahmen verwendet werden. Weitere Informationen zu Bedingungsanweisungen finde Sie unter **Verwenden von Bedingungsanweisungen** auf Seite 75.

Der „Math“-Schritt beschäftigt sich primär mit Zuweisungsausdrücken, bei denen die Ausgabe eines Ausdrucks einer Variable zugewiesen wird. Im Schritt werden mehrere Zuweisungsvorgänge unterstützt. Zudem kann die Ausgabe eines vorherigen Zuweisungsvorgangs darin verwendet werden.

Beispiele für Zuweisungsausdrücke

Im nachfolgenden Szenario x = 10 und z = 1000:

```
x=5+5
z=x*100
```

Im nachfolgenden Szenario wird der Bereich eines Vielecks basierend auf der Länge einer Seite und der Anzahl der Seiten berechnet.

```
AreaOfPolygon=
     ((SideLength^2)*NumberOfSides)/
     (4*Tan(pi/NumberOfSides));
```

Verwenden des Steuerelements „Felder“

Mit dem Steuerelement „Felder“ können Sie Eingabe- und Ausgabefeldtypen ändern. Sie können Feldtypen innerhalb dieses Steuerelements ändern, indem Sie auf den Dropdown-Pfeil in der Spalte „Typ“ klicken und eine Auswahl aus der Liste treffen, die folgende Optionen umfasst:

boolean Ein logischer Typ mit zwei Werten: wahr und falsch. Boolesche Variablen können in Bedingungsanweisungen zum Steuern des Flusses verwendet werden. Im folgenden Codebeispiel wird ein boolescher Ausdruck dargestellt:

```
if(x && y)
     z=1;
  else if(x)
     z=2;
  else if(y)
     z=3;
  else
     z=4;
```

double Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit doppelter Genauigkeit zwischen 2^{-1074} und $(2-2^{-52})\times2^{1023}$ enthält. In der E-Notation liegt der Wertebereich bei -1,79769313486232E+308 bis 1,79769313486232E+308.

float Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit einzelner Genauigkeit zwischen 2^{-149} und $(2-2^{23})\times2^{127}$ enthält. In der E-Notation liegt der Wertebereich bei -3,402823E+38 bis 3,402823E+38.

integer Ein numerischer Datentyp, der sowohl positive als auch negative ganze Zahlen zwischen -2^{31} (-2,147,483,648) und 2^{31}-1 (2,147,483,647) enthält.

long Ein numerischer Datentyp, der sowohl negative als auch positive ganze Zahlen zwischen -2^{63} (-9.223.372.036.854.775.808) und 2^{63}-1 (9.223.372.036.854.775.807) enthält.

Verwenden des Steuerelements „Vorschau“

Mit dem Steuerelement „Vorschau“ können Sie mathematische Ausdrücke testen. Im Bereich „Eingabedaten“ werden Felder aufgeführt; Sie können bestimmte Werte angeben, die an den Ausdruck übergeben werden sollen, und die Ausgabe im Bereich „Ergebnisse“ unter „Eingabedaten“ anzeigen.

Numerische Felder werden mit 0 (0,000 bei „Double“) und boolesche Werte mit „Falsch“ initialisiert. Felder vom Typ „Double“ und „Float“ sind auf vier Dezimalstellen begrenzt, Felder vom Typ „Integer“ und „Long“ haben keine Dezimalstellen.

Record Combiner

Record Combiner kombiniert zwei oder mehr Datensätze aus mehreren Streams in einem einzigen Datensatz. Record Combiner kann über einen oder mehrere Schritteingabeports verfügen. So können Sie beispielsweise eine Gruppe von Datensätzen aus einer Schritteingabe (Port) mit einer anderen Gruppe aus einer zweiten Schritteingabe (Port 2) kombinieren, um die Datensätze in einem einzigen Datensatz zusammenzuführen. Wenn Sie einen mittlere Schritt löschen, werden die Ports nicht erneut aufeinander folgend nummeriert.

> **Anmerkung:** „Record Combiner“ gibt erst einen Datensatz für die Ausgabe aus, wenn jeder seiner Eingabeports einen Datensatz empfangen hat. Er muss so viele Datensätze kombinieren, wie Eingabeports vorhanden sind, bevor ein Datensatz ausgegeben wird.

Sie können angeben, welcher Port beibehalten werden soll, wenn die Eingabestreams Felder mit gleichem Namen enthalten. Wenn Sie beispielsweise Datensätze aus zwei Streams kombinieren und beide Streams ein Feld mit dem Namen „AccountNumber“ enthalten, können Sie festlegen, welches Feld eines Streams beibehalten werden soll. Wählen Sie dazu den Record Combiner-Eingabeport aus, der dem beizubehaltenden Stream entspricht. Daraufhin werden die Daten aus dem Feld „AccountNumber“ des anderen Streams verworfen.

Record Joiner

Record Joiner führt einen SQL-ähnlichen `JOIN`-Vorgang aus, um Datensätze aus unterschiedlichen Streams basierend auf einer Beziehung zwischen Feldern in den Streams zu kombinieren. Mithilfe von Record Joiner können Sie Datensätze aus mehreren Dateien, Datenbanken oder beliebigen Upstream-Kanälen im Datenfluss zusammenführen. Dabei müssen Sie mindestens zwei Eingabekanäle mit Record Joiner verbinden. Die Ergebnisse des `JOIN`-Vorgangs werden anschließend in einen Ausgabekanal geschrieben. Optional können Datensätze, die nicht die Zusammenführungsbedingung erfüllen, in einen separaten Ausgabekanal geschrieben werden.

> **Anmerkung:** Vor der Verwendung von Record Joiner müssen Sie den SQL `JOIN`-Vorgang gut verstanden haben. Weitere Informationen finden Sie unter **wikipedia.org/wiki/Join_(SQL)**.

Join-Definition

Option	Beschreibung
Linker Port	Der Port, dessen Datensätze Sie als die linke Tabelle im `JOIN`-Vorgang verwenden möchten. Alle anderen Eingabeports werden als rechte Tabellen im `JOIN`-Vorgang verwendet. **Anmerkung:** „Linke“ und „rechte“ Tabelle sind SQL `JOIN`-Begriffe. Vor der Verwendung von Record Joiner müssen Sie den SQL `JOIN`-Vorgang gut verstanden haben. Weitere Informationen finden Sie unter **wikipedia.org/wiki/Join_(SQL)**.
Join-Typ	Der `JOIN`-Vorgangstyp, den Sie ausführen möchten. Zur Auswahl stehen: **Left Outer** Gibt alle Datensätze aus dem linken Port zurück, selbst wenn es keine Übereinstimmungen zwischen dem linken Port und den anderen Ports gibt. Diese Option gibt alle Datensätze aus dem linken Port UND jeden Datensatz zurück, der in einem beliebigen anderen Port übereinstimmt. **Vollständig** Gibt alle Datensätze aus allen Ports zurück. **Inner** Gibt nur die Datensätze zurück, die zwischen dem linken Port und einem anderen Port eine Übereinstimmung aufweisen. Wenn Sie z. B. über vier Eingabequellen verfügen und Port 1 der linke Port ist, gibt ein Inner-Join Datensätze zurück, die übereinstimmende Felder zwischen Port 1 und Port 2, Port 1 und Port 3, und Port 1 und Port 4 aufweisen.
Join-Felder	Das Feld oder die Felder aus dem linken Port, die mit den Daten in einem Feld eines anderen Ports übereinstimmen müssen, damit die Datensätze zusammengeführt werden können. **Anmerkung:** Es können nur Felder als Join-Felder verwendet werden, die den Datentyp „string“, „integer“, „date“ oder „datetime“ aufweisen.

Option	Beschreibung
Daten aus dem linken Port werden sortiert	Gibt an, ob die Datensätze im linken Port bereits nach dem in **Join-Felder** angegebenen Feld sortiert sind. Wenn die Datensätze bereits sortiert sind, kann die Aktivierung dieses Kästchens die Leistung verbessern. Wenn Sie dieses Kästchen nicht aktivieren, sortiert Record Joiner entsprechend dem in **Join-Felder** angegebenen Feld, bevor der Join-Vorgang ausgeführt wird. Wenn Sie mehrere Join-Felder angeben, müssen die Datensätze unter Verwendung der Feldsortierung in **Join-Felder** sortiert werden. Sie haben z. B. zwei Join-Felder: Amount Region Die Datensätze müssen zunächst nach dem Feld „Betrag“ und anschließend nach dem Feld „Region“ sortiert werden. **Wichtig:** Wenn Sie diese Option auswählen, und die Datensätze nicht sortiert sind, erhalten Sie falsche Ergebnisse von Record Joiner. Wählen Sie diese Option nur aus, wenn Sie sicher sind, dass die Datensätze im linken Port bereits sortiert sind.
Join-Definitionen	Beschreibt die Join-Bedingung, die zur Bestimmung verwendet wird, ob ein Datensatz aus dem linken Port mit einem Datensatz aus einem der anderen Ports zusammengeführt werden soll. Beispiel: `port1.Name = port2.Name` Dies gibt an, dass wenn der Wert im Feld „Name“ eines Datensatzes aus Port 1 mit dem Wert im Feld „Name“ eines Datensatzes aus Port 2 übereinstimmt, die beiden Datensätze zusammengeführt werden. Klicken Sie zum Ändern der Join-Bedingung auf **Ändern**. Wählen Sie ein Feld aus dem rechten Port aus, dessen Daten mit den Daten im Join-Feld des linken Ports übereinstimmen müssen, damit die Datensätze zusammengeführt werden können. Wenn Sie das Feld im linken Port ändern möchten, klicken Sie auf **Abbrechen** und ändern Sie es im Feld **Join-Felder**. Wenn die Datensätze im rechten Port nach dem Join-Feld sortiert sind, aktivieren Sie das Kästchen **Daten aus dem rechten Port werden sortiert**. Die Aktivierung dieses Kästchens kann die Leistung verbessern. **Wichtig:** Wenn Sie die Option **Daten aus dem rechten Port werden sortiert** auswählen, und die Datensätze nicht sortiert sind, erhalten Sie falsche Ergebnisse von Record Joiner. Wählen Sie diese Option nur aus, wenn Sie sicher sind, dass die Datensätze im rechten Port bereits sortiert sind.

Feldauflösung

Auf dieser Registerkarte wird angegeben, welche Portdaten im zusammengeführten Datensatz verwendet werden, wenn der gleiche Feldname in mehr als einem Eingabeport vorhanden ist. Wenn Sie beispielsweise ein Join für zwei Datenquellen durchführen, und jede Quelle ein Feld mit dem Namen „DateOfBirth“ enthält, können Sie festlegen, welche Portdaten im Feld „DateOfBirth“ des zusammengeführten Datensatzes verwendet werden.

Wenn es Felder mit dem gleichen Namen aber unterschiedlichen Daten gibt, und Sie die Daten beider Felder im zusammengeführten Datensatz beibehalten möchten, müssen Sie eines der Felder

umbenennen, bevor die Daten an Record Joiner gesendet werden. Sie können den Transformer-Schritt verwenden, um die Felder umzubenennen.

Handhabung von nicht zusammengeführten Datensätzen

Damit ein Datensatz in die Record Joiner-Ausgabe einbezogen werden kann, muss er die Join-Bedingung erfüllen. Oder es muss ein Join-Typ ausgewählt werden, der sowohl zusammengeführte Datensätze als auch die Datensätze zurückgibt, die nicht die Join-Bedingung erfüllen. Ein vollständiger Join gibt z. B. alle Datensätze aus allen Eingabeports zurück, egal ob ein Datensatz die Join-Bedingung erfüllt. Im Falle eines Join-Typs, der nicht alle Datensätze aus allen Ports zurückgibt (wie z. B. der Left Outer-Join oder ein Inner-Join), werden nur Datensätze in die Record Joiner-Ausgabe einbezogen, die die Join-Bedingung erfüllen.

Verwenden Sie zur Erfassung der Datensätze, die nicht im Ergebnis des Join-Vorgangs enthalten sind, den Ausgabeport **not_joined**. Die Ausgabe aus diesem Port enthält alle Datensätzen, die nicht in den regulären Ausgabeport einbezogen wurden. Der Ausgabeport **not_joined** ist das weiße Dreieck auf der rechten Seite des „Record Joiner“-Schrittes, wie hier dargestellt:

Datensätzen, die aus diesem Port stammen, wurde das Feld **InputPortIndex** hinzugefügt. Dieses Feld enthält die Zahl des Record Joiner-Eingabeports, aus dem der Datensatz stammt. Dadurch können Sie die Quelle des Datensatzes identifizieren.

Anmerkung:

- Um in diesem Schritt eine optimale Leistung zu erreichen, stellen Sie sicher, dass zwei unabhängige Datensatzströme zusammengeführt werden, um eine konsolidierte Ausgabe zu generieren.
- Wenn ein einzelner Pfad zuerst über einen Broadcaster oder einen Conditional Router verzweigt wird und dann über einen Record Joiner wieder zusammengeführt wird, könnte der Fluss hängen bleiben. Wenn zwischen Verzweigung und Zusammenführung mehrere Schritte vorhanden sind, verwenden Sie den Sorter so dicht am Record Joiner wie möglich.

Sorter

Der Sorter sortiert Datensätze mithilfe der von Ihnen angegebenen Felder. Sie können beispielsweise Datensätze nach Namen, Städten oder einem anderen Feld in Ihrem Datenfluss sortieren.

Sortieren von Datensätzen mithilfe des Sorters

Im Sorter-Schritt können Sie Datensätze mithilfe von angegebenen Feldern sortieren.

1. Ziehen Sie in den Steuerungsschritten den Sorter auf die Arbeitsfläche, um ihn an einer beliebigen Stelle im Datenfluss zu platzieren.
2. Doppelklicken Sie auf den Sorter.
3. Klicken Sie auf **Hinzufügen**.
4. Klicken Sie auf den nach unten zeigenden Pfeil in der Spalte **Feldname**, und wählen Sie das Feld aus, nach dem Sie sortieren möchten.

   Anmerkung: Die Liste der verfügbaren Felder basiert auf den Feldern, die in den vorherigen Schritten im Datenfluss verwendet wurden.

5. Wählen Sie in der Spalte **Sortierung** aus, ob Sie in aufsteigender oder absteigender Reihenfolge sortieren möchten.
6. Wählen Sie in der Spalte **Typ** den Datentyp des Feldes aus.

   Anmerkung: Wenn Ihre eingehenden Daten nicht im Zeichenfolgenformat sind, wird die Spalte **Typ** deaktiviert.

bigdecimal	Ein numerischer Datentyp, der eine Genauigkeit von 38 Dezimalstellen unterstützt. Verwenden Sie diesen Datentyp für Daten in mathematischen Berechnungen, die einen hohen Grad von Genauigkeit erfordern, speziell solche mit Finanzdaten. Der Datentyp „bigdecimal“ unterstützt genauere Berechnungen als der Datentyp „double“.
double	Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit doppelter Genauigkeit zwischen 2^{-1074} und $(2-2^{-52})\times2^{1023}$ enthält. In der E-Notation liegt der Wertebereich bei -1,79769313486232E+308 bis 1,79769313486232E+308.
float	Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit einzelner Genauigkeit zwischen 2^{-149} und $(2-2^{23})\times2^{127}$ enthält. In der E-Notation liegt der Wertebereich bei -3,402823E+38 bis 3,402823E+38.
integer	Ein numerischer Datentyp, der sowohl positive als auch negative ganze Zahlen zwischen -2^{31} (-2,147,483,648) und $2^{31}-1$ (2,147,483,647) enthält.
long	Ein numerischer Datentyp, der sowohl negative als auch positive ganze Zahlen zwischen -2^{63} (-9.223.372.036.854.775.808) und $2^{63}-1$ (9.223.372.036.854.775.807) enthält.
string	Eine Folge von Zeichen.

7. Aktivieren Sie das Kästchen in der Spalte **Kürzen**, um vor dem Sortieren Leerzeichen vor und nach dem Wert zu entfernen. Die Kürzungsoption ändert nicht den Wert des Feldes. Sie kürzt nur den Wert für das Sortieren. Beachten Sie, dass wenn Ihre eingehenden Daten nicht im Zeichenfolgenformat sind, die Spalte **Typ** deaktiviert wird.

8. Wählen Sie in der Spalte **NULL behandeln als** die Optionen **Größte** oder **Kleinste** aus, um die Platzierung der Nullwerte in der sortierten Liste anzugeben. Die Platzierung hängt von der Kombination der ausgewählten Optionen in den Feldern **Sortierung** und **NULL behandeln als** ab, wie in der unteren Tabelle dargestellt:

Sortierung	NULL behandeln als	Platzierung von Nullwerten in der sortieren Liste.
Aufsteigend	Größte	Ende der Liste
Aufsteigend	Kleinste	Anfang der Liste
Absteigend	Größte	Anfang der Liste
Absteigend	Kleinste	Ende der Liste

9. Wiederholen Sie dies, bis Sie alle Felder für die Sortierung hinzugefügt haben.
10. Mithilfe der Optionen **Nach oben** oder **Nach unten** können Sie die Sortierreihenfolge wie gewünscht ändern. Dadurch können Sie zunächst nach einem Feld und die resultierende Sortierung nach einem anderen Feld sortieren.
11. Wenn Sie die standardmäßigen Sortierleistungsoptionen überschreiben möchten, die von Ihrem Administrator definiert wurden, klicken Sie auf **Erweitert**, aktivieren Sie das Kästchen **Optionen für Sortierleistung außer Kraft setzen** und konfigurieren Sie die folgenden Optionen:

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:

```
(NumberOfRecords × 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 × MaxNumberOfTempFiles ÷ 2) >=
TotalNumberOfRecords
```

12. Klicken Sie auf **OK**.

Anmerkung: Sie können die Sortierkriterien wie gewünscht entfernen, indem Sie eine Zeile markieren und auf **Entfernen** klicken.

Splitter

Ein Splitter konvertiert hierarchische Daten in flache Daten. Splitter haben einen Eingabeport und einen Ausgabeport, der Daten vom Splitter zum nächsten Schritt weiterleitet. Sie können das Splitter-Feature z. B. auf eine Liste von Informationen anwenden und jedes diskrete Informationselement in seine eigene Datenzeile extrahieren. So kann Ihre Eingabe beispielsweise Orientierungspunkte innerhalb einer bestimmten Entfernung eines Längengrad-/Breitengradpunktes beinhalten. Und der Splitter könnte jeden Orientierungspunkt in eine separate Datenzeile setzen.

Verwenden des Splitter-Schrittes

1. Klicken Sie in den Steuerungsschritten auf **Splitter**, und ziehen Sie ihn auf die Arbeitsfläche. Platzieren Sie ihn an einer beliebigen Stelle im Datenfluss und verbinden Sie ihn mit Eingabe- und Ausgabeschritten.
2. Doppelklicken Sie auf den Splitter. Das Dialogfeld **Splitter-Optionen** wird angezeigt.

3. Klicken Sie auf die Dropdown-Liste **Unterteilen bei**, um andere Listentypen für diesen Schritt anzuzeigen. Klicken Sie auf den Listentyp, den der Splitter erstellen soll. Das Dialogfeld **Splitter-Optionen** passt sich entsprechend Ihrer Auswahl an und zeigt die für diesen Listentyp verfügbaren Felder an.

   Alternativ können Sie auf die Ellipsenschaltfläche (**...**) neben der Dropdown-Liste **Unterteilen bei** klicken. Daraufhin wird das Dialogfeld **Feldschema** angezeigt, das das Schema für die in den Splitter eingehenden Daten enthält. Die Listentypen werden fett dargestellt, gefolgt von den einzelnen Listen für jeden Typ. Außerdem wird das Format für diese Felder (string, double usw.) angezeigt. Klicken Sie auf den Listentyp, den der Splitter erstellen soll, und klicken Sie anschließend auf **OK**. Das Dialogfeld **Splitter-Optionen** passt sich entsprechend Ihrer Auswahl an und zeigt die für diesen Listentyp verfügbaren Felder an.

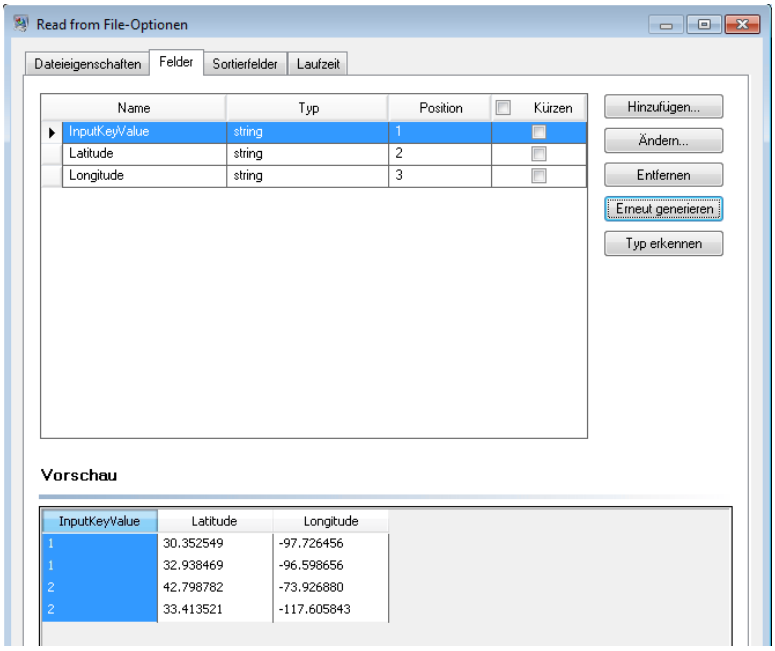
4. Durch Klicken auf **Kopfzeilendatensatz ausgeben** wird der ursprüngliche Datensatz mit der extrahierten Unterteilungsliste zurückgegeben.
5. Durch Klicken auf **Nur wenn Eingabeliste leer ist** wird nur der ursprüngliche Datensatz zurückgegeben, wenn keine Unterteilungsliste für diesen Datensatz vorhanden ist.
6. Wählen Sie aus, welche Felder der Splitter in die Ausgabe einbeziehen soll, indem Sie das Kästchen **Einschließen** für diese Felder aktivieren.
7. Klicken Sie auf **OK**.

Splitter-Beispiel

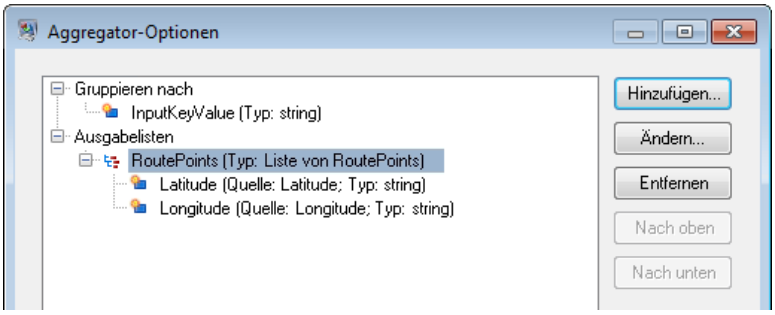
Das folgende Beispiel nutzt die Ausgabe aus einem Routenführungsschritt, der Wegbeschreibungen enthält, und setzt jede Wegbeschreibung (oder jedes Listenelement) in eine Datenzeile. Der Datenfluss sieht wie folgt aus:

Der Datenfluss führt das Feature folgendermaßen aus:

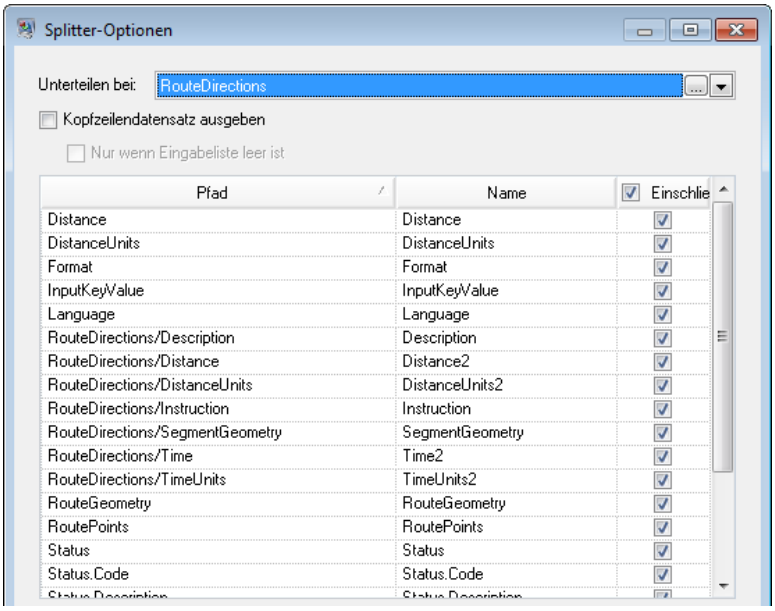
1. Der Schritt **Read from File** enthält Längengrade, Breitengrade und Eingabeschlüsselwerte, um Ihnen die Identifizierung der einzelnen Punkte zu erleichtern.



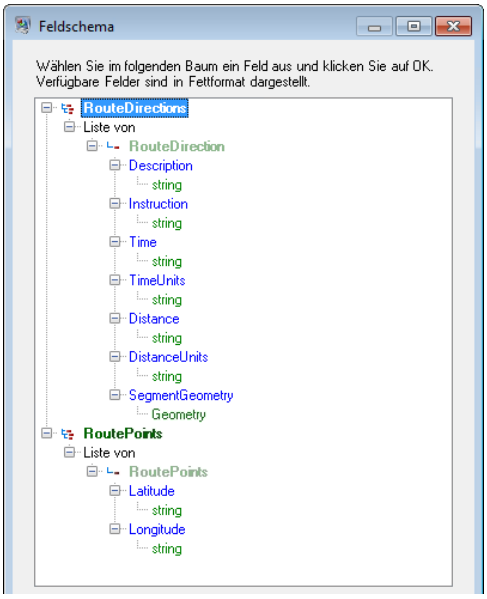
2. Der Aggregator-Schritt erstellt Daten aus dem „Read from File“-Schritt in einem Schema (eine strukturierte Datenhierarchie) und identifiziert die Gruppe von Längen- und Breitengraden als eine Liste von Routenpunkten. Dies ist ein erforderlicher Schritt, damit der nächste Schritt ordnungsgemäß funktioniert.



3. Der Schritt **Get Travel Directions** des Location Intelligence-Moduls erstellt Wegbeschreibungen von einem Standort zu einem anderen mithilfe der Routenpunkte aus Schritt 2.
4. Im **Splitter**-Schritt wird festgelegt, dass die Daten am Feld „Wegbeschreibung“ aufgeteilt werden sollen und dass die Ausgabelisten alle möglichen Felder aus dem „Get Travel Directions“-Schritt enthalten sollen.



Das Schema ist wie folgt strukturiert, wobei die Routenbeschreibungen und die Routenpunkte die verfügbaren Listentypen für diesen Auftrag sind:



5. Der **Write to File**-Schritt schreibt die Ausgabe in eine Datei.

Stream Combiner

Stream Combiner kombiniert zwei oder mehr Streams von Datensätzen aus mehreren Schritten. Stream Combiner besitzt einen oder mehrere Schritteingangsports. So können Sie beispielsweise

eine Gruppe von Datensätzen aus einem Schritt und eine andere Gruppe aus einem zweiten Schritt in einem einzelnen Stream zusammenführen.

Stream Combiner beinhaltet keine Einstellungen.

Transformer

Der „Transformer“-Schritt ändert Feldwerte und Formatierung. Sie können mehrere Transformationen für ein Feld ausführen, solange die Eingabe- und Ausgabefeldnamen identisch sind.

Allgemeine Transformationen

Feld erstellen Verwendet Werte aus vorhandenen Feldern und/oder konstante Werte, um Feldwerte entweder zu ersetzen oder ein neues Feld zu erstellen. Angenommen, Sie haben ein Feld mit Namen „City“ und möchten die Wortgruppe „City of“ zu den Werten im Feld „City“ hinzufügen. Sie würden eine Vorlage wie diese erstellen:

```
City of ${City}
```

Im Feld **An-Feld** würden Sie das Feld „City“ auswählen. Dies hat den Effekt, dass die vorhandenen Werte im Feld „City“ durch einen mithilfe der Vorlage erstellten Werte ersetzt werden. Beispiel: Wenn der Wert im Feld „City“ Chicago ist, wäre der neue Wert „City of Chicago“.

Manchen Zeichen muss ein umgekehrter Schrägstrich („\“) vorangestellt werden, um eine gültige Vorlage zu erzeugen. So muss etwa einem einzelnen Anführungszeichen wie folgt ein umgekehrter Schrägstrich vorangestellt werden: `\'`. Unter **groovy-lang.org/syntax.html** finden Sie eine Liste der Zeichen, die als Escape-Zeichen einen umgekehrten Schrägstrich erfordern.

Kopieren Kopiert den Wert aus einem Feld in ein anderes.

Benutzerdefiniert Ermöglicht Ihnen, mithilfe der Sprache Groovy eigene Transformationen zu definieren. Weitere Informationen finden Sie unter **Erstellen einer benutzerdefinierten Transformation** auf Seite 94.

Für Benutzer des Location Intelligence-Moduls können benutzerdefinierte Transformationen auf Geodaten-Datasets zugreifen. Siehe dazu den Abschnitt „Schritte“ im *Spectrum Spatial-Handbuch* auf **support.pb.com**.

Rename Ändert den Namen eines Feldes. Sie können eine Auswahl aus einer Liste von Feldnamen treffen, die bereits im Datenfluss vorhanden sind, oder Sie können den gewünschten Namen eingeben.

Status Ändert das Feld „Status“ entweder in einen Erfolgs- oder Fehlerwert. Beim Fehlerwert können auch eine optionale Beschreibung und ein Code festgelegt werden.

Transformationen formatieren

Groß-/Kleinbuchstaben	Ändert die Schreibweise in Groß- oder Kleinbuchstaben.
Maskieren	Wendet Zeichen auf ein Feld an oder entfernt sie daraus. Weitere Informationen finden Sie unter **Verwenden einer Maskentransformation** auf Seite 99.
Auffüllen	Fügt Zeichen links oder rechts vom Feldwert hinzu.

Zeichenfolgentransformationen

Leerzeichen minimieren	Entfernt Leerzeichen am Anfang und Ende des Feldes. Ersetzt auch eine beliebige Folge von Leerzeichen (wie mehrere zusammenhängende Leerzeichen) durch ein einzelnes Leerzeichen.
Teilzeichenfolge entfernen	Entfernt alle Vorkommen einer Zeichenfolge aus einem Feld. Sie könnten beispielsweise „CA“ aus dem Feld „StateProvince“ entfernen.
Teilzeichenfolge	Kopiert eine zusammenhängende Folge von Zeichen aus einem Feld in ein anderes.
Kürzen	Entfernt angegebene Zeichen von links, rechts oder beiden Seiten eines Feldes. Beachten Sie, dass bei dieser Transformation die Groß-/Kleinschreibung beachtet wird.
Abschneiden	Entfernt eine angegebene Anzahl von Zeichen von der linken oder rechten Seite eines Feldes.

Listentransformationen

Dieses Feature hilft Ihnen, vorgefertigte Transformationen zu erstellen, die auf Listen angewendet werden können, beispielsweise Eingabe von „Read from XML“.

Folgen Sie zum Definieren von Listentransformationen den angegebenen Schritten:

1. Wählen Sie einen Vorgang für die Listentransformationen aus. Eingabefelder erscheinen in einer Strukturansicht auf der rechten Seite.
2. Wählen Sie in der Struktur ein gültiges Feld aus, auf das der Vorgang angewendet werden soll. Die Eigenschaften des Vorgangs werden unter der Strukturansicht der Eingabefelder angezeigt.
3. Geben Sie die Vorgangseigenschaften ein, und klicken Sie auf „Hinzufügen“. Die Transformation wird zur Liste im übergeordneten Fenster, also im Fenster „Transformer-Optionen“, hinzugefügt.

Feld erstellen	Ermöglicht das Erstellen eines Felds unterhalb des vom Benutzer ausgewählten Listentypfelds. Wenn beispielsweise in einer Liste namens „football“ zwei Vereine vorhanden sind, „Knitters“ und „Lambs“, kann der Benutzer einen neuen Verein namens „Irons“ hinzufügen, sodass in der Liste drei Vereine vorhanden sind.
Sortieren	Führt Sortiervorgänge mit Werten durch, die im ausgewählten Feld vorhanden sind. In einer komplexen Liste muss der Benutzer das Schlüsselelement für die Sortierung angeben. Bei einer einfachen Liste wird die Sortierung anhand der in der Liste vorhandenen Elemente durchgeführt. Der Benutzer kann als Sortierreihenfolge entweder aufsteigend oder absteigend auswählen. Wenn die Liste im Beispiel „football“ drei Vereine enthält, muss der Benutzer unter „clubs“ das Feld „club name“ auswählen,

um die Vereine nach Namen zu sortieren. Die aktuell eingetragenen Vereine werden als „Irons“, „Knitters“ und „Lambs“ aufgelistet, wenn als Sortierreihenfolge aufsteigend festgelegt wurde, oder andersherum, wenn als Sortierreihenfolge absteigend festgelegt wurde. Wenn der Benutzer jetzt die Liste der Spieler sortiert haben möchte, muss er das Feld „player“ auswählen und eine Sortierreihenfolge für das Feld definieren.

Summe Führt eine Summierung aller Werten durch, die im ausgewählten Feld vorhanden sind. Die Ausgabe wird in einem vom Benutzer angegebenen Feld gespeichert. Wenn der Benutzer beispielsweise die Gesamtpunktzahl jedes Vereins anzeigen möchte, muss er unter „Tournament“ das Feld „points“ auswählen und den Ausgabefeldnamen angeben.

Kopieren Führt einen Kopiervorgang vom ausgewählten Feld zum durch den Benutzer angegebenen Feld durch. Wenn der Benutzer ein Feld zum Kopieren auswählt, werden dieses Feld und alle untergeordneten Felder (falls vorhanden) zum angegebenen neuen Feld kopiert. Dieser Vorgang findet auf derselben Hierarchieebene statt.

Rename Führt einen Umbenennungsvorgang mit dem ausgewählten Feld zum durch den Benutzer angegebenen neuen Namen durch.

Der folgende XML-Code bietet eine Referenz für das Feature „Listentransformation“:

```
<?xml version="1.0"?>
<sports_details>
  <sports name="football">
 <clubs>
   <club name="Knitters">
    <player>Samuel</player>
    <player>Messi</player>
    <player>kaka</player>
    <player>Alan</player>
    <coach>Stuart</coach>
    <Tournament name="Football League">
   <result>won</result>
   <points>4</points>
    </Tournament>
    <Tournament name="UEFA">
   <result>draw</result>
   <points>2</points>
    </Tournament>
   </club>
   <club name="Lambs">
    <player>Ronaldo</player>
    <player>Neymar</player>
    <player>Zlatan</player>
    <player>Mesut</player>
    <coach>Ivan</coach>
    <Tournament name="Airtel League">
   <result>draw</result>
```

```
   <points>2</points>
    </Tournament>
    <Tournament name="Champions League">
   <result>lost</result>
   <points>0</points>
    </Tournament>
   </club>
   <club name="Irons">
    <player>Scott</player>
    <player>Paul</player>
    <player>John</player>
    <player>Andrew</player>
    <coach>Jeff</coach>
    <Tournament name="CAF">
   <result>won</result>
   <points>4</points>
    </Tournament>
    <Tournament name="Copa America">
   <result>won</result>
   <points>4</points>
    </Tournament>
   </club>
</clubs>
 </sports>
 <sports name="badminton">
<clubs>
   <club name="Shuttlers">
    <player>Saina</player>
    <player>Viktor</player>
    <player>Chen</player>
    <player>Srikanth</player>
    <coach>Jan</coach>
    <Tournament name="Olympic Games">
   <result>won</result>
   <points>4</points>
    </Tournament>
    <Tournament name="Commonwealth Games">
   <result>won</result>
   <points>4</points>
    </Tournament>
   </club>
   <club name="Choppers">
    <player>Wang</player>
    <player>Sindhu</player>
    <player>Carolina</player>
    <player>Li Xuerui</player>
    <coach>Ratchanok</coach>
    <Tournament name="World Junior">
   <result>draw</result>
   <points>2</points>
    </Tournament>
    <Tournament name="Uber Cup">
   <result>draw</result>
```

```
   <points>2</points>
    </Tournament>
   </club>
   <club name="Lobbers">
    <player>Nozomi</player>
    <player>Chou</player>
    <player>Marc</player>
    <player>Lin</player>
    <coach>Kevin</coach>
    <Tournament name="World Senior">
   <result>won</result>
   <points>4</points>
    </Tournament>
    <Tournament name="Thomas Cup">
   <result>won</result>
   <points>4</points>
    </Tournament>
   </club>
 </clubs>
  </sports>
</sports_details>
```

Ändern der Sortierung von Transformationen

Wenn mehrere Transformationen für ein bestimmtes Ausgabefeld ausgeführt werden sollen, können Sie die Sortierung ihrer Ausführung definieren.

> **Anmerkung:** Wenn Sie ein einzelnes Feld zwei verschiedenen Ausgabefeldern zuordnen (z. B. „ValidateAddress.City“ sowohl „Output.City1“ als auch „Output.City2“) und jedem Feld Transformationen hinzufügen, muss die Transformation für das zweite Feld zuerst ausgeführt werden. Sie müssen die Ausführungsreihenfolge der Transformationen ändern, um die zweite Feldtransformation (Output.City2) zuerst auszuführen.

1. Doppelklicken Sie auf den „Transformer“-Schritt. Das Dialogfeld „Transformer-Optionen“ wird angezeigt.
2. Wählen Sie eine Transformation aus und verwenden Sie die Schaltflächen „Nach oben“ und „Nach unten“, um die Sortierung der Transformationen zu ändern. Die oberste Transformation wird zuerst ausgeführt.

   > **Anmerkung:** Abhängige Transformationen können nicht über primäre Transformationen verschoben werden (die Transformationen, auf die die abhängigen Transformationen angewiesen sind).

3. Klicken Sie auf **OK**.

Erstellen einer benutzerdefinierten Transformation

Der „Transformer“-Schritt enthält vordefinierte Transformationen, die eine Vielzahl von gängigen Datentransformationen ausführen. Wenn die vordefinierten Transformationen Ihre Anforderungen nicht erfüllen, können Sie mit Groovy ein benutzerdefiniertes Transformationsskript schreiben. Diese Schritte beschreiben, wie mit Groovy einfache benutzerdefinierte Transformationen erstellt werden. Eine vollständige Dokumentation zu Groovy finden Sie unter **groovy-lang.org**.

1. Fügen Sie dem Datenfluss im Enterprise Designer einen „Transformer“-Schritt hinzu.
2. Doppelklicken Sie auf den „Transformer“-Schritt.
3. Klicken Sie auf **Hinzufügen**.
4. Klicken Sie unter **Allgemein** auf **Benutzerdefiniert**.
5. Geben Sie in das Feld **Name der benutzerdefinierten Transformation** einen Namen für die zu erstellende Transformation ein. Der Name muss eindeutig sein.
6. Klicken Sie auf **Skript-Editor**.

   Dieser Editor bietet eine Vielzahl von Features, die Ihnen die Entwicklung Ihrer Transformation erleichtern, darunter Codevervollständigung und Paletten mit Funktionen und Feldern.

Aufgabe	Anleitungen
So fügen Sie eine Funktion hinzu:	Doppelklicken Sie im Bereich **Funktionen** auf die Funktion, die Sie hinzufügen möchten. **Anmerkung:** Die im Editor aufgelisteten Funktionen sind Funktionen, die bereitgestellt werden, um das Schreiben benutzerdefinierter Transformationsskripts einfacher zu machen. Sie führen Funktionen aus, deren Realisierung sonst mehrere Zeilen Groovy-Code erfordern würde. Es handelt sich dabei nicht um Groovy-Standardfunktionen.

So rufen Sie den Wert aus einem Datenflussfeld ab:	Doppelklicken Sie im Bereich **Eingabefelder** auf das gewünschte Eingabefeld. Folgendes wird Ihrem Skript hinzugefügt:
	
	`data['FieldName']`
	
	Wenn Sie beispielsweise den Wert aus dem Feld „CurrentBalance“ abrufen möchten, würde Folgendes hinzugefügt:
	
	`data['CurrentBalance']`

So legen Sie den Wert eines Datenflussfeldes fest:	Geben Sie im Skript-Editor den folgenden Code ein: `data['FieldName']=NewValue`

Aufgabe	Anleitungen
	Um beispielsweise das Feld „Day“ auf den Wochentag zu setzen, der im Feld „PurchaseDate“ enthalten ist: `data['Day']=dayOfWeek(data['PurchaseDate'])` In diesem Beispiel wird die Funktion `dayOfWeek()` verwendet, um den Tag aus dem Datumswert im Feld „PurchaseDate“ abzurufen, und das Ergebnis in das Feld „Day“ geschrieben. **Tipp:** Sie können im Bereich **Ausgabefelder** auf den Namen des Ausgabefeldes doppelklicken, um die Feldreferenz dem Skript hinzuzufügen.

So erstellen Sie ein neues Feld mit einem numerischen Datentyp:	Geben Sie im Skript-Editor den folgenden Code ein:
	`data['FieldName'] = new constructor;`
	Dabei kann es sich bei *Konstruktor* um eines der folgenden Elemente handeln:
	java.lang.Double(*Zahl*)
	Erstellt ein Feld mit dem Datentyp „Double“.
	java.lang.Float(*Zahl*)
	Erstellt ein Feld mit dem Datentyp „Float“.
	java.lang.Integer(*Zahl*)
	Erstellt ein Feld mit dem Datentyp „Integer“. Sie können ein neues Integer-Feld auch erstellen, indem Sie eine ganze Zahl angeben. Beispiel: Damit wird ein Integer-Feld mit dem Wert 23 erstellt:
	`data['MyNewField'] = 23;`
	java.lang.Long(*Zahl*)
	Erstellt ein Feld mit dem Datentyp „Long“.
	Um beispielsweise ein neues Feld mit Namen „Transactions“ mit dem Datentyp „Double“ und dem Wert 23,10 zu erstellen, würden Sie Folgendes angeben:
	`data['Transactions'] = new com.java.lang.Double(23.10);`

So erstellen Sie ein neues Feld mit dem Datentyp	Geben Sie im Skript-Editor den folgenden Code ein: `data['FieldName'] = new constructor;`

Aufgabe	Anleitungen

„Date“ oder „Time“:	Dabei kann es sich bei *Konstruktor* um eines der folgenden Elemente handeln:
	com.pb.spectrum.api.datetime.Date(*Jahr*,*Monat*,*Tag*)
	Erstellt ein Feld mit dem Datentyp „Date“. 23. Dezember 2013 wäre beispielsweise:
	`2013,12,23`
	com.pb.spectrum.api.datetime.Time(*Stunde*,*Minute*,*Sekunde*)
	Erstellt ein Feld mit dem Datentyp „Time“. 16:15 Uhr wäre beispielsweise:
	`16,15,0`
	.
	com.pb.spectrum.api.datetime.DateTime(*Jahr*,*Monat*,*Tag*,*Stunde*,*Minute*,*Sekunde*)
	Erstellt ein Feld mit dem Datentyp „DateTime“. 16:15 Uhr am 23. Dezember 2013 wäre beispielsweise:
	`2013,12,23,16,15,0`
	Um beispielsweise ein neues Feld mit Namen „TransactionDate“ mit dem Datentyp „Date“ und dem Wert 23. Dezember 2013 zu erstellen, würden Sie Folgendes angeben:
	`data['TransactionDate'] = new com.pb.spectrum.api.datetime.Date(2013,12,23);`

So erstellen Sie ein neues Feld mit einem Datentyp „boolean“:	Geben Sie im Skript-Editor den folgenden Code ein:
	`data['FieldName'] = true or false;`
	Beispiel: Um ein Feld mit Namen „IsValidated“ zu erstellen und es auf „false“ (falsch) zu setzen, würden Sie Folgendes angeben:
	`data['IsValidated'] = false;`

Aufgabe	Anleitungen
So erstellen Sie ein neues Druckausgabefeld:	Erstellen Sie mit der Methode `factory.create()` neue Felder in einem Datensatz. Fügen Sie dann mithilfe des „leftShift“-Operators `<<` den neuen Datensatz an das Listenfeld an.

```
NewListField = []

NewRecord = factory.create()
NewRecord['NewField1'] = "Value"
NewRecord['NewField12'] = "Value"
...
NewListField << NewRecord

NewRecord = factory.create()
NewRecord['NewField1'] = "Value"
NewRecord['NewField12'] = "Value"
...
NewListField << NewRecord
data['ListOfRecords'] = NewListField
```

Beispiel: Hiermit wird ein neues Listenfeld mit Namen „addresses“ erstellt, das aus zwei „address“-Datensätzen besteht.

```
addresses = []
address = factory.create()
address['AddressLine1'] = "123 Main St"
address['PostalCode'] = "12345"
addresses << address

address = factory.create()
address['AddressLine1'] = "PO Box 350"
address['PostalCode'] = "02134"
addresses << address
data['Addresses'] = addresses
```

Sie können auch ein neues Listenfeld erstellen, das eine Liste von einzelnen Feldern statt einer Liste von Datensätzen enthält. Beispiel: Hiermit wird ein neues Listenfeld mit Namen „PhoneNumbers“ erstellt, das private und geschäftliche Telefonnummern enthält:

```
phoneNumbers = []
phoneNumbers << data['HomePhone']
phoneNumbers << data['WorkPhone']
data['PhoneNumbers'] = phoneNumbers
```

Aufgabe	Anleitungen

So verketten Sie Felder:	Verwenden Sie das Symbol `+`. So wird beispielsweise mit Folgendem das Feld „FirstName“ mit dem Feld „LastName“ zu einem Wert verkettet und dieser im Feld „FullName“ gespeichert.
	`String fullname = data['FirstName'] + ' ' + data['LastName'];`
	`data['FullName']=fullname;`
	Im folgenden Beispiel gibt es zwei Eingabefelder („AddressLine1“ und „AddressLine2“), die verkettet und in das Ausgabefeld „Address“ geschrieben werden.
	`address1 = data['AddressLine1'];`
	`address2 = data['AddressLine2'];`
	`data['Address']=address1+ ',' + address2;`

So parsen Sie ein Feld:	Identifizieren Sie ein Trennzeichen und verwenden Sie dann `substring`, um das Feld zu parsen. Im folgenden Beispiel gilt: Wenn das Feld „PostalCode“ größer ist als fünf Zeichen, werden die fünfstellige Postleitzahl und der +4-Teil getrennt und in separate Felder im Ausgabesatz geschrieben.
	`if (data['PostalCode'].length() > 5)`
	`{`
	`  String postalCode = data['PostalCode'];`
	`  int separatorPosition = postalCode.indexOf('-');`
	`  String zip = postalCode.substring(0, separatorPosition);`
	`  String plusFour = postalCode.substring(`
	`  separatorPosition + 1,`
	`  postalCode.length();`
	`  data['Zip']=zip;`
	`  data['PlusFour']=plusFour;`
	`}`

So führen Sie eine bedingte Verarbeitung aus:	Verwenden Sie eine `if`- oder `switch`-Anweisung. Dies sind die gängigsten Konstrukte für bedingte Verarbeitung. Weitere Informationen finden Sie unter **groovy-lang.org**.
	In diesem Beispiel wird das Feld „AddressCity“ auf die erste Adresszeile und den Ortsnamen gesetzt, wenn die Stadt Austin ist.
	`city = data['City'];`
	`address1 = data['AddressLine1']`
	`if(city.equals('Austin'))`
	`data['AddressCity']=address1 +',' + city;`

Aufgabe	Anleitungen
So führen Sie Schleifen aus:	Verwenden Sie die Schleife `for`. Dies sollte das einzige Konstrukt für Schleifen sein, das Sie benötigen. Weitere Informationen zu Schleifen bzw. zur Syntax finden Sie unter **groovy-lang.org**.

So vergrößern Sie Daten:	Definieren Sie eine Konstante und verwenden Sie das Verkettungszeichen `+`. Beispiel: Das folgende Skript fügt das Wort „Incorporated“ an das Feld „FirmName“ an.
	`firmname = data['FirmName'];`
	`constant = 'Incorporated';`
	`if(firmname.length() > 0)`
	`data['FirmName']=firmname + ' ' + constant;`

So greifen Sie auf eine zur Laufzeit angegebene Option zu:	Wenn für den Datenfluss Laufzeitoptionen aktiviert sind, können Sie mit folgender Syntax auf die dem Datenfluss zur Laufzeit übergebenen Einstellungen zugreifen:
	`options.get("optionName")`
	Wenn Sie etwa auf eine Option namens `casing` zugreifen wollten, würden Sie im benutzerdefinierten Transformationsskript Folgendes angeben:
	`options.get("casing")`

7. Nachdem Sie mit der Eingabe des Skripts fertig sind, klicken Sie auf die Schaltfläche „X“ im Fenster, um den Editor zu schließen.
8. Wählen Sie im Feld **Eingabefelder** das bzw. die Felder aus, auf das bzw. die Sie die Transformation anwenden möchten.
9. Geben Sie im Feld **Ausgabefelder** das Feld an, in das Sie die Ausgabe der Transformation schreiben möchten. Falls erforderlich, können Sie ein neues Feld definieren, indem Sie rechts neben dem Feld **Ausgabefelder** auf die Schaltfläche **Hinzufügen** klicken.
10. Wenn Sie fertig sind, klicken Sie unten im Fenster auf die Schaltfläche **Hinzufügen**.
11. Klicken Sie auf **OK**.

Verwenden einer Maskentransformation

Sie können den „Transformer“-Schritt verwenden, um eine Maskentransformation auf ein Feld anzuwenden. Eine Maskentransformation wendet Zeichen auf ein Feld an oder entfernt Zeichen aus einem Feld. Dazu wird ein angegebenes Muster verwendet. Beispiel: Mit einer

Maskentransformation könnten Sie eine Zeichenfolge von Zahlen wie `8003685806` wie folgt als Telefonnummer formatieren: `(800) 368 5806`.

1. Ziehen Sie im Enterprise Designer einen „Transformer“-Schritt auf die Arbeitsfläche und verbinden Sie ihn an der gewünschten Stelle.
2. Doppelklicken Sie auf den „Transformer“-Schritt.
3. Klicken Sie auf **Add** (Hinzufügen).
4. Erweitern Sie **Formatierung** und wählen Sie **Maskieren** aus.
5. Wählen Sie den gewünschten Maskentyp aus.

Übernehmen	Fügt Zeichen zu einem Feld hinzu, um die Zeichenfolge in einem neuen Muster zu bilden.
Entfernen	Extrahiert ein Muster von Zeichen aus einer Zeichenfolge.

6. Geben Sie im Feld **Maskenzeichenfolge** das Muster an, das Sie entweder beim Hinzufügen oder Entfernen von Zeichen verwenden möchten.

   Es gibt zwei Typen von Zeichen, die Sie verwenden, wenn Sie die Maskenzeichenfolge angeben: Literal- und Maskenzeichen.

   Literalzeichen sind Zeichen, die in der Zeichenfolge vorhanden sind. Wenn eine Maske „Entfernen“ verwendet wird, muss das Eingabezeichen exakt dem Literalzeichen entsprechen. Wenn das der Fall ist, wird es aus der Eingabe entfernt. Ebenso werden die Literalzeichen der Eingabe an der durch die Maskendefinition angegebenen Position hinzugefügt, wenn die Maske „Anwenden“ verwendet wird.

   Der andere Typ von Zeichen, den Sie in einer Maskenzeichenfolge verwenden können, ist ein Maskenzeichen. Ein Maskenzeichen gibt den Typ des Zeichens an, das sich an einer bestimmten Position der Eingabezeichenfolge befinden kann. Wenn Sie beispielsweise eine Eingabe haben, bei der das erste Zeichen eine Zahl ist, muss das erste Maskenzeichen `#` sein. Alles in der Eingabe, was diesem Maskenzeichen entspricht, wird in der Ausgabe beibehalten.

   In der folgenden Tabelle werden die Maskenzeichen aufgelistet, die Sie im Feld **Maskenzeichenfolge** verwenden können:

   Tabelle 8: Maskenzeichen

Zeichen	Definition
#	Beliebige Zahl.
'	Escape-Zeichen; für beliebige der speziellen Formatierungszeichen verwendet.

Zeichen	Definition
U	Beliebiges Zeichen. Alle Kleinbuchstaben werden Großbuchstaben zugeordnet.
L	Beliebiges Zeichen. Alle Großbuchstaben werden Kleinbuchstaben zugeordnet.
A	Beliebiges Zeichen oder beliebige Zahl.
?	Beliebiges Zeichen.
*	Alles.
H	Beliebiges Hexadezimalzeichen (0–9, a–f oder A–F).

7. Klicken Sie auf **Add** (Hinzufügen).
8. Klicken Sie auf **OK**.

Beispiele für Maskentransformationen

Dies ist eine Maske „Anwenden“, die eine Formatierung auf eine Zeichenfolge anwendet. Weil „(“ und „)“ und <Leerzeichen> Literale sind, werden sie der Ausgabe hinzugefügt. Alle Zahlen werden beibehalten, weil # ein Maskenzeichen ist.

Eingabe: `8003685806`
Maskenzeichenfolge: `(###) ### ####`
Ausgabe: `(800) 368 5806`

Das folgende Beispiel zeigt eine Maske „Entfernen“, die den Bindestrich in der Postleitzahl entfernt.

Eingabe: `60510-1135`
Maskenzeichenfolge: `*****-****`
Ausgabe: `605101135`

Unique ID Generator

Der „Unique ID Generator“-Schritt erstellt einen eindeutigen Schlüssel, der einen bestimmten Datensatz angibt. Eine eindeutige ID ist für Data Warehouse-Initiativen wichtig, in denen Transaktionen möglicherweise nicht alle Namens- und Adressdaten enthalten, aber demselben Datensatz/Kontakt zugeschrieben werden müssen. Eine eindeutige ID kann auf der Ebene Einzelperson, Haushalt, Unternehmen und/oder Haus implementiert werden. Unique ID Generator stellt eine Vielzahl von Algorithmen bereit, um eindeutige IDs zu erstellen.

Die eindeutige ID basiert entweder auf einer sequenziellen Zahl oder einem Datums-/Uhrzeitstempel. Außerdem können Sie optional eine Vielzahl von Algorithmen verwenden, um an die ID anzufügende Daten zu generieren, und dadurch die Wahrscheinlichkeit erhöhen, dass die ID eindeutig ist. Die IDs des Typs Sequenzielle Zahl oder Datums-/Uhrzeitstempel sind erforderlich und können nicht aus der generierten ID entfernt werden.

Mit Unique ID Generator kann mithilfe eines der Algorithmen zur Schlüsselgenerierung ein nicht eindeutiger Schlüssel generiert werden. Im nicht eindeutigen Modus können Sie Schlüssel erstellen, die zum Vergleichen verwendet werden sollen. Das kann in einem Data Warehouse nützlich sein, wo Sie bereits Schlüssel zu einer Dimension hinzugefügt haben und einen Schlüssel für neue Datensätze generieren möchten, um festzustellen, ob die neuen Datensätze mit einem vorhandenen Datensatz übereinstimmen.

Das folgende Beispiel zeigt, dass jedem Datensatz in der Eingabe eine sequenzielle Datensatz-ID in der Ausgabe zugewiesen wird.

Datensatz	RecordID
John Smith	0
Mary Smith	1
Jane Doe	2
John Doe	3

Der „Unique ID“-Schritt erzeugt ein Feld namens „RecordID“, das die eindeutige ID enthält. Sie können das Feld „RecordID“ nach Bedarf umbenennen.

Definieren einer eindeutigen ID

Standardmäßig erstellt der „Unique ID Generator“-Schritt eine sequenzielle ID, wobei der erste Datensatz die ID 0 hat, der zweite Datensatz die ID 1, der dritte Datensatz die ID 2 usw. Wenn Sie ändern möchten, wie die eindeutige ID generiert wird, gehen Sie wie folgt vor.

1. Klicken Sie im Schritt „Unique ID Generator“ auf der Registerkarte **Regeln** auf **Ändern**.
2. Wählen Sie die Methode, die Sie zum Generieren der eindeutigen ID verwenden möchten.

Optionen	Beschreibung
Sequenzielles numerisches Tag, beginnend bei	Weist jedem Datensatz einen inkrementellen numerischen Wert zu, der mit der angegebenen Zahl beginnt. Wenn Sie 0 angeben, hat der erste Datensatz die ID 0, der zweite Datensatz die ID 1 usw.

Optionen	Beschreibung
Sequenzielles numerisches Tag, beginnend beim Wert in einem Datenbankfeld	

Optionen	Beschreibung
	Weist jedem Datensatz einen inkrementellen numerischen Wert zu, der mit dem aus dem Datenbankfeld gelesenen Maximalwert beginnt. Diese Zahl wird dann um 1 erhöht und dem ersten Datensatz zugewiesen. Beispiel: Wenn die aus dem Datenbankfeld gelesene Zahl 30 ist, hat der erste Datensatz die ID 31, der zweite Datensatz die ID 32 usw.
Verbindung	Wählen Sie die gewünschte Datenbankverbindung aus. Die Auswahlmöglichkeiten variieren, je nachdem, welche Verbindungen im Verbindungs-Manager der Management Console definiert sind. Wenn Sie eine neue Datenbankverbindung herstellen oder eine vorhandene Verbindung ändern oder löschen müssen, klicken Sie auf **Verwalten**. Wenn Sie eine Datenbankverbindung hinzufügen oder ändern, füllen Sie diese Felder aus: **Verbindungsname** Geben Sie einen Namen für die Verbindung ein. Sie können den Namen frei wählen. **Datenbanktreiber** Wählen Sie den passenden Datenbanktyp aus. **Verbindungsoptionen** Geben Sie Host, Port, Instanz, Benutzername und Kennwort an, die für die Verbindung zur Datenbank verwendet werden sollen.
Tabellenansicht	Gibt die Tabelle oder Ansicht in der Datenbank an, die Sie abfragen möchten.
Datenbankfeld	Wählen Sie eine Spalte aus der Liste aus, um einen eindeutigen Schlüssel zu generieren. Die unterstützten Datentypen zum Generieren der eindeutigen ID sind: **long** Ein numerischer Datentyp, der sowohl negative als auch positive ganze Zahlen zwischen -2^{63} (-9.223.372.036.854.775.808) und $2^{63}-1$ (9.223.372.036.854.775.807) enthält. **integer** Ein numerischer Datentyp, der sowohl positive als auch negative ganze Zahlen zwischen -2^{31} (-2,147,483,648) und $2^{31}-1$ (2,147,483,647) enthält. **bigdecimal** Ein numerischer Datentyp, der eine Genauigkeit von 38 Dezimalstellen unterstützt. Verwenden Sie diesen Datentyp für Daten in mathematischen Berechnungen, die einen hohen Grad von Genauigkeit erfordern, speziell solche mit Finanzdaten. Der Datentyp „bigdecimal“ unterstützt genauere Berechnungen als der Datentyp „double“.

Optionen	Beschreibung	
	double	Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit doppelter Genauigkeit zwischen 2^{-1074} und $(2-2^{-52})\times2^{1023}$ enthält. In der E-Notation liegt der Wertebereich bei -1,79769313486232E+308 bis 1,79769313486232E+308.
	float	Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit einzelner Genauigkeit zwischen 2^{-149} und $(2-2^{23})\times2^{127}$ enthält. In der E-Notation liegt der Wertebereich bei -3,402823E+38 bis 3,402823E+38.
Datums-/Uhrzeitstempel	Erstellt einen eindeutigen Schlüssel basierend auf dem Datums- und Uhrzeitstempel statt der sequenziellen Nummerierung.	
UUID	Erstellt einen universell eindeutigen 32-stelligen Bezeichnerschlüssel für jeden Datensatz. Die Ziffern im Schlüssel werden in fünf Gruppen durch Bindestriche getrennt angezeigt, und zwar im Format 8-4-4-4-12, also insgesamt 36 Zeichen (32 alphanumerische Zeichen und vier Bindestriche). Beispiel: 123e4567-e89b-12d3-a456-432255330000	
Aus	Wählen Sie diese Option nur aus, wenn Sie mithilfe eines Algorithmus einen nicht eindeutigen Schlüssel generieren möchten.	

3. Klicken Sie auf **OK**.

Verwenden von Algorithmen zum Vergrößern einer eindeutigen ID

Der „Unique ID Generator“ generiert eine eindeutige ID für jeden Datensatz entweder durch sequenzielle Nummerierung jedes Datensatzes oder durch Generierung eines Datums-/Uhrzeitstempels für jeden Datensatz. Sie können optional Algorithmen verwenden, um zusätzliche Information an die eindeutige ID des Typs Sequenziell oder Datum/Uhrzeit anzufügen und dadurch eine komplexere eindeutige ID zu erstellen, die mit größerer Wahrscheinlichkeit wirklich eindeutig ist.

1. Klicken Sie im Schritt „Unique ID Generator“ auf **Hinzufügen**.
2. Wählen Sie im Feld **Algorithmus** den Algorithmus aus, den Sie zum Generieren zusätzlicher Informationen in der ID verwenden möchten. 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 Schreibwiese 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 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.

3. Wählen Sie im Feld **Feldname** das Feld, auf das Sie den Algorithmus anwenden möchten. Wenn Sie beispielsweise den Soundex-Algorithmus und ein Feld namens „City“ wählen, wird die ID generiert, indem der Soundex-Algorithmus auf die Daten im Feld „City“ angewendet wird.
4. Wenn Sie den Algorithmus „SubString“ auswählen, geben Sie den Teil des Feldes an, den Sie in der Teilzeichenfolge verwenden möchten:
   a) Geben Sie im Feld **Startposition** die Position im Feld an, an der die Teilzeichenfolge beginnen soll.
   b) Wählen Sie im Feld **Länge** die Anzahl von Zeichen ab der Startposition aus, die Sie in die Teilzeichenfolge einfügen möchten.

   Angenommen, Sie haben die folgenden Daten in einem Feld namens „LastName“:

   ```
   Augustine
   ```

   Wenn Sie 3 als Startposition und 6 als Endposition angeben, ergibt die Teilzeichenfolge:

   ```
   gustin
   ```

5. Aktivieren Sie das Kästchen **Sonderzeichen entfernen**, um vor der Anwendung des Algorithmus alle nicht numerischen und nicht alphabetischen Zeichen wie Bindestriche, Leerzeichen und sonstige Sonderzeichen aus dem Feld zu entfernen.
6. Für die Algorithmen „Consonant“ und „SubString“ können Sie die Daten im Feld vor der Anwendung des Algorithmus sortieren. Aktivieren Sie dazu das Kästchen **Eingabe sortieren**. Sie können dann wählen, ob entweder die Zeichen im Feld oder Begriffe im Feld in alphabetischer Reihenfolge sortiert werden sollen.
7. Klicken Sie auf **OK**, um Ihre Einstellungen zu speichern.
8. Wiederholen Sie diese Schritte nach Bedarf, wenn Sie zusätzliche Algorithmen hinzufügen möchten, um eine komplexere ID zu erzeugen.

Anmerkung: Die eindeutige Schlüsseldefinition wird immer in einer anderen Farbe angezeigt und kann nicht gelöscht werden.

Definieren einer nicht eindeutigen ID

Mit Unique ID Generator kann mithilfe eines der Algorithmen zur Schlüsselgenerierung ein nicht eindeutiger Schlüssel generiert werden. Im nicht eindeutigen Modus können Sie Schlüssel erstellen,

die zum Vergleichen verwendet werden sollen. Das kann in einem Data Warehouse nützlich sein, wo Sie bereits Schlüssel zu einer Dimension hinzugefügt haben und einen Schlüssel für neue Datensätze generieren möchten, um festzustellen, ob die neuen Datensätze mit einem vorhandenen Datensatz übereinstimmen.

1. Klicken Sie im Schritt „Unique ID Generator“ auf der Registerkarte **Regeln** auf **Ändern**.
2. Wählen Sie **Aus** aus.

   Das deaktiviert den Teil „Eindeutige ID“ der ID-Generierungsregeln. Ist diese Option deaktiviert, wird nur der in den folgenden Schritten gewählte Algorithmus verwendet, um die ID zu erstellen. Das bedeutet, dass alle Datensätze mit denselben Daten in den Feldern, die Sie zum Generieren der ID verwenden, dieselbe ID haben. Sie können dann die ID zum Vergleichen verwenden.

3. Klicken Sie auf **OK**.
4. Klicken Sie bei der Warnungseingabeaufforderung auf **Ja**.
5. Klicken Sie im Schritt „Unique ID Generator“ auf **Hinzufügen**.
6. Wählen Sie im Feld **Algorithmus** den Algorithmus aus, den Sie zum Generieren zusätzlicher Informationen in der ID verwenden möchten. 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 Schreibwiese 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 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.

7. Wählen Sie im Feld **Feldname** das Feld, auf das Sie den Algorithmus anwenden möchten. Wenn Sie beispielsweise den Soundex-Algorithmus und ein Feld namens „City“ wählen, wird die ID generiert, indem der Soundex-Algorithmus auf die Daten im Feld „City“ angewendet wird.
8. Wenn Sie den Algorithmus „SubString“ auswählen, geben Sie den Teil des Feldes an, den Sie in der Teilzeichenfolge verwenden möchten:
   a) Geben Sie im Feld **Startposition** die Position im Feld an, an der die Teilzeichenfolge beginnen soll.
   b) Wählen Sie im Feld **Länge** die Anzahl von Zeichen ab der Startposition aus, die Sie in die Teilzeichenfolge einfügen möchten.

   Angenommen, Sie haben die folgenden Daten in einem Feld namens „LastName“:

   ```
   Augustine
   ```

   Wenn Sie 3 als Startposition und 6 als Endposition angeben, ergibt die Teilzeichenfolge:

```
gustin
```

9. Aktivieren Sie das Kästchen **Sonderzeichen entfernen**, um vor der Anwendung des Algorithmus alle nicht numerischen und nicht alphabetischen Zeichen wie Bindestriche, Leerzeichen und sonstige Sonderzeichen aus dem Feld zu entfernen.
10. Für die Algorithmen „Consonant“ und „SubString“ können Sie die Daten im Feld vor der Anwendung des Algorithmus sortieren. Aktivieren Sie dazu das Kästchen **Eingabe sortieren**. Sie können dann wählen, ob entweder die Zeichen im Feld oder Begriffe im Feld in alphabetischer Reihenfolge sortiert werden sollen.
11. Klicken Sie auf **OK**, um Ihre Einstellungen zu speichern.
12. Wiederholen Sie diese Schritte nach Bedarf, wenn Sie zusätzliche Algorithmen hinzufügen möchten, um eine komplexere ID zu erzeugen.

Anmerkung: Die eindeutige Schlüsseldefinition wird immer in einer anderen Farbe angezeigt und kann nicht gelöscht werden.

Flussausgabe

Verwenden Sie zum Definieren der Ausgabe aus einem Datenfluss den Schritt zur Datenladung. Bei einer Datenladung handelt es sich um den letzten Schritt in einem Datenfluss. Hierdurch wird definiert, was mit der Ausgabe aus dem Datenfluss geschehen soll. Eine Datenladung kann auch andere Aktionen am Ende eines Datenflusses durchführen, z. B. einen Datenfluss ausführen.

Ausgabe aus einem Auftrag

Die Ausgabe aus einem Auftrag kann in eine Datei oder eine Datenbank geschrieben werden. Spectrum™ Technology Platform kann Daten in viele Dateiformate und Datenbanktypen schreiben. Die Datenladungstypen, in die Sie schreiben können, hängen von den lizenzierten Modulen ab. Weitere Informationen finden Sie in den Lösungshandbüchern für Ihre Module unter **support.pb.com/spectrum**.

Ausgabe aus einem Dienst

Ausgabedaten aus einem Dienst werden in einem Output-Schritt definiert. In diesem Schritt werden die Felder definiert, die der Dienst als Antwort auf eine Webservice-Anforderung oder einen API-Aufruf zurückgibt.

Definieren der Dienstausgabe

Der Output-Schritt definiert die Ausgabefelder, die vom Service oder Unterfluss zurückgegeben werden. Führen Sie die folgenden Schritte aus, um die Serviceausgabe zu definieren.

1. Doppelklicken Sie auf das Symbol „Ausgabe“ auf der Arbeitsfläche. Das Dialogfeld **Ausgabeoptionen** wird angezeigt. Wenn Sie das Dialogfeld „**Ausgabeoptionen**“ zum ersten Mal öffnen, wird eine Liste mit unter „Input“ definierten Feldern angezeigt.
2. Möchten Sie ein neues Feld zur Feldliste hinzufügen, klicken Sie auf **Hinzufügen**. Das Dialogfeld **Benutzerdefiniertes Feld hinzufügen** wird angezeigt. Sie können ein benutzerdefiniertes Feld auch ändern oder löschen.
3. Klicken Sie erneut auf **Hinzufügen**.
4. Geben Sie den Feldnamen in das Textfeld ein.
5. Wählen Sie den **Datentyp** aus und klicken Sie auf **OK**. Die folgenden Datentypen werden unterstützt:

bigdecimal	Ein numerischer Datentyp, der eine Genauigkeit von 38 Dezimalstellen unterstützt. Verwenden Sie diesen Datentyp für Daten in mathematischen Berechnungen, die einen hohen Grad von Genauigkeit erfordern, speziell solche mit Finanzdaten. Der Datentyp „bigdecimal“ unterstützt genauere Berechnungen als der Datentyp „double“.
boolean	Ein logischer Typ mit zwei Werten: wahr und falsch.
bytearray	Ein Array (Liste) von Bytes. **Anmerkung:** Bytearray wird nicht als Eingabe für einen REST-Dienst unterstützt.
date	Ein Datentyp, der einen Monat, einen Tag und ein Jahr enthält. Beispiel: 2012-01-30 oder 30. Januar 2012. Sie können ein standardmäßiges Datumsformat in der Management Console angeben.
datetime	Ein Datentyp, der Tag, Monat, Jahr und Stunden, Minuten und Sekunden enthält. Beispiel: 30.01.2012 18:15:00.
double	Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit doppelter Genauigkeit zwischen 2^{-1074} und $(2-2^{-52})\times2^{1023}$ enthält. In der E-Notation liegt der Wertebereich bei -1,79769313486232E+308 bis 1,79769313486232E+308.
float	Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit einzelner Genauigkeit zwischen 2^{-149} und $(2-2^{23})\times2^{127}$ enthält. In der E-Notation liegt der Wertebereich bei -3,402823E+38 bis 3,402823E+38.
integer	Ein numerischer Datentyp, der sowohl positive als auch negative ganze Zahlen zwischen -2^{31} (-2,147,483,648) und $2^{31}-1$ (2,147,483,647) enthält.

list Streng genommen ist eine Liste kein Datentyp. Wenn jedoch ein Feld hierarchische Daten enthält, wird es als „list“-Feld behandelt. In Spectrum™ Technology Platform ist eine Liste eine Sammlung von Daten und besteht aus mehreren Werten. So kann etwa ein Feld „Name“ eine Liste von Namenswerten enthalten. Das kann in einer XML-Struktur wie folgt aussehen:

```
<Names>
    <Name>John Smith</Name>
    <Name>Ann Fowler</Name>
</Names>
```

Beachten Sie, dass sich der Spectrum™ Technology Platform-Datentyp „list“ vom Datentyp „list“ des XML-Schemas unterscheidet. Der XML-Datentyp „list“ ist ein einfacher Datentyp, der aus mehreren Werten besteht, während der Spectrum™ Technology Platform-Datentyp „list“ einem komplexen XML-Datentyp ähnelt.

long Ein numerischer Datentyp, der sowohl negative als auch positive ganze Zahlen zwischen -2^{63} (-9.223.372.036.854.775.808) und $2^{63}-1$ (9.223.372.036.854.775.807) enthält.

string Eine Folge von Zeichen.

time Ein Datentyp, der die Tageszeit enthält. Beispiel: 21:15:59 oder 9:15:59 PM.

Sie können bei Bedarf auch einen neuen, benutzerdefinierten Datentyp hinzufügen, und dieser neue Typ kann eine Liste eines beliebigen definierten Datentyps sein. Sie könnten beispielsweise eine Liste von Namen (Zeichenfolge) oder einen neuen Datentyp von Adressen mit „AddressLine1“ (Zeichenfolge), „City“ (Zeichenfolge), „StateProvince“ (Zeichenfolge) und „PostalCode“ (Zeichenfolge) definieren. Nachdem Sie das Feld erstellt haben, können Sie den Datentyp anzeigen, indem Sie auf das Dialogfeld „Eingabeoptionen“ zugreifen und auf die Schaltfläche in der Spalte „Datentyp“ klicken. Das Dialogfeld **Datentypdetails** wird angezeigt. Es enthält die Struktur des Feldes.

6. Klicken Sie erneut auf **OK**.
7. Klicken Sie auf das Kästchen neben **Verfügbar machen**, um das Kästchen für alle Felder in der Feldliste zu aktivieren. Durch die Auswahl eines Feldes in der Feldliste wird dieses im Datenfluss für Schrittvorgänge verfügbar gemacht. Klicken Sie erneut auf das Kästchen, um das Kästchen für alle Felder zu deaktivieren. Durch Deaktivieren des Kästchens bei mindestens einem Feld in der Feldliste und Klicken auf **OK** wird das Feld aus der Feldliste gelöscht.

   Anmerkung: Wenn Sie hierarchische Daten in den Eingabefeldern definieren, können Sie keine Daten importieren oder vertikal anzeigen.

8. Klicken Sie auf **OK**, um zur Arbeitsfläche zurückzukehren.

DeDefinieren eines Webservice-Datentyps

Im Feld **Datentypname** können Sie die Schnittstellen von WSDL (SOAP) und WADL (REST) für den Dienst, den Sie erstellen, steuern. Der Name des Elements „Spalten“ wird durch den Namen festgelegt, den Sie diesem Schritt im Dienst zuweisen. Der Name des Elements „Spalte“ wird durch den Text festgelegt, den Sie hier eingeben.

> **Anmerkung:** Bei WSDL betrifft dies sowohl Anforderungen als auch Antworten, bei WADL sind jedoch nur Antworten betroffen.

Bevor Sie diesen Schritt benennen und Text in dieses Feld eingeben, könnte Ihr Code folgendermaßen aussehen:

```
<Rows>
     <Row>
          <FirstName>John</FirstName>
          <LastName>Doe</LastName>
     </Row>
     <Row>
          <FirstName>Jane</FirstName>
    <LastName>Doe></LastName>
     </Row>
</Rows>
```

Nachdem Sie diesen Schritt benannt und Text in dieses Feld eingeben haben, könnte Ihr Code folgendermaßen aussehen:

```
<Names>
     <Name>
          <FirstName>John</FirstName>
          <LastName>Doe</LastName>
     </Name>
     <Name>
          <FirstName>Jane</FirstName>
          <LastName>Doe></LastName>
     </Name>
</Names>
```

Ausführen eines externen Programms

Ein „Execute Program“-Schritt ruft eine ausführbare Datei auf, z. B. ein Programm oder einen Befehl mit Befehlszeile, wenn ein Datensatz eingeht. So verwenden Sie einen „Execute Program“-Schritt in Ihrem Datenfluss:

Optionen

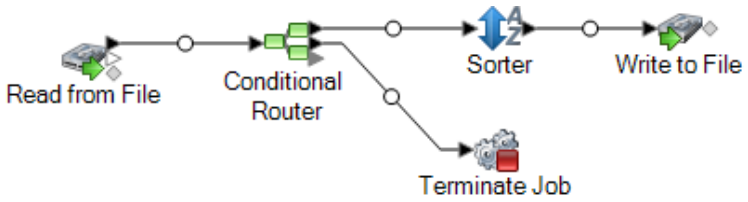
Option	Beschreibung
Befehlszeile	Der ausführbare Name und die ausführbaren Argumente (falls zutreffend). Bei den Argumenten kann es sich um im Datenfluss verfügbare Daten handeln. Klicken Sie für den Zugriff auf diese Daten auf die Schaltfläche „[...]“ (Durchsuchen). Sie können aus den folgenden drei Kontexten auswählen: ID des aktuellen Auftrags, Name des aktuellen Auftrags oder Name des aktuellen Benutzers. Sie können auch aus den verfügbaren Feldern auswählen. Beispiel: Auftragsstatus und Auftragskommentar.
Timeout	Gibt an, ob die Ausführung abgebrochen werden soll, wenn der Befehl innerhalb eines bestimmten Zeitraums keine Antwort sendet. Zur Auswahl stehen: **Kein Timeout** – Bricht die Ausführung nicht ab, wenn der Befehl nicht antwortet. **Timeout in Millisekunden** – Bricht den Ausführungsversuch ab, wenn der Befehl innerhalb der angegebenen Anzahl an Millisekunden keine Antwort zurückgibt.
Umgebungsvariablen	Optional: Gibt die Umgebungsvariablen an, die beim Ausführen des Befehls verwendet werden sollen. Klicken Sie auf **Hinzufügen**, um eine Umgebungsvariable hinzuzufügen. Geben Sie das entsprechende Schlüsselwort in das Feld **Schlüssel** ein. Ein Beispiel könnte „JAVA_HOME“ sein. Geben Sie den entsprechenden Wert in das Feld **Wert** ein. Ein Beispiel könnte `C:\Java\jre7` sein. Alternativ können Sie ein Feld aus dem Dialogfeld „Feldliste“ auswählen, indem Sie auf die Schaltfläche „[...]“ (Durchsuchen) klicken. Sie können aus den folgenden drei Kontexten auswählen: ID des aktuellen Auftrags, Name des aktuellen Auftrags oder Name des aktuellen Benutzers. Sie können auch aus den verfügbaren Feldern auswählen. Beispiel: Auftragsstatus und Auftragskommentar.

Beenden eines Auftrags basierend auf einer Bedingung

Der „Auftrag beenden“-Schritt wird in Verbindung mit „Conditional Router“ verwendet, um einen Auftrag zu beenden, wenn bestimmte Kriterien innerhalb eines Datensatzes gefunden werden. Wenn ein Datensatz an „Auftrag beenden“ gesendet wird, endet der Auftrag.

Anmerkung: „Auftrag beenden“ ist nicht in Diensten oder Unterflüssen verfügbar.

Um „Auftrag beenden“ zu verwenden, fügen Sie einen „Conditional Router“- und einen „Auftrag beenden“-Schritt zu Ihrem Datenfluss hinzu. Verbinden Sie dann die Schritte, und konfigurieren Sie „Conditional Router“. „Conditional Router“ sollte so konfiguriert werden, dass der Schritt die Kriterien enthält, die die Auftragsbeendigung auslösen sollen. Wenn ein Datensatz gefunden wird, der die Kriterien erfüllt, wird er an „Auftrag beenden“ übergeben und der Auftrag beendet. Dabei wird eine Meldung erzeugt, die Folgendes besagt: „Auftrag beendet durch Schritt: <Schrittbeschriftung>“.



Löschen von Datensätzen

Der „Write to Null“-Schritt löscht Datensätze. Die Datensätze werden gezählt, aber gelöscht. Verwenden Sie diesen Schritt, wenn es Datensätze gibt, die Sie nach Abschluss des Datenflusses nicht beibehalten möchten.

Eingebettete Datenflüsse

Ein eingebetteter Datenfluss verringert die Anzahl der Schritte, die auf der Arbeitsfläche des Enterprise Designer gleichzeitig angezeigt werden, indem Schritte gruppiert werden. Die gruppierten Schritte erscheinen dann als einzelner Schritt. Sie können eingebettete Datenflüsse wie folgt verwenden:

- Vereinfachen Sie das Layout komplexer Datenflüsse, indem Sie Schritte gruppieren, sodass sie auf der Arbeitsfläche als ein Schritt angezeigt werden.
- Verarbeiten Sie Datensätze in Gruppen, indem Sie das Iterationsfeature verwenden.
- Verwenden Sie einen Wert in einem Feld, um Schrittoptionen im eingebetteten Datenfluss festzulegen.

Sie können eine unbegrenzte Anzahl an eingebetteten Datenflüssen zu einem Datenfluss hinzufügen und eingebettete Datenflüsse innerhalb von eingebetteten Datenflüssen platzieren.

Unterschiede zwischen eingebetteten Datenflüssen und Unterflüssen

Es gibt zwei wesentliche Unterschiede zwischen einem eingebetteten Datenfluss und einem Unterfluss. Erstens steht die Iterationsverarbeitung nur in eingebetteten Datenflüssen zur Verfügung. Über Iterationen können Sie Datensatzgruppen zusammen verarbeiten, um beispielsweise Datensätze zum Verarbeiten oder zum Festlegen von auf Feldwerten basierten Schrittoptionen zu aggregieren.

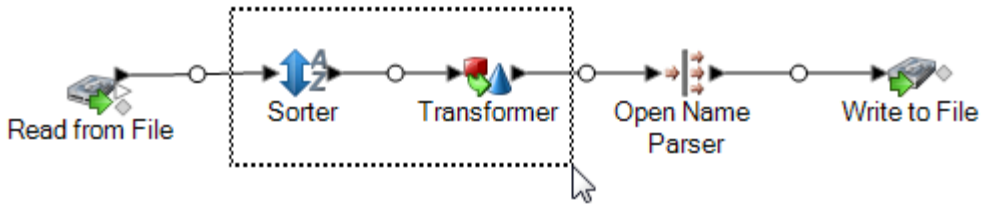
Der zweite Unterschied zwischen einem eingebetteten Datenfluss und einem Unterfluss besteht darin, dass ein eingebetteter Datenfluss nicht in mehr als einem Datenfluss verwendet werden kann. Wenn Sie einen Teil eines Datenflusses in mehreren Datenflüssen wiederverwenden möchten, erstellen Sie einen Unterfluss anstatt eines eingebetteten Datenflusses. Eingebettete Datenflüsse können in Unterflüsse konvertiert werden, wenn Sie sich entscheiden, einen eingebetteten Datenfluss in anderen Datenflüssen zu verwenden. Wenn Sie allerdings einen eingebetteten Datenfluss in einen Unterfluss konvertieren, werden seine Iterationsoptionen entfernt.

Gruppieren von Schritten in einem eingebetteten Datenfluss

Ein eingebetteter Datenfluss gruppiert Schritte in einem einzelnen Schritt, was Ihnen ermöglicht, das Layout komplexer Datenflüsse zu vereinfachen und mit Feldwerten Verarbeitungsoptionen festzulegen.

1. Fügen Sie in Ihrem Datenfluss die Schritte hinzu, die Sie in einen eingebetteten Datenfluss konvertieren möchten.
2. Wählen Sie die Schritte aus, die Sie in einen eingebetteten Datenfluss konvertieren möchten, indem Sie klicken und ein Kästchen um die Schritte ziehen.

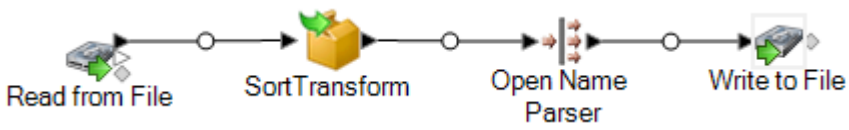
   Beispiel: Dies würde die Schritte „Sorter“ und „Transformer“ auswählen, um sie in einen eingebetteten Datenfluss zu konvertieren.



   Anmerkung: Berichtsschritte dürfen nicht in einen eingebetteten Datenfluss eingefügt werden.

3. Klicken Sie mit der rechten Maustaste auf die ausgewählten Schritte, und wählen Sie **In eingebettetem Datenfluss gruppieren** aus.
4. Geben Sie einen Namen für den eingebetteten Datenfluss ein. Der Name wird als Beschriftung für den eingebetteten Datenflussschritt auf der Arbeitsfläche verwendet.
5. Klicken Sie auf **OK**.

Die ausgewählten Schritte werden jetzt in einem eingebetteten Datenfluss gruppiert. Das folgende Beispiel zeigt einen eingebetteten Datenfluss namens „SortTansform“.

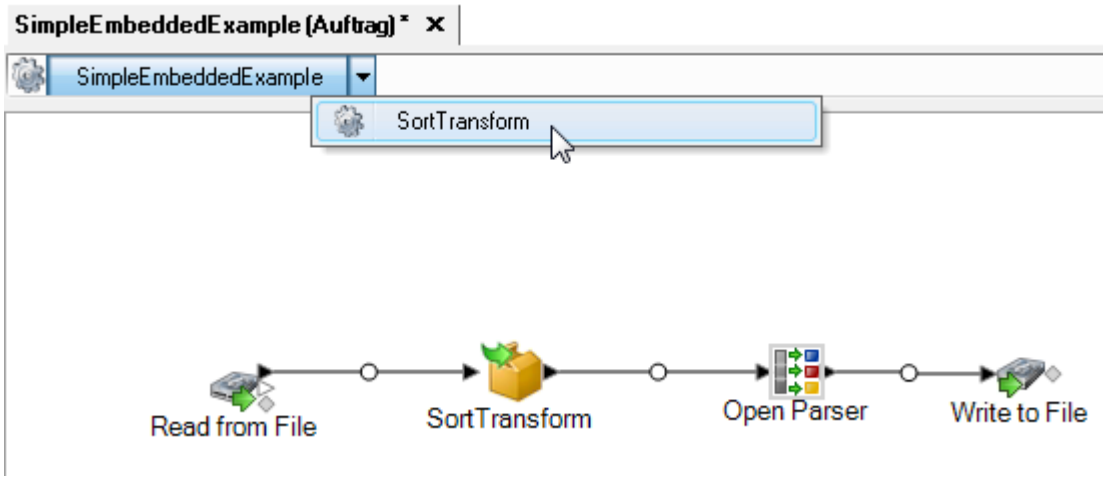


Bearbeiten eines eingebetteten Datenflusses

Ein eingebetteter Datenfluss gruppiert Schritte in einem einzelnen Schritt, was Ihnen ermöglicht, das Layout komplexer Datenflüsse zu vereinfachen und mit Feldwerten Verarbeitungsoptionen festzulegen.

1. Klicken Sie mit der rechten Maustaste auf das Symbol, und wählen Sie **Diesen eingebetteten Datenfluss bearbeiten** aus.

   Tipp: Sie können alternativ auch die Breadcrumb-Links oberhalb des Datenflusses verwenden, um eingebettete Datenflüsse zu öffnen. Im Folgenden wird beispielsweise ausgeführt, wie Sie einen eingebetteten Datenfluss mit dem Namen „SortTransform“ öffnen:



   Wenn Sie einen eingebetteten Datenfluss öffnen, wird ein Input- und Output-Schritt angezeigt. Diese Schritte stellen die Eingabe in den eingebetteten Datenfluss aus dem übergeordneten Datenfluss und die Ausgabe in den übergeordneten Datenfluss aus dem eingebetteten Datenfluss dar.

2. Ändern Sie den eingebetteten Datenfluss nach Bedarf.

   Anmerkung: Eine Datenflussüberprüfung ist in eingebetteten Datenflüssen nicht verfügbar.

Alle Änderungen, die Sie an einem eingebetteten Datenfluss vornehmen, werden beim nächsten Speichern des übergeordneten Datenflusses gespeichert.

Verwenden von Iterationen mit einem eingebetteten Datenfluss

Iterationseinstellungen geben an, wie ein eingebetteter Datenfluss eingehende Datensätze verarbeiten soll. Standardmäßig verarbeiten eingebettete Datenflüsse genau wie jeder andere Schritt im

Datenfluss jeden Datensatz einzeln. Wenn Sie jedoch Iterationen verwenden, können Sie Datensatzgruppen zusammen verarbeiten. Dies kann für Vorgänge wie Vergleiche oder Berechnungen basierend auf Datensatzgruppen statt auf dem gesamten Satz an Eingabedaten nützlich sein. Sie können auch Iterationen verwenden, um Schrittoptionen basierend auf den Daten in jedem Datensatz festzulegen.

Es gibt zwei Arten von Iterationen: nach Datensatz oder nach Gruppe. Bei Iterationen nach Datensatz verarbeitet ein eingebetteter Datenfluss jeweils einen Datensatz gleichzeitig und das Ergebnis wird an den nächsten, auf den eingebetteten Datenfluss folgenden Schritt gesendet. Die Iteration nach Datensatz ist nützlich, wenn Sie Schrittoptionen mithilfe von Feldwerten von Datensatz zu Datensatz festlegen möchten.

Bei Iterationen nach Gruppe werden Datensätze nach einem Schlüsselfeld gruppiert und der eingebettete Datenfluss verarbeitet jede Gruppe. Alle Datensätze in einer Gruppe werden in einer Iteration verarbeitet, anschließend wird die Gruppe in den nächsten, auf den eingebetteten Unterfluss folgenden Schritt geschrieben. Verwenden Sie Iterationen nach Gruppe, um Gruppen von verbundenen Datensätzen zu verarbeiten und um Schrittoptionen festzulegen, die bei der Verarbeitung der Gruppe von Datensätzen verwendet werden sollen. Sie möchten Datensätze beispielsweise nach Kunden-ID gruppieren, um eine Analyse des Datensatzes jedes Kunden durchzuführen, damit Sie ermitteln können, welches Geschäft von welchem Kunden am häufigsten besucht wird.

Sie sollten die Auswirkungen auf die Leistung berücksichtigen, wenn Sie Iterationen verwenden. Beim Start jeder neuen Iteration entsteht ein Mehraufwand während der Initialisierung des eingebetteten Datenflusses. Dieser Mehraufwand kann bedeutsame Auswirkungen haben, insbesondere bei eingebetteten Datenflüssen innerhalb anderer eingebetteter Datenflüsse. Wenn beispielsweise ein eingebetteter Datenfluss 1.000 Mal durchlaufen wird und innerhalb eines anderen Datenflusses enthalten ist, der ebenfalls 1.000 Male durchlaufen wird, würde die Gesamtsumme der Iterationen 1.000.000 betragen. Die Iterationen nach Datensatz haben größere Auswirkungen auf die Leistung, da jeder Datensatz eine neue Iteration auslöst.

1. Erstellen Sie einen eingebetteten Datenfluss mit dem Schritt oder den Schritten, den/die Sie für Iterationen verwenden.

   Anmerkung: Es gibt einige Einschränkungen im Hinblick auf Elemente, die in eingebetteten Datenflüssen mit aktivierten Iterationen eingeschlossen werden können:

   - Der „Stream Combiner“-Schritt darf nicht der erste Schritt in einem eingebetteten Datenfluss sein, bei dem Iterationen aktiviert sind.
   - Der eingebettete Datenfluss darf keine Datenladung enthalten, die in eine Datei auf dem Client schreibt. Datenladungen innerhalb eines eingebetteten Datenflusses müssen in eine Datei auf dem Spectrum™ Technology Platform-Server oder auf einem Dateiserver geschrieben werden.

2. Klicken Sie doppelt auf das Symbol des eingebetteten Datenflusses.
3. Aktivieren Sie das Kästchen **Iteration aktivieren**.

4. Wenn mehr als ein Eingangskanal mit dem eingebetteten Datenfluss verbunden ist, müssen Sie über das Feld **Port** den Port auswählen, dessen Datensätze Sie zum Durchlaufen der Iterationen verwenden möchten.

   Sie haben beispielsweise die beiden Eingangsports A und B und möchten immer Iterationen durchführen, wenn sich ein Schlüsselfeld ändert. Wenn Sie Port B für die Iterationen auswählen, wird der eingebettete Datenfluss jedes Mal eine neue Iteration starten, wenn ein Schlüsselfeld in den Datensätzen aus Port B geändert wird. Alle Datensätze aus dem anderen Port (Port A) werden in den eingebetteten Datenfluss gelesen, zwischengespeichert und für jede Iteration verwendet.

5. Wählen Sie den Iterationstyp aus, den Sie ausführen möchten.

Immer durchlaufen, wenn sich ein Schlüsselfeld ändert	Bei diesem Iterationstyp verarbeitet der eingebettete Datenfluss Gruppen von Datensätzen mit demselben Wert in einem oder mehreren Feldern. Wenn der eingebettete Datenfluss die Verarbeitung der Gruppe von Datensätzen fertigstellt, wird der eingebettete Datenfluss zurückgesetzt und eine neue Gruppe von Datensätzen wird verarbeitet. Verwenden Sie diesen Iterationstyp, um eingebettete Datenflüsse zu erstellen, die Gruppen von Datensätzen verarbeiten und anschließend jede Datensatzgruppe einzeln ausgeben. **Tipp:** Wenn Sie diesen Iterationstypen auswählen, können Sie die Leistung verbessern, indem Sie vor dem eingebetteten Datenfluss einen Sorter-Schritt platzieren und die Datensätze nach Schlüsselfeld sortieren.
Pro Datensatz durchlaufen	Bei dieser Art Iteration verarbeitet der Datenfluss jeweils einen Datensatz. Nach jedem Abschluss der Verarbeitung eines eingebetteten Datenflusses für einen Datensatz wird das Ergebnis an die Ausgabe gesendet und ein neuer Datensatz wird verarbeitet. Eingebettete Datenflüsse, die pro Datensatz durchlaufen, verarbeiten jeden Datensatz wie eine neue Datenflussausführung.

6. Sollten Sie **Immer durchlaufen, wenn sich ein Schlüsselfeld ändert** auswählen, müssen Sie **Beim Vergleichen von Werten Groß-/Kleinschreibung ignorieren** aktivieren, wenn Sie bei der Auswertung von Schlüsselfeldwerten zur Bestimmung von Datensatzgruppen Unterschiede in der Schreibweise ignorieren möchten.
7. Geben Sie ein oder mehrere Schlüsselfelder an.
   a) Klicken Sie auf **Hinzufügen**.
   b) Wählen Sie das Feld aus, das Sie als Schlüsselfeld verwenden möchten.
   c) Wenn Sie anhand des Feldwertes eine Schrittoption innerhalb des eingebetteten Datenflusses festlegen möchten, müssen Sie den Namen der festzulegenden Option angeben.
   d) Klicken Sie auf **OK**.
   e) Fügen Sie bei Bedarf weitere Schlüsselfelder hinzu.

Wenn Sie mehr als ein Schlüsselfeld haben und die Option **Immer durchlaufen, wenn sich ein Schlüsselfeld ändert** auswählen, müssen die Datensätze denselben Wert in allen Schlüsselfeldern enthalten, um gruppiert zu werden.

Gruppierung eines eingebetteten Datenflusses aufheben

Wenn Sie die Gruppierung eines eingebetteten Datenflusses aufheben, werden die Schritte des eingebetteten Datenflusses im übergeordneten Datenfluss an der Stelle des eingebetteten Datenflusses platziert. Sämtliche Iterationseinstellungen, die Sie möglicherweise für den eingebetteten Datenfluss konfiguriert haben, werden entfernt.

Um die Gruppierung der Schritte eines eingebetteten Datenflusses aufzuheben, klicken Sie mit der rechten Maustaste auf den eingebetteten Datenfluss und wählen Sie **Gruppierung dieses eingebetteten Datenflusses aufheben** aus.

Umwandeln eines eingebetteten Datenflusses in einen Unterfluss

Wenn Sie einen eingebetteten Datenfluss erneut in einem anderen Datenfluss verwenden möchten, müssen Sie den eingebetteten Datenfluss in einen Unterfluss umwandeln. Dies liegt daran, dass eingebettete Datenflüsse nicht in anderen Datenflüssen verwendet werden können. Sobald Sie einen eingebetteten Datenfluss in einen Unterfluss umgewandelt haben, kann er wie jeder andere Unterfluss verwendet werden.

> **Anmerkung:** Wenn Iterationen beim eingebetteten Datenfluss aktiviert sind, werden die Iterationseinstellungen entfernt, wenn er in einen Unterfluss umgewandelt wird. Unterflüsse unterstützen Iterationen nicht.

1. Öffnen Sie im Enterprise Designer den Datenfluss mit dem eingebetteten Datenfluss, den Sie in einen Unterfluss umwandeln möchten.
2. Klicken Sie mit der rechten Maustaste auf den eingebetteten Datenfluss, und wählen Sie **Schritt in Unterfluss umwandeln** aus.
3. Geben Sie einen Namen für den neuen Unterfluss ein, und klicken Sie auf **OK**.

Der eingebettete Datenfluss wird in einen Unterfluss umgewandelt und in der Palette des Enterprise Designer im Ordner „Benutzerdefinierte Schritte“ verfügbar gemacht.

Berichte

Spectrum™ Technology Platform bietet Berichtsfunktionalitäten für Aufträge. Sie können Standardberichte verwenden, die in einigen Modulen enthalten sind, oder Ihre eigenen Berichte entwerfen. Wenn ein Bericht in einem Datenfluss enthalten ist, wird der gesamte Datenfluss ausgeführt. Nach Abschluss dieses Vorgangs werden die Berichtsschritte im Datenfluss ausgeführt. Anschließend werden die Berichte im gewünschten Format, z. B. PDF gespeichert.

Hinzufügen eines Standardberichts zu einem Auftrag

Ein Standardbericht ist ein vorkonfigurierter Bericht, der in einem Spectrum™ Technology Platform-Modul enthalten ist. Beispiel: Das Location Intelligence-Modul enthält den Zusammenfassungsbericht „Point In Polygon“, der die Ergebnisse von „Point In Polygon“-Berechnungen zusammenfasst, etwa die Anzahl der Vieleckübereinstimmungen, die für den Auftrag verwendete Datenbank und weitere Informationen.

Die folgenden Schritte beschreiben, wie ein Standardbericht zu einem Auftrag hinzugefügt wird.

1. Klicken Sie im Enterprise Designer auf der linken Seite des Fensters unter „Palette“ auf **Berichte**. Eine Liste der verfügbaren Berichte wird angezeigt.
2. Ziehen Sie den gewünschten Bericht auf die Arbeitsfläche. Sie müssen das Berichtssymbol mit nichts verbinden.
3. Doppelklicken Sie auf den Bericht.
4. Wählen Sie die Schritte aus, die Sie zum Bericht beitragen möchten.
5. Klicken Sie auf die Registerkarte **Parameter**.
6. Deaktivieren Sie das Kästchen **Standard-Berichterstellungsoptionen verwenden** und wählen Sie das passenden Ausgabeformat aus, wenn Sie ein anderes Format als PDF (wie HTML oder TXT) angeben möchten.

Festlegen von Berichtsoptionen für einen Auftrag

Berichte liefern zusammenfassende Informationen über einen Auftrag, etwa die Anzahl der verarbeiteten Datensätze, die für den Auftrag verwendeten Einstellungen usw. Berichtsoptionen geben an, wie die von einem Auftrag generierten Berichte verarbeitet werden sollen, etwa das Ausgabeformat und Archivierungsoptionen. Standardwerte für Berichtsoptionen werden in der Management Console angegeben, aber Sie können die Standardoptionen für einen Auftrag im Enterprise Designer überschreiben.

Die folgenden Schritte beschreiben, wie Berichtsoptionen für einen Auftrag angegeben werden.

1. Öffnen Sie den Auftrag im Enterprise Designer, und rufen Sie **Bearbeiten** > **Auftragsoptionen** auf.
2. Klicken Sie auf die Registerkarte **Berichterstellung**.
3. Deaktivieren Sie das Kästchen **Globale Berichterstellungsoptionen verwenden**.
4. Wählen Sie das Format aus, in dem Berichte gespeichert werden sollen. Berichte können als HTML, PDF oder Text gespeichert werden.
5. Wählen Sie den Speicherort für die Berichte aus.

Berichte im Auftragsverlauf speichern	Speichert Berichte als Bestandteil des Auftragsverlaufs auf dem Server. So können Benutzer der Management Console und des Enterprise Designer Berichte bequem anzeigen, da sie im Ausführungsverlauf verfügbar sind.
Berichte in Datei speichern	Speichert Berichte in einer Datei an einem von Ihnen angegebenen Speicherort. Dies ist hilfreich, wenn Sie Berichte für Personen freigeben möchten, die keine Benutzer von Spectrum™ Technology Platform sind. Dies ist auch hilfreich, wenn Sie an einem anderen Speicherort ein Berichtsarchiv erstellen möchten. Zum Anzeigen von Berichten, die auf diese Art gespeichert wurden, können Sie ein beliebiges Tool verwenden, mit dem der Bericht in dem entsprechenden Format geöffnet werden kann, z. B. einen PDF-Viewer für PDF-Berichte oder einen Webbrowser für HTML-Berichte.

6. Wenn Sie **Berichte in Datei speichern** ausgewählt haben, füllen Sie die folgenden Felder aus.

Berichtsspeicherort	Der Ordner, in dem Berichte gespeichert werden sollen.
An Berichtsnamen anfügen	Gibt Variableninformationen an, die in den Dateinamen einbezogen werden sollen. Sie können mindestens eine der folgenden Variablen auswählen: **Auftrags-ID** Eine eindeutige ID, die einer Auftragsausführung zugewiesen ist. Bei der ersten Ausführung eines Auftrags in Ihrem System hat der Auftrag die ID 1. Bei der zweiten Ausführung eines Auftrags, entweder desselben Auftrags oder eines anderen Auftrags, hat der Auftrag die ID 2 usw. **Schritt** Der im Berichtsschritt im Enterprise Designer angegebene Name des Schrittes, der Daten zum Bericht beigetragen hat. **Date** Das Datum, der Monat und das Jahr, an bzw. in dem der Bericht erstellt wurde.
Vorhandene Berichte überschreiben	Ersetzt vorherige Berichte, die den gleichen Dateinamen haben, durch den neuen Bericht. Wenn Sie diese Option nicht auswählen und ein Bericht vorhanden ist, der den gleichen Namen hat wie der neue Bericht, wird der Auftrag erfolgreich ausgeführt, der neue Bericht wird jedoch

nicht gespeichert. Im Ausführungsverlauf wird ein Kommentar angezeigt, der angibt, dass der Bericht nicht gespeichert wurde.

7. Klicken Sie auf **OK**.

Wenn Sie Ihren Auftrag ausführen, enthält der Ausführungsverlauf eine Spalte, die zeigt, ob es Berichte gibt, die dem Auftrag zugeordnet sind. Ein leeres Symbol steht für keine Berichte, ein Dokumentensymbol für einen Bericht und mehrere Dokumentensymbole für mehrere Berichte. Sie können die Auftragsdetails verwenden, um den Bericht anzuzeigen, zu speichern oder zu drucken.

Anmerkung: Um einen Bericht zu löschen, klicken Sie mit der rechten Maustaste auf das Berichtssymbol auf der Arbeitsfläche und wählen Sie **Löschen** aus.

Anzeigen von Berichten

Um Berichte anzuzeigen, führen Sie zuerst den Auftrag und dann eine der folgenden Aktionen aus:

- Im Enterprise Designer wird das Fenster **Ausführungsdetails** angezeigt, wenn Sie Ihren Auftrag ausführen. Wählen Sie den Bericht aus, die Sie anzeigen möchten.
- Klicken Sie in der Management Console im Knoten „Ausführung“ auf **Verlauf** und wählen Sie den anzuzeigenden Auftrag aus. Klicken Sie danach auf **Details**.

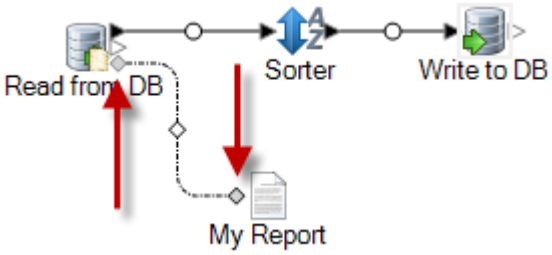
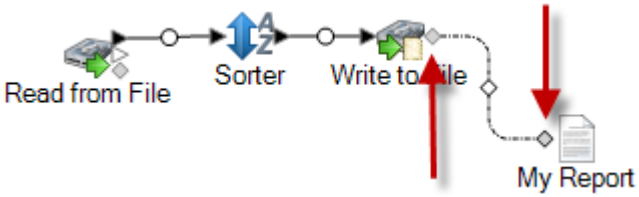
Verwenden von benutzerdefinierten Berichten

Spectrum™ Technology Platform-Module werden mit Berichten geliefert, die für die grundlegende Berichterstellung nützlich sind. Wenn Sie aber Berichtsanforderungen haben, die von den Standardberichten nicht erfüllt werden, können Sie eigene benutzerdefinierte Berichte erstellen und sie in Ihren Datenfluss einschließen.

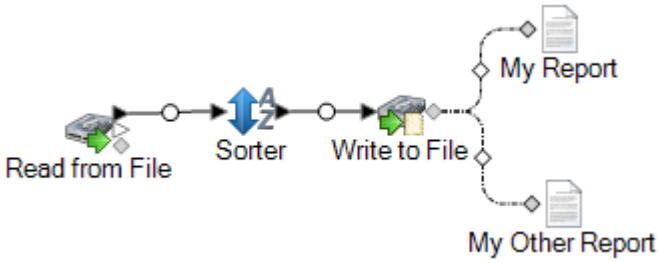
1. Erstellen Sie die Berichtsvorlage mit einem Berichtsentwurfstool Ihrer Wahl. Ihr Entwurfstool muss in der Lage sein, den Bericht im JasperReports-Format (.jrxml) zu exportieren.
2. Kopieren Sie Ihre JRXML-Datei in den Ordner `server\app\import` auf dem Spectrum™ Technology Platform-Server.

   Innerhalb von wenigen Sekunden wird die Berichtsvorlage ins System importiert und im Enterprise Designer bereitgestellt.

3. Öffnen Sie im Enterprise Designer den Auftrag, zu dem Sie Ihren benutzerdefinierten Bericht hinzufügen möchten.
4. Klicken Sie auf der linken Seite des Fensters unter „Palette“ auf **Berichte**.
5. Ziehen Sie Ihren benutzerdefinierten Bericht auf die Arbeitsfläche.
6. Geben Sie die Datenquelle für den Bericht an, indem Sie eine der folgenden Aktionen ausführen:

Option	Bezeichnung
Bericht über Eingabe des Datenflusses erstellen	Verbinden Sie den Bericht mit dem Quellschritt, über die Sie Berichte erstellen möchten. Verwenden Sie dazu den rautenförmigen Berichtsport wie in dieser Abbildung:  . Der Bericht basiert auf den Eingabedaten des Datenflusses und berücksichtigt die Verarbeitung im Datenfluss nicht.
Bericht über Ausgabe des Datenflusses erstellen	Verbinden Sie den Bericht mit dem Datenladeschritt, über den Sie Berichte erstellen möchten. Verwenden Sie dazu den rautenförmigen Berichtsport wie in dieser Abbildung:  Der Bericht basiert auf den Ausgabedaten des Datenflusses und berücksichtigt den Effekt des Datenflusses auf die Daten nicht.
So verwenden Sie eine in der Berichtsvorlage eingebettete Abfrage	Wenn die Berichtsvorlage eine eingebettete SQL-Abfrage im Element `<queryString>` der JRXML-Datei enthält, doppelklicken Sie auf das Berichtssymbol und aktivieren Sie das Kästchen **Eingebettete Abfrage verwenden**. Wählen Sie dann die für die Abfrage zu verwendende Datenbankverbindung aus. **Anmerkung:** Wenn Sie eine Datenbankverbindung definieren müssen, öffnen Sie Management Console und wechseln Sie zu **Ressourcen** und dann zu **Verbindungen**.

Sie können mehrere Berichte mit einer Quelle oder ein Laden von Daten verbinden (siehe Abbildung):



7. Wenn der Bericht benutzerdefinierte Parameter enthält:
   a) Doppelklicken Sie auf das Berichtssymbol auf der Arbeitsfläche.
   b) Geben Sie auf der Registerkarte **Parameter** die Werte an, die Sie für die benutzerdefinierten Parameter des Berichts verwenden möchten.
8. Optional: Klicken Sie bei Bedarf mit der rechten Maustaste auf den Kanal und ordnen Sie die Felder aus der Quelle oder der Datenladung den Feldern im Bericht zu.

Überlegungen zur Leistung

Designrichtlinien für eine optimale Leistung

Das sorgfältige Designen Ihrer Datenflüsse zur Leistungsoptimierung ist das Wichtigste, um eine gute Leistung von Spectrum™ Technology Platform zu erreichen. Diese Richtlinien beschreiben die Techniken, die Sie zur Optimierung der Datenflussleistung anwenden können.

Minimieren der Schrittanzahl

Spectrum™ Technology Platform erreicht durch eine parallele Verarbeitung eine hohe Leistung. Jeder Schritt in einem Fluss wird asynchron in einem eigenen Thread ausgeführt. Es ist jedoch möglich, die Prozessoren bei der Ausführung von bestimmten Datenflusstypen zu überlasten. Wenn dies passiert, verbringt das System mehr Zeit mit der Verwaltung der Threads als mit der eigentlichen Arbeit. Wir haben Datenflüsse mit 130 einzelnen Schritten gesehen, die sehr schlecht auf kleineren Servern mit einem oder zwei Prozessoren ausgeführt werden.

Die erste Überlegung beim Designen von Datenflüssen ist also eine gute Leistung, um so viele Schritte wie nötig aber so wenig Schritte wie möglich zu verwenden. Einige Beispiele zur Verwendung von mehr Schritten als nötig sind:

- Die Verwendung von mehreren Conditional Routers, wenn einer ausreichend wäre
- Die Definition von mehreren Transformer-Schritten anstelle der Kombination von Transformationen in einem einzelnen Schritt

Glücklicherweise ist es in der Regel möglich, diese Datenflüsse neu zu designen, um redundante oder nicht benötigte Schritte zu entfernen und die Leistung zu verbessern.

Ziehen Sie bei komplexen Flüssen die Verwendung von eingebetteten Flüssen oder Unterflüssen in Betracht, um die Elemente auf der Arbeitsfläche zu verringern und so die Ansicht und auch die Navigation durch den Fluss zu vereinfachen. Die Verwendung von eingebetteten Flüssen hat keine Auswirkung auf die Laufzeitleistung. Jedoch wird die Arbeit mit Flüssen im Enterprise Designer vereinfacht. Die Verwendung von Unterflüssen zur Vereinfachung komplexer Flüsse kann die Leistung vom Enterprise Designer beim Bearbeiten von Flüssen verbessern.

Verringern der Datensatzlänge

Da Daten zwischen gleichzeitig ausgeführten Schritten übermittelt werden, ist die Länge der Eingabedatensätze eine weitere Überlegung. Im Allgemeinen dauert die Eingabe mit einer großen Datensatzlänge länger als bei einer Eingabe mit einer kürzeren Datensatzlänge, da einfach mehr Daten gelesen, geschrieben und sortiert werden müssen. Datenflüsse mit mehreren Sortiervorgängen profitieren besonders von einer kürzeren Datensatzlänge. Bei sehr großen Datensatzlängen kann es besser sein, unnötige Felder aus der Eingabe zu entfernen, bevor der Spectrum™ Technology Platform-Auftrag ausgeführt wird, anstatt sie an die resultierende Ausgabedatei anzuhängen.

Verwenden einer angemessenen Sortierung

Eine weitere Überlegung ist die Minimierung der Sortiervorgänge. Die Sortierung ist oftmals zeitaufwendiger als andere Vorgänge und kann zu einem Problem werden, wenn die Anzahl und Größe der Eingabedatensätze zunimmt. Viele Spectrum™ Technology Platform-Schritte erfordern oder bevorzugen jedoch sortierte Eingabedaten. Das Universal Addressing-Modul und Enterprise Geocoding-Modul werden z. B. ausgeführt, wenn die Eingabe nach Land und Postleitzahl sortiert ist. Schritte wie „Intraflow Match“ und „Interflow Match“ erfordern eine sortierte Eingabe nach dem Feld „Gruppieren nach“. In manchen Fällen können Sie eine externe Sortieranwendung nutzen, um Eingabedaten vorzusortieren. Dies kann schneller als die Sortierung im Spectrum™ Technology Platform-Datenfluss sein.

Schritt-Laufzeitleistungsoptionen

Laufzeitleistungsoptionen steuern, wie einzelne Schritte in einem Datenfluss ausgeführt werden, und stellen Einstellungen bereit, mit deren Hilfe Sie die Leistung Ihres Datenflusses verbessern können. Welche Einstellungen Ihnen zur Verfügung stehen, hängt davon ab, wie Ihre Spectrum™ Technology Platform-Umgebung konfiguriert wurde.

- Die Option **Lokal** ist die Standardeinstellung, mit der Schritte auf dem lokalen Spectrum™ Technology Platform-Server ausgeführt werden und eine Laufzeitinstanz nutzen. Die Einstellung für Laufzeitinstanzen kann vergrößert werden, wodurch Parallelverarbeitung genutzt und die Leistung verbessert wird.
- Die Option **Verteilt** wird normalerweise in einer Clusterumgebung verwendet, in der ein Load Balancer und mehrere Spectrum™ Technology Platform-Server installiert werden müssen.

- Die Option **Remote** kann verwendet werden, falls Ihre Umgebung aus mehreren Spectrum™ Technology Platform-Servern besteht, aber nicht für verteilte Verarbeitung konfiguriert ist. Diese Option ermöglicht Ihnen, die Verarbeitung eines Schrittes von einem anderen Server ausführen zu lassen.

Datenbank-Poolgröße und Laufzeitinstanzen

In den meisten Spectrum™ Technology Platform-Umgebungen werden mehrere Flüsse zur selben Zeit ausgeführt, unabhängig davon, ob es sich dabei um Batchaufträge handelt oder um Dienste, die auf Werbservice- oder API-Anforderungen antworten. Um die gleichzeitige Verarbeitung zu optimieren, können Sie die Einstellung „Datenbank-Poolgröße“ verwenden, die die Anzahl der gleichzeitigen Anforderungen, die eine Spectrum-Datenbank verarbeitet, begrenzt und die Einstellung „Laufzeitinstanzen“, die die Anzahl der Instanzen eines Flussschritts steuert, die gleichzeitig ausgeführt werden. Diese beiden Einstellungen sollten gemeinsam optimiert werden, um eine optimale Leistung zu erreichen.

Datenbank-Poolgröße

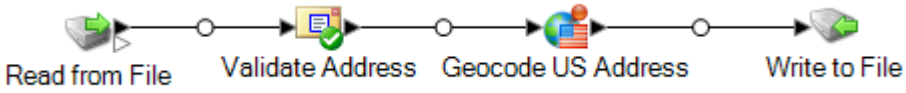
Spectrum-Datenbanken enthalten Verweisdaten, die in bestimmten Schritten verwendet werden, wie postalische Daten zum Validieren von Adressen oder Geocoding-Daten zum geographischen Referenzieren von Adressen. Diese Datenbanken können konfiguriert werden, mehrere gleichzeitige Anforderungen aus Datenflussschritten oder Diensten, die diese verwenden, zu akzeptieren und dadurch die Leistung dieser Datenfluss- oder Dienstanforderungen zu verbessern. Die Datenbank-Poolgröße legt die maximale Anzahl an gleichzeitigen Anforderungen fest, die eine Spectrum-Datenbank verarbeitet. Standardmäßig haben Spectrum-Datenbanken eine Poolgröße von 4, was bedeutet, dass die Datenbank gleichzeitig vier Anfragen verarbeiten kann.

Die optimale Poolgröße ist je nach Modul unterschiedlich. Im Allgemeinen erzielen Sie die besten Ergebnisse, wenn Sie die Poolgröße auf einen Wert zwischen der Hälfte und der doppelten Anzahl der Server-CPUs festlegen. Für die meisten Module entspricht die optimale Poolgröße der Anzahl der CPUs. Wenn Ihr Server beispielsweise über vier CPUs verfügt, können Sie mit einer Poolgröße zwischen 2 (die Hälfte der vorhandenen CPUs) und 8 (das Doppelten der vorhandenen CPUs) experimentieren, wobei die optimale Größe möglicherweise bei 4 liegt (Anzahl der vorhandenen CPUs).

Beim Ändern der Poolgröße müssen Sie außerdem die Anzahl von Laufzeitinstanzen berücksichtigen, die im Datenfluss der Schritte, die auf die Datenbank zugreifen, angegeben ist. Betrachten Sie beispielsweise einen Datenfluss mit einem „Geocode US Address“-Schritt der zur Verwendung einer Laufzeitinstanz konfiguriert ist. Wenn Sie die Poolgröße für die Geocoding-Datenbank USA auf 4 festlegen, verbessert sich die Leistung nicht, da nur eine Laufzeitinstanz vorhanden ist und nur eine Anforderung gleichzeitig an die Datenbank gestellt wird. Wenn Sie jedoch die Anzahl der Laufzeitinstanzen von „Geocode US Address“ auf 4 erhöhen würden, könnte sich die Leistung verbessern, da dann vier Instanzen von „Geocode US Address“ gleichzeitig auf die Datenbank zugreifen und damit die volle Poolgröße nutzen würden.

Laufzeitinstanzen

Jeder Schritt in einem Datenfluss läuft asynchron in seinem eigenen Thread und ist unabhängig von jedem anderen Schritt. Dies ermöglicht eine parallele Verarbeitung von Schritten in einem Datenfluss, wodurch Sie mehr als eine Laufzeitinstanz für einen Schritt nutzen können. Das ist in Datenflüssen hilfreich, bei denen einige Schritte die Daten schneller als andere verarbeiten. Dies kann zu einer unausgewogenen Arbeitsverteilung unter den Threads führen. Stellen Sie sich z. B. einen Datenfluss aus den folgenden Schritten vor:



Je nach Konfiguration der Schritte könnte es sein, dass der „Validate Address“-Schritt Datensätze schneller als der „Geocode US Address“-Schritt verarbeitet. In diesem Fall werden an einem Punkt während der Datenflussausführung alle Datensätze durch „Validate Address“ verarbeitet worden sein, aber „Geocode US Address“ ist noch mitten in der Verarbeitung. Zur Verbesserung der Leistung dieses Datenflusses ist es notwendig, die Leistung des langsamsten Schrittes zu verbessern; in diesem Fall „Geocode US Address“. Eine Möglichkeit wäre die Angabe von mehreren Laufzeitinstanzen des Schrittes. Die Einstellung der Anzahl von Laufzeitinstanzen auf zwei bedeutet, dass es zwei Instanzen dieses Schrittes gibt, von denen jede in ihrem eigenen Thread läuft und Datensätze verarbeitet.

Als allgemeine Regel gilt, dass die Anzahl der Laufzeitinstanzen mindestens der Anzahl der Instanzen der Remote-Komponente entsprechen sollte. Weitere Informationen zu Remote-Komponenten finden Sie im *Administratorhandbuch*. Obwohl das Festlegen von mehreren Laufzeitinstanzen die Leistung verbessern kann, kann eine zu hohe Einstellung die Systemressourcen überlasten, wodurch sich die Leistung wieder verringert.

> **Anmerkung:** Die Verwendung mit mehreren Laufzeitinstanzen verbessert nur die Leistung beim Ausführen von Aufträgen oder beim Ausführen von Dienstanforderungen mit mehr als einem Datensatz.

Optimierungsverfahren

Experimentieren Sie mit verschiedenen Einstellungen für „Datenbank-Poolgröße“ und „Laufzeitinstanzen“, um diejenigen zu finden, die die verfügbaren Serverressourcen maximieren, ohne die Ressourcen zu überladen und eine Leistungsverschlechterung zu verursachen.

> **Anmerkung:** Sie sollten die Datenfluss-Poolgröße optimieren, bevor Sie die Datenbank-Poolgröße anpassen. Informationen zum Optimieren der Datenfluss-Poolgröße finden Sie unter **Datenfluss-Poolgröße**.

1. Beginnen Sie damit, Testdaten zu suchen, die Sie zum Testen der verschiedenen Einstellungen verwenden können. Das Test-Dataset sollte mindestens so groß sein, dass die Ausführungszeit gemessen werden kann und eine Überprüfung auf Konsistenz durchgeführt werden kann. Die Testdaten sollten außerdem für die tatsächlichen Daten, die Sie verarbeiten möchten, repräsentativ sein. Wenn Sie einen Leistungstest für das Geocoding durchführen, müssen Sie

sicherstellen, dass Ihre Testdaten für alle zu geocodierenden Länder eine übereinstimmende Anzahl von Datensätzen enthalten.

2. Wenn Sie einen Dienst oder Datenfluss testen, für den die Verwendung einer Datenbankressource erforderlich ist, z. B. Postdatenbanken oder Geocoding-Datenbanken, müssen Sie sicherstellen, dass Sie die aktuelle Version der Datenbank installiert haben.
3. Erstellen Sie aus den vorbereiteten Testdaten und den aktuell installierten Datenbankressourcen einen einfachen Datenfluss, der Daten aus einer Datei liest, sie in dem zu optimierenden Schritt verarbeitet und in eine Datei schreibt. Beispiel: Erstellen Sie zum Testen von Leistungseinstellungen für Validate Address einen aus „Read from File“, „Validate Address“ und „Write to File“ bestehenden Datenfluss.
4. Legen Sie die Poolgröße der Datenbankressource auf 1 fest:
   a. Öffnen Sie die Management Console.
   b. Öffnen Sie **Ressourcen** > **Spectrum-Datenbanken**.
   c. Wählen Sie die Datenbankressource aus, die Sie optimieren möchten, und klicken Sie auf die Schaltfläche „Ändern“ .
   d. Geben Sie in das Feld **Poolgröße** 1 ein.
   e. Klicken Sie auf **OK**.
5. Legen Sie die Schritt-Laufzeitinstanzen auf 1 fest:
   a. Öffnen Sie den Datenfluss im Enterprise Designer.
   b. Doppelklicken Sie auf den Schritt, für den Sie mehrere Laufzeitinstanzen einstellen möchten.
   c. Klicken Sie auf **Laufzeit**.

      Anmerkung: Nicht alle Schritte können mehrere Laufzeitinstanzen verwenden. Wenn die Schaltfläche **Laufzeit** nicht im unteren Bereich des Schrittfensters vorhanden ist, kann der Schritt nicht mehrere Laufzeitinstanzen verwenden.

   d. Wählen Sie **Lokal** aus und geben Sie 1 an.
   e. Klicken Sie auf **OK**, um das Fenster **Laufzeitleistung** zu schließen. Klicken Sie anschließend auf **OK**, um den Schritt zu schließen.
6. Berechnen Sie eine Basisleistung, indem Sie den Datenfluss mehrmals ausführen und die Durchschnittswerte aufzeichnen für:
   - Verstrichene Zeit
   - CPU-Auslastung
   - Speicherauslastung

   Tipp: Sie können mit der JMX Console die Leistung überwachen. Weitere Informationen finden Sie unter **Überwachen der Leistung mit der JMX Console**.
7. Führen Sie bei einem zu unterstützenden Anwendungsfall mehrere Instanzen des Auftrags gleichzeitig aus. Zeichnen Sie die verstrichene Zeit, die CPU-Auslastung und die Speicherauslastung für die einzelnen Szenarien auf.

Tipp: Zur gleichzeitigen Ausführung mehrerer Instanzen eines Auftrags können Sie eine Dateiüberwachung verwenden. Weitere Informationen finden Sie unter **Auslösen eines Flusses mit einer Steuerungsdatei** auf Seite 169.

8. Erhöhen Sie die Einstellung für die Poolgröße der Datenbankressource und die Schritt-Laufzeitinstanzen.
9. Starten Sie den Server neu.
10. Führen Sie den Datenfluss erneut aus und zeichnen Sie dabei die verstrichene Zeit, die CPU-Auslastung und die Speicherauslastung auf.
11. Fahren Sie so lange mit der Erhöhung der Poolgröße der Datenbankressource und der Schritt-Laufzeitinstanzen fort, bis Sie eine Verringerung der Leistung feststellen können.
12. Wiederholen Sie beim Testen der Geocodingleistung diese Prozedur mit der Eingabe eines einzelnen Landes und mehrerer Länder.

Verteilte Verarbeitung

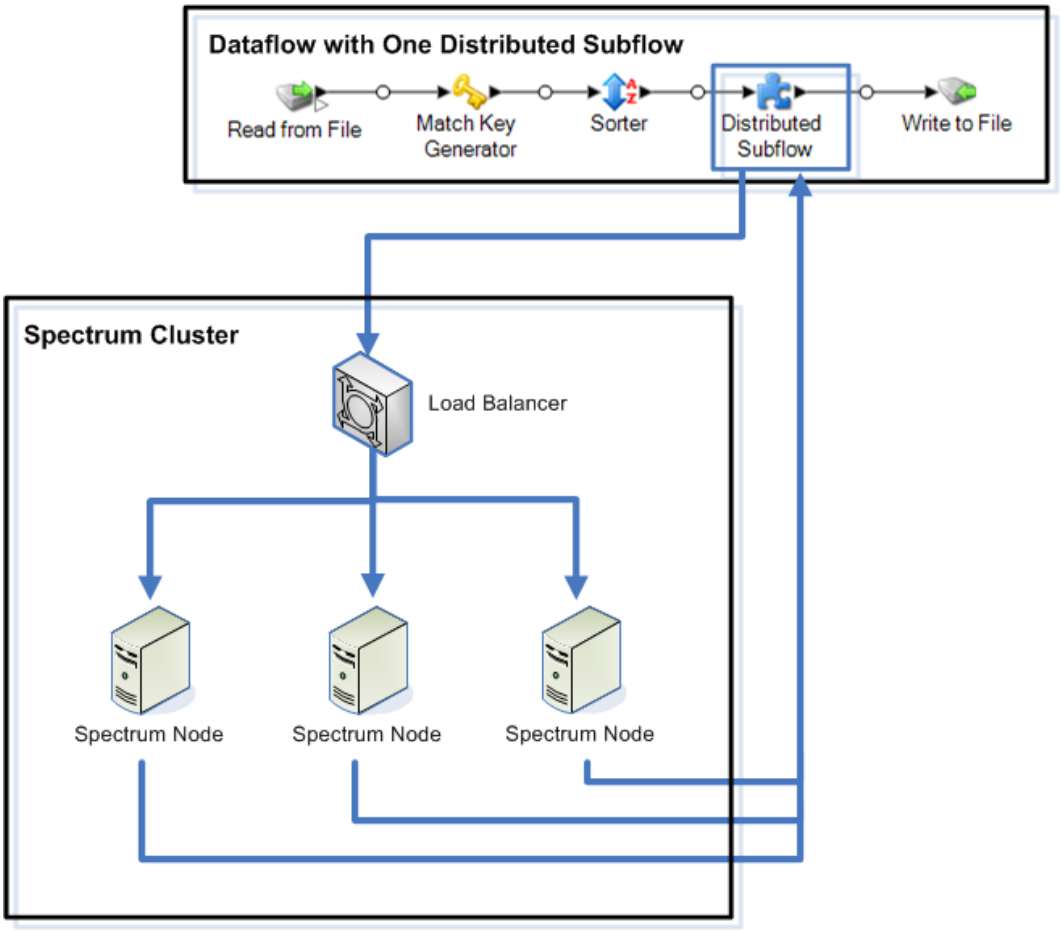
Wenn Sie über einen sehr komplexen Auftrag verfügen oder sehr große Datasets mit Millionen von Datensätzen verarbeiten müssen, können Sie die Datenflussleistung verbessern, indem Sie die Verarbeitung des Datenflusses auf mehrere Instanzen des Spectrum™ Technology Platform-Servers auf einem oder mehreren physischen Servern verteilen.

Die skalierbarste Lösung für eine verteilte Verarbeitung ist die Installation von Spectrum™ Technology Platform in einem Cluster. Anweisungen für die Installation und Konfiguration eines Clusters finden Sie im *Installationshandbuch*.

Anmerkung: Während es ebenfalls möglich ist, die verteilte Verarbeitung auf einem einzelnen Spectrum™ Technology Platform-Server auszuführen, beschreiben die folgenden Informationen die verteilte Verarbeitung in einem Cluster. Wenn Sie einen einzelnen Server verwenden, wird die verteilte Unterflussverarbeitung in Mikrobatches aufgeteilt, die nacheinander auf einem Server anstatt durch das Cluster verarbeitet werden.

Sobald Ihre Cluster-Umgebung eingerichtet ist, können Sie die verteilte Verarbeitung in einem Datenfluss erstellen, indem Sie Unterflüsse für die Teile des Datenflusses erstellen, die Sie auf mehrere Server verteilen möchten. Spectrum™ Technology Platform verwaltet automatisch die Verarbeitungsverteilung, nachdem Sie nur wenige Konfigurationsoptionen für den Unterfluss festlegen.

Das folgende Diagramm zeigt die verteilte Verarbeitung:



Beim Einlesen von Datensätzen in den Unterfluss werden die Daten in Batches gruppiert. Diese Batches werden anschließend in das Cluster geschrieben und automatisch auf einen Knoten im Cluster verteilt, der das Batch verarbeitet. Diese Verarbeitung wird als Mikrofluss bezeichnet. Ein Unterfluss kann konfiguriert werden, um mehrere, gleichzeitig zu verarbeitende Mikroflüsse zuzulassen, die potenziell die Leistung des Datenflusses verbessern. Wenn die verteilte Instanz die Verarbeitung eines Mikroflusses abgeschlossen hat, schickt sie die Ausgabe zurück in den übergeordneten Datenfluss.

Je mehr Spectrum™ Technology Platform-Knoten Sie haben, desto mehr Mikroflüsse können gleichzeitig verarbeitet werden. Dadurch können Sie Ihre Umgebung nach Bedarf skalieren, um die erforderliche Leistung zu erhalten.

Nach der Einrichtung ist eine Cluster-Umgebung einfach zu pflegen, da alle Knoten im Cluster automatisch ihre Konfiguration synchronisieren. Das bedeutet, dass die von Ihnen über die Management Console angewendeten Einstellungen und die im Enterprise Designer erstellten Datenflüsse automatisch für alle Instanzen verfügbar sind.

Entwerfen eines Datenflusses für eine verteilte Verarbeitung

Die verteilte Verarbeitung nimmt Teile Ihres Datenflusses und verteilt die Verarbeitung dieser Teile auf ein Cluster von Spectrum™ Technology Platform-Servern. Ihr Datenfluss könnte beispielsweise

Geocoding ausführen, und Sie könnten die Geocoding-Verarbeitung auf mehrere Spectrum™ Technology Platform-Knoten in einem Cluster verteilen, um die Leistung zu verbessern.

1. Entscheiden Sie, welche Schritte Ihres Datenflusses Sie verteilen möchten. Erstellen Sie anschließend einen Unterfluss, der die zu verteilenden Schritte enthält.

   Verwenden Sie nicht die folgenden Schritte in einem Unterfluss, der für eine verteilte Verarbeitung verwendet wird:

   - Sorter
   - Unique ID Generator
   - Record Joiner
   - Interflow Match

   Die folgenden Schrittsätze müssen zusammen in einem Unterfluss für eine verteilte Verarbeitung verwendet werden:

   - Vergleichsschritte (Intraflow Match und Transactional Match) sowie Konsolidierungsschritte (Filter, Best of Breed und Duplicate Synchronization).
   - Aggregator und Splitter

   Schließen Sie keine anderen Unterflüsse innerhalb des Unterflusses ein (geschachtelte Unterflüsse).

   Beachten Sie Folgendes, wenn Sie Vergleichsvorgänge in einem Unterfluss ausführen, der für eine verteilte Verarbeitung verwendet wird:

   - Die Sortierung muss im Auftrag und nicht im Unterfluss erfolgen. Sie müssen die Sortierung im Schritt deaktivieren und die Sortierung auf Auftragsebene ausführen.
   - „Match Analysis“ wird nicht in einem verteilten Unterfluss unterstützt.
   - Sammlungsnummern werden innerhalb einer Mikrofluss-Batch-Gruppe erneut verwendet.

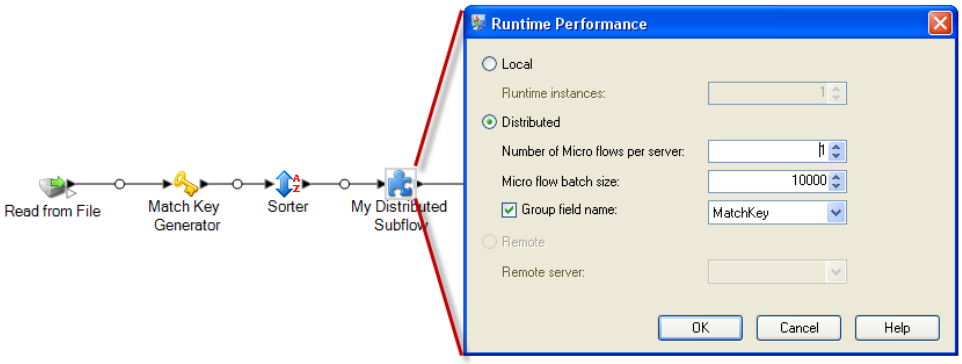
   Die Verwendung eines „Write Exception“-Schrittes in einem Unterfluss könnte unerwartete Ergebnisse generieren. Stattdessen sollten Sie diesen Schritt Ihrem Datenfluss auf Auftragsebene hinzufügen.

2. Sobald Sie Ihren Unterfluss für den Teil des zu verteilenden Datenflusses erstellt haben, fügen Sie den Unterfluss dem übergeordneten Datenfluss hinzu und verbinden Sie ihn mit einem vor- und nachgestellten Schritt. Unterflüsse für eine verteilte Verarbeitung könnten nur einen Eingabeport haben.
3. Klicken Sie mit der rechten Maustaste auf den Unterfluss, und wählen Sie **Optionen** aus.
4. Wählen Sie **Verteilt** aus.
5. Geben Sie die Anzahl der Mikroflüsse an, die an jeden Server geschickt werden sollen.
6. Geben Sie die Anzahl der Datensätze an, die in jedem Mikrofluss-Batch enthalten sein sollen.
7. Optional: (Optional) Aktivieren Sie **Feldname gruppieren**, und wählen Sie den Namen des Feldes aus, nach dem die Mikrofluss-Batches gruppiert werden sollen.

   Wenn Sie ein Gruppenfeld bereitstellen, könnte Ihre Batchgröße größer als die im Feld **Mikrofluss-Batchgröße** angegebene Zahl sein, da eine Gruppe nicht auf mehrere Batches

aufgeteilt wird. Wenn Sie beispielsweise eine Batchgröße von 100 angeben, die gleiche Gruppe jedoch 108 Datensätze enthält, umfasst dieses Batch 108 Datensätze. Wenn Sie eine Batchgröße von 100 angeben und eine neue Gruppe von 28 Datensätzen mit der gleichen ID bei Datensatz 80 beginnt, sind gleichermaßen 108 Datensätze in diesem Batch enthalten.

Das folgende Beispiel zeigt einen Datenfluss, für den ein Unterfluss mit dem Namen „My Distributed Subflow“ konfiguriert wurde, um im verteilten Modus ausgeführt zu werden:



Ausführen eines Schrittes auf einem Remoteserver

Wenn Ihr Systemadministrator Remoteserver in der Management Console aktiviert hat, können Sie Schritte in Ihrem Datenfluss ihre Verarbeitung auf einem Remoteserver ausführen lassen. Die Verwendung von Remoteservern kann die Leistung verbessern, indem die Datenflussverarbeitung auf mehrere Spectrum™ Technology Platform-Server verteilt wird.

Ihr Systemadministrator kann bereits bestimmte Schritte angegeben haben, die auf einem Remoteserver ausgeführt werden sollen. Wenn ein Schritt bereits an einen Remoteserver weitergeleitet wird, sehen Sie im Enterprise Designer in der linken oberen Ecke des Schrittsymbols auf der Arbeitsfläche einen roten Stern.

Diese Schritte beschreiben, wie die Remoteverarbeitung für einen Schritt in einem Datenfluss konfiguriert wird.

1. Öffnen Sie den Datenfluss im Enterprise Designer.
2. Doppelklicken Sie auf den Schritt, den Sie an einen Remoteserver weiterleiten möchten.
3. Klicken Sie auf **Laufzeit**.

   Das Dialogfeld **Laufzeitleistung** wird angezeigt.
4. Klicken Sie auf **Remote** und wählen Sie den Remoteserver aus, an den Sie den Prozess für diesen Schritt weiterleiten möchten.
5. Klicken Sie auf **OK**.

Fehlerbehebung bei Remoteserverfehlern

In diesem Abschnitt werden mögliche Fehler behandelt, die bei der Verwendung von Remoteservern auftreten können.

Modul nicht lizenziert

Der Remoteserver muss über die Lizenz für das Modul und den Ausführungsmodus verfügen, die Sie versuchen auszuführen (Batch oder Echtzeit). Die Lizenz auf dem Remoteserver kann von der Lizenz auf dem lokalen Server abweichen. Melden Sie sich über die Management Console bei dem Remoteserver an und überprüfen Sie, ob die richtige Lizenz installiert wurde. Sie müssen sich mit einem Konto mit Administratorberechtigungen anmelden, um Lizenzinformationen anzeigen zu können.

Remoteserver nicht verfügbar

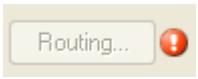
Wenn der Remoteserver nicht ausgeführt wird oder aus einem anderen Grund nicht erreichbar ist, sind die Remote-Services im Enterprise Designer und der Management Console nicht mehr verfügbar. Im unteren Bereich des Bildschirms wird Ihnen in der Statusleiste ein gelbes Warnsymbol angezeigt:

Klicken Sie auf dieses Symbol, um eine Fehlermeldung anzuzeigen, die beschreibt, welche Server nicht verfügbar sind.

Zusätzlich werden im Enterprise Designer alle Schritte, in denen ein Remoteschritt verwendet wird, durch ein Symbol ersetzt, das Ihnen den nicht mehr zur Verfügung stehenden Schritt anzeigt:

Weiterleitung wurde geändert

Wenn Sie einen Service löschen oder dessen Bereitstellung zurücknehmen, der lokal und remote installiert wurde und über einen Remoteserver weitergeleitet wurde, und anschließend in der Management Console auf diesen Dienst klicken, wird Ihnen neben der Schaltfläche „Weiterleitung“ auf der Registerkarte „Optionen“ dieses Services ein Indikator für Weiterleitungsänderungen angezeigt (ein blinkendes Ausrufezeichen). Dieser Indikator gibt an, dass die Weiterleitung bei diesem Service geändert wurde.



Optimieren von Schritten

Optimieren von Vergleichen

Vergleichen ist typischerweise einer der am zeitaufwändigsten Vorgänge in jeder Implementierung von Datenqualität. Deshalb ist es wichtig sicherzustellen, dass Vergleichen so effizient wie möglich abläuft. Es gibt immer ein Gleichgewicht zwischen Übereinstimmungen und Leistung. Wenn jeder

Datensatz in einer Datei mit jedem anderen Datensatz verglichen wird, können Sie ziemlich sicher sein, dass alle Übereinstimmungen identifiziert werden. Dieser Ansatz ist jedoch untragbar, da das Datenvolumens immens anwächst. Zum Beispiel würde bei einer Eingabedatei mit 1 Million Datensätzen das Vergleichen von jedem Datensatz mit jedem anderen Datensatz nahezu 1 Billion Vergleiche erfordern, um jede Vergleichsregel zu berücksichtigen.

Da die meisten Datensätze in einer Datei keine Übereinstimmung ergeben, liegt der allgemeine Ansatz zur Lösung dieses Problems darin, einen Vergleichsschlüssel zu definieren und nur die Datensätze zu vergleichen, die über denselben Vergleichsschlüssel verfügen. Die angemessene Definition eines Vergleichsschlüssels ist die entscheidendste Variable, die sich auf die Leistung der Vergleichs-Engine auswirkt. Sie müssen die Arbeitsweise der Vergleichs-Engine und die verfügbaren Optionen verstehen, um einen angemessenen Vergleichsschlüssel zu definieren.

Die Standardvergleichsmethode führt einen vollständigen Vergleich des Datensatzes in einer Vergleichswarteschlange durch, um die maximale Anzahl von Übereinstimmungen zu identifizieren. Daher ist dies oft die zeitaufwändigste Methode, einen Vergleich durchzuführen. Bei der Standardvergleichsmethode wird der erste Datensatz in der Vergleichswarteschlange zum Kopfdubletten-Datensatz. Der nächste Datensatz wird damit verglichen und bei einer Übereinstimmung als Dublette ausgegeben. Wenn keine Übereinstimmung vorliegt, wird der Datensatz als Kopfdublette hinzugefügt und der nächste Datensatz wird mit den beiden aktiven Kopfdubletten verglichen. Betrachten Sie die folgende Vergleichswarteschlange:

Eindeutige ID	Vergleichsschlüssel
1	123A
2	123A
3	123A
4	123A
5	123A
6	123A
7	123A
8	123A
9	123A
10	123A

Als Erstes würde Datensatz 2 mit Datensatz 1 verglichen werden. Angenommen, es besteht keine Übereinstimmung, dann würde Datensatz 2 als Kopfdublette hinzugefügt werden. Als Nächstes würde Datensatz 3 mit den Datensätzen 1 und 2 verglichen werden und so weiter. Wenn keine übereinstimmenden Datensätze gefunden werden, beträgt die Gesamtzahl der Vergleiche 45. Wenn einige Datensätze übereinstimmen, ist die Anzahl der Vergleiche geringer. Bei einer Vergleichswarteschlange mit gegebener Größe N beträgt die maximale Anzahl an Vergleichen N×(N-1)÷2. Bei einer kleinen Warteschlange sind keine Auswirkungen auf die Leistung festzustellen, doch werden sie bei zunehmender Warteschlangengröße deutlich größer. Eine Warteschlangengröße von 100 könnte beispielsweise 4.450 Vergleiche nach sich ziehen, und eine Warteschlangengröße von 500 könnte zu 124.750 Vergleichen führen.

Definieren eines angemessenen Vergleichsschlüssels

Berücksichtigen Sie beim Definieren eines angemessenen Vergleichsschlüssels Folgendes:

- Es ist am wichtigsten, nicht zu vergessen, dass die meisten Datensätze keine Übereinstimmung aufweisen. Daher sollten Sie nur Datensätze vergleichen, bei denen wahrscheinlich eine Übereinstimmung vorliegt.
- Nur Datensätze mit demselben Vergleichsschlüssel werden verglichen.
- Es ist wichtig, die Leistung zu berücksichtigen:
  - Der Vergleichsschlüssel legt die Größe der Vergleichswarteschlange fest.
  - Bei einer gegebenen Anzahl von Datensätzen verdoppelt sich die Ausführungszeit, wenn sich die Größe der Vergleichswarteschlange verdoppelt.
  - Ein „enger“ Vergleichsschlüssel führt zu höherer Leistung. Ein „enger“ Vergleichsschlüssel ist spezifisch und enthält mehr Zeichen aus möglicherweise mehr Feldern.
  - Ein „lockerer“ Vergleichsschlüssel führt möglicherweise zu mehr Übereinstimmungen. Ein „lockerer“ Vergleichsschlüssel ist weniger spezifisch und enthält weniger Zeichen aus möglicherweise weniger Feldern.

Finden eines Gleichgewichts zwischen Leistung und Vergleichsergebnissen

Berücksichtigen Sie die Vergleichsregel und die Datendichte, um ein gutes Gleichgewicht zwischen Leistung und Ergebnissen zu finden.

- Berücksichtigen Sie die Vergleichsregeln:
  - Sie könnten Felder, die eine genaue Übereinstimmung erfordern, in den Vergleichsschlüssel einschließen.
  - Erstellen Sie einen angemessenen Schlüssel für die Vergleichsregel. Beispielsweise ist für eine phonetische Vergleichsregel ein phonetischer Vergleichsschlüssel wahrscheinlich angemessen.
  - Ein Vergleichsschlüssel besteht oft aus Teilen aller zu vergleichenden Felder.
  - Seien Sie sich der Auswirkungen fehlender Daten bewusst.
- Berücksichtigen Sie die Datendichte:

  - Beispielsweise wäre bei einem Adressenvergleich der Vergleichsschlüssel wahrscheinlich enger, wenn alle Datensätze sich ein einer einzelnen Stadt anstatt in einem nationalen Dataset befinden würden.
  - Berücksichtigen Sie die größte Vergleichswarteschlange, nicht nur eine mittlere. Überprüfen Sie den Zusammenfassungsbericht der Vergleiche, um die größte Vergleichswarteschlange herauszufinden.
- Bei der Verwendung von Transactional Match gelten dieselben Überlegungen für die SELECT-Anweisung in Candidate Finder.

Schnellvergleichsschlüssel

In einer typischen Datei stimmen die meisten Datensatzdubletten genau oder fast genau überein. Wenn Sie einen Schnellvergleichsschlüssel definieren, kann die Vergleichen-Engine zu Beginn einen Vergleich der Schnellvergleichsschlüssel durchführen, um festzustellen, dass zwei Datensätze Dubletten sind. Dies kann die Leistung deutlich erhöhen, da nicht alle Vergleichsregeln auf Feldebene ausgewertet werden müssen.

Methoden von Intraflow Match

Die standardmäßige Vergleichsmethode von Intraflow Match vergleicht alle Datensätze, die denselben Vergleichsschlüssel aufweisen. Bei einer Vergleichswarteschlangengröße von N führt die Standardmethode N-1 bis N×(N-1) Vergleiche durch. Wenn alle Datensätze übereinstimmen, ist die Anzahl der Vergleiche N-1. Wenn keine Datensätze übereinstimmen, ist die Anzahl der Vergleiche N×(N-1). Üblicherweise liegt die Anzahl der Vergleiche irgendwo im oberen Teil dieses Bereichs.

Ziehen Sie in Betracht, die Vergleichsmethode „Gleitendes Fenster“ statt der Standardmethode zu verwenden, wenn Ihre Priorität auf Leistung liegt. Die Vergleichsmethode „Gleitendes Fenster“ vergleicht jeden Datensatz mit den nächsten F Datensätzen, wobei „F“ für die Fenstergröße steht. Bei einer gegebenen Dateigröße N führt die Methode „Gleitendes Fenster“ nicht mehr als N×F Vergleiche durch. Dies kann zu höherer Leistung führen, jedoch könnten einige Übereinstimmungen nicht gefunden werden.

OpOptimieren von Candidate Finder

Candidate Finder wählt Kandidatendatensätze aus einer Datenbank aus, um sie einem Vergleich über Transactional Match zur Verfügung zu stellen. Da Transactional Match den Kopfdubletten-Datensatz mit allen Kandidatendatensätzen, die von Candidate Finder zurückgegeben wurden, vergleicht, verhält sich die Leistung von Transactional Match proportional zur Anzahl der Vergleiche.

Sie können die Leistung von Candidate Finder jedoch auf verschiedene Weisen verbessern. Ein Datenbankadministrator oder ein Entwickler mit umfassenden Kenntnissen zum Datenbankschema und zu den Indizes sollte die SQL SELECT-Anweisung in Candidate Finder optimieren, um eine maximale Leistung von Candidate Finder zu erhalten. Eines der häufigsten Leistungsprobleme liegt in einer Abfrage mit einem JOIN, die einen vollständigen Tabellenscan erfordert. Ziehen Sie in diesem Fall in Betracht, einen Index hinzuzufügen oder anstelle eines JOIN ein UNION zu verwenden.

Als allgemeine Regel sollten SQL-Abfragen von qualifizierten Personen untersucht und optimiert werden.

Optimieren von Transformationen

Der Transformer-Schritt bietet eine Reihe von vordefinierten Vorgängen, die für Eingabedaten ausgeführt werden können. Im Allgemeinen werden diese vordefinierten Transformationen schneller als benutzerdefinierte Transformationen ausgeführt, da sie bereits kompiliert sind. Bei der Definition einer großen Anzahl von Transformationen wird eine benutzerdefinierte Transformation häufig schneller ausgeführt. Zum Kürzen einer Anzahl von Feldern wird die folgende benutzerdefinierte Transformation beispielsweise schneller ausgeführt als neun separate Kürzungstransformationen.

```
data['AddressLine1'] = (data['AddressLine1'] != null) ?
data['AddressLine1'].trim() : null;
data['AddressLine2'] = (data['AddressLine2'] != null) ?
data['AddressLine2'].trim() : null;
data['AddressLine3'] = (data['AddressLine3'] != null) ?
data['AddressLine3'].trim() : null;
data['AddressLine4'] = (data['AddressLine4'] != null) ?
data['AddressLine4'].trim() : null;
data['City'] = (data['City'] != null) ? data['City'].trim() : null;
data['StateProvince'] = (data['StateProvince'] != null) ?
data['StateProvince'].trim() : null;
data['PostalCode'] = (data['PostalCode'] != null) ?
data['PostalCode'].trim() : null;
data['LastName'] = (data['LastName'] != null) ? data['LastName'].trim()
 : null;
data['FirstName'] = (data['FirstName'] != null) ?
data['FirstName'].trim() : null;
```

Optimieren von „Write to DB“

Der „Write to DB“-Schritt wird standardmäßig übergeben, nachdem jede Zeile in die Tabelle eingefügt wurde. Aktiveren Sie jedoch zur Verbesserung der Leistung die Option **Übergabe als Batch**. Durch Aktivierung dieser Option erfolgt die Übergabe nach der angegebenen Anzahl von Datensätzen. Je nach Datenbank kann dies die Schreibleistung wesentlich verbessern.

Berücksichtigen Sie bei der Auswahl einer Batch-Größe Folgendes:

- **Dateneingangsrate für Schritt „Write to DB“:** Wenn Daten mit einer geringeren Rate eingehen, als die Datenbank verarbeiten kann, erzielen Sie durch Ändern der Batchgröße keine Verbesserung der Datenflussgesamtleistung. So profitieren beispielsweise Datenflüsse mit Adressenüberprüfung oder Geocoding nicht von größeren Batches.
- **Netzwerkverkehr:** Bei langsamen Netzwerken bewirkt eine Erhöhung der Batchgröße auf eine mittlere Größe (1.000 bis 10.000) eine bessere Leistung.
- **Lade- und/oder Verarbeitungsgeschwindigkeit der Datenbank:** Bei Datenbanken mit hoher Verarbeitungsleistung sorgen größere Batches für bessere Leistung.

- **Mehrere Laufzeitinstanzen:** Wenn Sie mehrere Laufzeitinstanzen des Schrittes „Write to DB“ verwenden, verursachen große Batches einen hohen Speicherverbrauch. Verwenden Sie daher eine kleine oder mittlere Batchgröße (100 bis 10.000).
- **Datenbank-Rollbacks:** Immer, wenn eine Anweisung fehlschlägt, wird der vollständige Batch rückgängig gemacht. Je größer der Batch, desto länger dauert es, die Rollback-Operation auszuführen.

Optimieren der Adressenüberprüfung

„Validate Address“ bietet die beste Leistung, wenn die Eingabedatensätze nach Postleitzahl sortiert sind. Das liegt daran, wie die Referenzdaten in den Arbeitsspeicher geladen werden. Bei einer sortierten Eingabe läuft die Verarbeitung um ein Vielfaches schneller als bei einer unsortierten Eingabe. Da nicht alle Datensätze immer über Daten im Postleitzahlfeld verfügen, ist die folgende Sortierung zu empfehlen:

1. Country (nur bei der Verarbeitung von Datensätzen mehrerer Länder erforderlich)
2. PostalCode
3. StateProvince
4. City

Optimieren von Geocoding

Geocoding-Schritte bieten die beste Leistung, wenn die Eingabedatensätze nach Postleitzahl sortiert sind. Das liegt daran, wie die Referenzdaten in den Arbeitsspeicher geladen werden. Bei einer sortierten Eingabe läuft die Verarbeitung um ein Vielfaches schneller als bei einer unsortierten Eingabe. Da nicht alle Datensätze immer über Daten im Postleitzahlfeld verfügen, ist die folgende Sortierung zu empfehlen:

1. PostalCode
2. StateProvince
3. City

Sie können Geocoding-Schritte zudem optimieren, indem Sie mit verschiedenen Vergleichsmodi experimentieren. Der Vergleichsmodus steuert, wie der Geocoding-Schritt feststellt, ob es sich bei einem Geocoding-Ergebnis um eine hohe Übereinstimmung handelt. Ziehen Sie in Betracht, die Einstellung für den Vergleichsmodus auf **Unscharf** festzulegen. Prüfen Sie anschließend, ob die Ergebnisse Ihren Anforderungen entsprechen. Der Modus **Unscharf** weist im Allgemeinen eine bessere Leistung auf als andere Vergleichsmodi.

Optimieren von Geocode US Address

Der „Geocode US Address“-Schritt verfügt über diverse Optionen, die sich auf die Leistung auswirken. Diese Optionen werden in der folgenden Datei festgelegt:

`SpectrumLocation\server\modules\geostan\java.properties`

egm.us.multimatch.max.records	Gibt die maximale Anzahl zurückzugebender Übereinstimmungen an. Eine kleinere Zahl führt zu besserer Leistung auf Kosten von Übereinstimmungen.

egm.us.multimatch.max.processing	Gibt die Anzahl durchzuführender Suchvorgänge an. Eine kleinere Zahl führt zu besserer Leistung auf Kosten von Übereinstimmungen.
FileMemoryLimit	Steuert, in welchem Umfang Referenzdaten zu Beginn in den Arbeitsspeicher geladen werden.

Datenflussversionen

Das Feature „Version“ im Enterprise Designer ermöglicht Ihnen, einen Revisionsverlauf Ihrer Datenflüsse beizubehalten. Sie können vorherige Versionen eines Datenflusses anzeigen, ältere Versionen zur Ausführung verfügbar machen und einen Verlauf Ihrer Änderungen beibehalten – für den Fall, dass Sie jemals zu einer vorherigen Version eines Datenflusses zurückkehren müssen.

Speichern einer Datenflussversion

Es gibt zwei Methoden, um eine Version Ihres Datenflusses im Enterprise Designer zu speichern:

- Machen Sie Ihren Datenfluss verfügbar. Jedes Mal, wenn Sie einen Datenfluss verfügbar machen, entweder durch Auswahl von **Datei** > **Verfügbar machen/Verfügbarkeit aufheben und speichern** oder durch Klicken auf die Glühbirne in der Symbolleiste, speichert der Enterprise Designer automatisch eine Version des Datenflusses.
- Speichern Sie manuell eine Version im Bereich **Versionen** im Enterprise Designer.

  Anmerkung: Es wird keine Datenflussversion erstellt, wenn Sie einen Datenfluss einfach nur speichern.

Die folgenden Schritte beschreiben, wie manuell eine Version im Bereich **Versionen** im Enterprise Designer gespeichert wird.

1. Öffnen Sie im Enterprise Designer den Datenfluss.
2. Falls der Bereich **Versionen** nicht sichtbar ist, wählen Sie **Ansicht** > **Versionen** aus.
3. Stellen Sie sicher, dass die letzte gespeicherte Version in der Liste **Versionen** ausgewählt ist. Dies ist die Version oben in der Liste.
4. Klicken Sie auf das grüne Plussymbol im Bereich **Versionen**.

Eine neue Version des Datenflusses wird gespeichert und im Bereich **Versionen** hinzugefügt.

Anzeigen einer Datenflussversion

Sie können eine vorherige Version eines Datenflusses anzeigen. Das ermöglicht Ihnen, zu sehen, wie ein Datenfluss in der Vergangenheit entworfen wurde, bevor zuletzt Änderungen vorgenommen wurden. Vorherige Versionen können nur angezeigt, aber nicht geändert werden. Um eine vorherige Version zu ändern, muss sie zuerst zur zuletzt gespeicherten Version heraufgestuft werden.

1. Öffnen Sie im Enterprise Designer den Datenfluss.
2. Falls der Bereich **Versionen** nicht sichtbar ist, wählen Sie **Ansicht** > **Versionen** aus.
3. Wählen Sie die Version aus, die Sie anzeigen möchten.

Die ausgewählte Version wird auf der Datenfluss-Arbeitsfläche angezeigt.

Bearbeiten einer Datenflussversion

Sie können eine vorherige Version eines Datenflusses bearbeiten, indem Sie sie zur zuletzt gespeicherten Version heraufstufen. Durch Heraufstufen einer Datenflussversion wird sie zur zuletzt gespeicherten Version verschoben und dadurch zur Bearbeitung verfügbar.

> **Anmerkung:** Bevor Sie diese Schritte ausführen, beachten Sie, dass die vorhandene zuletzt gespeicherte Version durch die Version überschrieben wird, die Sie heraufstufen und bearbeiten. Wenn Sie eine Kopie der vorhandenen zuletzt gespeicherten Version beibehalten möchten, speichern Sie sie vor dem Heraufstufen der älteren Version als Version.

1. Öffnen Sie im Enterprise Designer den Datenfluss.
2. Falls der Bereich **Versionen** nicht sichtbar ist, wählen Sie **Ansicht** > **Versionen** aus.
3. Wählen Sie die Version aus, die Sie bearbeiten möchten.
4. Klicken Sie auf das Symbol „Heraufstufen“.

Die ausgewählte Version wird zur zuletzt gespeicherten Version heraufgestuft. Sie können den Datenfluss jetzt bearbeiten.

Bearbeiten von Versionseigenschaften

Wenn Sie eine Datenflussversion speichern, erhält sie eine Standardversionsnummer. Sie können die Versionsnummer ändern und Kommentare hinzufügen, um die Änderungen oder den Zweck der Version zu dokumentieren.

1. Öffnen Sie im Enterprise Designer den Datenfluss.
2. Falls der Bereich **Versionen** nicht sichtbar ist, wählen Sie **Ansicht** > **Versionen** aus.
3. Wählen Sie die Version aus, die Sie bearbeiten möchten.
4. Klicken Sie auf das Symbol „Eigenschaften“:

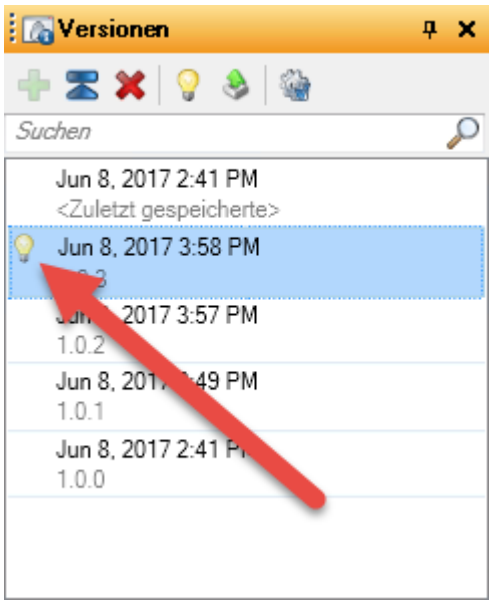
5. Geben Sie in das Feld **Name** einen Namen für die Version ein. Sie können Versionsnummern oder einen beliebigen aussagekräftigen Namen verwenden. Sie können den Namen frei wählen.
6. In das Feld **Kommentar** können Sie einen längeren Kommentar eingeben, der den Zweck der Version der vorgenommenen Änderungen ausführlicher beschreibt. Sie müssen allerdings keinen Kommentar hinzufügen.
7. Klicken Sie auf **OK**.

Verfügbarmachen einer Version

Wenn Sie mehrere Versionen eines Datenflusses gespeichert haben, können Sie wählen, welche Version zur Ausführung verfügbar gemacht werden soll.

1. Öffnen Sie im Enterprise Designer den Datenfluss.
2. Falls der Bereich **Versionen** nicht sichtbar ist, wählen Sie **Ansicht** > **Versionen** aus.
3. Wählen Sie im Bereich **Versionen** die Version des Datenflusses aus, die Sie verfügbar machen möchten.
4. Wählen Sie **Datei** > **Verfügbar machen/Verfügbarkeit aufheben und Speichern** aus.

Die ausgewählte Version wird jetzt verfügbar gemacht und ist zur Ausführung verfügbar. Die mit einer Glühbirne gekennzeichnete Version ist die Version, die verfügbar gemacht wurde (siehe Abbildung):



Wenn ein Datenfluss verfügbar gemacht wird, gibt die Glühbirnenschaltfläche in der Symbolleiste des Enterprise Designer an, dass der Datenfluss verfügbar gemacht wurde (siehe Abbildung):

Die Glühbirne gibt an, dass der Datenfluss verfügbar gemacht wurde, auch wenn Sie gerade eine andere Version als die verfügbar gemachte Version anzeigen. Wenn Sie auf die Glühbirne klicken, während Sie eine nicht verfügbar gemachte Version anzeigen, wird die verfügbar gemachte Version in die derzeit angezeigte Version geändert. Wenn Sie auf die Glühbirne klicken, während Sie die verfügbar gemachte Version anzeigen, wird die Verfügbarkeit des Datenflusses aufgehoben.

3 - Überprüfen und Testen

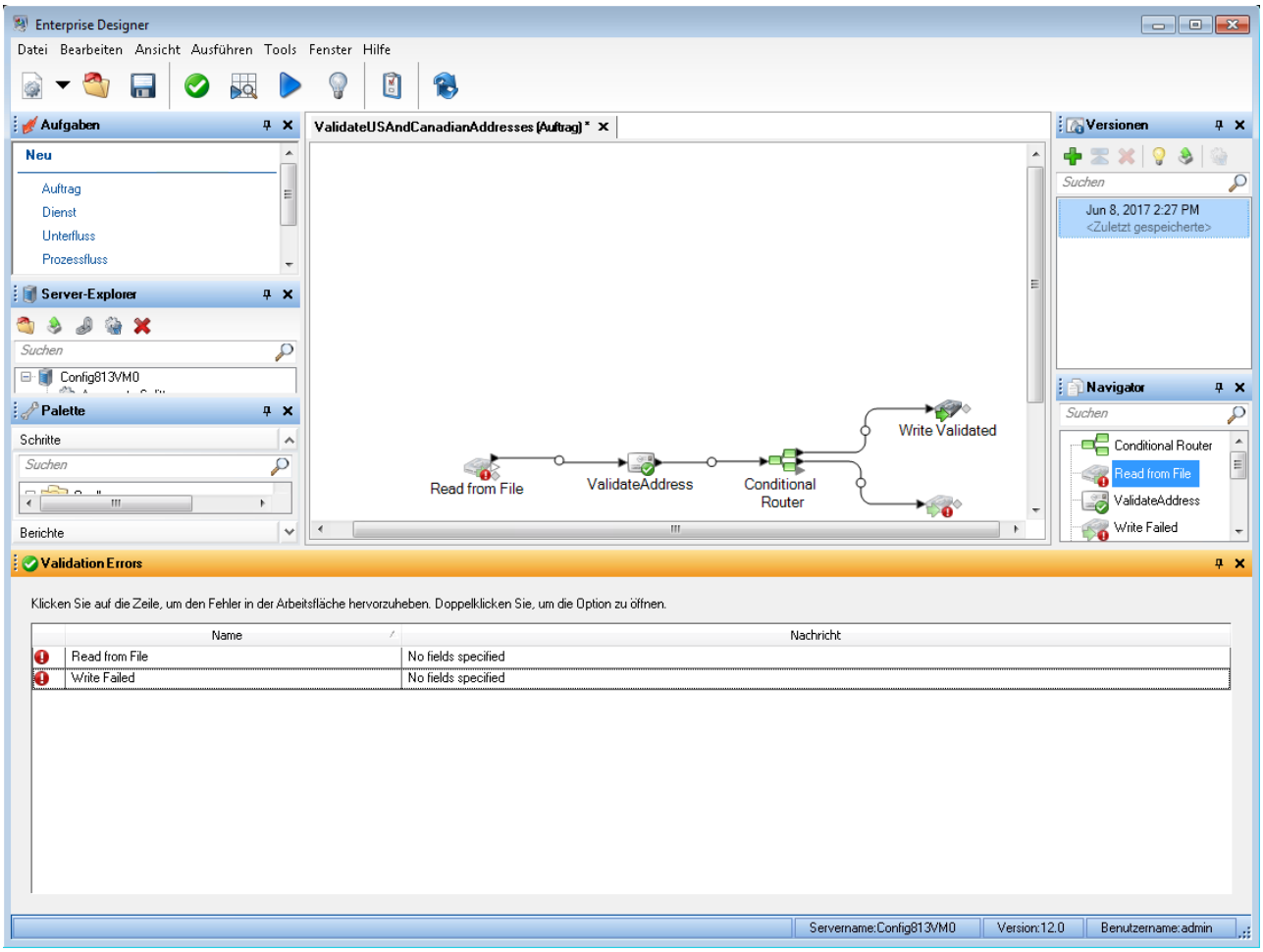
In this section

Prüfen eines Flusses auf Fehler 146
Überprüfen eines Datenflusses 147
Testen eines Dienstes mit Management Console 152

Prüfen eines Flusses auf Fehler

Der Enterprise Designer prüft einen Fluss automatisch auf Fehler, wenn Sie einen Fluss ausführen, eine Überprüfung durchführen, einen Fluss verfügbar machen oder einen verfügbar gemachten Fluss speichern. Sie können auch eine Fehlerüberprüfung durchführen, indem Sie auf die Schaltfläche „Überprüfung“ klicken.

Wenn ein Fehler gefunden wird, wird der Bereich **Überprüfung** auf der unteren Seite des Fensters „Enterprise Designer“ angezeigt. Klicken Sie auf einen Fehler, um ihn auf der Arbeitsfläche hervorzuheben. Klicken Sie doppelt auf einen Fehler, um das Fenster „Optionen“ des Elements mit dem Fehler zu öffnen.



Überprüfen eines Datenflusses

Verwenden Sie das Überprüfungstool im Enterprise Designer, um die Auswirkung Ihres Datenflusses auf die Eingabedaten an anderen Punkten im Datenfluss anzuzeigen. Durch die Überprüfung können Sie bestätigen, dass der Datenfluss die gewünschten Auswirkungen auf Ihre Daten hat, Probleme isolieren oder defekte Datensätze identifizieren.

Anmerkung: Eine Datenflussüberprüfung ist in eingebetteten Datenflüssen nicht verfügbar.

1. Legen Sie die Daten für die Überprüfung fest.

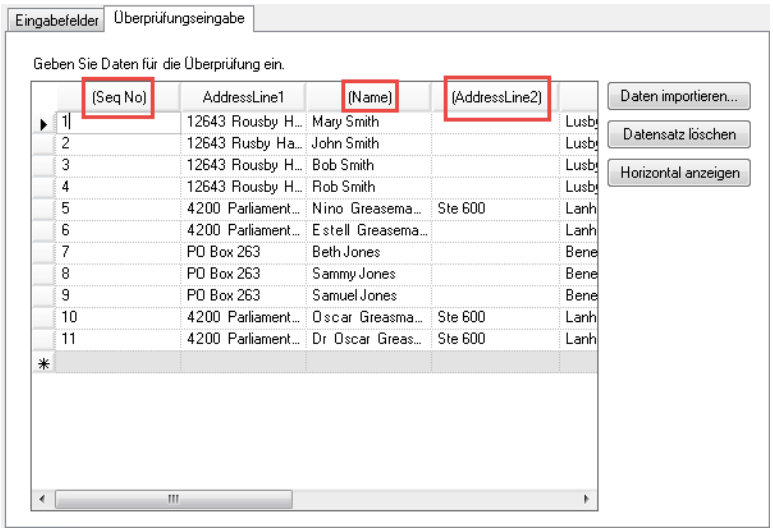
   Die ausgewählten Daten müssen die tatsächlichen Daten gut repräsentieren. Oder wenn Sie ein bestimmtes Problem untersuchen, müssen es die Daten sein, die das Problem verursachen. Es gibt zwei Wege, die Daten für die Überprüfung festzulegen. Dies hängt davon ab, ob Sie einen Dienst oder einen Auftrag überprüfen.

Szenario	Beschreibung
So legen Sie die Überprüfungsdaten für einen Auftrag fest:	Wenn Sie einen Auftrag überprüfen, sind die Daten für die Überprüfung die im Quellschritt festgelegten Daten. Das Überprüfungstool kann maximal 50 Datensätze verarbeiten. Dies sind standardmäßig die ersten 50 Datensätze in der Eingabedatei oder -datenbank. Wenn Sie Daten verwenden möchten, die an anderer Stelle als beim ersten Datensatz beginnen, doppelklicken Sie auf den „Read from File“-Schritt, und füllen Sie auf der Registerkarte „Laufzeit“ das Feld **Startdatensatz** aus.
So legen Sie die Überprüfungsdaten für einen Dienst fest:	Dienstdatenflüsse verwenden einen Input-Schritt, um die Eingabe in den Datenfluss zu definieren. Da ein Input-Schritt keinen Zugriff auf Daten hat, wenn Sie den Datenfluss bearbeiten, müssen Sie die Überprüfungsdaten im Input-Schritt auf der Registerkarte **Überprüfungsdaten** definieren. Sie können maximal 50 Datensätze festlegen. Es gibt einige Wege, wie Sie die Überprüfungsdaten in einen Input-Schritt eingeben. • Wenn Sie einige wenige Datensätze für die Überprüfung verwenden möchten, können Sie die Daten manuell eingeben. **Tipp:** Wenn Sie die eingegebenen Überprüfungsdaten speichern möchten, um sie in einem anderen Schritt zu verwenden, können Sie die Daten in eine Datendatei exportieren, indem Sie auf **Daten exportieren** klicken.

Szenario	Beschreibung

- Wenn Sie über Daten in einer CSV- oder TXT-Datei verfügen, können Sie die Daten durch Klicken auf **Daten importieren** importieren. Die Daten müssen eines der folgenden Begrenzungszeichen verwenden:
  - \t
  - |
  - ,
  - ;
- Sie können Daten mit Begrenzungszeichen aus einer anderen Anwendung kopieren und Sie in den Editor mit den Überprüfungsdaten einfügen.

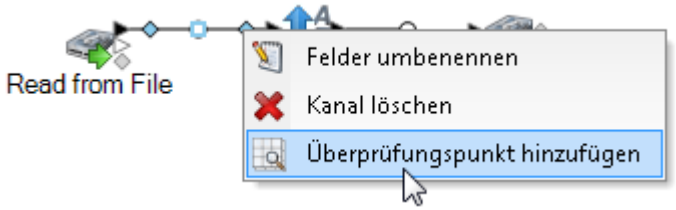
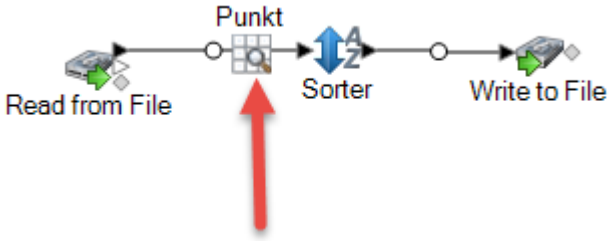
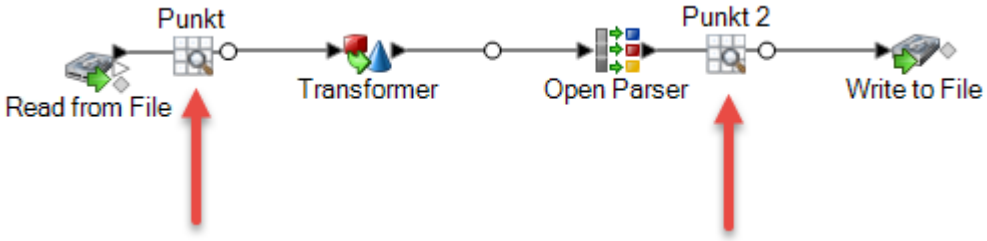
Die Registerkarte **Überprüfungseingabe** gibt die Durchlaufdaten an, indem der Feldname wie folgt in Klammern gesetzt wird:

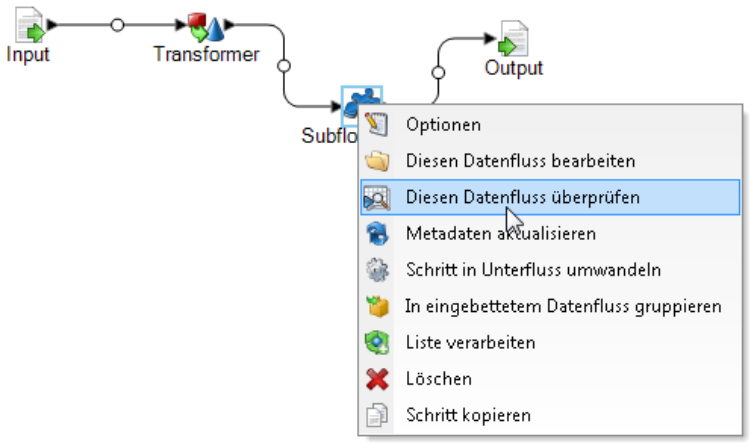


Anmerkung: Bestimmte Feldtypen beinhalten Einschränkungen bei der Überprüfung:

- Double- und Float-Felder dürfen nur numerische Daten enthalten. Das Feld kann bis zu 16 Ziffern und 6 Dezimalstellen enthalten. In der Überprüfung wird keine exponentielle Darstellung unterstützt.
- Integer- und Long-Felder dürfen nur numerische Daten enthalten.

2. Gibt die Punkte im Datenfluss an, wo Sie die Daten anzeigen möchten.

Szenario	Beschreibung
So fügen Sie einen Überprüfungspunkt zu einem Kanal hinzu:	Klicken Sie mit der rechten Maustaste links neben den Knoten „Umbenennen“ in einem Kanal, und wählen Sie die Option **Überprüfungspunkt hinzufügen** aus.  Dem Auftrag wird ein Punkt hinzugefügt: 
Vergleichen von Datensätzen an zwei Punkten in einem Datenfluss	Fügen Sie zwei Überprüfungspunkte an den Punkten im Datenfluss hinzu, die Sie vergleichen möchten:  **Tipp:** Wenn Sie Überprüfung verwenden, um ein Problem zu identifizieren, überprüfen Sie die äußeren Punkte im Datenfluss zuerst und bewegen Sie sich anschließend nach innen, um einzuschränken, wo sich ein Problem befinden könnte.
So überprüfen Sie einen in einen Auftrag oder	Klicken Sie mit der rechten Maustaste auf den Unterflussschritt, und wählen Sie die Option **Diesen Datenfluss überprüfen** aus:

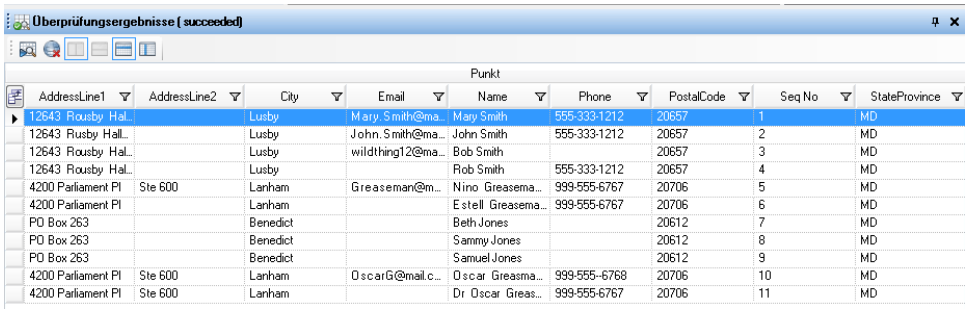
Szenario	Beschreibung
Dienst eingebetteten Unterfluss:	

Die Eingabedaten (in einem Auftrag) oder die Überprüfungsdaten (in einem Dienst) werden automatisch an den Unterfluss geleitet, so dass Sie die Überprüfungsdaten nicht in den Input-Schritt des Unterflusses eingeben müssen.

Anmerkung: Wenn Sie einen Unterfluss überprüfen, wird die verfügbar gemachte Version des Unterflusses angezeigt. Wenn Sie Änderungen am Unterfluss vornehmen und die Überprüfung erneut ausführen möchten, müssen Sie die neue Version verfügbar machen.

3. Wählen Sie **Ausführen** > **Aktuellen Fluss überprüfen** aus, oder klicken Sie in der Symbolleiste auf die Schaltfläche **Aktuellen Fluss überprüfen**.

   Wenn Sie einen Überprüfungspunkt angegeben haben, zeigt der Bereich **Überprüfungsergebnisse** die überprüften Daten in einer horizontalen Ansicht an. Sie können das Layout der Ansicht mithilfe der Symbolleistensymbole über der Tabelle ändern. Wenn Ihre Überprüfungsdaten hierarchisch sind, können Sie nicht vertikal angezeigt werden.

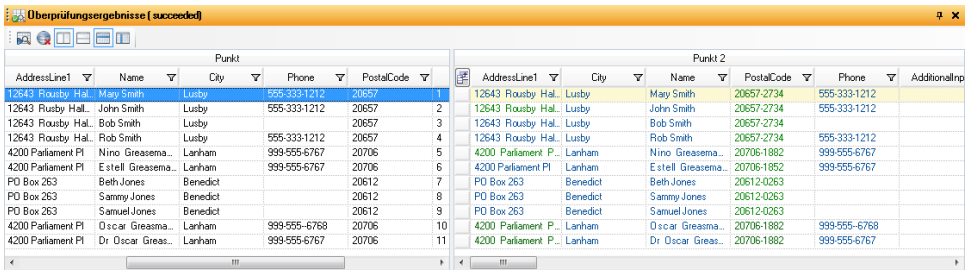


AddressLine1	AddressLine2	City	Email	Name	Phone	PostalCode	Seq No	StateProvince
12643 Rousby Hal...		Lusby	Mary.Smith@ma...	Mary Smith	555-333-1212	20657	1	MD
12643 Rusby Hall...		Lusby	John.Smith@ma...	John Smith	555-333-1212	20657	2	MD
12643 Rousby Hal...		Lusby	wildthing12@ma...	Bob Smith		20657	3	MD
12643 Rousby Hal...		Lusby		Rob Smith	555-333-1212	20657	4	MD
4200 Parliament Pl	Ste 600	Lanham	Greaseman@m...	Nino Greasema...	999-555-6767	20706	5	MD
4200 Parliament Pl		Lanham		Estell Greasema...	999-555-6767	20706	6	MD
PO Box 263		Benedict		Beth Jones		20612	7	MD
PO Box 263		Benedict		Sammy Jones		20612	8	MD
PO Box 263		Benedict		Samuel Jones		20612	9	MD
4200 Parliament Pl	Ste 600	Lanham	OscarG@mail.c...	Oscar Greasma...	999-555-6768	20706	10	MD
4200 Parliament Pl	Ste 600	Lanham		Dr Oscar Greas...	999-555-6767	20706	11	MD

Anmerkung: Datums- und Zeitdaten werden in dem Format angezeigt, das in den Typkonvertierungsoptionen festgelegt wurde.

Tipp: Sie können einen Überprüfungspunkt verschieben, indem Sie ihn in einen anderen Kanal ziehen. Die Überprüfungsdaten werden automatisch aktualisiert.

Wenn Sie zwei Überprüfungspunkte angegeben haben, zeigt der Bereich **Überprüfungsergebnisse** die Datensätze an, wie sie an den beiden Punkten vorhanden sind. Der linke Bereich enthält den Überprüfungspunkt, der sich im Datenfluss am weitesten links befindet, und der rechte Bereich enthält den Überprüfungspunkt, der sich im Datenfluss am weitesten rechts befindet. Klicken Sie auf einen Datensatz im rechten Bereich, um den entsprechenden Datensatz im linken Bereich hervorzuheben und festzustellen, wie sich der Datensatz zwischen den beiden Überprüfungspunkten verändert hat.



Jede Spalte steht für ein Feld im Datenfluss. Spalten sind in alphabetischer Reihenfolge angeordnet. Neue Felder, die zwischen den Überprüfungspunkten hinzugefügt wurden, sind im rechten Bereich nach den ursprünglichen Spalten aufgeführt. Sie können die Spalten über Klicken und Ziehen in die gewünschte Reihenfolge bringen.

Folgende Situationen beeinflussen die Darstellung der Überprüfungsergebnisse für zwei Überprüfungspunkte:

- Wenn sich ein Sorter-Schritt zwischen den beiden Überprüfungspunkten befindet, werden die Datensätze in den Überprüfungsergebnissen in der Reihenfolge angezeigt, die vor dem Sorter-Schritt bestand. Die Sortierung des zweiten Überprüfungspunkts wird ignoriert, damit Sie die zusammengehörigen Datensätze beider Überprüfungspunkte nebeneinander vergleichen können.
- Wenn sich zwischen den beiden Überprüfungspunkten Schritte befinden, die neue Datensätze erstellen, wie beispielsweise ein Aggregator-Schritt, verfügen die für den zweiten Überprüfungspunkt angezeigten Datensätze nicht über einen entsprechenden Datensatz für den ersten Überprüfungspunkt.
- Datensätze, die am zweiten, aber nicht am ersten Überprüfungspunkt vorhanden sind, werden am Ende der Liste der Datensätze für den zweiten Überprüfungspunkt angezeigt.

4. Wenn Sie einen Datenfluss aktualisieren oder ändern, klicken Sie auf **Ausführen** > **Aktuellen Fluss überprüfen**, um die Überprüfungsergebnisse zu aktualisieren.
5. Wenn Sie den Bereich „Überprüfungsergebnisse“ schließen, gehen Sie Überprüfungsdaten verloren. Wenn Sie einen Auftrag schließen, gehen gleichermaßen die Überprüfungspunkte und Überprüfungsdaten verloren. So speichern Sie die Überprüfungsergebnisse in einer Datei:
   a) Wählen Sie im Grid der Überprüfungsergebnisse die Zeilen aus, die Sie speichern möchten. Sie können alle Daten auswählen, indem Sie mit der rechten Maustaste in einen der Bereiche klicken und die Option **Alle auswählen** auswählen.

b) Wählen Sie aus dem Kontextmenü die Option **Kopieren** aus.
c) Öffnen Sie die Anwendung, in der Sie die Daten speichern möchten (z. B. Microsoft Excel oder Notepad).
d) Fügen Sie die Daten in die Anwendung ein.
e) Speichern Sie die Datei.

Testen eines Dienstes mit Management Console

Management Console stellt ein Vorschaufeature bereit, das Ihnen erlaubt, Testdaten an einen Dienst zu senden und die Ergebnisse anzuzeigen.

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. Wechseln Sie zum Menü **Dienste** und klicken Sie auf das Modul mit dem zu testenden Dienst.
3. Klicken Sie auf den zu testenden Dienst.
4. Klicken Sie auf **Vorschau**.
5. Geben Sie die gewünschten Eingabedaten für Ihren Test ein. Um Daten aus einer Datei zu importieren, klicken Sie auf die Schaltfläche „Importieren“ .
6. Klicken Sie auf **Vorschau ausführen**.

4 - Ausführen eines Flusses

In this section

Ausführen eines Auftrags- oder Prozessflusses 154
Verfügbarmachen eines Dienstes 178
Laufzeitoptionen 181
Konfigurieren einer E-Mail-Benachrichtigung für einen Datenfluss 184

Ausführen eines Auftrags- oder Prozessflusses

Ausführen eines Datenflusses im Enterprise Designer

Die folgenden Schritte beschreiben, wie ein Auftrag oder Prozessfluss manuell ausgeführt wird.

1. Wählen Sie im Enterprise Designer **Datei** > **Öffnen** aus, und öffnen Sie den auszuführenden Datenfluss.
2. Überprüfen Sie einen Datenfluss vor dem Ausführen, um sicherzustellen, dass er keine Fehler enthält. Wählen Sie dazu **Ausführen** > **Überprüfen** aus.
3. Wählen Sie **Ausführen** > **Aktuellen Fluss ausführen** aus.

Ausführen eines Auftrags über die Befehlszeile

Bevor Sie einen Auftrag über die Befehlszeile ausführen können, muss er verfügbar gemacht werden. Öffnen Sie dazu den Auftrag im Enterprise Designer und wählen Sie **Datei** > **Verfügbar machen/Verfügbarkeit aufheben und speichern** aus.

Um einen Auftrag über die Befehlszeile auszuführen, müssen Sie das Dienstprogramm „Job Executor“ auf dem System installieren, auf dem Sie den Auftrag ausführen möchten. Job Executor ist über die Spectrum™ Technology Platform-Begrüßungsseite auf dem Spectrum™ Technology Platform-Server verfügbar (z. B. http://MeinServer:8080).

Usage

`java -jar jobexecutor.jar -u` *Benutzer-ID* `-p` *Kennwort* `-j` *Auftrag* `[Optionale Argumente]`

Erforderlich	Argument	Beschreibung
Nein	`-?`	Gibt Nutzungsinformationen aus.
Nein	`-d` *delimiter*	Legt das Instanz/Status-Begrenzungszeichen fest. Dieses erscheint nur in synchroner Ausgabe.
Nein	`-e`	Verwendet eine sichere HTTPS-Verbindung für die Kommunikation mit dem Spectrum™ Technology Platform-Server.
Nein	`-f` *property file*	Gibt einen Pfad zu einer Auftragseigenschaftsdatei an. Eine

Erforderlich	Argument	Beschreibung
		Auftragseigenschaftsdatei enthält Job Executor-Argumente. Weitere Informationen zu Auftragseigenschaftsdateien finden Sie unter **Verwenden einer Auftragseigenschaftsdatei** auf Seite 162.
Nein	`-h` *host name*	Gibt Namen oder die IP-Adresse des Spectrum™ Technology Platform-Servers an.
Nein	`-i` *poll interval*	Gibt an, wie häufig auf abgeschlossene Aufträge geprüft werden sollen (in Sekunden). Gilt nur im synchronen Modus.
Ja	`-j` *job name*	Eine durch Kommas getrennte Liste von Aufträgen, die ausgeführt werden sollen. Bei Auftragsnamen wird die Groß-/Kleinschreibung beachtet. Aufträge werden in der aufgelisteten Reihenfolge gestartet.
Nein	`-n` *email list*	Gibt eine durch Kommas getrennte Liste zusätzlicher E-Mail-Adressen für konfigurierte Auftragsbenachrichtigungen an.

Nein	`-o` *property file*	Gibt einen Pfad zu einer Datenflussoptions-Eigenschaftsdatei an. Verwenden Sie eine Datenflussoptions-Eigenschaftsdatei, um Optionen für Schritte im Datenfluss festzulegen. Um Datenflussoptionen über eine Eigenschaftsdateien festzulegen, müssen Sie den Datenfluss konfigurieren, Schrittoptionen zur Laufzeit verfügbar zu machen. Weitere Informationen finden Sie unter **Hinzufügen von Datenfluss-Laufzeitoptionen** auf Seite 181.
		
		So kann etwa eine Datenflussoptions-Eigenschaftsdatei für einen Datenfluss, der einen „Assign GeoTAX Info“-Schritt enthält, wie folgt aussehen:
		
		`OutputCasing=U`
		`UseStreetLevelMatching=N`
		`TaxKey=T`
		`Database.GTX=gsl`

Ja	`-p` *password*	Das Kennwort des Benutzers.

Erforderlich	Argument	Beschreibung
Nein	`-r`	Geben Sie dieses Argument an, um einen detaillierten Bericht über den Auftrag zurückzugeben. Diese Option funktioniert nur, wenn Sie auch `-w` angeben. Der Bericht enthält die folgenden Informationen: • **Position 1** – Name des Auftrags • **Position 2** – Prozess-ID des Auftrags • **Position 3** – Status • **Position 4** – Startdatum/-uhrzeit (MM/TT/JJJJ HH:MM:SS) • **Position 5** – Enddatum/-uhrzeit (MM/TT/JJJJ HH:MM:SS) • **Position 6** – Anzahl der erfolgreichen Datensätze • **Position 7** – Anzahl der fehlgeschlagenen Datensätze • **Position 8** – Anzahl der falsch formatierten Datensätze • **Position 9** – derzeit nicht verwendet Beispiel: `MySimpleJob|4|succeeded|04/09/2010 14:50:47|04/09/2010 14:50:47|100|0|0|` Die Informationen sind das durch Trennzeichen getrennt, das im Argument `-d` angegeben ist.
Nein	`-s` *port*	Der Socket (Port), an dem der Spectrum™ Technology Platform-Server ausgeführt wird. Der Standardwert ist 8080.
Nein	`-t` *timeout*	Legt den Timeout (in Sekunden) für den synchronen Modus fest. Der Standardwert ist 3600. Der Maximalwert ist 2147483. Dies ist ein globaler, aggregierter Timeout und stellt die maximale Zeit dar, die gewartet wird, bis alle erzeugten Aufträge abgeschlossen sind.
Ja	`-u` *user name*	Der Anmeldename des Benutzers.
Nein	`-v`	Gibt eine ausführliche Ausgabe zurück.
Nein	`-w`	Führt Job Executor im synchronen Modus aus. Das bedeutet, dass Job Executor weiter

Erforderlich	Argument	Beschreibung
		ausgeführt wird, bis der Auftrag abgeschlossen ist. Wenn Sie `-w` nicht angeben, wird Job Executor nach Start des Auftrags beendet, wenn der Auftrag keine Dateien vom Server liest oder nicht in Dateien auf dem Server schreibt. In diesem Fall wird Job Executor weiter ausgeführt, bis alle lokalen Dateien verarbeitet wurden, und danach beendet.
Nein	*StageName*=*Protocol*:*FileName*	Überschreibt die Eingabe- oder Ausgabedatei, die in „Read from File“ oder „Write to File“ angegeben ist. Weitere Informationen finden Sie unter **Überschreiben von Auftragsdateispeicherorten** auf Seite 157.
Nein	*StageName*:schema=*Protocol*:*SchemaFile*	Überschreibt die Dateilayoutdefinition, die in „Read from File“ oder „Write to File“ angegeben ist, durch eine in der Schemadatei definierte. Weitere Informationen finden Sie unter **Überschreiben des Dateiformats in der Befehlszeile** auf Seite 160.

Anwendungsbeispiel für Job Executor

Das folgende Beispiel zeigt Befehlszeilenaufruf und -ausgabe:

```
D:\spectrum\job-executor>java -jar jobexecutor.jar -u user123
 -p "MeinKennwort" -j validateAddressJob1 -h
spectrum.beispiel.com -s 8888 -w -d "%" -i 1 -t 9999

validateAddressJob1%105%succeeded
```

In diesem Beispiel gibt die Ausgabe an, dass der Auftrag mit Namen „validateAddressJob1“ (mit Bezeichner 105) ohne Fehler ausgeführt wurde. Andere mögliche Ergebnisse sind „fehlgeschlagen“ oder „wird ausgeführt“.

Überschreiben von Auftragsdateispeicherorten

Wenn Sie einen Auftrag in der Befehlszeile mit Job Executor oder mit der Administrationsumgebung ausführen, können Sie die im Quellschritt des Datenflusses (wie „Read from File“) angegebene Eingabedatei sowie die in der Datenladung des Datenflusses (wie „Write to File“) angegebene Ausgabedatei überschreiben.

Um dies im Job Executor zu tun, geben Sie Folgendes am Ende des Befehlszeilenbefehls des Job Executor an:

StageName=Protocol:Dateiname

Verwenden Sie in der Administrationsumgebung das Argument `--l` im Befehl `job execute`:

`--l` *StageName=Protocol:Dateiname*

Dabei gilt Folgendes:

StageName

Die Schrittbeschriftung unter dem Symbol des Schrittes im Datenfluss im Enterprise Designer. Hat der Schritt beispielsweise die Beschriftung „Read from File“, würden Sie `Read from File` als Schrittnamen angeben.

Um einen Schritt innerhalb eines eingebetteten Datenflusses oder Unterflusses anzugeben, stellen Sie dem Schrittnamen als Präfix den Namen des eingebetteten Datenflusses oder Unterflusses voran, gefolgt von einem Punkt und schließlich dem Schrittnamen:

EmbeddedOrSubflowName.StageName

Beispiel: Um einen Schritt mit Namen „Write to File“ in einem Unterfluss mit Namen „Unterfluss1“ anzugeben, würden Sie Folgendes angeben:

```
Unterfluss1.Write to File
```

Um einen Schritt in einem eingebetteten Datenfluss anzugeben, der sich innerhalb eines anderen eingebetteten Datenflusses befindet, fügen Sie den übergeordneten Datenfluss hinzu und verwenden Sie jeweils einen Punkt als Trennzeichen. Beispiel: Wenn sich „Eingebetteter Datenfluss 2“ innerhalb von „Eingebetteter Datenfluss 1“ befindet und Sie den „Write to File“-Schritt im „Eingebetteten Datenfluss 2“ angeben möchten, würden Sie Folgendes angeben:

```
Eingebetteter Datenfluss 1.Eingebetteter Datenfluss 2.Write to
File
```

Protokoll

Ein Kommunikationsprotokoll. Zur Auswahl stehen:

file Verwendet das Protokoll file, wenn die Datei auf demselben Computer ist wie der Spectrum™ Technology Platform-Server. Geben Sie beispielsweise unter Windows Folgendes an:

`„file:C:/MeineDatei.txt“`

Geben Sie unter UNIX oder Linux Folgendes an:

`„file:/testfiles/MeineDatei.txt“`

esclient Verwendet das Protokoll „esclient“, wenn sich die Datei auf dem Computer befindet, auf dem Sie den Auftrag ausführen, wenn es sich dabei um einen anderen Computer handelt als den, auf dem der Spectrum™ Technology Platform-Server ausgeführt wird. Verwenden Sie das folgende Format:

`esclient:`*`Computername/Pad zur Datei`*

Beispiel:

```
esclient:MeinComputer/Testdateien/MeineDatei.txt
```

Anmerkung: Wenn Sie den Auftrag auf dem Server selbst ausführen, können Sie entweder das Protokoll file oder esclient verwenden, erzielen aber wahrscheinlich mit file eine bessere Leistung.

Falls der Hostname des Spectrum™ Technology Platform-Servers nicht aufgelöst werden kann, können Sie den Fehler „Fehler beim Zugriff auf Datei“ erhalten. Um dieses Problem zu beheben, öffnen Sie diese Datei auf dem Server: *SpectrumLocation*`/server/app/conf/spectrum-container.properties`. Legen Sie für die Eigenschaft `spectrum.runtime.hostname` die IP-Adresse des Servers fest.

esfile Verwenden Sie das Protokoll „esfile“, wenn sich die Datei auf einem Dateiserver befindet. Der Dateiserver muss in der Management Console als Ressource definiert sein. Verwenden Sie das folgende Format:

```
esfile://Dateiserver/Pfad zur Datei
```

Beispiel:

```
esfile://MeinServer/Testdateien/MeineDatei.txt
```

Dabei ist „MeinServer“ eine in der Management Console definierte FTP-Dateiserverressource.

webhdfs Verwenden Sie das Protokoll „webhdfs“, wenn sich die Datei auf einem Hadoop Distributed File Server befindet. Der HDFS-Dateiserver muss in der Management Console als Ressource definiert sein. Verwenden Sie das folgende Format:

```
webhdfs://Dateiserver/Pfad zur Datei
```

Beispiel:

```
webhdfs://MeinServer/Testdateien/MeineDatei.txt
```

Dabei ist „MeinServer“ eine in der Management Console definierte HDFS-Dateiserverressource.

Dateiname

Der vollständige Pfad zu der Datei, die Sie als Eingabe oder Ausgabe verwenden möchten.

Anmerkung: Sie müssen in Dateipfaden Schrägstriche (/) angeben, keine umgekehrten Schrägstriche.

Um mehrere Überschreibungen anzugeben, trennen Sie jede Überschreibung durch ein Komma.

Beispiel für das Überschreiben einer Datei

Der folgende Job Executor-Befehl würde die Datei `C:/myfile_input.txt` als Eingabedatei für den Schritt „Read from File“ verwenden und die Datei `C:/myfile_output.txt` als Ausgabedatei für den Schritt „Write to File“.

```
java -jar jobexecutor.jar -j Job1 -u Bob1234 -p "" "Read from
File"="file:C:/myfile_input.txt" "Write to
File"="file:C:/myfile_output.txt"
```

Überschreiben des Dateiformats in der Befehlszeile

Wenn Sie einen Auftrag mit dem Job Executor oder der Administrationsumgebung ausführen, können Sie das Dateilayout (bzw. das Schema) der Datei überschreiben, die im „Read from File“- und „Write to File“-Schritt des Datenflusses angegeben ist.

Um dies im Job Executor zu tun, geben Sie Folgendes am Ende des Befehlszeilenbefehls des Job Executor an:

StageName`:schema=`*Protocol*`:`*Schemadatei*

Verwenden Sie in der Administrationsumgebung das Argument `--l` im Befehl `job execute`:

`--l`*StageName*`:schema=`*Protocol*`:`*Schemadatei*

Dabei gilt Folgendes:

StageName

Die Schrittbeschriftung unter dem Symbol des Schrittes im Datenfluss im Enterprise Designer. Hat der Schritt beispielsweise die Beschriftung „Read from File“, würden Sie `Read from File` als Schrittnamen angeben.

Um einen Schritt innerhalb eines eingebetteten Datenflusses oder Unterflusses anzugeben, stellen Sie dem Schrittnamen als Präfix den Namen des eingebetteten Datenflusses oder Unterflusses voran, gefolgt von einem Punkt und schließlich dem Schrittnamen:

EmbeddedOrSubflowName`.`*StageName*

Beispiel: Um einen Schritt mit Namen „Write to File“ in einem Unterfluss mit Namen „Unterfluss1“ anzugeben, würden Sie Folgendes angeben:

```
Unterfluss1.Write to File
```

Um einen Schritt in einem eingebetteten Datenfluss anzugeben, der sich innerhalb eines anderen eingebetteten Datenflusses befindet, fügen Sie den übergeordneten Datenfluss hinzu und verwenden Sie jeweils einen Punkt als Trennzeichen. Beispiel: Wenn sich „Eingebetteter Datenfluss 2“ innerhalb von „Eingebetteter Datenfluss 1“ befindet und Sie den „Write to File“-Schritt im „Eingebetteten Datenfluss 2“ angeben möchten, würden Sie Folgendes angeben:

```
Eingebetteter Datenfluss 1.Eingebetteter Datenfluss 2.Write to
File
```

Protokoll

Ein Kommunikationsprotokoll. Zur Auswahl stehen:

file Verwendet das Protokoll file, wenn die Datei auf demselben Computer ist wie der Spectrum™ Technology Platform-Server. Geben Sie beispielsweise unter Windows Folgendes an:

```
„file:C:/MeineDatei.txt"
```

Geben Sie unter UNIX oder Linux Folgendes an:

```
„file:/testfiles/MeineDatei.txt"
```

esclient Verwendet das Protokoll „esclient", wenn sich die Datei auf dem Computer befindet, auf dem Sie den Auftrag ausführen, wenn es sich dabei um einen anderen Computer handelt als den, auf dem der Spectrum™ Technology Platform-Server ausgeführt wird. Verwenden Sie das folgende Format:

```
esclient:Computername/Pad zur Datei
```

Beispiel:

```
esclient:MeinComputer/Testdateien/MeineDatei.txt
```

> **Anmerkung:** Wenn Sie den Auftrag auf dem Server selbst ausführen, können Sie entweder das Protokoll file oder esclient verwenden, erzielen aber wahrscheinlich mit file eine bessere Leistung.

Falls der Hostname des Spectrum™ Technology Platform-Servers nicht aufgelöst werden kann, können Sie den Fehler „Fehler beim Zugriff auf Datei" erhalten. Um dieses Problem zu beheben, öffnen Sie diese Datei auf dem Server: `SpectrumLocation/server/app/conf/spectrum-container.properties`. Legen Sie für die Eigenschaft `spectrum.runtime.hostname` die IP-Adresse des Servers fest.

esfile Verwenden Sie das Protokoll „esfile", wenn sich die Datei auf einem Dateiserver befindet. Der Dateiserver muss in der Management Console als Ressource definiert sein. Verwenden Sie das folgende Format:

```
esfile://Dateiserver/Pfad zur Datei
```

Beispiel:

```
esfile://MeinServer/Testdateien/MeineDatei.txt
```

Dabei ist „MeinServer" eine in der Management Console definierte FTP-Dateiserverressource.

webhdfs Verwenden Sie das Protokoll „webhdfs", wenn sich die Datei auf einem Hadoop Distributed File Server befindet. Der HDFS-Dateiserver muss in der Management Console als Ressource definiert sein. Verwenden Sie das folgende Format:

```
webhdfs://Dateiserver/Pfad zur Datei
```

Beispiel:

```
webhdfs://MeinServer/Testdateien/MeineDatei.txt
```

Dabei ist „MeinServer“ eine in der Management Console definierte HDFS-Dateiserverressource.

SchemaFile

Der vollständige Pfad zu der Datei, die das gewünschte Layout definiert.

Anmerkung: Sie müssen in Dateipfaden Schrägstriche (/) angeben, keine umgekehrten Schrägstriche.

Um eine Schemadatei zu erstellen, definieren Sie das in „Read from File“ oder „Write to File“ gewünschte Layout und klicken Sie dann auf die Schaltfläche **Exportieren**, um eine XML-Datei zu erstellen, die das Layout definiert.

Anmerkung: Sie dürfen den Datentyp eines Feldes in einer Schemadatei nicht überschreiben, wenn Sie den Job Executor verwenden. Der Wert im Element <Typ>, das ein untergeordnetes Element des Elementes <Feldschema> ist, muss dem im „Read from File“- oder „Write to File“-Schritt des Datenflusses angegebenen Typ des Feldes entsprechen.

Beispiel für das Überschreiben eines Dateiformats

Der folgende Job Executor-Befehl würde die Datei `C:/myschema.xml` als Layoutdefinition für die durch den „Read from File“-Schritt gelesene Datei verwenden.

```
java -jar jobexecutor.jar -j Job1 -u Bob1234 -p "" "Read from File":schema="file:C:/myschema.xml"
```

Verwenden einer Auftragseigenschaftsdatei

Eine Auftragseigenschaftsdatei enthält Argumente, die die Ausführung von Aufträgen steuern, wenn Sie Job Executor oder die Administrationsumgebung verwenden, um einen Auftrag auszuführen. Verwenden Sie eine Auftragseigenschaftsdatei, wenn Sie Argumente wiederverwenden möchten, indem Sie ein einzelnes Argument in der Befehlszeile angeben (`-f`), statt jedes Argument einzeln in der Befehlszeile anzugeben.

Um eine Eigenschaftsdatei zu erstellen, erstellen Sie eine Textdatei mit einem Argument pro Zeile. Beispiel:

```
d %
h spectrum.mydomain.com
i 30
j validateAddressJob1
u user
p password
s 8888
```

```
t 9999
w true
```

Die Auftragseigenschaftsdatei kann diese Argumente enthalten:

Erforderlich	Argument	Beschreibung
Nein	`?`	Gibt Nutzungsinformationen aus.
Nein	`d` *delimiter*	Legt das Instanz/Status-Begrenzungszeichen fest. Dieses erscheint nur in synchroner Ausgabe.
Nein	`e`	Verwendet eine sichere HTTPS-Verbindung für die Kommunikation mit dem Spectrum™ Technology Platform-Server.
Nein	`h` *hostname*	Gibt Namen oder die IP-Adresse des Spectrum™ Technology Platform-Servers an.
Nein	`i` *pollinterval*	Gibt an, wie häufig auf abgeschlossene Aufträge geprüft werden sollen (in Sekunden). Gilt nur im synchronen Modus.
Ja	`j` *jobname*	Eine durch Kommas getrennte Liste von Aufträgen, die ausgeführt werden sollen. Bei Auftragsnamen wird die Groß-/Kleinschreibung beachtet. Aufträge werden in der aufgelisteten Reihenfolge gestartet.
Nein	`n` *emaillist*	Gibt eine durch Kommas getrennte Liste zusätzlicher E-Mail-Adressen für konfigurierte Auftragsbenachrichtigungen an.
Ja	`p` *password*	Das Kennwort des Benutzers.
Nein	`r`	Gibt eine durch Trennzeichen getrennt Liste mit den folgenden Informationen über den zur Standardausgabe geschriebenen Auftrag zurück: • **Position 1** – Name des Auftrags • **Position 2** – Prozess-ID des Auftrags • **Position 3** – Status • **Position 4** – Startdatum/-uhrzeit (MM/TT/JJJJ HH:MM:SS) • **Position 5** – Enddatum/-uhrzeit (MM/TT/JJJJ HH:MM:SS) • **Position 6** – Anzahl der erfolgreichen Datensätze • **Position 7** – Anzahl der fehlgeschlagenen Datensätze • **Position 8** – Anzahl der falsch formatierten Datensätze • **Position 9** – derzeit nicht verwendet Die Informationen sind das durch Trennzeichen getrennt, das im Argument „-d“ angegeben ist. Beispiel: `MySimpleJob|4|succeeded|04/09/2010 14:50:47|04/09/2010 14:50:47|100|0|0|`
Nein	`s` *port*	Der Socket (Port), an dem der Spectrum™ Technology Platform-Server ausgeführt wird. Der Standardwert ist 8080.

Erforderlich	Argument	Beschreibung
Nein	`t` *timeout*	Legt den Timeout (in Sekunden) für den synchronen Modus fest. Der Standardwert ist 3600. Der Maximalwert ist 2147483. Dies ist ein globaler, aggregierter Timeout und stellt die maximale Zeit dar, die gewartet wird, bis alle erzeugten Aufträge abgeschlossen sind.
Ja	`u` *username*	Der Anmeldename des Benutzers.
Nein	`v`	Gibt eine ausführliche Ausgabe zurück.
Nein	`w`	Gibt an, auf den Abschluss von Aufträgen in einem synchronen Modus zu warten.

Sowohl Befehlszeilenargumente als auch Eigenschaftsdatei verwenden

Eine Kombination des Befehlszeileneintrags mit dem Eigenschaftsdateieintrag ist ebenfalls gültig. Beispiel:

```
java -jar jobexecutor.jar -f /dcg/job.properties -j job1
```

In diesem Fall haben Befehlszeilenparameter vor den in der Eigenschaftsdatei angegebenen Argumenten Vorrang. Im obigen Beispiel hätte der Auftrag „job1“ Vorrang vor einem in der Eigenschaftsdatei angegebenen Auftrag.

Ausführen eines Prozessflusses über die Befehlszeile

Verwenden Sie den Process Flow Executor für die Ausführung eines Prozesses über die Befehlszeile. Sie können den Process Flow Executor über die Spectrum™ Technology Platform-Begrüßungsseite (zum Beispiel http://myserver:8080) installieren.

> **Anmerkung:** Zudem können Sie die Administrationsumgebung verwenden, um Prozessflüsse über die Befehlszeile auszuführen.

Verwendung

`java -jar pflowexecutor.jar -r`*Prozessflussname* `-u` *Benutzer-ID* `-p` *Kennwort* `[Optionale Argumente]`

Erforderlich	Argument	Beschreibung
Nein	`-?`	Gibt Nutzungsinformationen aus.
Nein	`-d` *DelimiterCharacter*	Legt ein Begrenzungszeichen fest, mit dem in der Befehlszeile angezeigte Statusinformationen voneinander getrennt werden sollen, wenn Sie den Befehl ausführen. Standardmäßig wird das Begrenzungszeichen „|“ verwendet. Unter Verwendung des Standardzeichens würde bei der Ausführung eines Prozessflusses

Erforderlich	Argument	Beschreibung
		mit dem Namen „MyProcessflow“ beispielsweise Folgendes in der Befehlszeile angezeigt: `MyProcessflow|1|Succeeded`
Nein	`-e`	Verwenden Sie eine HTTPS-Verbindung für die Kommunikation mit dem Spectrum™ Technology Platform-Server. **Anmerkung:** Wenn Sie eine Datei angeben, die dieses Argument überschreibt, darf dies nicht das zuletzt angegebene Argument sein.
Nein	`-f` *PropertyFile*	Gibt einen Pfad zu einer Eigenschaftsdatei an. Weitere Informationen zu Eigenschaftsdateien finden Sie unter **Verwenden der Eigenschaftsdatei eines Prozessflusses** auf Seite 166.
Nein	`-h` *HostName*	Gibt Namen oder die IP-Adresse des Spectrum™ Technology Platform-Servers an.
Nein	`-i` *PollInterval*	Gibt an, wie häufig auf abgeschlossene Aufträge geprüft werden sollen (in Sekunden). Der Standardwert ist „5“.
Ja	`-p` *Password*	Das Kennwort des Benutzers. Erforderlich.
Ja	`-r` *ProcessFlowNames*	Eine durch Kommas getrennte Liste von Prozessflüssen, die ausgeführt werden sollen. Erforderlich. **Anmerkung:** Wenn Sie eine Datei angeben, die dieses Argument überschreibt, darf dies nicht das zuletzt angegebene Argument sein.
Nein	`-s` *Port*	Der Socket (Port), an dem der Spectrum™ Technology Platform-Server ausgeführt wird. Der Standardwert ist 8080.
Nein	`-t` *Timeout*	Diese Option ist veraltet und wird ignoriert.
Ja	`-u` *UserName*	Der Anmeldename des Benutzers. Erforderlich.
Nein	`-v` *Verbose*	Gibt eine ausführliche Ausgabe zurück, wobei *Ausführlich* für einen der folgenden Werte steht: **true** Gibt eine ausführliche Ausgabe zurück. **false** Gibt keine ausführliche Ausgabe zurück. **Anmerkung:** Wenn Sie eine Datei angeben, die dieses Argument überschreibt, darf dies nicht das zuletzt angegebene Argument sein.
Nein	`-w` *WaitToComplete*	Diese Option ist veraltet und wird ignoriert.

Erforderlich	Argument	Beschreibung
Nein	*StageName=FileName*	Überschreibt die im Auftrag angegebene Eingabe- oder Ausgabedatei. Weitere Informationen finden Sie unter **Überschreiben von Dateispeicherorten von Prozessflüssen**.

Beispiele

Im Folgenden wird ein Standardeintrag in einer Befehlszeile dargestellt, mit Prozessflussnamen, Benutzer-ID und Kennwort:

```
java -jar pflowexecutor.jar -r MyFlow1 -u Bob1234 -p
"mypassword1"
```

In dem folgenden Beispiel werden die gleichen Informationen wie oben dargestellt, jedoch mit zusätzlichen Argumenten:

```
java -jar pflowexecutor.jar -r Flow1 -u Bob1234 -p
"mypassword1" -h spectrum.example.com -s 8888 -w -d "%" -i
1 -t 9999
```

In dem folgenden Beispiel werden der Aufruf und die Ausgabe der Befehlszeile dargestellt.

```
D:\spectrum\pflow-executor>java -jar pflowexecutor.jar -u
Bob1234 -p "mypassword1" -r
validateAddressFlow1 -h spectrum.example.com -s 8888 -w -d
"%" -i
1 -t 9999
validateAddressJob1%111%succeeded
```

In diesem Beispiel wurde der Prozessfluss mit dem Namen „validateAddressFlow1“ ausgeführt (mit dem Bezeichner 111). Es sind keine Fehler aufgetreten. Andere mögliche Ergebnisse sind „fehlgeschlagen“ oder „wird ausgeführt“.

Verwenden der Eigenschaftsdatei eines Prozessflusses

Eine Eigenschaftsdatei enthält Argumente, die Sie bei der Angabe des Pfades zur Eigenschaftsdatei mit dem Argument `-f` im Process Flow Executor wiederverwenden können. Die Eigenschaftsdatei muss mindestens den Prozessfluss (`r`), die Benutzer-ID (`u`) und das Kennwort (`p`) enthalten.

1. Öffnen Sie einen Texteditor.
2. Geben Sie in jeder Zeile ein Argument an, wie im folgenden Beispiel dargestellt. Eine Liste der Argumente finden Sie unter **Ausführen eines Prozessflusses über die Befehlszeile** auf Seite 164.

Anmerkung: Sie können mit einer Eigenschaftsdatei keine Eingabe- oder Ausgabedateien überschreiben. Eingabe- und Ausgabedateien können nur mit Befehlszeilenargumenten überschrieben werden.

```
d=%
h=myserver.mydomain.com
i=30
u=user
p=password
r=MyFlow1
s=8888
```

3. Speichern Sie die Datei mit der Dateierweiterung .properties (zum Beispiel „Beispiel.properties“).
4. Geben Sie bei der Ausführung des Process Flow Executor den Pfad zur Eigenschaftsdatei mit dem Argument `-f` an. Eine Kombination des Befehlszeileneintrags mit dem Eigenschaftsdateieintrag ist ebenfalls gültig. Befehlszeilenargumente haben Vorrang vor den in der Eigenschaftsdatei angegebenen Argumenten.
   ```
   java -jar pflowexecutor.jar -f /dcg/flow.properties -r MyFlow2
   ```
   Im obigen Beispiel hätte der Prozessfluss „MyFlow2“ Vorrang vor einem in der Eigenschaftsdatei angegebenen Prozessfluss.

Planen eines Datenflusses

Ein Datenflusszeitplan führt einen Auftrag oder einen Prozessfluss zu einer bestimmten Zeit automatisch aus. Sie können planen, dass ein Datenfluss einmal ausgeführt werden soll, oder eine wiederkehrende Ausführung einrichten.

1. Machen Sie den Auftrag oder den Prozessfluss verfügbar, wenn dies nicht bereits geschehen ist.

   Sie können Aufträge und Prozessflüsse verfügbar machen, indem Sie den Auftrag oder den Prozessfluss im Enterprise Designer öffnen und **Datei** > **Verfügbar machen/Verfügbarkeit aufheben und speichern** auswählen.

2. Öffnen Sie die Management Console.
3. Öffnen Sie **Ausführung**, und klicken Sie anschließend auf **Planung**.
4. Klicken Sie auf **Hinzufügen**, um einen neuen Zeitplan zu erstellen. Wenn Sie einen vorhandenen Zeitplan ändern möchten, wählen Sie den Zeitplan aus und klicken Sie auf **Ändern**.
5. Wählen Sie im Fenster „**Aufgabe hinzufügen**“ oder „**Aufgabe ändern**“ die Einstellungen für diese Aufgabe aus.
   - **Aufgabenname** – Der Name, den Sie dieser geplanten Aufgabe geben möchten. Dies ist der Name, der in der Auflistung der Aufgaben angezeigt wird.

- **Flusstyp** – Wählen Sie den Prozesstyp aus, den Sie planen: einen Auftrag oder einen Prozessfluss.
- **Flussname** – Wählen Sie den Auftrag oder den Prozessfluss aus, den Sie planen möchten. Hier sind nur Aufträge und Prozessflüsse verfügbar, die gespeichert und verfügbar gemacht wurden. Wenn der gewünschte Auftrag oder Prozessfluss nicht angezeigt wird, müssen Sie den Auftrag oder Prozessfluss im Enterprise Designer öffnen und anschließend **Datei** > **Verfügbar machen/Verfügbarkeit aufheben und speichern** auswählen.
- **Aufgabe aktivieren** – Aktivieren Sie dieses Kästchen, um den Auftrag oder den Prozessfluss zur angegebenen Zeit auszuführen. Deaktivieren Sie dieses Kästchen, um den Zeitplan auszusetzen.
- **Zeitplan** – Geben Sie das Datum und die Uhrzeit an, an dem bzw. zu der der Auftrag oder der Prozessfluss ausgeführt werden soll.

6. Wenn der Fluss Dateien für die Eingabe oder Ausgabe verwendet, müssen sich diese Dateien auf dem Spectrum™ Technology Platform-Server oder auf einem Dateiserver befinden, der in der Management Console als externe Ressource definiert ist. Dies gilt sowohl für Aufträge als auch für Auftragsaktivitäten innerhalb eines Prozessflusses. Gehen Sie wie folgt vor, wenn in einem Quell- oder Datenladeschritt auf eine Datei auf einem Client-Computer verwiesen wird:

Option	Bezeichnung
Option 1: Ändern Sie den Datenfluss	Verschieben Sie die Datei auf den Spectrum™ Technology Platform-Server oder auf einen Dateiserver und ändern Sie den Datenfluss anschließend: 1. Öffnen Sie den Fluss im Enterprise Designer. 2. Doppelklicken Sie auf den Quell- oder Datenladeschritt. 3. Klicken Sie im Feld **Dateiname** auf die Schaltfläche „Durchsuchen“. 4. Klicken Sie auf **Remote-Computer** und wählen Sie anschließend die gewünschte Datei aus. **Anmerkung:** Wenn Sie den Enterprise Designer auf demselben Computer ausführen wie den Spectrum™ Technology Platform-Server, macht es den Anschein, dass das Klicken auf „Remote-Computer“ und auf „Arbeitsplatz“ keinen Unterschied ergibt. Sie müssen die Datei jedoch über den Remote-Computer auswählen, damit das System die Datei als Datei des Spectrum™ Technology Platform-Servers erkennt.
Option 2: Überschreiben Sie	Sie können die im Quellschritt des Datenflusses (z. B. „Read from File“) angegebene Eingabedatei und die im Datenladeschritt des

Option	Bezeichnung
den Speicherort der Datenflussdatei bei der Ausführung dieses Zeitplans	Datenflusses (z. B. „Write to File“) angegebene Ausgabedatei überschreiben. 1. Klicken Sie auf **Optionen**. 2. Wählen Sie unter **Schrittdatei-Speicherorte** den Schritt aus, in dem auf eine lokale Datei verwiesen wird. 3. Klicken Sie auf **Ändern** und wählen Sie die Datei auf dem Spectrum™ Technology Platform-Server aus.

7. Wenn der Auftrag oder der Prozessfluss in einem wiederkehrenden Zeitplan ausgeführt werden soll, aktivieren Sie das Kästchen **Aufgabenwiederholung**, klicken Sie anschließend auf die Schaltfläche **Wiederholung** und füllen Sie die Felder aus.
8. Wenn der Datenfluss für E-Mail-Benachrichtigungen konfiguriert wurde, können Sie zusätzliche Empfänger für die Benachrichtigungen angeben, die bei der Ausführung des Datenflusses gesendet werden.
   a) Klicken Sie auf **Optionen**.
   b) Klicken Sie unter „Benachrichtigung“ auf **Hinzufügen**.
   c) Geben Sie die E-Mail-Adresse ein, an die die Benachrichtigung gesendet werden soll. Beispiel: me@mycompany.com.
   d) Klicken Sie auf **OK**.

   Anmerkung: Damit E-Mail-Benachrichtigungen versendet werden können, müssen Benachrichtigungen in der Management Console konfiguriert worden sein. Überprüfen Sie zudem, ob der Datenfluss so konfiguriert wurde, dass er Benachrichtigungen unterstützt. Öffnen Sie dazu den Datenfluss im Enterprise Designer, und wählen Sie **Bearbeiten** > **Benachrichtigungen** aus.
9. Klicken Sie auf **OK**.

Auslösen eines Flusses mit einer Steuerungsdatei

Ein Fluss kann automatisch ausgeführt werden, wenn in einem überwachten Verzeichnis eine Steuerungsdatei erkannt wird. Dieses Feature ist in Situationen hilfreich, in denen ein anderer Prozess abgeschlossen werden muss, bevor der Fluss ausgeführt werden kann. Beispiel: Sie verfügen über einen Fluss, für den in einem anderen Geschäftsprozess eine Eingabedatei generiert werden muss. Sie können den anderen Prozess so einrichten, dass eine Steuerungsdatei in einem Ordner platziert wird, und Spectrum™ Technology Platform so konfigurieren, dass ein Fluss ausgeführt wird, wenn diese Steuerungsdatei angezeigt wird.

Anmerkung: Stellen Sie sicher, dass die Steuerungsdatei erst dann in dem überwachten Ordner platziert wird, wenn alle für den Fluss erforderlichen Dateien vorhanden sind und verarbeitet werden können.

1. Machen Sie den Fluss verfügbar, wenn dies nicht bereits geschehen ist.

   Sie können einen Fluss verfügbar machen, indem Sie ihn im Enterprise Designer öffnen und **Datei** > **Verfügbar machen/Verfügbarkeit aufheben und speichern** auswählen.

2. Öffnen Sie die Management Console.
3. Öffnen Sie **Flüsse** > **Zeitpläne**.
4. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
5. Geben Sie im Feld **Name** den Namen ein, den Sie diesem Zeitplan geben möchten. Dies ist der Name, der in der Auflistung der Zeitpläne angezeigt wird.
6. Geben Sie im Feld **Fluss** den Auftrag oder den Prozessfluss ein, die Sie ausführen möchten. Hier sind nur Aufträge und Prozessflüsse verfügbar, die gespeichert und verfügbar gemacht wurden.
7. Nach der Angabe eines Flusses werden unter dem Feld **Fluss** zusätzliche Felder angezeigt, und zwar ein Feld für jeden Quellschritt des Flusses (z. B. „Read from File“) und jeden Datenladeschritt (z. B. „Write to File“). In diesen Feldern werden die Dateien angezeigt, die bei der Ausführung des Flusses durch diesen Zeitplan verwendet werden. Der Fluss verwendet standardmäßig die in den Flussquellen und -datenladungen angegebenen Dateien. Sie können verschiedene Dateien angeben, die bei der Ausführung dieses Zeitplans verwendet werden sollen, indem Sie den Dateipfad durch den Pfad zu einer anderen Datei ersetzen. Wenn Ihr Fluss beispielsweise den Schritt „Read from File“ enthält, die Daten aus `C:\FlowInput\Customers.csv` einliest, Sie bei der Ausführung dieses Zeitplans aber Daten aus `C:\FlowInput\UpdatedCustomers.csv` verwenden möchten, würden Sie `C:\FlowInput\UpdatedCustomers.csv` im Feld „Read from File“ angeben.

   Anmerkung: Um die Dateien in den Quell- und Datenladeschritten zu ändern, müssen Sie für den gesicherten Entitätstyp **Ressourcen – Dateiserver** über eine Berechtigung zum Lesen verfügen.

Beachten Sie, dass sich die Dateien bei der Auslösung eines Flusses durch einen Zeitplan auf dem Spectrum™ Technology Platform-Server oder auf einem Dateiserver befinden müssen, der in der Management Console als externe Ressource definiert ist. Dies gilt sowohl für Aufträge als auch für Auftragsaktivitäten innerhalb eines Prozessflusses. Gehen Sie wie folgt vor, wenn in einem Quell- oder Datenladeschritt auf eine Datei auf einem Client-Computer verwiesen wird:

Option	Bezeichnung
Option 1: Ändern Sie den Datenfluss	Verschieben Sie die Datei auf den Spectrum™ Technology Platform-Server oder auf einen Dateiserver und ändern Sie den Datenfluss anschließend: 1. Öffnen Sie den Datenfluss im Enterprise Designer.

Option	Bezeichnung
	2. Doppelklicken Sie auf den Quell- oder Datenladeschritt. 3. Klicken Sie im Feld **Dateiname** auf die Schaltfläche „Durchsuchen“. 4. Klicken Sie auf **Remote-Computer** und wählen Sie anschließend die gewünschte Datei aus. **Anmerkung:** Wenn Sie den Enterprise Designer auf demselben Computer ausführen wie den Spectrum™ Technology Platform-Server, macht es den Anschein, dass das Klicken auf „Remote-Computer“ und auf „Arbeitsplatz“ keinen Unterschied ergibt. Sie müssen die Datei jedoch über den Remote-Computer auswählen, damit das System die Datei als Datei des Spectrum™ Technology Platform-Servers erkennt.
Option 2: Überschreiben Sie den Speicherort der Datenflussdatei bei der Ausführung dieses Zeitplans	Sie können die Dateiverweise in dem Fluss überschreiben, wenn dieser Zeitplan ausgeführt wird. Ersetzen Sie dazu die Standarddatei, die in den einzelnen Quell- und Datenladungsfeldern angezeigt wird, durch den Pfad zu einer Datei auf dem Spectrum™ Technology Platform-Server oder einer in der Management Console definierten Dateiserverressource.

8. Wählen Sie im Feld **Auslöser** die Option **Steuerungsdatei** aus.
9. Geben Sie im Feld **Steuerungsdatei** den vollständigen Pfad und den Namen der Steuerungsdatei an, die den Fluss auslöst. Sie können entweder einen genauen Dateinamen angeben oder das Sternchen (*) als Platzhalter verwenden. `*.trg` würde beispielsweise den Fluss auslösen, wenn eine Datei mit einer `.trg`-Erweiterung im Ordner erscheint.

   Die Anwesenheit einer Steuerungsdatei gibt an, dass alle für den Fluss erforderlichen Dateien vorhanden sind und im Fluss verwendet werden können.

   Bei der Steuerungsdatei kann es sich um eine leere Datei handeln. Bei Aufträgen kann die Steuerungsdatei Überschreibungen von Dateipfaden angeben, die in den Schritten „Write to File“ oder „Read from File“ konfiguriert wurden. Um mithilfe einer Steuerungsdatei die Dateipfade zu überschreiben, geben Sie den Namen des „Read from File“- oder „Write from File“-Schrittes zusammen mit der Eingabe- oder Ausgabedatei als letzte Argumente wie folgt an:

```
stagename=filename
```

Beispiel:

```
Read\ from\ File=file:C:/myfile_input.txt
Write\ to\ File=file:C:/myfile_output.txt
```

Der in der Steuerungsdatei angegebene Schrittname muss mit der Schrittbeschriftung übereinstimmen, die unter dem Symbol des Schrittes im Datenfluss angezeigt wird. Wenn der Eingabeschritt beispielsweise mit „Read From File“ beschriftet ist, würden Sie Folgendes angeben:

```
Read\ From\ File=file:C:/inputfile.txt
```

Wenn der Eingabeschritt mit „Illinois Customers“ beschriftet ist, würden Sie Folgendes angeben:

```
Illinois\ Customers=file:C:/inputfile.txt
```

Stellen Sie beim Überschreiben eines Speicherorts des „Read from File“- oder „Write to File“-Schrittes sicher, dass folgende Richtlinien befolgt werden:

- Starten Sie den Pfad mit dem Protokoll „file:“. Geben Sie beispielsweise unter Windows „`file:C:/myfile.txt`“ und unter Unix oder Linux „`"file:/testfiles/myfile.txt"`“ an.
- Bei den Inhalten der Datei muss eine ASCII-basierte Zeichencodierung verwendet werden, die mit ISO-8559-1 (Latin-1) kompatibel ist.
- Sie müssen in Dateipfaden Schrägstriche (/) angeben, keine umgekehrten Schrägstriche.
- Leerzeichen in Schrittnamen müssen mit einem umgekehrten Schrägstrich geschützt werden.
- Bei Schrittnamen muss die Groß-/Kleinschreibung beachtet werden.

  Anmerkung: Wenn in mehreren Zeitplänen ein Steuerungsdateiauslöser verwendet wird, ist es wichtig, dass in den einzelnen Zeitpläne unterschiedliche Steuerungsdateien überwacht werden. Andernfalls könnte dieselbe Steuerungsdatei mehrere Aufträge oder Prozessflüsse auslösen und zu unerwartetem Verhalten führen. Zu Organisationszwecken wird empfohlen, alle erforderlichen Dateien und die Steuerungsdatei in ein dediziertes Verzeichnis einzureihen.

10. Geben Sie im Feld **Abfrageintervall** an, wie häufig geprüft werden soll, ob eine Steuerungsdatei vorhanden ist. Wenn Sie beispielsweise 10 angeben, wird bei der Überwachung alle 10 Sekunden geprüft, ob sich die Steuerungsdatei im überwachten Ordner befindet.

    Der Standardwert beträgt 60 Sekunden.

11. Geben Sie im Feld „**Arbeitsordner**“ einen Ordner an, in dem sich die Steuerungsdatei temporär während der Ausführung des Flusses befindet. Spectrum™ Technology Platform kopiert die Datei vor dem Ausführen des Flusses aus dem überwachten Ordner in den Arbeitsordner. Dadurch wird der überwachte Ordner gelöscht und verhindert, dass der Fluss erneut von derselben Steuerungsdatei gestartet wird.
12. Geben Sie im Feld **Arbeitsordneroptionen** an, was mit den Dateien im Arbeitsordner geschehen soll, wenn die Ausführung des Flusses abgeschlossen ist.

Beibehalten	Die Dateien werden mit ihrem aktuellen Namen an ihrem aktuellen Speicherort beibehalten. Wenn Sie diese Option auswählen, werden die Dateien im Arbeitsordner bei jeder Ausführung des Zeitplans überschrieben.
Verschieben nach	Verschiebt die Dateien aus dem Arbeitsordner in einen von Ihnen angegebenen Ordner. Dadurch können Sie die Dateien, die im Arbeitsordner enthalten waren, beibehalten, indem Sie sie an einen anderen Speicherort verschieben, sodass sie bei der nächsten Ausführung der Dateiüberwachung nicht überschrieben werden. Sie können die Dateien über diese Option auch in einen anderen überwachten Ordner verschieben, um einen nachgelagerten Prozess auszulösen, z. B. einen anderen Datenfluss oder einen anderen Prozess.
Mit Zeitstempel umbenennen	Fügt einen Zeitstempel zu dem Dateinamen im Arbeitsordner hinzu. Dadurch können Sie eine Kopie der Dateien im Arbeitsordner beibehalten, da die umbenannte Datei einen eindeutigen Namen haben wird und folglich nicht überschrieben wird, wenn das nächste Mal bei der Überwachung ein Datenfluss ausgeführt wird.
Löschen	Löscht die Dateien aus dem Arbeitsordner, nachdem die Ausführung des Flusses abgeschlossen wurde.

13. Wenn der Fluss so konfiguriert wurde, dass E-Mail-Benachrichtigungen gesendet werden, können Sie zusätzliche Empfänger für die Benachrichtigungen angeben, die bei der Ausführung dieses Zeitplans gesendet werden. Die hier von Ihnen angegebenen Empfänger erhalten zusätzlich zu den in den Benachrichtigungseinstellungen des Flusses angegebenen Empfängern Benachrichtigungen. Um einen Fluss so zu konfigurieren, dass Benachrichtigungen gesendet werden, müssen Sie den Fluss im Enterprise Designer öffnen und zu **Bearbeiten** > **Benachrichtigungen** navigieren.
14. Klicken Sie auf **Speichern**.

Beispiel: überwachter Ordner und Arbeitsordner

Angenommen, Sie betreiben eine Autowerkstatt. Sie möchten den Kunden des vorherigen Tages jeden Tag per E-Mail einen Gutschein mit einem Rabatt für künftige Leistungen schicken. Hierfür verfügen Sie über einen Datenfluss, der die Liste der Kunden des Tages enthält, sicherstellt, dass die Namen in der richtigen Groß-/Kleinschreibung eingegeben wurden, und die Adressen überprüft. Die Liste der Kunden wird jeden Abend von einem anderen System generiert. Dieses andere System generiert eine Datei mit der Kundenliste. Diese Datei wird dann von Ihnen als Eingabe für den Datenfluss verwendet.

Das System, das die Kundenliste generiert, platziert diese in einem Ordner mit dem Namen `DailyCustomerReport`. Zudem platziert es, wenn es fertig ist, eine leere Auslöserdatei in dem Ordner. Sie konfigurieren Spectrum™ Technology Platform für die Überwachung dieses Ordners und geben dabei folgende Datei als Auslöserdatei an:

`C:\DailyCustomerReport\*.trg`

Dadurch weiß Spectrum™ Technology Platform, dass der Datenfluss immer dann ausgeführt werden soll, wenn eine Datei mit einer .trg-Erweiterung in diesem Ordner angezeigt wird. Sie könnten auch einen bestimmten Dateinamen angeben, in diesem Beispiel verwenden wir jedoch einen Platzhalter.

Wenn im Ordner `DailyCustomerReport` eine .trg-Datei erkannt wird, muss Spectrum™ Technology Platform diese vor der Ausführung des Datenflusses in einen anderen Ordner verschieben. Die Datei muss verschoben werden, da sie andernfalls im nächsten Abfrageintervall erneut erkannt würde und dies dazu führen würde, dass der Datenfluss erneut ausgeführt wird. Die Datei wird also in einen „Arbeitsordner" verschoben, in dem sie während der Ausführung des Datenflusses bleibt. Sie wählen als Arbeitsordner `C:\SpectrumWorkingFolder` aus.

Nachdem der Datenfluss die Kundenliste fertig verarbeitet hat, möchten Sie die Auslöserdatei an einen anderen Speicherort verschieben, an dem sie einen anderen Prozess für die Abfrage auslöst. Sie wählen also die Option **Verschieben nach** aus und wählen einen Ordner mit dem Namen `C:\DailyBilling` aus.

In diesem Beispiel wird die Auslöserdatei in `C:\DailyCustomerReport` gestartet und anschließend in den Arbeitsordner `C:\SpectrumWorkingFolder` verschoben. Nachdem die Ausführung des Datenflusses abgeschlossen ist, wird die Auslöserdatei zum Einleiten des Abrechnungsprozesses nach `C:\DailyBilling` verschoben.

Anzeigen von Flussstatus und -verlauf

Sie können in der Management Console und im Enterprise Designer den Verlauf eines Auftrags, eines Prozessflusses und einer Dienstausführung anzeigen.

In der Management Console

Um den Flussstatus und -verlauf in der Management Console anzuzeigen, öffnen Sie **Flüsse** > **Verlauf**. Auf der Registerkarte **Flüsse** wird der Verlauf von Aufträgen und Prozessflüssen und auf der Registerkarte **Transaktionen** der Verlauf von Diensten angezeigt.

> **Anmerkung:** Beim Flussverlauf spiegelt die Anzahl der Datensätze, die angezeigt wird, wenn Sie den Mauszeiger über die Spalte **Ergebnisse** bewegen, die Gesamtzahl der Datensätze wider, die von allen Datenflussdatenladungen ausgegeben wurden. Diese Anzahl kann sich von der Anzahl der Eingabedatensätze unterscheiden, wenn im Datenfluss Datensätze kombiniert, Datensätze aufgeteilt oder neue Datensätze erstellt werden.

Der Transaktionsverlauf ist standardmäßig deaktiviert, da eine Aktivierung des Transaktionsverlaufs negative Auswirkungen auf die Leistung haben kann. Wenn Sie den Transaktionsverlauf sehen möchten, müssen Sie die Protokollierung des Transaktionsverlaufs aktivieren, indem Sie auf den

Schalter **Transaktionsprotokollierung** klicken. Zum Anzeigen der Benutzeraktivität sollten Sie das Prüfprotokoll verwenden, auf das Sie über **System** > **Protokolle** zugreifen können.

Die Liste des Flussverlaufs wird automatisch alle 30 Sekunden aktualisiert. Wenn sie eher aktualisiert werden soll, klicken Sie auf die Schaltfläche „Aktualisieren“ .

Im Enterprise Designer

Um den Flussstatus und -verlauf im Enterprise Designer anzuzeigen, öffnen Sie **Anzeigen** > **Ausführungsverlauf**.

Die Liste des Flussverlaufs wird automatisch alle 30 Sekunden aktualisiert. Wenn Sie beim Anzeigen des Ausführungsverlaufs eine Verlangsamung feststellen, deaktivieren Sie das Kästchen **Automatische Aktualisierung**.

Über die Registerkarte **Aufträge** wird der Auftragsstatus überwacht. Zudem werden aktive Aufträge angehalten, fortgesetzt oder abgebrochen und abgeschlossene Aufträge gelöscht.

> **Anmerkung:** Die auf der Registerkarte **Aufträge** angezeigte Anzahl der Datensätze spiegelt die Gesamtzahl der Datensätze wider, die von allen Datenflussdatenladungen ausgegeben wurden. Diese Anzahl kann sich von der Anzahl der Eingabedatensätze unterscheiden, wenn im Datenfluss Datensätze kombiniert, Datensätze aufgeteilt oder neue Datensätze erstellt werden.

- In der Spalte „**Erfolgreich**“ wird die Gesamtzahl der Datensätze angezeigt, die von allen Datenflussdatenladungen ausgegeben wurden und im Feld „Status“ einen leeren Wert aufweisen.
- In der Spalte „**Fehlgeschlagen**“ wird die Gesamtzahl der Datensätze angezeigt, die von den Datenflussdatenladungen ausgegeben wurden und im Feld „Status“ den Wert „F“ aufweisen.
- In der Spalte „**Falsch formatiert**“ wird die Gesamtzahl der Datensätze angezeigt, die aus allen Fehlerports des Quellenschrittes stammen.

Die Registerkarte **Prozessflüsse** dient zum Überwachen des Prozessfluss-Status und zum Abbrechen aktiver Prozessflüsse sowie zum Löschen abgeschlossener Prozessflüsse. Wenn Sie auf das Pluszeichen neben einem bestimmten Prozessfluss klicken, werden Ihnen Informationen zum Aktivitätsstatus des Prozessflusses angezeigt. In diesem Bereich sind folgende Informationen enthalten:

ActivityName	Umfasst die Namen aller Aktivitäten, erfolgreich ausgeführte Aktivitäten inbegriffen, aus denen der Prozessfluss besteht.	
State	Der Status der Aktivität („Fehlgeschlagen“, „Erfolgreich“, „Wird ausgeführt“, „Abgebrochen“).	
ReturnCode	Ein Code, der das Ergebnis des Prozessflusses angibt. Zur Auswahl stehen:	
	1	Der Prozessfluss ist fehlgeschlagen.
	0	Der Prozessfluss wurde erfolgreich abgeschlossen.
	-1	Der Prozessfluss wurde abgebrochen.
	Weitere Zahlen	Wenn der Prozess die Aktivität „Programm ausführen“ enthält, gibt das externe Programm unter Umständen eigene Rückgabecodes zurück. Alle Werte in der Spalte

	„Rückgabecode“, die von den Werten 1, 0 und -1 abweichen, stammen von dem externen Programm. Eine Erläuterung der Rückgabecodes des externen Programms finden Sie in der Dokumentation zum externen Programm.
Gestartet	Das Datum und die Uhrzeit des Aktivitätsstarts.
Finished	Das Datum und die Uhrzeit des Aktivitätsabschlusses.
Kommentar	Alle der Aktivität zugeordneten Kommentare.

Herunterladen des Flussverlaufs

Sie können die auf der Seite „Verlauf“ in der Management Console angezeigten Informationen in eine Microsoft Excel-Datei herunterladen.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Flüsse** > **Verlauf**.
3. Klicken Sie auf **Transaktionsverlauf**, um Verlaufsinformationen für Dienste herunterzuladen. Zeigen Sie die Registerkarte **Flüsse** weiter an, um den Verlauf für Aufträge und Prozessflüsse herunterzuladen.
4. Klicken Sie auf die Schaltfläche „Herunterladen“ .

   Tipp: Wenn Sie nur bestimmte Einträge aus der Verlaufsliste herunterladen möchten, ändern Sie die Filtereinstellungen so, dass nur der gewünschte Verlauf heruntergeladen wird.

Festlegen eines Standards für falsch formatierte Datensätze

Ein falsch formatierter Datensatz ist ein Eingabedatensatz, der von Spectrum™ Technology Platform nicht geparst werden kann. Wenn die Eingabedaten für einen Auftrag einen falsch formatierten Datensatz enthalten, wird der Auftrag standardmäßig beendet. Sie können diese Einstellung ändern, um weitere falsch formatierte Datensätze oder sogar eine unbegrenzte Anzahl falsch formatierter Datensätze zuzulassen. In dieser Prozedur wird beschrieben, wie ein Standardschwellenwert für falsch formatierte Datensätze für Aufträge in Ihrem System festgelegt wird.

Anmerkung: Sie können den Standardschwellenwert für falsch formatierte Datensätze überschreiben, indem Sie den Auftrag im Enterprise Designer öffnen und zu **Bearbeiten** > **Auftragsoptionen** navigieren.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Flüsse** > **Standardwerte**.
3. Klicken Sie auf **Falsch formatierte Datensätze**.
4. Wählen Sie eine der folgenden Optionen aus:

Aufträge mit diesen vielen falsch formatierten Datensätzen beenden	Wählen Sie diese Option aus, um Aufträge zu beenden, wenn die Eingabedaten mindestens einen falsch formatierten Datensatz enthalten. Geben Sie die Anzahl der falsch formatierten Datensätze ein, bei der Sie eine Beendigung des Auftrags auslösen möchten. Der Standardwert ist 1.
Flüsse mit falsch formatierten Datensätzen nicht beenden	Wählen Sie diese Option aus, um eine unbegrenzte Anzahl falsch formatierter Datensätze in den Eingabedaten zuzulassen.

Festlegen von Standardwerten für Berichte

Berichte werden nach Aufträgen generiert, die einen Berichtsschritt enthalten. Berichte können Verarbeitungszusammenfassungen beinhalten, z. B. zu der Anzahl der im Auftrag verarbeiteten Datensätze oder zu postüblichen Formularen wie dem Formular USPS CASS 3553. Einige Module enthalten vordefinierte Berichte. Sie können auch benutzerdefinierte Berichte erstellen. Durch das Festlegen von Standardwerten für Berichte werden die Standardeinstellungen zum Speichern von Berichten erstellt. Die Standardeinstellungen können für einen Auftrag oder für einen bestimmten Bericht mit einem Auftrag mit dem Enterprise Designer überschrieben werden.

In dieser Prozedur wird beschrieben, wie die Standardoptionen für die Berichterstellung für Ihr System festgelegt werden.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Flüsse** > **Standardwerte**.
3. Klicken Sie auf **Berichte**.
4. Wählen Sie das Format aus, in dem Berichte gespeichert werden sollen. Berichte können als HTML, PDF oder Text gespeichert werden.
5. Wählen Sie den Speicherort für die Berichte aus.

Berichte im Auftragsverlauf speichern	Speichert Berichte als Bestandteil des Auftragsverlaufs auf dem Server. So können Benutzer der Management Console und des Enterprise Designer Berichte bequem anzeigen, da sie im Ausführungsverlauf verfügbar sind.
Berichte in Datei speichern	Speichert Berichte in einer Datei an einem von Ihnen angegebenen Speicherort. Dies ist hilfreich, wenn Sie Berichte für Personen freigeben möchten, die keine Benutzer von Spectrum™ Technology Platform sind. Dies ist auch hilfreich, wenn Sie an einem anderen Speicherort ein Berichtsarchiv erstellen möchten. Zum Anzeigen von Berichten, die auf diese Art gespeichert wurden, können Sie ein beliebiges Tool verwenden, mit dem der Bericht in dem entsprechenden Format geöffnet werden kann, z. B. einen PDF-Viewer für PDF-Berichte oder einen Webbrowser für HTML-Berichte.

6. Wenn Sie **Berichte in Datei speichern** ausgewählt haben, füllen Sie die folgenden Felder aus.

Berichtsspeicherort	Der Ordner, in dem Berichte gespeichert werden sollen.
An Berichtsnamen anfügen	Gibt Variableninformationen an, die in den Dateinamen einbezogen werden sollen. Sie können mindestens eine der folgenden Variablen auswählen: **Auftrags-ID** Eine eindeutige ID, die einer Auftragsausführung zugewiesen ist. Bei der ersten Ausführung eines Auftrags in Ihrem System hat der Auftrag die ID 1. Bei der zweiten Ausführung eines Auftrags, entweder desselben Auftrags oder eines anderen Auftrags, hat der Auftrag die ID 2 usw. **Schritt** Der im Berichtsschritt im Enterprise Designer angegebene Name des Schrittes, der Daten zum Bericht beigetragen hat. **Date** Das Datum, der Monat und das Jahr, an bzw. in dem der Bericht erstellt wurde.
Vorhandene Berichte überschreiben	Ersetzt vorherige Berichte, die den gleichen Dateinamen haben, durch den neuen Bericht. Wenn Sie diese Option nicht auswählen und ein Bericht vorhanden ist, der den gleichen Namen hat wie der neue Bericht, wird der Auftrag erfolgreich ausgeführt, der neue Bericht wird jedoch nicht gespeichert. Im Ausführungsverlauf wird ein Kommentar angezeigt, der angibt, dass der Bericht nicht gespeichert wurde.

Verfügbarmachen eines Dienstes

Verfügbarmachen eines Dienstes als Webservice

Spectrum™ Technology Platform-Dienste können als REST- und/oder SOAP-Webservices verfügbar gemacht werden. So machen Sie einen Dienst als Webservice auf Ihrem Server verfügbar:

1. Öffnen Sie den Enterprise Designer.
2. Öffnen Sie den Dienst, den Sie als Webservice verfügbar machen möchten.
3. Wählen Sie die Optionen **Bearbeiten** > **Webservice-Optionen** aus.
4. Aktivieren Sie das Kästchen **Als SOAP-Webdienst verfügbar machen**, um den Dienst als einen SOAP-Webdienst verfügbar zu machen.

5. Aktivieren Sie das Kästchen **Als REST-Webdienst verfügbar machen**, um den Dienst als einen REST-Webdienst verfügbar zu machen, und folgen Sie den nachstehenden Schritten.
   a) Wenn Sie den Standardendpunkt überschreiben möchten, legen Sie den zu verwendenden Endpunkt im Feld **Pfad** fest.

      Die Angabe eines Pfades ist optional. Standardmäßig lautet der Endpunkt eines REST-Webservices wie folgt:

      `http://server:port/rest/service_name/results.qualifier`

      Wenn Sie einen anderen Endpunkt verwenden möchten, wird der von Ihnen festgelegte Pfad hinter dem Dienstnamen hinzufügt. Wenn Sie z. B. `Americas/Shipping` im Feld **Pfad** festlegen, sieht Ihr JSON-Endpunkt wie folgt aus:

      `http://myserver:8080/rest/MyService/`**`Americas/Shipping`**`/results.json`

      Sie können Felder und Optionen aus dem Datenfluss als Variablennamen im Pfad verwenden, indem Sie auf das Dropdown-Menü **Variable einfügen** klicken und das Feld oder die Option auswählen. Die Variable wird im Pfad mithilfe der Schreibweise `${Option.Name}` für Datenflussoptionen und `${Daten.Name}` für Datenflussfelder dargestellt.

   b) REST-Webservices unterstützen standardmäßig die GET-Methode und geben Daten im XML- und JSON-Format zurück. Sie können zusätzliche HTTP-Methoden und Ausgabeformate definieren, indem Sie auf **Hinzufügen** klicken, um eine Ressource zum Webdienst hinzuzufügen.

      Beim Hinzufügen einer Ressource können Sie die HTTP-Methode auswählen (**GET** oder **POST**). Die unterstützten Datenformate sind unten aufgeführt. Ihnen könnten nicht alle diese Formate zur Verfügung stehen, da einige Formate nur verfügbar sind, wenn Sie bestimmte Module auf Ihrem Spectrum™ Technology Platform-Server installiert haben.

      XML Das standardmäßige XML-Format. Verwenden Sie dieses Format, wenn Sie XML als das Format für Anforderungen und Antworten verwenden möchten und es kein spezifisches XML-Format für die Art der zu verarbeitenden Daten gibt.

      JSON Das standardmäßige JSON-Format. Verwenden Sie dieses Format, wenn Sie JSON als das Format für Anforderungen und Antworten verwenden möchten und es kein spezifisches JSON-Format für die Art der zu verarbeitenden Daten gibt.

      GeoJSON Ein spezielles JSON-Format, das für Dienste zur Handhabung von geografischen Daten geeignet ist. Es werden nur „geometry“ und die folgenden systemeigenen Plattformtypen unterstützt:

      - boolean
      - double
      - float
      - integer
      - bigdecimal

- long
- string
- date
- time
- datetime
- timespan

Wenn Sie versuchen, einen Fluss mit einem anderen Typ verfügbar zu machen, können Sie GeoJSON nicht festlegen. (Beim Design wird ein Fehler angezeigt.) GeoJSON ermöglicht außerdem nur eine einzige Geometrie. Wenn die Ausgabe mehrere Geometriefelder enthält, sucht das System nach einem Feld mit dem Namen „geometry“ gefolgt von einem Feld mit dem Namen „obj“. Wenn diese Felder nicht vorhanden sind, wird das erste Geometriefeld ausgewählt.

c) Klicken Sie auf **OK**.

Die neue Ressource wird zum Webservice hinzugefügt.

6. Klicken Sie auf **OK**, wenn Sie die Konfiguration der Webdienstoptionen beendet haben.
7. Klicken Sie in der Symbolleiste auf die graue Glühbirne, um den Dienst verfügbar zu machen.

   Wenn ein Datenfluss verfügbar gemacht wird, gibt die Glühbirnenschaltfläche in der Symbolleiste des Enterprise Designer an, dass der Datenfluss verfügbar gemacht wurde (siehe Abbildung):

Um zu überprüfen, ob der Dienst nun als Webservice verfügbar ist, öffnen Sie eine der folgenden URLs:

- Für REST:`http://server:port/rest`
- Für SOAP:`http://server:port/soap`

Dabei steht *server* für den Namen oder die IP-Adresse Ihres Spectrum™ Technology Platform-Servers und *port* für den HTTP-Kommunikationsport.

Verfügbarmachen eines Dienstes für den API-Zugriff

Damit ein Dienst über die Spectrum™ Technology Platform-API zugänglich ist, müssen Sie den Dienst verfügbar machen.

1. Öffnen Sie den Dienst im Enterprise Designer.
2. Klicken Sie in der Symbolleiste auf die graue Glühbirne, um den Dienst verfügbar zu machen.

   Wenn ein Datenfluss verfügbar gemacht wird, gibt die Glühbirnenschaltfläche in der Symbolleiste des Enterprise Designer an, dass der Datenfluss verfügbar gemacht wurde (siehe Abbildung):

Laufzeitoptionen

Hinzufügen von Datenfluss-Laufzeitoptionen

Über Datenfluss-Laufzeitoptionen können Sie das Verhalten von Schritten steuern, wenn Sie den Datenfluss ausführen. Dies ist nützlich, wenn Sie die Möglichkeit haben möchten, das Verhalten des Datenflusses zu ändern, während er ausgeführt wird. Sie können beispielsweise während der Ausführung des Datenflusses eine Quelldatenbank für einen „Read from DB“-Schritt angeben, die anstelle der im „Read from DB“-Schritt im Datenfluss angegebenen Datenbank verwendet werden soll.

Diese Prozedur beschreibt, wie Sie Optionen verfügbar machen können, die zur Laufzeit festgelegt werden können. Nach Durchführen dieser Prozedur können Sie mithilfe folgender Techniken Datenflussoptionen zur Laufzeit festlegen:

- Bei Aufträgen können Sie Laufzeitoptionen angeben, indem Sie eine Eigenschaftsdatei für Datenflussoptionen und das Argument `-o` vom Job Executor verwenden.
- Bei Diensten können Sie Laufzeitoptionen als API-Optionen angeben.
- Bei Diensten, die als Webservice verfügbar gemacht sind, können Sie Laufzeitoptionen als Parameter in der Anforderung angeben.
- Bei Unterflüssen werden die Laufzeitoptionen vom übergeordneten Datenfluss übernommen und mit einem der oben genannten Mittel verfügbar gemacht, je nach Typ des übergeordneten Datenflusses (Auftrag, Dienst oder als Webservice verfügbar gemachter Dienst).

Sie können Laufzeitoptionen wie folgt zu einem Datenfluss hinzufügen:

1. Öffnen Sie den Datenfluss im Enterprise Designer.
2. Wenn Sie Laufzeitoptionen für einen Schritt in einem eingebetteten Datenfluss konfigurieren möchten, öffnen Sie den eingebetteten Datenfluss.
3. Klicken Sie in der Symbolleiste auf das Symbol „Datenflussoptionen“ oder klicken Sie auf **Bearbeiten** > **Datenflussoptionen**. Das Dialogfeld **Datenflussoptionen** wird angezeigt.
4. Klicken Sie auf **Add** (Hinzufügen). Das Dialogfeld **Datenflussoptionen definieren** wird angezeigt.
5. Geben Sie im Feld **Optionsname** den Namen ein, den Sie für diese Option verwenden möchten. Dies ist der Optionsname, der zur Laufzeit angegeben werden muss, um diese Option festzulegen.

6. Im Feld **Beschriftung** können Sie eine andere Beschriftung angeben oder den Optionsnamen als Beschriftung beibehalten.
7. Geben Sie eine Beschreibung der Option in das Feld **Beschreibung** ein.
8. Wählen Sie im Feld **Ziel**, ob diese Option auf alle Schritte im Datenfluss oder nur auf bestimmte Schritte angewendet werden soll.

   Ausgewählte(r) Schritt/e

   Wählen Sie diese Option aus, wenn die Option nur auf die angegebenen Schritte angewendet werden soll.

   Alle Schritte

   Wählen Sie diese Option aus, wenn die Option auf alle Schritte im Datenfluss angewendet werden soll.

   Enthält Transformationen

   Wählen Sie diese Option aus, wenn die Laufzeitoption allen benutzerdefinierten Transformationen in Transformer-Schritten im Datenfluss zur Verfügung gestellt werden soll. Wenn Sie diese Option wählen, können Sie mit folgender Syntax auf den zur Laufzeit im Groovy-Skript einer benutzerdefinierten Transformation angegebenen Wert zugreifen:

   ```
   options.get("optionName")
   ```

   Wenn Sie etwa auf eine Option namens `casing` zugreifen wollten, würden Sie im benutzerdefinierten Transformationsskript Folgendes angeben:

   ```
   options.get("casing")
   ```

9. Wenn Sie im Feld **Ziel** die Option **Ausgewählte(r) Schritt/e** auswählen, zeigt die Tabelle **Datenflussoptionen zu Schritten zuordnen** eine Liste der Schritte im Datenfluss. Wählen Sie die Option aus, die Sie als Datenflussoption verfügbar machen möchten. Sie sehen, wie die Felder **Standardwert** und **Zulässige Werte** mit Daten ausgefüllt werden, wenn Sie das erste Element auswählen.

   Anmerkung: Sie können mehrere Optionen auswählen, sodass die Datenflussoption mehrere Schrittoptionen steuern kann. Wenn Sie das tun, muss jede der ausgewählten Schrittoptionen zulässige Werte gemeinsam haben. Beispiel: Eine Option hat die Werte Y und N. Jede der zusätzlichen Optionen muss entweder Y oder N in ihrer Wertegruppe enthalten, und Sie können nur den gemeinsamen Wert zur Laufzeit zur Verfügung stellen. Wenn Sie also eine Option mit den Werten Y und N auswählen, können Sie keine Option mit den Werten E, T, M und L auswählen, eine Option mit den Werten P, S und N hingegen schon, weil beide Optionen „N“ als Wert gemeinsam haben. Allerdings wäre dann nur N ein verfügbarer Wert für diese Option, nicht aber Y, P oder S.

10. Wenn Sie die Werte beschränken möchten, die sich zur Laufzeit angeben lassen, bearbeiten Sie die Optionen im Feld **Zulässige Werte**. Klicken Sie dazu auf das Symbol rechts neben dem Feld.

11. Wenn Sie den Standardwert ändern möchten, geben Sie im Feld **Standardwert** einen anderen Wert an.

    Anmerkung: Für einen Dienst können Sie die Standardwerte nur ändern, bevor der Dienst erstmals verfügbar gemacht wird. Nachdem der Dienst verfügbar gemacht wurde, können Sie die Standardwerte nicht mehr im Enterprise Designer ändern. Stattdessen müssen Sie die Management Console verwenden. Weitere Informationen finden Sie unter **Angeben von Standarddienstoptionen** auf Seite 183.

12. Klicken Sie auf **OK**.
13. Fügen Sie auf Wunsch weitere Optionen hinzu.
14. Klicken Sie im Dialogfeld **Datenflussoptionen** auf **OK**, sobald Sie damit fertig sind.
15. Wenn Sie eine Laufzeitoption zu einem eingebetteten Datenfluss hinzugefügt haben, müssen Sie die Laufzeitoption im übergeordneten Datenfluss sowie in allen vorherigen Datenflüssen definieren, um die Optionen zur Laufzeit zur Verfügung zu stellen. Öffnen Sie dazu den Datenfluss, der den eingebetteten Datenfluss enthält, und machen Sie die gerade erstellte Option verfügbar. Öffnen Sie ggf. den diesem Datenfluss übergeordneten Datenfluss und definieren Sie die Option auch dort usw., bis die Datenflussoption in allen vorherigen Datenflüssen definiert ist.

    Angenommen, Datenfluss A enthält einen eingebetteten Datenfluss B. Dieser wiederum enthält einen eingebetteten Datenfluss C. Daraus ergäbe sich folgende eingebettete Datenflusshierarchie: A > B > C. Wenn Sie eine Option namens „Groß-/Kleinschreibung“ in einem Schritt im eingebetteten Datenfluss „C“ verfügbar machen wollten, würden Sie den eingebetteten Datenfluss „C“ öffnen und sie dort definieren. Dann würden Sie den eingebetteten Datenfluss B öffnen und die Option dort definieren. Schließlich würden Sie Datenfluss A öffnen und die Option dort definieren und sie so zur Laufzeit verfügbar machen.

Der Datenfluss ist jetzt so konfiguriert, dass Optionen zur Laufzeit angegeben werden dürfen.

Angeben von Standarddienstoptionen

Standarddienstoptionen steuern das Standardverhalten der einzelnen Dienste auf Ihrem System. Sie können für jede Option in einem Dienst einen Standardwert angeben. Die Standardoptionseinstellung wird wirksam, wenn in einem API-Aufruf oder einer Webservice-Anforderung ein Wert für eine bestimmte Option nicht explizit definiert ist. Standarddienstoptionen sind auch die Einstellungen, die standardmäßig verwendet werden, wenn Sie unter Verwendung dieses Dienstes einen Datenfluss im Enterprise Designer erstellen.

Anmerkung: Für einen Dienst können Sie die Standardwerte nur ändern, bevor der Dienst erstmals verfügbar gemacht wird. Nachdem der Dienst verfügbar gemacht wurde, können Sie die Standardwerte nicht mehr im Enterprise Designer ändern. Stattdessen müssen Sie die Management Console verwenden.

1. Öffnen Sie die Management Console.
2. Klicken Sie auf **Dienste**.
3. Aktivieren Sie das Kästchen neben dem gewünschten Dienst und klicken Sie anschließend auf die Schaltfläche „Bearbeiten“ .
4. Legen Sie die Optionen für den Dienst fest. Informationen zu den Optionen für den Dienst finden Sie im Lösungshandbuch für das Dienstmodul.
5. Klicken Sie auf **Speichern**.

Löschen von Datenfluss-Laufzeitoptionen

Datenfluss-Laufzeitoptionen ermöglichen Ihnen, Schrittoptionen zur Laufzeit festzulegen. Die Schrittoptionen können festgelegt werden, wenn der Auftrag über einen Prozessfluss oder über das Befehlszeilenwerkzeug „Job Executor“ aufgerufen wird. Diese Prozedur beschreibt, wie man eine Laufzeitoption eines Datenflusses löscht, sodass die Option nicht mehr zur Laufzeit festgelegt werden kann.

1. Öffnen Sie den Auftrag, Dienst oder Unterfluss.
2. Klicken Sie auf das Symbol **Datenflussoptionen** oder klicken Sie auf **Bearbeiten** > **Datenflussoptionen**. Das Dialogfeld **Datenflussoptionen** wird angezeigt.
3. Markieren Sie die Option, die Sie löschen möchten, und klicken Sie auf **Entfernen**.

Konfigurieren einer E-Mail-Benachrichtigung für einen Datenfluss

Ein Datenfluss kann so konfiguriert werden, dass eine E-Mail-Benachrichtigung mit Informationen zum Auftragsstatus gesendet wird. Beispiel: Sie möchten eine E-Mail-Warnung senden, wenn ein Datenfluss fehlschlägt. Die E-Mail-Benachrichtigung kann Informationen wie den Flussnamen, die Start- und die Endzeit, die Anzahl der verarbeiteten Datensätze usw. enthalten.

> **Anmerkung:** Bevor Sie die Benachrichtigung für einen Datenfluss einrichten können, muss ein Mailserver in der Management Console konfiguriert werden. Weitere Informationen finden Sie im *Spectrum™ Technology Platform-Administratorhandbuch*.

1. Wählen Sie, wenn ein Datenfluss oder ein Prozessfluss im Enterprise Designer geöffnet ist, **Bearbeiten** > **Benachrichtigungen** aus.
2. Klicken Sie auf **Hinzufügen**.
3. Geben Sie im Feld **Benachrichtigung senden an** die E-Mail-Adresse ein, an die Benachrichtigungen gesendet werden sollten.

4. Wählen Sie die Ereignisse aus, zu denen Sie Benachrichtigungen erhalten möchten.
5. Geben Sie im Feld **Betreff** den Text ein, der in der Betreffzeile der E-Mail angezeigt werden soll.
6. Geben Sie im Feld **Nachricht** den Text ein, der im Hauptteil der E-Mail angezeigt werden soll.

   Sie können auswählen, ob Informationen zum Auftrag in die E-Mail einbezogen werden sollen, indem Sie auf **Hier klicken, um ein Feld in den Betreff oder die Nachricht einzufügen** klicken. Zu den Beispielen für Informationen zum Auftrag zählen die Start- und die Endzeit sowie die Anzahl der fehlgeschlagenen Datensätze.

7. Klicken Sie auf **Vorschau**, wenn Sie eine Vorschau der Benachrichtigung anzeigen möchten.
8. Klicken Sie auf **OK**. Das Dialogfeld „**Benachrichtigungen**“ wird mit der neuen aufgeführten Benachrichtigung wieder angezeigt.
9. Klicken Sie auf **OK**.

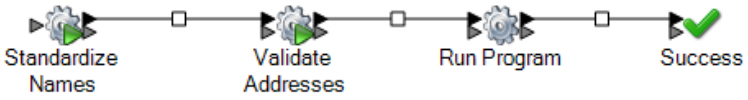
5 - Kombinieren von Flüssen zu einem Prozessfluss

In this section

Einführung in Prozessflüsse 187
Entwerfen von Prozessflüssen 187

Einführung in Prozessflüsse

In einem Prozessfluss werden eine Reihe von Aktivitäten ausgeführt, wie z. B. Aufträge und externe Anwendungen. Jede Aktivität im Prozessfluss wird nach Beendigung der vorherigen Aktivität ausgeführt. Prozessflüsse sind nützlich, wenn Sie mehrere Datenflüsse nacheinander ausführen möchten oder wenn Sie ein externes Programm ausführen möchten. In einem Prozessfluss könnte beispielsweise ein Auftrag zum Standardisieren von Namen und Überprüfen von Adressen ausgeführt werden und anschließend eine externe Anwendung aufgerufen werden, um die Datensätze in die richtige Reihenfolge zu bringen, damit Portoermäßigungen geltend gemacht werden können. Ein solcher Prozessfluss würde wie folgt aussehen:



In diesem Beispiel wurden die Aufträge „Namen standardisieren“ und „Adressen überprüfen“ auf dem Spectrum™ Technology Platform-Server verfügbar gemacht. Über die Aktivität „Programm ausführen“ wird eine externe Anwendung aufgerufen und die Aktivität „Erfolg“ gibt das Ende des Prozessflusses an.

Entwerfen von Prozessflüssen

Erstellen eines Prozessflusses

Verwenden Sie den Enterprise Designer für die Erstellung eines Prozessflusses, um eine Sequenz von Aktivitäten zu erstellen, die Aufträge und/oder externe Anwendungen ausführen.

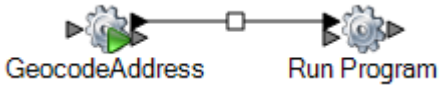
1. Öffnen Sie den Enterprise Designer.
2. Wählen Sie **Datei** > **Neu** > **Prozessfluss** aus.
3. Fügen Sie die erste Aktion hinzu, die im Prozessfluss ausgeführt werden soll. Sie können eine der folgenden Aktionen ausführen:
   - Ziehen Sie zum Ausführen eines Auftrags das Auftragssymbol aus dem Ordner **Aktivitäten** in die Palette auf der Arbeitsfläche.
   - Ziehen Sie zum Ausführen eines externen Programms das Symbol „Programm ausführen“ aus dem Ordner **Aktivitäten** in die Palette auf der Arbeitsfläche.

4. Fügen Sie die zweite Aktion hinzu, die im Prozessfluss ausgeführt werden soll.

   Sie können einen Auftrag hinzufügen, indem Sie das Auftragssymbol auf die Arbeitsfläche ziehen, oder ein externes Programm hinzufügen, indem Sie das Symbol „Programm ausführen“ auf die Arbeitsfläche ziehen.

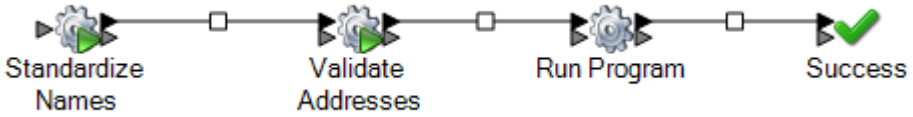
5. Verbinden Sie die zwei Aktivitäten miteinander, indem Sie rechts vom ersten Symbol auf das graue Dreieck klicken und es zum Herstellen einer Verbindung auf das graue Symbol links vom zweiten Symbol ziehen.

   Wenn Sie beispielsweise über einen Prozessfluss verfügen, der zunächst einen Auftrag mit dem Namen „GeocodeAddress“ ausführt und anschließend ein externes Programm ausführt, würde Ihr Prozessfluss wie folgt aussehen:

6. Fügen Sie nach Bedarf zusätzliche Aktivitäten hinzu.
7. Wenn Sie alle Aktivitäten hinzugefügt haben, die im Prozessfluss ausgeführt werden sollen, ziehen Sie die Aktivität „Erfolg“ auf die Arbeitsfläche und verbinden Sie diese mit der letzten Aktivität im Prozessfluss.

   Der folgende Prozessfluss enthält beispielsweise zwei Aufträge („Namen standardisieren“ und „Adressen überprüfen“) und die Aktivität „Programm ausführen“. Am Ende dieses Prozessflusses befindet sich die Aktivität „Erfolg“:

8. Doppelklicken Sie auf die Aktivitäten, die Sie auf der Arbeitsfläche platziert haben, um ihre Laufzeitoptionen zu konfigurieren. Sie können auch auf die Verbindung zwischen Aktivitäten doppelklicken, um Übergangsoptionen zu konfigurieren.

Variable zum Verweis auf eine Datei verwenden

In einem Prozessfluss sind Variablen nützlich, wenn Sie möchten, dass mehrere Aktivitäten im Prozessfluss auf dieselbe Datei verweisen. Mithilfe einer Variablen können Sie die Datei an einer Stelle definieren und dann in allen Downstream-Aktivitäten, die auf diese Datei verweisen müssen, stattdessen auf die Variable verweisen. Sollte sich die Datei je ändern, können Sie die Variablendefinition ändern und müssen nicht alle Downstream-Aktivitäten ändern.

Wenn Sie eine Auftragsaktivität zu einem Prozessfluss hinzufügen, erstellt die Aktivität automatisch Variablen für jede Quelle und jedes Laden von Daten im Datenfluss. Wenn Sie im Prozessfluss

Dateien verwenden möchten, die nicht in der Quelle oder dem Laden von Daten eines Auftrags definiert werden, können Sie Variablen erstellen.

Wenn Sie eine Aktivität „Programm ausführen“ hinzufügen, werden standardmäßig keine Variablen erstellt. Wenn Sie mit einer Aktivität „Programm ausführen“ Variablen verwenden möchten, müssen Sie diese erstellen.

Die folgenden Schritte beschreiben, wie Sie eine Variable in einer Auftragsaktivität oder einer Aktivität „Programm ausführen“ erstellen.

1. Öffnen Sie den Prozessfluss im Enterprise Designer.
2. Doppelklicken Sie auf die Auftragsaktivität oder Aktivität „Programm ausführen“, in der Sie die Variable definieren möchten.

   Anmerkung: Aktivitäten können nur auf eine Variable verweisen, wenn diese in einer vorher ausgeführten Aktivität definiert wurde. Achten Sie daher darauf, die Variable einer Aktivität zu definieren, die vor den Aktivitäten liegt, in denen Sie die Variable verwenden möchten.

3. Klicken Sie auf die Registerkarte **Variablen**.
4. Erstellen Sie die Variable.

Option	Bezeichnung	
Neue Variable für eine Eingabedatei erstellen	Klicken Sie neben **Eingaben** auf **Hinzufügen**. Geben Sie im Feld **Name** einen Namen für die Variable ein. Dies ist der Name, auf den Downstream-Aktivitäten verweisen. Wählen Sie im Feld **Speicherort** eine der folgenden Optionen aus:	
	Im Auftrag angegebene Datei verwenden	Wählen Sie diese Option aus, um die Datei zu verwenden, die im Quellschritt des Auftrags definiert ist. Diese Option ist nur verfügbar, wenn Sie eine Variable für eine Auftragsaktivität definieren.
	Datei auf dem Server suchen	Wählen Sie diese Option aus, wenn Sie eine Datei auswählen möchten, um sie dieser Variablen zuzuweisen.
	Auf eine Datei einer Upstream-Aktivität verweisen	Wählen Sie diese Option aus, wenn Sie eine Datei verwenden möchten, die von einem Upstream-Schritt zu einer vorhandenen Variablen zugewiesen wurde.
Neue Variable für eine	Klicken Sie neben **Eingaben** auf **Hinzufügen**. Geben Sie im Feld **Name** einen Namen für die Variable ein. Dies ist der Name, auf den	

Option	Bezeichnung	
Ausgabedatei erstellen	Downstream-Aktivitäten verweisen. Wählen Sie im Feld **Speicherort** eine der folgenden Optionen aus:	
	Datei auf dem Server suchen	Wählen Sie diese Option aus, wenn Sie eine Datei auswählen möchten, um sie dieser Variablen zuzuweisen.
	Vom Server verwaltete temporäre Datei	Wählen Sie diese Option aus, wenn diese Variable auf eine temporäre Datei verweisen soll, die nach Bedarf automatisch erstellt und gelöscht wird. Diese Option ist in den Fällen hilfreich, in denen eine Datei in einem Prozessfluss nur als Zwischenschritt verwendet wird und nach Abschluss des Prozessflusses nicht mehr benötigt wird.

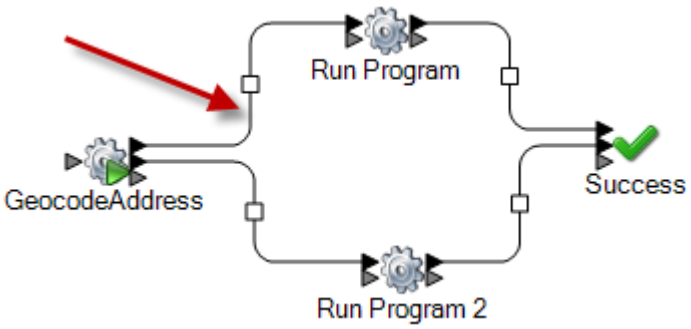
5. Klicken Sie auf **OK**, um das Fenster **Variable hinzufügen** zu schließen.
6. Klicken Sie auf **OK**, um das Fenster mit den Aktivitätsoptionen zu schließen.
7. In einer Downstream-Aktivität auf die Variable verweisen:
   a) Doppelklicken Sie auf die Aktivität, die auf die Variable verweisen soll.
   b) Wählen Sie den Eingabeschritt aus, die auf die Variable verweisen soll, und klicken Sie auf **Ändern**.
   c) Wählen Sie im Feld **Speicherort** Reference an upstream activity's file... aus.

Bedingte Logik zu einem Prozessfluss hinzufügen

Sie können bedingte Logik zu einem Prozessfluss hinzufügen, sodass unterschiedliche Aktivitäten basierend auf dem Rückgabecode einer vorherigen Aktivität ausgeführt werden. Sie könnten beispielsweise eine Aktivität ausführen, wenn ein Auftrag einen Rückgabecode von 1 zurückgibt, und eine andere Aktivität ausführen, wenn ein Auftrag einen Rückgabecode von 0 zurückgibt. So können Sie Ihren Prozessfluss um eine bedingte Verzweigung erweitern.

1. Öffnen Sie den Prozessfluss im Enterprise Designer.
2. Doppelklicken Sie auf den Übergang zwischen den beiden Aktivitäten des Flusses.

   Ein *Übergang* ist die Linie, die die beiden Aktivitäten verbindet. Beispielsweise ist die hier dargestellte Linie zwischen der Aktivität „GeocodeAddress“ und der Aktivität „Programm ausführen“ ein Übergang:



Das Fenster **Übergangsoptionen** wird angezeigt.

3. Wählen Sie den Übergangstyp aus, den Sie hinzufügen möchten.

Einfach	Wählen Sie diese Option aus, wenn Sie möchten, dass dieser Pfad im Prozessfluss immer ausgeführt wird.
Bedingt	Wählen Sie diese Option aus, wenn Sie möchten, dass dieser Pfad im Prozessfluss nur ausgeführt wird, wenn die Upstream-Aktivität einen bestimmten Rückgabecode, Rückgabecodes oder einen Bereich von Rückgabecodes zurückgibt.
Andernfalls	Wählen Sie diese Option aus, wenn Sie möchten, dass dieser Pfad im Prozessfluss nur ausgeführt wird, wenn die Bedingungen der anderen Übergänge, die von der Aktivität weg führen, nicht erfüllt werden.

Anmerkung: Unter den Übergängen, die von einer Aktivität weg führen, kann nur ein Übergang **Andernfalls** vorhanden sein.

4. Klicken Sie auf **OK**.
5. Um zu konfigurieren, welche Übergänge eine Aktivität auslösen, klicken Sie mit der rechten Maustaste auf die Aktivität, wählen Sie **Eingabemodi** aus und wählen Sie dann eine der folgenden Optionen aus:

Erste Eingabe	Der erste Übergang, der diese Aktivität erreicht, ob durch einen Übergang **Einfach**, **Bedingt** oder **Andernfalls**, löst die Ausführung der Aktivität aus. Andere Übergänge werden ignoriert.
Alle Eingaben	Die Aktivität wird nicht ausgeführt, bis nicht alle eingehenden Übergänge dieser Aktivität aktiviert wurden.

6. Um zu konfigurieren, welche ausgehenden Übergänge einer Aktivität aktiviert werden, klicken Sie mit der rechten Maustaste auf die Aktivität, wählen Sie **Ausgabemodi** aus und wählen Sie dann eine der folgenden Optionen aus:

Erste Ausgabe	Der erste Übergang, dessen Bedingung erfüllt wird, wird aktiviert. Andere Übergänge werden ignoriert, auch wenn ihre Bedingungen erfüllt werden.
Alle Ausgaben	Alle Übergänge, deren Bedingungen erfüllt werden, werden aktiviert.

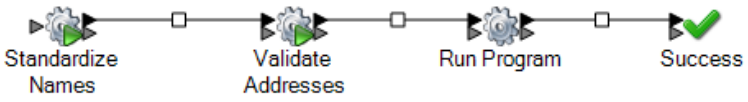
Löschen eines Prozessflusses

1. Öffnen Sie **Datei** > **Verwalten**. Das Dialogfeld „**Verwalten**“ wird angezeigt.
2. Klicken Sie mit der rechten Maustaste auf den Prozessfluss, den Sie löschen möchten, und wählen Sie **Löschen** aus.
3. Klicken Sie auf **OK**.

Aktivitäten

Auftrag

Eine Auftragsaktivität führt einen Auftrag als Teil eines Prozessflusses aus. Im folgenden Beispiel wird ein Prozessfluss dargestellt, in dem zwei Auftragsaktivitäten ausgeführt werden: „Namen standardisieren“ und „Adressen überprüfen“.



Um eine Auftragsaktivität zu einem Prozessfluss hinzuzufügen, ziehen Sie die Auftragsaktivität aus dem Ordner „Aktivitäten“ in die Palette auf der Arbeitsfläche.

> **Anmerkung:** Damit ein Auftrag für die Verwendung in einem Prozessfluss verfügbar ist, muss der Auftrag verfügbar gemacht werden. Wenn ein von Ihnen gewünschter Auftrag nicht verfügbar gemacht wurde, öffnen Sie den Auftrag im Enterprise Designer, und wählen Sie **Datei** > **Verfügbar machen** aus.

Doppelklicken Sie auf die Auftragsaktivität, um die Registerkarte **Optionen** und die Registerkarte **Variablen** zu konfigurieren.

Registerkarte „Optionen“

Auf der Registerkarte „Optionen“ werden die für den Auftrag zur Verfügung stehenden Laufzeitoptionen angezeigt. Sie können die Laufzeitoptionen so ändern, dass die von Ihnen hier angegebenen Optionen bei einer Ausführung des Auftrags in diesem Prozessfluss für den Auftrag verwendet werden. Wenn beispielsweise von einer der Datenflussoptionen des Auftrags die für die Einheiten verwendete Distanz gesteuert wird und standardmäßig Meilen verwendet werden, könnten Sie diese Option statt bei der Ausführung des Auftrags über diesen Prozessfluss hier überschreiben, sodass die Distanz in Kilometern zurückgegeben wird.

Registerkarte „Variablen“

Auf der Registerkarte „Variablen“ werden die Quell- und Datenladeschritte im Datenfluss angezeigt. Sie können die im Datenfluss angegebenen Eingabe- und Ausgabedateien überschreiben, sodass bei der Ausführung des Auftrags in diesem Prozessfluss verschiedene Eingabe- und/oder Ausgabedateien verwendet werden.

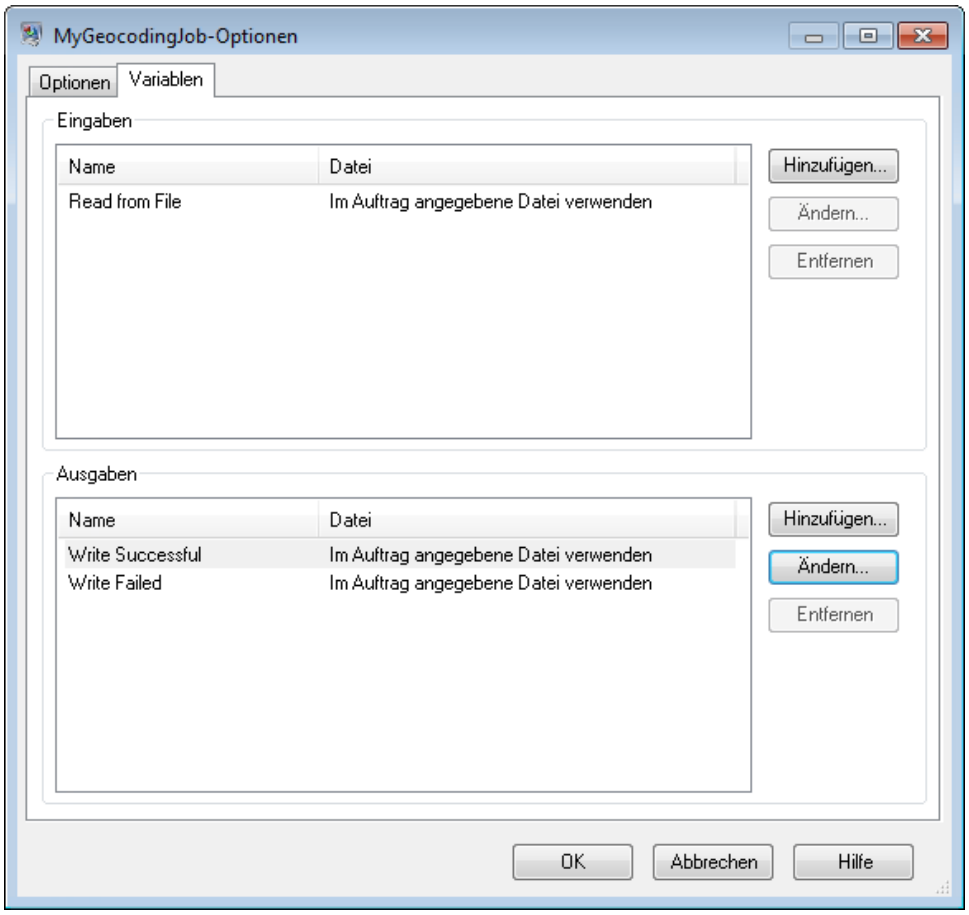
Überschreiben von Eingabe- und Ausgabedateien

Bei einem in einem Prozessfluss ausgeführten Auftrag werden standardmäßig die im Quell- und Datenladeschritt des Auftrags definierten Eingabe- und Ausgabedateien verwendet. Sie können jedoch die im Auftrag definierten Eingabe- und Ausgabedateien überschreiben, sodass bei der Ausführung des Prozessflusses statt der im Quell- oder Datenladeschritt des Auftrags angegebenen Datei die von Ihnen in der Auftragsaktivität des Prozessflusses angegebene Eingabe- oder Ausgabedatei verwendet wird. Sie können Eingabe- und Ausgabedateien überschreiben, indem Sie eine bestimmte Datei angeben oder eine Variable verwenden, mit der Sie auf eine in der Upstream-Aktivität definierte Datei verweisen.

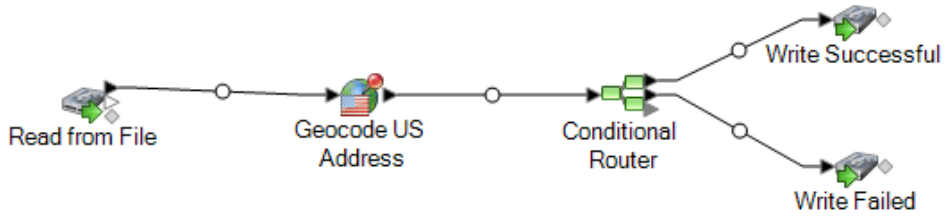
1. Öffnen Sie den Prozessfluss im Enterprise Designer.
2. Doppelklicken Sie auf die Auftragsaktivität, bei der Sie eine Eingabe- oder Ausgabedatei überschreiben möchten.
3. Klicken Sie auf die Registerkarte **Variablen**.

   Auf der Registerkarte **Variablen** entsprechen die unter **Eingaben** aufgeführten Variablen den Quellschritten des Auftrags. Die unter **Ausgaben** aufgeführten Variablen entsprechen den Datenladeschritten des Auftrags.

   Angenommen, Sie verfügen über einen Prozessfluss, der eine Auftragsaktivität für einen Auftrag mit dem Namen „MyGeocodingJob“ enthält. Auf der Registerkarte **Variablen** der Aktivitätsoptionen wird Ihnen Folgendes angezeigt:



Jede aufgeführte Variable entspricht dem Namen eines Datenladungs- oder Quellschrittes im Datenfluss von MyGeocodingJob. In diesem Beispiel werden auf der Registerkarte **Variablen** eine Quelle („Read from File“) und zwei Datenladungen („Write Failed“ und „Write Successful“) angezeigt. Wenn Sie den Datenfluss von MyGeocodingJob im Enterprise Designer öffnen würden, würde Ihnen etwas wie das Folgende angezeigt:



4. Wählen Sie auf der Registerkarte **Variablen** den Quellschritt oder Ausgabeschritt aus, der überschrieben werden soll, und klicken Sie auf **Ändern**.
5. Führen Sie eine der folgenden Aktionen aus:

Option	Bezeichnung
Zum Überschreiben einer Eingabedatei...	Wählen Sie im Feld **Speicherort** eine der folgenden Optionen aus: **Im Auftrag angegebene Datei verwenden**: Wählen Sie diese Option aus, um die Datei zu verwenden, die im Quellschritt des Auftrags definiert ist. **Datei auf dem Server suchen**: Wählen Sie diese Option aus, wenn Sie die im Auftrag definierte Datei überschreiben und eine andere von Ihnen ausgewählte Datei verwenden möchten. **Auf eine Datei einer Upstream-Aktivität verweisen**: Wählen Sie diese Option aus, wenn Sie die im Auftrag definierte Datei mit einer Datei überschreiben möchten, deren Name und Speicherort im „Read from File“- oder „Write to File“-Schritt einer Upstream-Aktivität oder in der Variablen einer Upstream-Aktivität definiert sind. Verwenden Sie diese Option, wenn die Ausgabedatei aus einer vorherigen Aktivität die Eingabe für diese Aktivität ist. Der Vorteil dieser Option besteht darin, dass wenn der „Write to File“-Schrittes einer Upstream-Aktivität je geändert wird, um auf eine andere Datei zu verweisen, diese Aktivität weiterhin auf die richtige Datei verweist. Ein weiterer Vorteil besteht darin, dass Ihnen der Dateipfad und der Name der Eingabedatei von Upstream-Aktivitäten nicht bekannt sein muss, um darauf verweisen zu können.
Zum Überschreiben einer Ausgabedatei...	Wählen Sie im Feld **Speicherort** eine der folgenden Optionen aus: **Datei auf dem Server suchen**: Wählen Sie diese Option aus, wenn Sie die im Auftrag definierte Datei überschreiben und eine andere von Ihnen ausgewählte Datei verwenden möchten. **Vom Server verwaltete temporäre Datei**: Wählen Sie diese Option aus, wenn diese Variable auf eine temporäre Datei verweisen soll, die nach Bedarf automatisch erstellt und gelöscht wird. Diese Option ist in den Fällen hilfreich, in denen eine Datei in einem Prozessfluss nur als Zwischenschritt verwendet wird und nach Abschluss des Prozessflusses nicht mehr benötigt wird.

6. Klicken Sie auf **OK**, um das Fenster „**Variable ändern**“ zu schließen.
7. Klicken Sie auf **OK**, um das Fenster mit den Aktivitätsoptionen zu schließen.

Während der Ausführung des Prozessflusses werden im Auftrag statt der im Auftrag selbst angegebenen Datei die von Ihnen in der Prozessflussaktivität angegebenen Dateien verwendet.

Cache löschen

Die Aktivität „Cache löschen“ löscht die globalen Cachedaten als Teil des Prozessflusses. Nicht der Cache wird gelöscht, sondern nur die Cachedaten.

1. Ziehen Sie die Aktivität **Cache löschen** auf die Arbeitsfläche.
2. Doppelklicken Sie auf „Cache löschen“.
3. Wählen Sie den Cache aus. Sie können auch mehrere Caches auswählen, um die zugehörigen Daten zu löschen.
4. Führen Sie den Prozessfluss aus.

SQL ausführen

Mit der Aktivität „SQL ausführen“ können Sie die SQL-Anweisungen vor und nach der Ausführung eines Datenflusses oder eines externen Programms ausführen.

Laden in Hive

Apache Hive ist eine auf Hadoop aufbauende Data Warehouse-Infrastruktur für Datenzusammenfassung, Abfragen und Analysen. Für die Abfrage der zugrunde liegenden Datenquelle über Hive wird die zugehörige eigene Abfragesprache, HiveQL, verwendet.

Hive unterstützt die nachfolgenden Hadoop-Dateiformate:

- TEXTFILE
- SEQUENCE FILE
- ORC
- RCFILE
- PARQUET
- AVRO

Anmerkung: Das Dateiformat AVRO wird in Hive ab Version 0.14 unterstützt.

Die Aktivität „**In Hive laden**“ ermöglicht es Ihnen, Daten über eine JDBC-Verbindung in eine Hive-Tabelle zu laden. Mit dieser Verbindung werden Daten aus einer bestimmten Hadoop-Datei gelesen und in eine vorhandene Tabelle einer ausgewählten Verbindung, oder eine neu erstellte Verbindung in der ausgewählten Verbindung, geladen.

Damit die Daten in eine neue Tabelle geladen werden können, muss das Schema der Tabelle definiert werden.

Anmerkung: Spectrum unterstützt keine hierarchischen Daten, auch wenn Hive dies tut.

Erstellen einer Hive-Verbindung

1. Öffnen Sie die Aktivität „**In Hive laden**“.
2. Geben Sie im Feld **Feldname** den Namen der Datei ein, die gelesen werden soll. Klicken Sie auf „Durchsuchen“ ([...]), um die zu lesende Datei auszuwählen.
3. Wählen Sie im Feld **Dateityp** das Format der zu lesenden Datei aus. Das ausgewählte Standarddateiformat lautet `Mit Begrenzungszeichen`.

   Wenn der ausgewählte **Dateityp** `Mit Begrenzungszeichen` oder `Sequenz` lautet, werden die Felder **Feldtrennzeichen** und **Datensatz-Trennzeichen** angezeigt. Andernfalls werden sie nicht angezeigt.
4. Wählen Sie im Feld **Feldtrennzeichen** das Zeichen aus, mit dem die aufeinanderfolgenden Felder eines Datensatzes voneinander getrennt werden.
5. Wählen Sie die im Feld **Verbindung** die Verbindung für die Hive-Datenbank aus, die Sie verwenden möchten.
   a) Klicken Sie auf **Verwalten**, um Verbindungen hinzuzufügen, zu ändern und zu löschen. Das Fenster „**Datenbankverbindungs-Manager**“ wird geöffnet.
   b) Klicken Sie auf **Hinzufügen**, um eine neue Verbindung zu erstellen, oder auf **Ändern**, um eine vorhandene Verbindung zu bearbeiten. Das Fenster „**Verbindungseigenschaften**“ wird geöffnet.
   c) Geben Sie den **Verbindungsnamen** ein.
   d) Wählen Sie im Feld **Datenbanktreiber** einen Hive-Datenbanktreiber für die Verbindung aus.
   e) Geben Sie alle Details der Verbindung an, sprich **Benutzer**, **Kennwort**, **Host**, **Port** und **Instanz**.
   f) Klicken Sie auf **Testen**, um die Verbindungsdetails zu testen.
   g) Klicken Sie auf **OK**, wenn der Verbindungstest erfolgreich war. Das Fenster „**Verbindungseigenschaften**“ wird geschlossen.
   h) Klicken Sie auf **OK**. Das Fenster „**Datenbankverbindungs-Manager**“ wird geschlossen.
6. Wählen Sie im Feld **Tabelle/Ansicht** die Tabelle aus, in die Sie schreiben möchten, oder geben Sie den Namen einer neuen Tabelle ein, die erstellt werden soll.

   Wenn Sie im Feld **Tabelle/Ansicht** eine neue Tabelle erstellen, wird das Kästchen **Extern** aktiviert. Andernfalls, wenn Sie eine vorhandene Tabelle auswählen, bleibt das Kästchen **Extern** deaktiviert.
7. Aktivieren Sie das Kästchen **Extern**, um die neue externe Tabelle der Hive-Datenbank zu erstellen.

   Wichtig: Bei externen Tabellen:

   1. Vorhandene Datensätze können nicht überschrieben werden, und neue Datensätze können nicht hinzugefügt werden. Sie können nur neue externe Tabellen erstellen und diese mit Datensätzen auffüllen.

2. Wenn Sie eine Datei aus einem bestimmten Ordner auswählen, werden automatisch alle Dateien aus diesem Ordner ausgewählt. Stellen Sie daher sicher, dass alle Dateien in diesem bestimmten Ordner das richtige Format haben.

Weitere Informationen zu externen Hive-Tabellen finden Sie **hier**.

8. Aktivieren Sie das Kästchen **Überschreiben**, um alle vorhandenen Datensätze der Tabelle zu überschreiben. Dadurch werden vorhandene Datensätze der ausgewählten Tabelle gelöscht und die aus der Datei gelesenen Datensätze zur Tabelle hinzugefügt.
9. Im Raster werden die Namen und Datentypen der Spalten der ausgewählten Tabelle angezeigt.

   Wenn Sie im Feld **Tabelle/Ansicht** eine neue Tabelle angegeben haben, verwenden Sie die Schaltflächen **Hinzufügen**, **Ändern** und **Entfernen** neben dem Raster, um Spalten zum Definieren der Tabelle hinzuzufügen, und geben Sie die entsprechenden Datentypen an. Verwenden Sie die Schaltflächen **Nach oben** und **Nach unten**, um die Reihenfolge der Tabellenspalten anzugeben.

   Anmerkung: Die Schaltflächen **Hinzufügen**, **Ändern**, **Entfernen**, **Nach oben** und **Nach unten** bleiben deaktiviert, wenn Sie im Feld **Tabelle/Ansicht** eine vorhandene Tabelle auswählen.

   Wichtig:

   1. Stellen Sie sicher, dass die Datentypen aller Felder in der Datei den Datentypen in den entsprechenden Tabellenspalten entsprechen, sofern nicht alle Datentypen vom Typ „Zeichenfolge“ sind. Andernfalls kann das Laden von Daten dazu führen, dass die Daten inkonsistent sind.
   2. Stellen Sie sicher, dass die Anzahl der Felder in der Datei der Anzahl der Tabellenspalten entspricht. Andernfalls werden die Daten in den zusätzlichen Feldern in der Datei gelöscht.
   3. Hive akzeptiert nur kleingeschriebene Tabellen- und Spaltennamen. Wenn Sie die Namen in Druckbuchstaben schreiben, wandelt Hive diese in Kleinbuchstaben um. Im resultierenden Schema werden alle Namen in Kleinbuchstaben angezeigt.

10. Klicken Sie auf **OK**.

    Anmerkung: Wenn Sie sich für die Erstellung einer neuen Tabelle entschieden haben und die zugehörigen Spalten definieren, wird die gleiche Tabelle zur Laufzeit erstellt. Die Aktivität **In Hive laden** dient nur zum Entwerfen der Tabellenstruktur. Zur Laufzeit wird die entworfene Tabelle erstellt und die aus der Datei gelesenen Daten werden in die Tabelle geschrieben.

Ausführen des Hadoop-Auftrags „MapReduce“

Die Aktivität **Hadoop-Auftrag MapReduce ausführen** ermöglicht es Ihnen, einen MapReduce-Auftrag durch Mapping der relevanten JAR-Datei in einem Hadoop-Cluster auszuführen. Sie können mit dieser Aktivität einen MapReduce-Auftrag des Spectrum™ Big Data Quality SDK oder einen externen MapReduce-Auftrag ausführen.

Anmerkung: Wenn der MapReduce-Auftrag fehlschlägt, wird eine Fehlermeldung zusammen mit dem Status des ausgeführten Auftrags angezeigt.

Feld	Beschreibung
Hadoop-Server	Die Liste der konfigurierten Hadoop-Server. Informationen zum Mapping von HDFS-Dateiservern über die Management Console finden Sie im *Administratorhandbuch*.
JAR-Pfad	Der Pfad der relevanten JAR-Datei für den auszuführenden Hadoop-Auftrag MapReduce. **Anmerkung:** Das JAR muss am Speicherort des externen Clients oder des Spectrum-Servers vorhanden sein. Es darf nicht im Hadoop-Cluster platziert werden.
Treiberklasse	Wählen Sie eine Option: **Standard** Um einen externen Auftrag auszuführen, indem Sie einfach nur den *Klassennamen* des Auftrags und die *Argumente* für den Auftrag eingeben, wählen Sie `Standard` aus. Wenn Sie `Standard` auswählen, werden die Felder **Klassenname** und **Argumente** angezeigt. **Konfigurieren** Um weitere Auftragseigenschaften eines beliebigen externen Auftrags einzugeben oder um einen beliebigen Spectrum Big Data Quality SDK-Auftrag auszuführen, wählen Sie `Konfigurieren` aus. Wenn Sie `Konfigurieren` auswählen, wird das Feld **Auftragstyp** angezeigt.
Auftragstyp	Wählen Sie eine Option: **Spectrum** Wählen Sie `Spectrum` aus, um einen beliebigen Spectrum Big Data Quality SDK-Auftrag auszuführen. Wenn Sie `Spectrum` auswählen, wird das Feld **Spectrum-Aufträge** angezeigt. **Allgemein** Wählen Sie `Allgemein` aus, um weitere Eigenschaften für beliebige externe Aufträge anzugeben.

Feld	Beschreibung
Spectrum-Aufträge	Wählen Sie einen Spectrum-Auftrag aus der Liste der Spectrum Big Data Quality SDK-Aufträge aus. Bei der Auswahl des gewünschten Spectrum-Auftrags: 1. Die Felder **Auftragsname**, **Klassenname** und **Argument** werden automatisch ausgefüllt. Alle automatisch ausgefüllten Felder außer **Klassenname** können bei Bedarf bearbeitet werden. **Wichtig:** Bei dem ausgewählten Spectrum-Auftrag darf der automatisch ausgefüllte **Klassenname** nicht bearbeitet werden, da der Auftrag ansonsten nicht ausgeführt werden kann. 2. Das Raster **Eigenschaften** wird automatisch mit den erforderlichen Konfigurationseigenschaften des ausgewählten Spectrum-Auftrags und deren Standardwerten ausgefüllt. Sie können weitere Eigenschaften hinzufügen oder importieren sowie die automatisch ausgefüllten Eigenschaften nach Bedarf ändern.
Klassenname	Der vollständig qualifizierte Name der Treiberklasse des Auftrags.

Feld	Beschreibung

Argumente	Die durch Leerzeichen getrennte Liste von Argumenten. Diese werden zur Laufzeit zur Ausführung des Auftrags an die Treiberklasse weitergeleitet.
	Beispiel:
	`23Dec2016 /home/Hadoop/EYInc.txt`
	1. Diese Variablen können als Argumente in der Argumentenliste weitergeleitet werden. Diese Argumente sind so definiert, dass sie Laufzeitwerte entweder im Quellschritt oder in diesem aktuellen Schritt des Prozessflusses akzeptieren.
	Wenn beispielsweise in der Ausgabe des vorherigen Schritts des Prozessflusses die Variable `SalesStartRange` definiert wurde, können Sie diese Variable in dieser Argumentenliste als `${SalesStartRange}` zusammen mit anderen erforderlichen Argumenten einschließen. Beispiel:
	`23Dec2016 /home/Hadoop/EYInc.txt ${SalesStartRange}`
	2. Wenn ein bestimmtes Argument ein Leerzeichen enthält, schließen Sie dieses Argument in doppelte Anführungszeichen ein.
	Beispiel: `"/home/Hadoop/Sales Records"`.
	Aufträge des Spectrum Big Data Quality SDK – Argumente:
	Um die Spectrum Big Data Quality SDK *MapReduce*-Aufträge auszuführen, übergeben Sie die verschiedenen Konfigurationsdateien als Liste von Argumenten. Jeder Argumentschlüssel akzeptiert den Pfad einer einzelnen Konfigurationseigenschaftsdatei, wobei jede Datei mehrere Konfigurationseigenschaften enthält.
	Die Syntax der Argumentenliste für Konfigurationseigenschaften ist:
	`[-config <Path to configuration file>] [-debug] [-input <Path to input configuration file>] [-conf <Path to MapReduce configuration file>] [-output <Path of output directory>]`
	Beispiel für einen MapReduce MatchKeyGenerator-Auftrag:
	`-config /home/hadoop/matchkey/mkgConfig.xml -input /home/hadoop/matchkey/inputFileConfig.xml -conf /home/hadoop/matchkey/mapReduceConfig.xml -output /home/hadoop/matchkey/outputFileConfig.xml`
	Anmerkung: Wenn derselbe Konfigurationseigenschaftsschlüssel sowohl im Feld **Argumente** als auch im Raster **Eigenschaften** angegeben ist, aber jeder Schlüssel auf eine andere Konfigurationsdatei verweist, hat die Datei Vorrang, auf die im Raster **Eigenschaften** verwiesen wird.
	Die Beispielkonfigurationseigenschaften werden mit dem Big Data Quality SDK geliefert und befinden sich unter `<Big Data Quality bundle>\samples\configuration`.

Registerkarte „Allgemein“

Feld	Beschreibung	Voraussetzung
Auftragsname	Der Name des Hadoop-Auftrags MapReduce.	Erforderlich
Eingabepfad	Der Pfad der Eingabedatei des Auftrags.	Erforderlich
Ausgabepfad	Der Pfad der Ausgabedatei des Auftrags.	Erforderlich
Ausgabe überschreiben	Zeigt an, ob der angegebene Ausgabepfad überschrieben werden muss, wenn er bereits vorhanden ist. **Anmerkung:** Wenn dieses Kästchen deaktiviert bleibt und der konfigurierte Ausgabepfad zur Laufzeit vorhanden ist, gibt Hadoop eine Ausnahme aus und der Prozessfluss wird abgebrochen.	Optional
Mapper-Klasse	Der vollständig qualifizierte Name der Klasse, die das Mapper-Feature für den Auftrag verarbeitet.	Erforderlich
Reducer-Klasse	Der vollständig qualifizierte Name der Klasse, die das Reducer-Feature für den Auftrag verarbeitet.	Optional
Combiner-Klasse	Der vollständig qualifizierte Name der Klasse, die das Combiner-Feature für den Auftrag verarbeitet.	Optional
Partitioner-Klasse	Der vollständig qualifizierte Name der Klasse, die das Partitioner-Feature für den Auftrag verarbeitet.	Optional
Reducer-Anzahl	Die Anzahl der für die Ausführung des Auftrags MapReduce verwendeten Reducer.	Optional
Eingabeformat	Das Format der Eingabedaten.	Erforderlich
Ausgabeformat	Das Format der Ausgabedaten.	Erforderlich
Ausgabeschlüsselklasse	Der Datentyp der Schlüssel in den Schlüssel-Wert-Paaren der Ausgabe.	Erforderlich
Ausgabewertklasse	Der Datentyp der Werte in den Schlüssel-Wert-Paaren der Ausgabe.	Erforderlich

Registerkarte „Eigenschaften“

Um zusätzliche Eigenschaften für die Ausführung des erforderlichen Auftrags anzugeben, definieren Sie auf dieser Registerkarte so viele Eigenschaften-/Wertepaare wie erforderlich. Sie können die erforderlichen Eigenschaften direkt nacheinander im Raster hinzufügen.

Alternativ können Sie Eigenschaften aus einer Datei importieren, indem Sie auf **Importieren** klicken. Navigieren Sie zum Speicherort der jeweiligen Eigenschaftsdatei, und wählen Sie die Datei im XML-Format aus. Die in der importierten Datei enthaltenen Eigenschaften werden in das Raster kopiert. Die Eigenschaftsdatei muss im XML-Format vorliegen und dieser Syntax folgen:

```
<configuration>
    <property>
        <name>key</name>
        <value>some_value</value>
        <description>A brief description of the
            purpose of the property key.</description>
    </property>
</configuration>
```

Sie können die Hadoop-Eigenschaftsdatei *mapred.xml* direkt importieren oder unter Verwendung dieses XML-Formats eigene Dateien erstellen.

Anmerkung:

1. Wenn dieselbe Eigenschaft dort und in der Management Console definiert wurde, überschreiben die dort definierten Werte die in der Management Console definierten Werte.
2. Wenn dieselbe Eigenschaft im Raster und auch in der importierten Eigenschaftsdatei vorhanden ist, dann überschreibt der aus der Datei importierte Wert den im Raster für dieselbe Eigenschaft vorhandenen Wert.
3. Sie können falls erforderlich mehrere Eigenschaftsdateien nacheinander importieren. Die in den einzelnen importierten Dateien enthaltenen Eigenschaften werden im Raster hinzugefügt.
4. Stellen Sie sicher, dass die Eigenschaftsdatei auf dem Spectrum™ Technology Platform-Server vorhanden ist.
5. Das Tag `<description>` ist für alle Eigenschaftsschlüssel in einer Konfigurationseigenschaftsdatei optional.

Registerkarte „Abhängigkeiten“

Fügen Sie auf dieser Registerkarte die Liste mit den Eingabedateien und JAR-Dateien hinzu, die zur Ausführung des Auftrags erforderlich sind.

Nachdem der Auftrag ausgeführt wurde, sind die dort hinzugefügten Referenzdateien und JAR-Referenzdateien im verteilten Cache des Auftrags verfügbar.

Referenzdateien

Um die verschiedenen Dateien hinzuzufügen, die als Eingabe für die Ausführung des Auftrags erforderlich sind, klicken Sie auf **Hinzufügen**, navigieren Sie zum

entsprechenden Speicherort auf Ihrem lokalen System oder Cluster und wählen Sie die gewünschte Datei aus.

Um eine beliebige, der Liste hinzugefügte Datei zu entfernen, wählen Sie diese Datei aus und klicken Sie auf **Entfernen**.

JAR-Referenzdateien

Um die JAR-Dateien hinzuzufügen, die für die Ausführung des Auftrags erforderlich sind, klicken Sie auf **Hinzufügen**, navigieren Sie zum entsprechenden Speicherort auf Ihrem lokalen System oder Cluster und wählen Sie die gewünschte Datei aus.

Um eine beliebige, der Liste hinzugefügte Datei zu entfernen, wählen Sie diese Datei aus und klicken Sie auf **Entfernen**.

Variablen

Eingaben
Wählen Sie in diesem Raster, aus den von einer Quellaktivität empfangenen Feldern, die Felder aus, die in dieser Aktivität verwendet werden sollen.

Ausgaben
Wählen Sie in diesem Raster die Felder aus, die in die Ausgabe für die Zielaktivität in diesem Prozessfluss eingeschlossen werden sollen.

Hadoop Pig ausführen

Die Aktivität „Hadoop Pig ausführen“ führt ein Apache Pig-Skript aus. Apache Pig ist eine Sprache auf hoher Ebene zum Ausdrücken von Datenanalyseprogrammen und verfügt über die Infrastruktur zum Auswerten dieser Programme. Pig-Programme können parallelisiert werden. Dadurch können in diesen Programmen sehr umfangreiche Datensets verarbeitet werden.

Mit „Hadoop Pig ausführen“ können Sie die Pig-Vorgänge auswählen, alle notwendigen Parameter eingeben und Ihr Pig-Skript automatisch vom System generieren lassen. Sie können das Pig-Skript auf einem beliebigen Hadoop-Server ausführen.

„Hadoop Pig ausführen“ kann nur auf Hadoop-Dateiservern ausgeführt werden. Apache Hadoop 1.x und 2.x werden unterstützt.

Gehen Sie wie folgt vor, um die Optionen für „Hadoop Pig ausführen“ festzulegen:

1. Ziehen Sie die Aktivität **Hadoop Pig ausführen** und legen Sie sie auf der Arbeitsfläche ab.
2. Machen Sie einen Rechtsklick auf die Aktivität **Hadoop Pig ausführen** und wählen Sie **Optionen** aus.
3. In den Servernamensfeldern wird der Hadoop-Server angegeben, auf dem sich die zu verarbeitende Datei befindet.
4. Klicken Sie auf die Schaltfläche „Durchsuchen“ ([...]), um zu der zu verarbeitenden Datei zu navigieren.
5. Wählen Sie den Dateityp aus. „Hadoop Pig ausführen“ unterstützt Sequenzdateien mit und ohne Begrenzungszeichen.

6. Wählen Sie das Begrenzungszeichen und das Textbegrenzungszeichen entsprechend aus.
7. Klicken Sie im Bereich „Felder“ auf **Hinzufügen** und fügen Sie die Felder hinzu, die in der zu verarbeitenden Datei enthalten sind. Bei Sequenzdateien gilt das erste Feld als Schlüssel und die anderen Felder gehören zu den Werten mit Begrenzungszeichen.
8. Wählen Sie wie gewünscht den Vorgang **Kürzen** aus. Bei dem Vorgang „Kürzen“ werden vor der Verarbeitung Leerzeichen im Eingabefeld gekürzt.
9. Öffnen Sie die Registerkarte „Vorgänge“. Klicken Sie auf **Hinzufügen**, um mit dem Hinzufügen der Pig-Vorgänge zu beginnen, die in der Datei ausgeführt werden sollen. Dadurch wird der Editor „Vorgänge“ geöffnet.
10. Wählen Sie einen Vorgang aus, der ausgeführt werden soll. Die verschiedenen Vorgänge lauten wie folgt:
    - Sortieren – Sortiert die Daten in alphabetischer Reihenfolge.
    - Filtern – Ermöglicht es Ihnen, die Daten Ihren Anforderungen entsprechend zu filtern.
    - Aggregieren – Ermöglicht es Ihnen, statistische Vorgänge zu den Daten auszuführen, wie z. B. „Summe“, „Anzahl“ und andere.
    - Verschieden – Wählen Sie alle eindeutigen Datensätze aus dem angegebenen Feld aus.
    - Begrenzen – Ermöglicht es Ihnen, die Anzahl der verarbeiteten Datensätze auf eine angegebene Anzahl zu begrenzen.
11. Verwenden Sie die Schaltflächen **Nach oben** und **Nach unten**, um die Reihenfolge der Vorgänge zu ändern.
12. Sobald Sie die Vorgänge ausgewählt haben und die erforderliche Eingabe für die Verarbeitung der Vorgänge gemacht haben, können Sie auf **Hinzufügen** klicken, um Ihre Auswahl zu speichern und zum Pig-Optionseditor zurückzukehren.
13. Das Pig-Skript wird automatisch basierend auf den ausgewählten Vorgängen generiert.

    Mit dem Editor können Sie das generierte Pig-Skript bei Bedarf mit Ihrem eigenen Skript überschreiben. Klicken Sie auf die Option **Skript bearbeiten** und geben Sie Ihr eigenes Skript im Textfeld des Pig-Skripts ein. Die Schaltfläche **Erneut generieren** ist in diesem Fall aktiviert. Wenn Sie das vom System generierte Skript wieder anzeigen möchten, klicken Sie im Abschnitt „Pig-Skript“ auf **Erneut generieren**, um das Pig-Skript zu generieren.
14. Sie können die Ausgabedatei unter der Registerkarte **Variablen** angeben. Die Ausgabedatei kann von den nachfolgenden Aktivitäten verwendet werden.
15. Klicken Sie auf **OK**, um das Pig-Skript zu speichern. Der Ausgabedateityp ist standardmäßig mit dem Eingabedateityp identisch. Sie können dies mithilfe des generierten Pig-Skripts ändern.

Hadoop Pig-Vorgänge

Die verschiedenen Pig-Vorgänge lauten wie folgt:

1. Sortieren: Sortiert die Daten in alphabetischer Reihenfolge. Der Sortiervorgang wird unter **Sortieren von Eingabedatensätzen** detailliert beschrieben.
2. Filtern: Ermöglicht es Ihnen, die Daten Ihren Anforderungen entsprechend zu filtern. Der Filtervorgang wird unter **Filtern von Eingabedatensätzen** detaillierter beschrieben.

3. Aggregieren: Ermöglicht es Ihnen, statistische Vorgänge zu den Daten auszuführen, wie z. B. „Summe“, „Anzahl“ und andere.

   Wählen Sie wie erwünscht die Aggregatvorgänge für die einzelnen Felder aus.

   - Summe: Berechnet die Summe der Werte im Feld.
   - Durchschnitt: Berechnet den Durchschnitt aller Werte im Feld.
   - Max.: Berechnet den maximalen Wert der Werte im Feld.
   - Min.: Berechnet den minimalen Wert der Werte im Feld.
   - Anzahl: Berechnet die Gesamtzahl der Werte im Feld.

     Anmerkung: Wenn Sie den Vorgang „Verschieden“ auswählen, werden nur die eindeutigen Werte gezählt.

4. Verschieden: Durch Auswahl dieser Option werden im Vorgang „Anzahl aggregieren“ nur eindeutige Werte im Feld gezählt.
5. Begrenzen: Geben Sie einen Wert ein, der größer als null ist, um die Anzahl der verarbeiteten Datensätze auf diesen Wert zu begrenzen.

Programm ausführen

Die Aktivität „Programm ausführen“ führt eine externe Anwendung als Teil eines Prozessflusses aus.

Tabelle 9: Optionen der Aktivität „Programm ausführen“

Name der Option	Beschreibung
Programmname	Der Pfad zu der ausführbaren Datei, die Sie ausführen möchten.
Argumente	Gibt Befehlszeilenargumente an, die an das im Feld **Programmname** angegebene Programm übergeben werden sollen. Trennen Sie mehrere Argumente mit Leerzeichen voneinander. Sie können auf der Registerkarte „Variablen“ definierte Variablen als Argumente verwenden, indem Sie auf **Variable einfügen** klicken. Weitere Informationen zu Variablen finden Sie unter **Variable zum Verweis auf eine Datei verwenden** auf Seite 188.
Timeout (in Sekunden)	Gibt eine Zeitspanne an, die gewartet werden soll, bis das im Feld **Programmname** angegebene Programm reagiert. Wenn das Programm in der angegebene Zeitspanne nicht reagiert, schlägt der Prozessfluss fehl.

Name der Option	Beschreibung
Umgebungsvariablen	Gibt Umgebungsvariablenwerte an, die beim Ausführen dieses Programms verwendet werden sollen. Wenn Sie hier Werte angeben, verwendet das Programm statt der in Ihrem System angegebenen Umgebungsvariablen diese Umgebungsvariablen. Andernfalls verwendet es die in Ihrem System angegebenen Umgebungsvariablen. Beachten Sie Folgendes: Wenn das von Ihnen aufgerufene Programm mehrere Umgebungsvariablen verwendet, müssen Sie entweder für alle Umgebungsvariablen oder für keine Umgebungsvariable Werte definieren. Durch die Angabe von Werten an dieser Stelle werden die Umgebungsvariablendefinitionen auf Ihrem System nicht geändert. Klicken Sie auf **Hinzufügen** und geben Sie den Namen der Variable im Feld **Variablenname** ein. Ein Beispiel könnte „JAVA_HOME“ sein. Geben Sie den Wert der Variable im Feld **Variablenwert** ein. Ein Beispiel könnte „C:\Program Files\Java\jdk1.6.0_17“ sein. Statt einen Wert einzugeben, können Sie auf **Variable einfügen** klicken, um den Wert auf eine Variable festzulegen, die auf der Registerkarte **Variablen** definiert ist.

Angeben von Eingabe- und Ausgabedateien

Sie können mit der Aktivität „Programm ausführen“ über einen Prozessfluss eine externe Anwendung aufrufen. Sie können die Datei angeben, die die Daten enthält, die an die externe Anwendung gesendet werden soll, sowie die Ausgabedatei, in die die externe Anwendung schreiben soll.

1. Doppelklicken Sie in einem Prozessfluss auf die Aktivität „Programm ausführen“.
2. Klicken Sie auf die Registerkarte **Variablen**.
3. Gehen Sie wie folgt vor, um eine Eingabedatei anzugeben:
   a) Klicken Sie im Abschnitt „**Eingaben**“auf die Schaltfläche **Hinzufügen**.
   b) Geben Sie im Feld **Name** einen aussagekräftigen Namen für diese Datei ein. Sie können den Namen frei wählen.
   c) Wählen Sie im Feld **Speicherort** eine der folgenden Optionen aus:

Datei auf dem Server suchen	Wählen Sie diese Option aus, wenn Sie nach der gewünschten Eingabedatei suchen und diese auswählen möchten.
Auf eine Datei einer Upstream-Aktivität verweisen	Wählen Sie diese Option aus, wenn Sie eine Datei verwenden möchten, die von einem Upstream-Schritt zu einer vorhandenen Variablen zugewiesen wurde.

   d) Klicken Sie auf **OK**.
4. Gehen Sie wie folgt vor, um eine Ausgabedatei anzugeben:

a) Klicken Sie im Abschnitt „**Ausgaben**“ auf die Schaltfläche **Hinzufügen**.
b) Geben Sie im Feld **Name** einen aussagekräftigen Namen für diese Datei ein. Sie können den Namen frei wählen.
c) Wählen Sie im Feld **Speicherort** eine der folgenden Optionen aus:

Datei auf dem Server suchen	Wählen Sie diese Option aus, wenn Sie nach der gewünschten Ausgabedatei suchen und diese auswählen möchten.
Vom Server verwaltete temporäre Datei	Wählen Sie diese Option aus, wenn die Ausgabe des Programms in eine temporäre Datei geschrieben werden soll, die nach Bedarf automatisch erstellt und gelöscht wird. Diese Option ist in den Fällen hilfreich, in denen eine Datei in einem Prozessfluss nur als Zwischenschritt verwendet wird und nach Abschluss des Prozessflusses nicht mehr benötigt wird.

d) Klicken Sie auf **OK**.

5. Klicken Sie auf **OK**, um das Fenster **Programmoptionen ausführen** zu schließen.

Verwenden einer Steuerungsdatei mit einem externen Programm

Sie können mit der Aktivität „Programm ausführen“ über einen Prozessfluss eine externe Anwendung aufrufen. Dabei können Sie eine Steuerungsdatei verwenden, die Konfigurationseinstellungen für die externe Anwendung enthält. Sie könnten beispielsweise VeriMove™ mit der Aktivität „Programm ausführen“ aufrufen und eine Steuerungsdatei angeben, die während der Ausführung verwendet werden soll.

1. Doppelklicken Sie in einem Prozessfluss auf die Aktivität „Programm ausführen“.
2. Klicken Sie auf die Registerkarte **Variablen**.
3. Klicken Sie im Abschnitt „**Steuerungsdateien**“ auf **Hinzufügen**.
4. Geben Sie der Steuerungsdatei im Feld **Name** einen Namen. Sie können den Namen frei wählen.
5. Geben Sie im Feld **Inhalte** die Inhalte der Steuerungsdatei an.

   Um mit einer unter **Eingaben** oder **Ausgaben** definierten Variable auf eine Datei zu verweisen, klicken Sie auf **Variable einfügen**.

6. Klicken Sie auf **OK**, um das Fenster **Steuerungsdatei hinzufügen** zu schließen.
7. Klicken Sie auf die Registerkarte „**Optionen**“.
8. Klicken Sie auf **Variable einfügen**.
9. Wählen Sie den Namen der von Ihnen erstellten Steuerungsdatei aus und klicken Sie anschließend auf **OK**.

   Eine Variable, die auf Ihre Steuerungsdatei verweist, wird zum Feld **Argumente** hinzugefügt.

10. Ändern Sie das Feld **Argumente** bei Bedarf so, dass die notwendigen Befehlszeilenargumente für die Angabe einer Steuerungsdatei verwendet werden. Weitere Informationen finden Sie in der Dokumentation der externen Anwendung.

11. Klicken Sie auf **OK**, um das Fenster **Programmoptionen ausführen** zu schließen.

Spark Sorter

Mit der Aktivität **Spark Sorter** können Sie riesige Datensatzmengen sortieren. Diese Aktivität verwendet zur Ausführung Apache Spark-Bibliotheken und läuft auf Ihrem Spectrum™ Technology Platform-Server.

Derzeit werden Dateien *mit Begrenzungszeichen*, die auf dem Spectrum™ Technology Platform-Server vorhanden sind, für das Einlesen der Eingabedatensätze akzeptiert.

Anmerkung: Dateien auf Remoteservern werden nicht unterstützt.

Feld	Beschreibung
Servername	Zeigt den Speicherort der Datei an, die Sie als Eingabe ausgewählt haben. Da die **Spark Sorter**-Aktivität nur Dateien akzeptiert, die sich auf dem Spectrum™ Technology Platform-Server befinden, wird in diesem Feld Spectrum™ Technology Platform angezeigt.
Dateiname	Gibt den Pfad zur Datei an. Klicken Sie auf die Ellipsenschaltfläche (...), um zur gewünschten Datei zu navigieren. Sie können ein Platzhalterzeichen verwenden, um Daten aus mehreren Dateien im Verzeichnis einzulesen. Die Platzhalterzeichen * und ? werden unterstützt. Sie können beispielsweise `*.csv` eingeben, um alle Dateien mit der Erweiterung `.csv` im Verzeichnis auszulesen. Um mehrere Dateien erfolgreich einzulesen, muss jede Datei über dasselbe Layout (die gleichen Felder an der gleichen Position) verfügen. Alle Datensätze, die nicht mit dem auf der Registerkarte **Felder** angegebenen Layout übereinstimmen, werden als falsch formatierte Datensätze behandelt. **Achtung:** Falls der Spectrum™ Technology Platform-Server unter Unix oder Linux ausgeführt wird, berücksichtigen Sie, dass bei Dateinamen und Pfaden auf diesen Plattformen die Groß-/Kleinschreibung beachtet wird.
Datensatztyp	Dies ist das Format der Datensätze in der Datei. Derzeit werden Dateiformate *mit Begrenzungszeichen* als Eingabe akzeptiert. **Mit Begrenzungszeichen**: Dies ist eine Textdatei, in der Datensätze durch ein Zeilenendzeichen (EOL, End of Line) wie Zeilenumbruch oder Zeilenvorschub (CR oder LF, Carriage Return oder Line Feed) getrennt sind und jedes Feld durch ein angegebenes Zeichen wie beispielsweise ein Komma getrennt ist.
Zeichencodierung	Die Zeichencodierung der Eingabedatei. Die Codierung `UTF-8` wird unterstützt. Weitere Informationen zu UTF finden Sie unter **unicode.org/faq/utf_bom.html**.

Feld	Beschreibung
Feldtrennzeichen	Gibt das Zeichen an, das zum Trennen von Feldern in einer Datei mit Begrenzungszeichen verwendet wird. Im folgenden Beispiel wird ein Pipe-Zeichen (|) im Datensatz als Feldtrennzeichen verwendet: `7200 13TH ST|MIAMI|FL|33144` Die folgenden Zeichen stehen zur Definition als Feldtrennzeichen zur Verfügung: • Leerzeichen • Tabulator • Komma • Punkt • Semikolon • Pipe Wenn in der Datei ein anderes Zeichen als Feldtrennzeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Begrenzungszeichen auszuwählen.
Textbegrenzungszeichen	Dies ist das Zeichen, das in einer Datei mit Begrenzungszeichen Textwerte umschließt. Im folgenden Beispiel werden im Datensatz doppelte Anführungszeichen (") als Textbegrenzungszeichen verwendet. `"7200 13TH ST"|"MIAMI"|"FL"|"33144"` Die folgenden Zeichen stehen zur Definition als Textbegrenzungszeichen zur Verfügung: • einfaches Anführungszeichen (') • doppeltes Anführungszeichen (") Wenn in der Datei ein anderes Textbegrenzungszeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Textbegrenzungszeichen auszuwählen.

Feld	Beschreibung
Datensatz-Trennzeichen	Gibt das Zeichen an, das zum Trennen von Datensätzen in einer zeilensequenziellen Datei oder einer Datei mit Begrenzungszeichen verwendet wird. Dieses Feld ist nicht verfügbar, wenn Sie das Kästchen **Standard-Zeilenende verwenden** aktivieren. Die verfügbaren Einstellungen für Datensatz-Trennzeichen sind: **Unix (U+000A)** – Die Datensätze sind durch ein Zeilenvorschubzeichen getrennt. Dies ist bei Unix-Systemen das standardmäßige Datensatz-Trennzeichen. **Macintosh (U+000D)** – Die Datensätze sind durch ein Zeilenumbruchzeichen getrennt. Dies ist bei Macintosh-Systemen das standardmäßige Datensatz-Trennzeichen. **Windows (U+000D U+000A)** – Die Datensätze sind durch ein Zeilenumbruchzeichen gefolgt von einem Zeilenvorschubzeichen getrennt. Dies ist bei Windows-Systemen das standardmäßige Datensatz-Trennzeichen. Wenn in der Datei ein anderes Datensatz-Trennzeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Datensatz-Trennzeichen auszuwählen.
Standard-Zeilenende verwenden	Gibt an, dass das standardmäßige Zeilenendzeichen (EOL) des Betriebssystems, auf dem der Spectrum™ Technology Platform-Server läuft, als Datensatz-Trennzeichen der Datei verwendet wird. Wählen Sie diese Option nicht aus, wenn in der Datei ein Zeilenendzeichen verwendet wird, das sich vom standardmäßigen Zeilenendzeichen des Betriebssystems des Servers unterscheidet. Wenn in der Datei beispielsweise ein Windows-Zeilenendzeichen verwendet wird, der Server aber auf Linux läuft, aktivieren Sie diese Option nicht. Wählen Sie stattdessen im Feld **Datensatz-Trennzeichen** die Option „Windows“ aus.

Erste Zeile ist Kopfzeilen-Datensatz	Gibt an, ob der erste Datensatz in einer Datei mit Begrenzungszeichen Überschriften und keine Daten enthält.
	
	Im folgenden Dateiausschnitt ist der erste Datensatz ein Kopfzeilen-Datensatz.
	
	`"AddressLine1"|"City"|"StateProvince"|"PostalCode"`
	`"7200 13TH ST"|"MIAMI"|"FL"|"33144"`
	`"One Global View"|"Troy"|"NY"|12180`

Ausgabe	Gibt den Pfad zur Ausgabedatei auf dem Spectrum™ Technology Platform-Server an. Klicken Sie auf die Ellipsenschaltfläche (...), um zum Ausgabeverzeichnis und zum gewünschten Dateinamen zu navigieren. **Achtung:** Falls der Spectrum™ Technology Platform-Server unter Unix oder Linux ausgeführt wird, berücksichtigen Sie, dass bei Dateinamen und Pfaden auf diesen Plattformen die Groß-/Kleinschreibung beachtet wird.

Feld	Beschreibung
Überschreiben	Zeigt an, dass die Ausgabedatei überschrieben werden muss, wenn eine Datei mit demselben Namen wie im Feld **Ausgabe** angegeben vorhanden ist.
Verketten	Zeigt an, dass alle Spark-Teildateien zu einer einzelnen Ausgabedatei am Speicherort der angegebenen **Ausgabe** zusammengefügt werden müssen.
Vorschau	Nachdem die Eingabedatei im Feld **Dateiname** ausgewählt wurde, werden im Raster **Vorschau** die ersten 100 Datensätze der vorhandenen Ausgabedatei angezeigt. Um alle einzelnen Spaltenwerte korrekt anzuzeigen, klicken Sie auf der Registerkarte **Felder** auf **Erneut generieren**.

Registerkarte „Felder“

Die Registerkarte **Felder** definiert die Namen, Typen und Positionen der Felder in der Datei. Weitere Informationen finden Sie unter:

Definieren von Feldern in einer Datei mit Begrenzungszeichen auf Seite 213

Registerkarte „Sortieren“

Die Registerkarte **Sortieren** definiert Felder, nach denen die Eingabedatensätze sortiert werden sollen, bevor sie in den Datenfluss gesendet werden. Weitere Informationen finden Sie unter **Sortieren von Datensätzen** auf Seite 215.

Registerkarte „Konfiguration“

Um zusätzliche Eigenschaften für die Ausführung des erforderlichen Auftrags anzugeben, definieren Sie auf dieser Registerkarte so viele Eigenschaften-/Wertepaare wie erforderlich. Sie können die erforderlichen Eigenschaften direkt nacheinander im Raster hinzufügen.

Alternativ können Sie Eigenschaften aus einer Datei importieren, indem Sie auf **Importieren** klicken. Navigieren Sie zum Speicherort der jeweiligen Eigenschaftsdatei, und wählen Sie die Datei im XML-Format aus. Die in der importierten Datei enthaltenen Eigenschaften werden in das Raster kopiert. Die Eigenschaftsdatei muss im XML-Format vorliegen und dieser Syntax folgen:

```
<configuration>
    <property>
        <name>key</name>
        <value>some_value</value>
        <description>A brief description of the
            purpose of the property key.</description>
    </property>
</configuration>
```

Anmerkung:

1. Wenn dieselbe Eigenschaft dort und in der Management Console definiert wurde, überschreiben die dort definierten Werte die in der Management Console definierten Werte.
2. Wenn dieselbe Eigenschaft im Raster und auch in der importierten Eigenschaftsdatei vorhanden ist, dann überschreibt der aus der Datei importierte Wert den im Raster für dieselbe Eigenschaft vorhandenen Wert.
3. Sie können falls erforderlich mehrere Eigenschaftsdateien nacheinander importieren. Die in den einzelnen importierten Dateien enthaltenen Eigenschaften werden im Raster hinzugefügt.
4. Stellen Sie sicher, dass die Eigenschaftsdatei auf dem Spectrum™ Technology Platform-Server vorhanden ist.
5. Das Tag `<description>` ist für alle Eigenschaftsschlüssel in einer Konfigurationseigenschaftsdatei optional.

Registerkarte „Laufzeit“

Feldname	Beschreibung
Dateiname	Zeigt den auf der ersten Registerkarte ausgewählten Dateinamen an.
Startdatensatz	Wenn Sie beim Einlesen der Datensätze in den Datenfluss Datensätze am Anfang der Datei überspringen möchten, geben Sie hier den ersten Datensatz an, der gelesen werden soll. Wenn Sie beispielsweise die ersten 50 Datensätze in einer Datei überspringen möchten, geben Sie 51 ein. Der 51. Datensatz ist dann der erste Datensatz, der in den Datenfluss eingelesen wird.
Alle Datensätze	Wählen Sie diese Option aus, wenn Sie alle Datensätze ab dem im Feld **Startdatensatz** definierten Datensatz bis zum Ende der Datei einlesen möchten.
Max. Anzahl Datensätze	Wählen Sie diese Option aus, wenn Sie nur eine bestimmte Anzahl von Datensätzen ab dem im Feld **Startdatensatz** angegebenen Datensatz einlesen möchten. Wenn Sie beispielsweise die ersten 100 Datensätze einlesen möchten, wählen Sie diese Option aus und geben Sie 100 an.

Definieren von Feldern in einer Datei mit Begrenzungszeichen

Die Registerkarte **Felder** definiert die Namen, Typen und Positionen der Felder in der Datei. Nach Definieren einer Eingabedatei auf der Registerkarte **Dateieigenschaften** können Sie die Felder definieren.

Wenn die Eingabedatei keinen Kopfzeilendatensatz enthält oder wenn Sie die Felder manuell definieren möchten, folgen Sie diesen Schritten auf der Registerkarte **Felder**:

1. Klicken Sie auf **Erneut generieren**, um Felder zu definieren, die bereits in der Eingabedatei vorhanden sind. Klicken Sie dann auf **Typ erkennen**. Damit wird automatisch der Datentyp für jedes Feld basierend auf den ersten 50 Datensätzen in der Datei festgelegt.
2. Klicken Sie auf **Hinzufügen**, um zusätzliche Felder in der Ausgabe hinzuzufügen.
3. Wählen Sie im Feld **Name** das Feld aus, das Sie hinzufügen möchten, oder geben Sie den Namen in das Feld ein.
4. Im Feld **Typ** können Sie den Datentyp `string` beibehalten, wenn Sie nicht vorhaben, mathematische Operationen auf die Daten anzuwenden. Wenn Sie allerdings solche Operationsarten anwenden möchten, wählen Sie einen entsprechenden Datentyp aus. Dann werden die Zeichenfolgendaten aus der Datei in einen Datentyp konvertiert, der die korrekte Manipulation der Daten im Datenfluss ermöglicht.

   Die Aktivität unterstützt diese Datentypen:

   bigdecimal Ein numerischer Datentyp, der eine Genauigkeit von 38 Dezimalstellen unterstützt. Verwenden Sie diesen Datentyp für Daten in mathematischen Berechnungen, die einen hohen Grad von Genauigkeit erfordern, speziell solche mit Finanzdaten. Der Datentyp „bigdecimal“ unterstützt genauere Berechnungen als der Datentyp „double“.

   double Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit doppelter Genauigkeit zwischen 2^{-1074} und $(2-2^{-52})\times2^{1023}$ enthält. In der E-Notation liegt der Wertebereich bei -1,79769313486232E+308 bis 1,79769313486232E+308.

   float Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit einzelner Genauigkeit zwischen 2^{-149} und $(2-2^{23})\times2^{127}$ enthält. In der E-Notation liegt der Wertebereich bei -3,402823E+38 bis 3,402823E+38.

   integer Ein numerischer Datentyp, der sowohl positive als auch negative ganze Zahlen zwischen -2^{31} (-2,147,483,648) und $2^{31}-1$ (2,147,483,647) enthält.

   long Ein numerischer Datentyp, der sowohl negative als auch positive ganze Zahlen zwischen -2^{63} (-9.223.372.036.854.775.808) und $2^{63}-1$ (9.223.372.036.854.775.807) enthält.

   string Eine Folge von Zeichen.

5. Geben Sie in das Feld **Position** die Position dieses Feldes innerhalb des Datensatzes ein.

   Beispiel: In dieser Eingabedatei ist „AddressLine1“ an Position 1, „City“ an Position 2, „StateProvince“ an Position 3 und „PostalCode“ an Position 4.

```
"AddressLine1"|"City"|"StateProvince"|"PostalCode"
"7200 13TH ST"|"MIAMI"|"FL"|"33144"
"One Global View"|"Troy"|"NY"|12180
```

6. Wenn Sie überzählige Leerzeichen am Anfang und Ende der Wertezeichenfolge eines Feldes entfernen möchten, aktivieren Sie das Kästchen **Kürzen**.

Sortieren von Datensätzen

Auf der Registerkarte **Sortieren** können Sie die Kriterien definieren, nach denen die verschiedenen Felder in den Eingabedatensätzen sortiert werden müssen, bevor sie in den Datenfluss gesendet werden.

1. Klicken Sie auf der Registerkarte **Sortieren** auf **Hinzufügen**.
2. Klicken Sie auf den Abwärtspfeil bei der Spalte **Feldname** und wählen Sie das Feld aus, nach dem sortiert werden soll. Die auswählbaren Felder hängen von den Feldern ab, die in dieser Eingabedatei definiert sind.
3. Wählen Sie in der Spalte **Sortierung** aus, ob die Felder Ascending oder Descending sortiert werden sollen.
4. Im Feld **Typ** können Sie den Datentyp `string` beibehalten, wenn Sie nicht vorhaben, mathematische oder Datum/Uhrzeit-Operationen auf die Daten anzuwenden. Wenn Sie allerdings diese Operationsarten anwenden möchten, wählen Sie einen entsprechenden Datentyp aus. Dann werden die `string`-Daten aus der Datei in einen Datentyp konvertiert, der die korrekte Manipulation der Daten im Datenfluss ermöglicht.
5. Wenn Sie überzählige Leerzeichen am Anfang und Ende der Wertezeichenfolge eines Feldes entfernen möchten, aktivieren Sie das Kästchen **Kürzen**.
6. Wählen Sie im Feld **NULL behandeln als** aus, ob `null`-Werte des jeweiligen Felds in einem Datensatz als `größter` oder als `kleinster` Wert aller Datensätze behandelt werden sollen.
7. Wiederholen Sie diese Schritte, bis Sie alle Eingabefelder hinzugefügt haben, die Sie zum Sortieren verwenden möchten. Ändern Sie die Reihenfolge der Sortierung. Markieren Sie dazu die Zeile für das Feld, das Sie verschieben möchten, und klicken Sie auf **Aufwärts** oder **Abwärts**.

Variablen

Eingaben
Wählen Sie in diesem Raster, aus den von einer Quellaktivität empfangenen Feldern, die Felder aus, die in dieser Aktivität verwendet werden sollen.

Ausgaben
Wählen Sie in diesem Raster die Felder aus, die in die Ausgabe für die Zielaktivität in diesem Prozessfluss eingeschlossen werden sollen.

Submit Spark Job

Die Aktivität **Submit Spark Job** ermöglicht die Ausführung eines beliebigen Spark-Auftrags auf einem Hadoop-Cluster oder auf einem Spark-Cluster. Mit dieser Aktivität können Sie entweder einen Spark-Auftrag des Spectrum™ Big Data Quality SDK oder einen beliebigen externen Spark-Auftrag ausführen.

Momentan können Sie einen Spark-Auftrag an einen der beiden Clustertypen übermitteln:

- YARN
- Spark

Bereitstellungsmodi

Für einen Spark-Auftrag können Sie den Bereitstellungsmodus *Cluster* oder *Client* verwenden. Diese Bereitstellungsmodi legen fest, ob der Treiber für den Spark-Auftrag auf dem *Cluster* oder auf dem *Client*-Spectrum™ Technology Platform ausgeführt wird.

Einfach ausgedrückt, können Sie einen Spark-Auftrag in einem beliebigen der Bereitstellungsmodi ausführen:

1. YARN-Clustermodus
2. YARN-Clientmodus
3. Spark-Clientmodus

Achtung: Der YARN- oder der Spark-*Client*modus wird zur Ausführung empfohlen, wenn der Spectrum-Server in der Clusterumgebung installiert ist und dort ausgeführt wird.

Feld	Beschreibung
Auftragsname	Der Name des Spark-Auftrags.
Hadoop-Server	Die Liste der konfigurierten Hadoop-Server. Informationen zum Mapping von HDFS-Dateiservern über die Management Console finden Sie im *Administratorhandbuch*.
JAR-Pfad	Der Pfad der relevanten JAR-Datei für den auszuführenden Spark-Auftrag. **Anmerkung:** Der JAR-Pfad muss auf ein Verzeichnis auf dem Spectrum-Servercomputer verweisen.
Auftragstyp	Wählen Sie eine Option: **Spectrum** — Wählen Sie `Spectrum` aus, um einen beliebigen Spectrum Big Data Quality SDK-Auftrag auszuführen. Wenn Sie `Spectrum` auswählen, wird das Feld **Spectrum-Aufträge** angezeigt. **Allgemein** — Wählen Sie `Allgemein` aus, um weitere Eigenschaften für beliebige externe Aufträge anzugeben.

Feld	Beschreibung
Spectrum-Aufträge	Wählen Sie einen Spectrum-Auftrag aus der Liste der Spectrum Big Data Quality SDK-Aufträge aus. Bei der Auswahl des gewünschten Spectrum-Auftrags: 1. Die Felder **Auftragsname**, **Klassenname** und **Argument** werden automatisch ausgefüllt. Alle automatisch ausgefüllten Felder außer **Klassenname** können bei Bedarf bearbeitet werden. **Wichtig:** Bei dem ausgewählten Spectrum-Auftrag darf der automatisch ausgefüllte **Klassenname** nicht bearbeitet werden, da der Auftrag ansonsten nicht ausgeführt werden kann. 2. Das Raster **Eigenschaften** wird automatisch mit den erforderlichen Konfigurationseigenschaften des ausgewählten Spectrum-Auftrags und deren Standardwerten ausgefüllt. Sie können weitere Eigenschaften hinzufügen oder importieren sowie die automatisch ausgefüllten Eigenschaften nach Bedarf ändern.
Klassenname	Der vollständig qualifizierte Name der Treiberklasse des Auftrags.

Feld	Beschreibung

Argumente	Die durch Leerzeichen getrennte Liste von Argumenten. Diese werden zur Laufzeit zur Ausführung des Auftrags an die Treiberklasse weitergeleitet.
	
	Beispiel:
	
	`23Dec2016 /home/Hadoop/EYInc.txt`
	
	1. Diese Variablen können als Argumente weitergeleitet werden, die definiert sind, zur Laufzeit Werte entweder im Quellschritt oder in diesem aktuellen Schritt des Prozessflusses zu akzeptieren.
	
	Wenn beispielsweise in der Ausgabe des vorherigen Schritts des Prozessflusses die Variable `SalesStartRange` definiert wurde, können Sie diese Variable in dieser Argumentenliste als `${SalesStartRange}` zusammen mit anderen erforderlichen Argumenten einschließen. Beispiel:
	
	`23Dec2016 /home/Hadoop/EYInc.txt ${SalesStartRange}`
	
	2. Wenn ein bestimmtes Argument ein Leerzeichen enthält, schließen Sie dieses Argument in doppelte Anführungszeichen ein. Beispiel: `"/home/Hadoop/Sales Records"`.
	
	Aufträge des Spectrum Big Data Quality SDK – Argumente:
	
	Um die Spectrum Big Data Quality SDK *Spark*-Aufträge auszuführen, übergeben Sie die verschiedenen Konfigurationsdateien als Liste von Argumenten. Jeder Argumentschlüssel akzeptiert den Pfad einer einzelnen Konfigurationseigenschaftsdatei, wobei jede Datei mehrere Konfigurationseigenschaften enthält.
	
	Die Syntax der Argumentenliste für Konfigurationseigenschaften ist:
	
	`[-config <Path to configuration file>] [-debug] [-input <Path to input configuration file>] [-conf <Path to Spark configuration file>] [-output <Path of output directory>]`
	
	Beispiel für einen Spark MatchKeyGenerator-Auftrag:
	
	`-config /home/hadoop/spark/matchkey/matchKeyGeneratorConfig.xml -input /home/hadoop/spark/matchkey/inputFileConfig.xml -output /home/hadoop/spark/matchkey/outputFileConfig.xml`
	
	Anmerkung: Wenn derselbe Konfigurationseigenschaftsschlüssel sowohl im Feld **Argumente** als auch im Raster **Eigenschaften** angegeben ist, aber jeder Schlüssel auf eine andere Konfigurationsdatei verweist, hat die Datei Vorrang, auf die im Raster **Eigenschaften** verwiesen wird.
	
	Die Beispielkonfigurationseigenschaften werden mit dem Big Data Quality SDK geliefert und befinden sich unter `<Big Data Quality bundle>\samples\configuration`.
	
	.

Allgemeine Eigenschaften

Feld	Beschreibung
Master	Wählen Sie eine der Optionen für die Ausführung des Spark-Auftrags aus: **YARN** Zum Starten und Verwalten des Spark-Auftrags über YARN. **Spark** Zum Starten und Verwalten des Spark-Auftrags über eine Spark-Anwendung.
Spark-URL	Die URL für den Zugriff auf der Spark-Cluster im Format `<hostname of Spark cluster>:<port of Spark cluster>`. Dieses Feld wird angezeigt, wenn Sie im Feld **Master** `Spark` auswählen.
Bereitstellungsmodus	Wählen Sie eine beliebige Option aus: **Client** Zur Ausführung des Spark-Auftragstreibers auf Client-Spectrum™ Technology Platform. **Cluster** Zur Ausführung des Spark-Auftragstreibers auf einem Cluster.

Feld	Beschreibung
Eigenschaften	

Feld	Beschreibung

Geben Sie im Raster in der Spalte **Eigenschaft** die Namen der Eigenschaften und in der Spalte **Wert** die Werte der zugehörigen Eigenschaften ein.

Abhängig vom Typ des **Masters** und des **Bereitstellungsmodus** sind einige Eigenschaften obligatorisch.

Obligatorische Eigenschaften für YARN	
`yarn.resourcemanager.hostname`	Die IP-Adresse des YARN ResourceManager.
`yarn.resourcemanager.address`	Die Adresse einschließlich der IP-Adresse und des Ports des YARN ResourceManager im Format `<Hostname>:<Port>`.

Eigenschaften für den Client-Bereitstellungsmodus		
`Spark.driver.host`	Die IP-Adresse des Computers, auf dem der Spark-Treiber ausgeführt werden soll.	Erforderlich
`spark.client.mode.temp.location`	Der Pfad des Ordners `temp` auf dem Spectrum-Server, der für die Universal Addressing-Aufträge verwendet werden soll: • Validate Address • Validate Address Global • Validate Address Loqate **Anmerkung:** Wir empfehlen dringend, diese Eigenschaft für die Universal Addressing-Aufträge zu verwenden, um sicherzustellen, dass der angegebene Ordner `temp` für	Optional

Feld	Beschreibung
Eigenschaften für den Client-Bereitstellungsmodus	Zwischenergebnisse verwendet wird.

Daher:

1. Für den YARN-Clustermodus sind die ersten zwei Eigenschaften obligatorisch.
2. Für den YARN-Clientmodus sind alle drei Eigenschaften obligatorisch.
3. Für den Spark-Clientmodus ist die dritte Eigenschaft obligatorisch.

Anmerkung: Sie können die oben aufgeführten obligatorischen Eigenschaften entweder bei der Erstellung der Verbindung in der Management Console oder in dieser Spark-Aktivität definieren. Wenn dieselben Eigenschaften sowohl in der Management Console als auch in der Spark-Auftragsaktivität definiert sind, dann werden die Werte angewendet, die in der Spark-Aktivität zugewiesen wurden.

Sie können zusätzlich zu diesen obligatorischen Eigenschaften so viele weitere Eigenschaften eingeben oder importieren, wie zur Ausführung des Auftrags erforderlich sind.

Feld	Beschreibung

Importieren	Um Eigenschaften aus einer Datei zu importieren, klicken Sie auf **Importieren**. Navigieren Sie zum Speicherort der jeweiligen Eigenschaftsdatei, und wählen Sie die Datei im XML-Format aus. Die in der importierten Datei enthaltenen Eigenschaften werden in das Raster **Eigenschaften** kopiert.
	
	Anmerkung:
	
	1. Wenn dieselbe Eigenschaft dort und in der Management Console definiert wurde, überschreiben die dort definierten Werte die in der Management Console definierten Werte.
	2. Die Eigenschaftsdatei muss im XML-Format vorliegen und dieser Syntax folgen:
	
	`<configuration>`
	`  <property>`
	`  <name>key</name>`
	`  <value>some_value</value>`
	`  <description>A brief description of the`
	`  purpose of the property key.</description>`
	`  </property>`
	`</configuration>`
	
	Erstellen Sie Ihre eigenen Eigenschaftsdateien unter Verwendung des oben angeführten XML-Formats.
	3. Wenn dieselbe Eigenschaft im Raster und auch in der importierten Eigenschaftsdatei vorhanden ist, dann überschreibt der aus der Datei importierte Wert den im Raster für dieselbe Eigenschaft vorhandenen Wert.
	4. Sie können falls erforderlich mehrere Eigenschaftsdateien nacheinander importieren. Die in den einzelnen importierten Dateien enthaltenen Eigenschaften werden im Raster hinzugefügt.
	5. Stellen Sie sicher, dass die Eigenschaftsdatei auf dem Spectrum™ Technology Platform-Server vorhanden ist.
	6. Das Tag `<description>` ist für alle Eigenschaftsschlüssel in einer Konfigurationseigenschaftsdatei optional.

Abhängigkeiten

Fügen Sie auf dieser Registerkarte die Liste mit den Eingabedateien und JAR-Dateien hinzu, die zur Ausführung des Auftrags erforderlich sind.

Nachdem der Auftrag ausgeführt wurde, sind die dort hinzugefügten Referenzdateien und JAR-Referenzdateien im verteilten Cache des Auftrags verfügbar.

Referenzdateien

Um die verschiedenen Dateien hinzuzufügen, die als Eingabe für die Ausführung des Auftrags erforderlich sind, klicken Sie auf **Hinzufügen**, navigieren Sie zum entsprechenden Speicherort auf Ihrem lokalen System oder Cluster und wählen Sie die gewünschte Datei aus.

Um eine beliebige, der Liste hinzugefügte Datei zu entfernen, wählen Sie diese Datei aus und klicken Sie auf **Entfernen**.

JAR-Referenzdateien

Um die JAR-Dateien hinzuzufügen, die für die Ausführung des Auftrags erforderlich sind, klicken Sie auf **Hinzufügen**, navigieren Sie zum entsprechenden Speicherort auf Ihrem lokalen System oder Cluster und wählen Sie die gewünschte Datei aus.

Um eine beliebige, der Liste hinzugefügte Datei zu entfernen, wählen Sie diese Datei aus und klicken Sie auf **Entfernen**.

Anmerkung: Der JAR-Pfad muss auf ein Verzeichnis auf dem Spectrum-Servercomputer verweisen.

Variablen

Eingaben
Wählen Sie in diesem Raster, aus den von einer Quellaktivität empfangenen Feldern, die Felder aus, die in dieser Aktivität verwendet werden sollen.

Ausgaben
Wählen Sie in diesem Raster die Felder aus, die in die Ausgabe für die Zielaktivität in diesem Prozessfluss eingeschlossen werden sollen.

Erfolg

Eine Aktivität vom Typ „Erfolg“ gibt das Ende eines Prozessflusses an. Ein Prozessfluss muss mindestens eine Aktivität vom Typ „Erfolg“ enthalten.

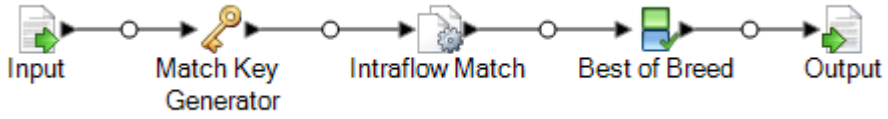
6 - Erstellen von wiederverwendbaren Flusskomponenten

In this section

Einführung in die Unterflüsse 226
Verwenden eines Unterflusses als Quelle 226
Verwenden eines Unterflusses in der Mitte eines Datenflusses 228
Verwenden eines Unterflusses als ein Laden von Daten 229
Ändern eines Unterflusses 230
Löschen eines Unterflusses 231
Verfügbarmachen eines Unterflusses und Aufhebung der Verfügbarkeit 231
Konvertieren eines Schrittes in einen Unterfluss 231

Einführung in die Unterflüsse

Ein Unterfluss ist ein Datenfluss, der innerhalb anderer Datenflüsse wiederverwendet werden kann. Unterflüsse sind nützlich, wenn Sie einen wiederverwendbaren Prozess erstellen möchten, der ohne großen Aufwand in Datenflüsse eingebunden werden kann. Sie könnten beispielsweise einen Unterfluss erstellen, der eine Deduplizierung ausführt und in jedem Schritt bestimmte Einstellungen nutzt, sodass Sie denselben Deduplizierungsprozess in mehreren Datenflüssen verwenden können. Zu diesem Zweck könnten Sie einen Unterfluss wie diesen erstellen:



Danach könnten Sie diesen Unterfluss in einem Datenfluss verwenden. Sie könnten beispielsweise den Deduplizierungsunterfluss innerhalb eines Datenflusses verwenden, der Geocoding ausführt, sodass die Daten vor der Geocoding-Operation dedupliziert werden:

In diesem Beispiel würden Daten aus einer Datenbank eingelesen und dann dem Deduplizierungsunterfluss übergeben, wo sie durch Match Key Generator, dann Intraflow Match, dann Best of Breed verarbeitet und schließlich aus dem Unterfluss heraus in den nächsten Schritt im übergeordneten Datenfluss, in diesem Fall „Geocode US Address“, gesendet würden. Unterflüsse werden durch ein Puzzleteil-Symbol im Datenfluss dargestellt (siehe oben).

Unterflüsse, die gespeichert und verfügbar gemacht werden, werden im Ordner **Benutzerdefinierte Schritte** im Enterprise Designer angezeigt.

Verwenden eines Unterflusses als Quelle

Sie können einen Unterfluss als ersten Schritt in einem Datenfluss verwenden, um Daten aus einer Quelle zu lesen oder sogar einige Verarbeitungen der Daten auszuführen, bevor sie an den übergeordneten Datenfluss weitergeleitet werden. Sie können einen Unterfluss als einfachen Quellschritt erstellen, die so konfiguriert ist, dass Sie sie in mehreren Datenflüssen wiederverwenden können. Oder Sie können einen komplexeren Unterfluss erstellen, der Daten liest und sie anschließend verarbeitet, bevor sie an den übergeordneten Datenfluss weitergeleitet werden.

1. Klicken Sie im Enterprise Designer auf **Datei** > **Neu** > **Datenfluss** > **Unterfluss**.

2. Ziehen Sie die entsprechende Datenquelle aus der Palette auf die Arbeitsfläche und konfigurieren Sie sie.

   Wenn der Unterfluss z. B. Daten aus einer kommagetrennten Datei lesen soll, ziehen Sie einen „Read from File“-Schritt auf die Arbeitsfläche.

3. Wenn der Unterfluss die Daten auf bestimmte Weise verarbeiten soll, bevor sie an den übergeordneten Datenfluss weitergeleitet werden, fügen Sie zusätzliche Schritte nach Bedarf hinzu, um die gewünschte Vorverarbeitung durchzuführen.
4. Fügen Sie am Ende des Datenfluss einen Output-Schritt hinzu und konfigurieren Sie ihn.

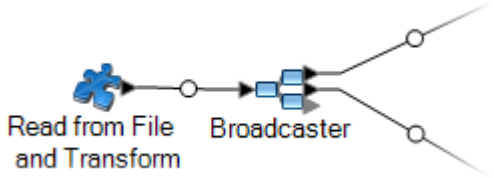
   Dadurch können die Daten aus dem Unterfluss an den übergeordneten Datenfluss gesendet werden.

   Wenn Sie z. B. einen Unterfluss erstellt haben, der Daten aus einer Datei liest und anschließend einen Transformer-Schritt zur Kürzung von Leerzeichen und zur Standardisierung von Klein- und Großschreibung verwendet, würde Ihr Unterfluss wie folgt aussehen:

5. Doppelklicken Sie auf den Output-Schritt, und wählen Sie die Felder aus, die an den übergeordneten Datenfluss weitergeleitet werden sollen.
6. Wählen Sie **Datei** > **Speichern** aus, und speichern Sie den Unterfluss.
7. Wählen Sie **Datei** > **Verfügbar machen** aus, um den Unterfluss in Datenflüsse einbeziehen zu können.
8. Ziehen Sie im Datenfluss, in den Sie den Unterfluss einbeziehen möchten, den Unterfluss aus der Palette in die Arbeitsfläche.
9. Verbinden Sie den Unterfluss mit dem gewünschten Datenflussschritt.

   Anmerkung: Da der Unterfluss einen Quellschritt anstelle eines Input-Schrittes enthält, zeigt das Unterflusssymbol nur einen Ausgabeport. Er kann nur als Quelle im Datenfluss verwendet werden.

   Der übergeordnete Datenfluss verwendet nun den Unterfluss, den Sie als Eingabe erstellt haben. Wenn Sie z. B. einen Unterfluss mit dem Namen „Read from File and Transform“ erstellt haben und den Unterfluss hinzufügen und ihn mit einem Broadcaster-Schritt verbinden, würde Ihr Datenfluss wie folgt aussehen:



Verwenden eines Unterflusses in der Mitte eines Datenflusses

Sie können einen Unterfluss in der Mitte eines Datenflusses verwenden, um Verarbeitungen durchzuführen, die Sie in anderen Datenflüssen erneut verwenden möchten. Dabei wird der Unterfluss zu einem benutzerdefinierten Schritt in Ihrem Datenfluss.

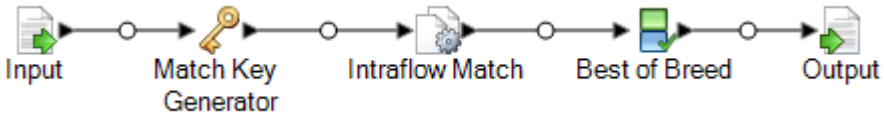
1. Klicken Sie im Enterprise Designer auf **Datei** > **Neu** > **Datenfluss** > **Unterfluss**.
2. Ziehen Sie einen Input-Schritt aus der Palette auf die Arbeitsfläche.

   Dadurch können die Daten aus dem übergeordneten Datenfluss an den Unterfluss übermittelt werden.

3. Doppelklicken Sie auf den Input-Schritt, und fügen Sie die Felder hinzu, die der Unterfluss vom Datenfluss erhalten soll, in dem er verwendet wird.
4. Fügen Sie nach der Konfiguration des Input-Schrittes zusätzliche Schritte nach Bedarf hinzu, um die gewünschte Verarbeitung auszuführen.
5. Fügen Sie am Ende des Datenflusses einen Output-Schritt hinzu.

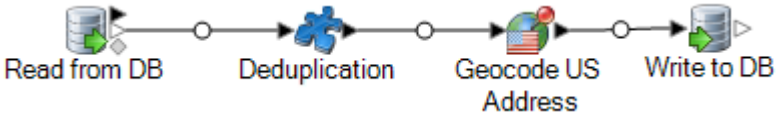
   Dadurch können die Daten aus dem Unterfluss an den übergeordneten Datenfluss gesendet werden.

   Sie könnten beispielsweise einen Unterfluss erstellen, der eine Deduplizierung ausführt und in jedem Schritt bestimmte Einstellungen nutzt, sodass Sie denselben Deduplizierungsprozess in mehreren Datenflüssen verwenden können. Zu diesem Zweck könnten Sie einen Unterfluss wie diesen erstellen:



6. Wählen Sie **Datei** > **Speichern** aus, und speichern Sie den Unterfluss.
7. Wählen Sie **Datei** > **Verfügbar machen** aus, um den Unterfluss in Datenflüsse einbeziehen zu können.
8. Ziehen Sie im Datenfluss, in den Sie den Unterfluss einbeziehen möchten, den Unterfluss aus der Palette in die Arbeitsfläche.
9. Verbinden Sie den Unterfluss mit dem gewünschten Datenflussschritt.

   Sie könnten beispielsweise den Deduplizierungsunterfluss innerhalb eines Datenflusses verwenden, der Geocoding ausführt, sodass die Daten vor der Geocoding-Operation dedupliziert werden:

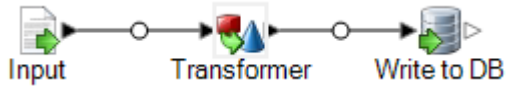


Verwenden eines Unterflusses als ein Laden von Daten

Sie können einen Unterfluss als letzten Schritt in einem Datenfluss verwenden, um Daten in eine Datei oder Datenbank zu schreiben oder sogar einige Verarbeitungen der Daten auszuführen, bevor sie in das Ausgabeziel geschrieben werden. Sie können einen Unterfluss als einfachen Datenladeschritt erstellen, die so konfiguriert ist, dass Sie sie in mehreren Datenflüssen wiederverwenden können. Oder Sie können einen komplexeren Unterfluss erstellen, der Daten auf gewisse Weise verarbeitet, bevor sie in das Ausgabeziel geschrieben werden.

1. Klicken Sie im Enterprise Designer auf **Datei** > **Neu** > **Datenfluss** > **Unterfluss**.
2. Ziehen Sie einen Input-Schritt aus der Palette auf die Arbeitsfläche.
3. Doppelklicken Sie auf den Input-Schritt, und fügen Sie die Felder hinzu, die der Unterfluss vom Datenfluss erhalten soll, in dem er verwendet wird.
4. Fügen Sie nach der Konfiguration des Input-Schrittes zusätzliche Schritte nach Bedarf hinzu, um die gewünschte Nachverarbeitung auszuführen.
5. Fügen Sie am Ende des Datenflusses das entsprechende Laden von Daten hinzu.

   Wenn Sie z. B. einen Unterfluss erstellt haben, der einen Transformer-Schritt zur Kürzung von Leerzeichen und zur Standardisierung der Groß- und Kleinschreibung eines Feldes verwendet, und die Daten anschließend in eine Datenbank schreibt, würde der Unterfluss wie folgt aussehen:

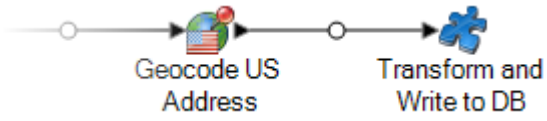
   

6. Wählen Sie **Datei** > **Speichern** aus, und speichern Sie den Unterfluss.
7. Wählen Sie **Datei** > **Verfügbar machen** aus, um den Unterfluss in Datenflüsse einbeziehen zu können.
8. Ziehen Sie im Datenfluss, in den Sie den Unterfluss einbeziehen möchten, den Unterfluss aus der Palette auf die Arbeitsfläche und verbinden Sie ihn mit dem letzten Schritt im Datenfluss.

   Anmerkung: Da der Unterfluss einen Schritt zum Laden von Daten anstelle eines Output-Schrittes enthält, zeigt das Unterflusssymbol nur einen Eingabeport. Er kann nur als ein Laden von Daten im Datenfluss verwendet werden.

   Der übergeordnete Datenfluss verwendet nun den Unterfluss, den Sie als ein Laden von Daten erstellt haben. Wenn Sie z. B. einen Unterfluss mit dem Namen „Transform and Write to DB“

erstellt haben und den Unterfluss hinzufügen und ihn mit einem „Geocode US Address“-Schritt verbinden, würde Ihr Datenfluss wie folgt aussehen:



Ändern eines Unterflusses

1. Öffnen Sie den Unterfluss im Enterprise Designer.
2. Berücksichtigen Sie eventuell vor dem Ändern des Unterflusses, welchen Einfluss die Änderung auf die Datenflüsse hat, die diesen Unterfluss verwenden. Wählen Sie zur Anzeige, welche Datenflüsse den Unterfluss nutzen, die Optionen **Tools** > **Verwendet von** aus.
3. Ändern Sie den Unterfluss wie gewünscht.

   Beachten Sie Folgendes:

   - Wenn Sie einen Input- oder Output-Schritt löschen oder einen zusätzlichen Input- oder Output-Schritt hinzufügen, zeigt der Enterprise Designer eine Warnmeldung an, die Sie daran erinnert, dass andere Datenflüsse den Unterfluss verwenden. Sie erhalten die Möglichkeit, diese Datenflüsse anzuzeigen. Wenn Sie den verwendbaren Schritt speichern, hebt der Enterprise Designer die Verfügbarkeit aller Datenflüsse auf, die von diesem Unterfluss verwendet werden.
   - Wenn Sie einen Unterfluss anderweitig ändern (z. B. durch Änderung eines Dateinamens oder der Schrittkonfigurationen), zeigt der Enterprise Designer eine Warnmeldung an, die Sie daran erinnert, dass andere Datenflüsse den Unterfluss verwenden. Sie erhalten die Möglichkeit, diese Datenflüsse anzuzeigen. Sie können fortfahren, ohne die Verfügbarkeit dieser Datenflüsse aufzuheben.

4. Wenn Sie die Änderungen abgeschlossen haben, wählen Sie **Datei** > **Speichern** aus.
5. Wählen Sie **Ansicht** > **Aktualisieren** aus, um die Änderungen im übergeordneten Datenfluss sichtbar zu machen.

   Anmerkung: Wenn Sie über mehr als eine Version des Unterflusses verfügen, denken Sie daran, dass die Version, die im übergeordneten Datenfluss verwendet wird, die verfügbar gemachte Version ist. Wenn Sie eine Änderung im Unterfluss vornehmen, stellen Sie sicher, dass die aktuelle Version verfügbar gemacht wird, damit Ihre Änderungen in den Datenflüssen in Kraft treten, die den Unterfluss verwenden.

Löschen eines Unterflusses

Wenn Sie einen verfügbar gemachten Unterfluss löschen möchten, zeigt der Enterprise Designer eine Warnmeldung an, die Sie erinnert, dass andere Datenflüsse den zu löschenden Unterfluss verwenden. Wenn Sie den Unterfluss dennoch löschen möchten, hebt der Enterprise Designer die Verfügbarkeit aller verbundenen Datenflüsse auf.

Verfügbarmachen eines Unterflusses und Aufhebung der Verfügbarkeit

Damit ein Unterfluss für die Nutzung innerhalb eines Datenflusses verfügbar ist, muss der Unterfluss verfügbar gemacht werden. Öffnen Sie zum Verfügbarmachen den Unterfluss im Enterprise Designer, und klicken Sie auf **Datei** > **Verfügbar machen/Verfügbarkeit aufheben und speichern**. Dadurch steht der Unterfluss für die Verwendung in anderen Datenflüssen zur Verfügung.

> **Anmerkung:** Wenn Sie über mehr als eine Version des Unterflusses verfügen, denken Sie daran, dass die Version, die im übergeordneten Datenfluss verwendet wird, die verfügbar gemachte Version ist. Wenn Sie eine Änderung im Unterfluss vornehmen, stellen Sie sicher, dass die aktuelle Version verfügbar gemacht wird, damit Ihre Änderungen in den Datenflüssen in Kraft treten, die den Unterfluss verwenden.

Öffnen Sie zur Aufhebung der Verfügbarkeit den Unterfluss im Enterprise Designer, und klicken Sie auf **Datei** > **Verfügbar machen/Verfügbarkeit aufheben und speichern**. Wenn Sie die Verfügbarkeit eines Unterflusses aufheben, zeigt der Enterprise Designer eine Warnmeldung an, die Sie erinnert, dass andere Datenflüsse den zu ändernden Unterfluss verwenden. Wenn Sie mit der Aufhebung der Verfügbarkeit fortfahren, hebt der Enterprise Designer die Verfügbarkeit aller Datenflüsse auf, die den Unterfluss verwenden.

Konvertieren eines Schrittes in einen Unterfluss

1. Erstellen Sie einen neuen Auftrag, Dienst oder Unterfluss.
2. Fügen Sie den Schritt hinzu, den Sie in den Auftrag, Dienst oder Unterfluss einbeziehen möchten.

3. Wenn Sie den Schritt zu diesem Zeitpunkt konfigurieren möchten, klicken Sie mit der rechten Maustaste auf den Schritt, und wählen Sie **Optionen** aus. Konfigurieren Sie anschließend die Schrittoptionen wie gewünscht, und klicken Sie auf **OK**.
4. Klicken Sie mit der rechten Maustaste auf den zu konvertierenden Schritt, und wählen Sie **Schritt in Unterfluss umwandeln** aus. Das Dialogfeld **Speichern unter** wird angezeigt.
5. Geben Sie den Namen ein, den Sie dem Unterfluss geben möchten, und klicken Sie auf **OK**. Speichern Sie anschließend den Dienst. Der Name muss im System eindeutig sein. Daraufhin finden drei Dinge statt:
   - Das System erstellt einen neuen Unterfluss, der Folgendes enthält:
     - den von Ihnen ausgewählten Schritt
     - eine Datenflusseingabe für jeden Eingangsport im Schritt
     - eine Datenflussausgabe für jeden Ausgangsport im Schritt
     - Verbindungen zwischen dem Schritt und seinen Ein- und Ausgaben
   - Das System ersetzt Ihren ausgewählten Schritt durch den neuen Unterfluss.
   - Das System macht den neuen Unterfluss verfügbar. Er wird im Server-Explorer und im Abschnitt „Benutzerdefinierte Schritte“ in der Toolbox angezeigt.

Nachdem Sie einen Unterfluss erstellt und ihn in anderen Datenflüssen verwendet haben, können Sie sehen, welche Datenflüsse den Unterfluss verwenden. Öffnen Sie den Unterfluss, und wählen Sie **Tools** > **Verwendet von** aus. (Sie können alternativ im Server-Explorer mit der rechten Maustaste auf den Unterfluss klicken und **Verwendet von** auswählen.) Dadurch wird eine Liste von Datenflüssen angezeigt, die den aktuellen Unterfluss verwenden. So können Sie sehen, welche Datenflüsse beeinflusst werden würden, wenn Sie den aktuellen Unterfluss ändern.

7 - Info zu Spectrum™ Technology Platform

In this section

Was ist Spectrum™ Technology Platform? 234
Architektur der Unternehmensdatenverwaltung 235
Architektur von Spectrum™ Technology Platform 239
Module und Komponenten 244

Was ist Spectrum™ Technology Platform?

Spectrum™ Technology Platform ist ein System, mit dem die Vollständigkeit, die Gültigkeit, die Konsistenz, die Zeitgenauigkeit und die Genauigkeit Ihrer Daten durch Datenstandardisierung, -überprüfung und -erweiterung verbessert werden. Durch die Sicherstellung, dass Ihre Daten genau, vollständig und auf dem neuesten Stand sind, kann Ihr Unternehmen Ihre Kunden besser verstehen und besser mit ihnen arbeiten.

Spectrum™ Technology Platform hilft beim Entwerfen und Implementieren von Geschäftsregeln für die Datenqualität, indem folgende Funktionen ausgeführt werden.

Parsing, Namensstandardisierung und Namensüberprüfung

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-/Extraktionsvorgänge verwendet werden sollen. Das Universal Name-Modul bietet diese Funktionalität.

Deduplizierung und Konsolidierung

Über das Identifizieren eindeutiger Entitäten können Sie Datensätze konsolidieren, Dubletten beseitigen und „Best of Breed“-Datensätze entwickeln. Ein „Best of Breed“-Datensatz ist ein aus Daten anderer Datensätze zusammengesetzter Datensatz. Das Advanced Matching-Modul und das Data Normalization-Modul bieten diese Funktionalität.

Adressenüberprüfung

Die Adressenüberprüfung wendet Regeln der entsprechenden Postbehörde an, um eine Adresse in eine Standardform zu überführen, und überprüft sogar, ob diese Adresse lieferfähig ist. Die Adressenüberprüfung kann Ihnen helfen, sich für Portoermäßigungen zu qualifizieren und die Lieferfähigkeit Ihrer Sendungen zu verbessern. Das Universal Addressing-Modul und das Address Now-Modul bieten diese Funktionalität.

Geocoding

Geocoding ist die Bestimmung der geografischen Koordinaten (Längen- und Breitengrad) einer Adresse. Geocoding kann u. a. zur Erstellung von Karten verwendet werden. Die zugrunde liegenden Standortdaten können z. B. auch bei der Entscheidungsfindung in Unternehmen helfen. In der Umkehrung des Prozesses können Sie auch einen Geocode eingeben – einen durch eine Breiten- und Längengradkoordinate dargestellten Punkt – und Adressinformationen zum Geocode erhalten. Das Enterprise Geocoding-Modul bietet diese Funktionalität.

Location Intelligence

Location Intelligence erstellt neue Informationen zu Ihren Daten, indem geografische Beziehungen bewertet, überprüft, analysiert und modelliert werden. Mit dem Einsatz von Location Intelligence können Sie Standorte überprüfen und Informationen in wertvolle Business Intelligence verwandeln. Das Location Intelligence-Modul bietet diese Funktionalität.

Master Data Management

Mit Master Data Management können Sie beziehungsorientierte Master Data-Ansichten Ihrer wichtigsten Datenbestände erstellen. Das Data Hub-Modul hilft Ihnen, Beeinflusser und nicht offensichtliche Beziehungen zu identifizieren, Betrug zu erkennen und die Qualität, Integration und Verfügbarkeit Ihrer Informationen zu verbessern.

Zuweisung von Steuerhoheit

Die Zuweisung von Steuerhoheit bestimmt die Steuerhoheiten, die für den Standort einer Adresse gelten. Das genaue Zuweisen von richtigen Steuerhoheiten kann das finanzielle Risiko und die gesetzliche Haftung verringern.

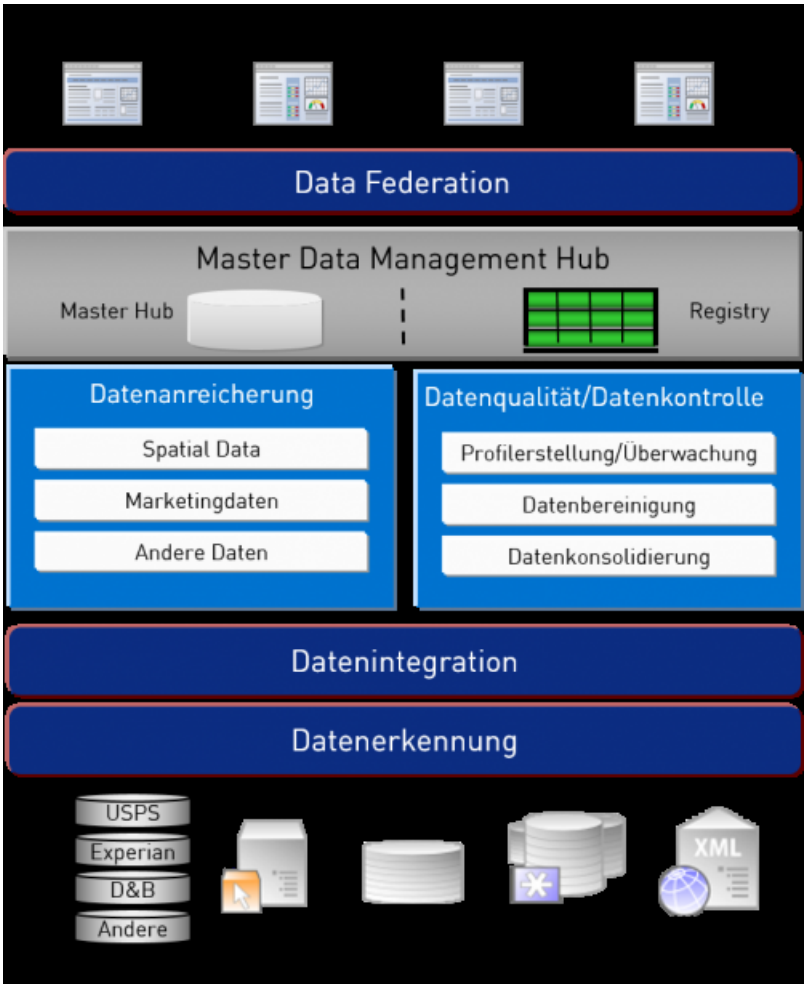
Die Software Spectrum™ Technology Platform von Pitney Bowes integriert die aktuellen Zuständigkeitsgrenzen über die exakten Straßenanschriften in Ihren Kundendatensätzen. So können Sie Ihren Datensätzen die korrekten Steuerinformationen bezüglich Bundesstaat, Bezirk, Township, Gemeinde und besonderer Steuergebiete hinzufügen. Einige Beispielanwendungen für die Zuweisung von Steuerhoheit sind:

- Umsatz- und Gebrauchssteuer
- persönliche Grundsteuer
- Versicherungsbeitragssteuer

Das Enterprise Tax-Modul bietet diese Funktionalität.

Architektur der Unternehmensdatenverwaltung

Mit Spectrum™ Technology Platform können Sie einen umfassenden Verwaltungsprozess für Unternehmensdaten aufbauen oder eine gezieltere Lösung anstreben. Das folgende Diagramm zeigt eine vollständige Lösung, die Quelldaten durch Datenanreicherungs- und Datenqualitätsprozesse leitet und einen Master Data Management Hub damit anreichert, der die Daten über ein einzelne Ansicht für mehrere Unternehmensanwendungen bereitstellt.



Datenerkennung

Datenerkennung ist ein Prozess, bei dem Ihre Datenquellen gescannt werden, um ein vollständiges Inventar Ihrer Datenlandschaft zu erstellen. Spectrum™ Technology Platform kann über eine Vielzahl an Techniken zur Datenprofilerstellung strukturierte Daten, unstrukturierte Daten und semistrukturierte Daten scannen. Die Ergebnisse des Scans werden verwendet, um automatisch eine Dokumentationsbibliothek zu generieren, die die Datenbestände Ihres Unternehmens beschreibt, und um eine Metadatendatenbank zu erstellen. Diese Dokumentation und die begleitende Metadatendatenbank bieten die Einsichten, die Sie benötigen, bevor Sie mit Projekten bezüglich Datenintegration, Datenqualität, Datenkontrolle oder Master Data Management beginnen.

Wenden Sie sich für weitere Informationen zum Spectrum™ Technology Platform Data Discovery-Modul an Ihren Kundenbetreuer.

Data Integration

Sobald Sie über ein Inventar Ihrer Datenlandschaft verfügen, müssen Sie sich überlegen, wie Sie auf die Daten, die Sie verwalten müssen, zugreifen. Spectrum™ Technology Platform kann direkt oder durch Integration in Ihre vorhandenen Datenzugriffstechnologien eine Verbindung mit Daten

in mehreren Quellen herstellen. Batch- und Echtzeitfunktionen für Datenintegration werden für verschiedenste Geschäftsanforderungen, einschließlich Data Warehousing, Datenqualität, Systemintegration und Migration, unterstützt. Spectrum™ Technology Platform kann auf Daten unter anderem in RDBMS-Datenbanken, Data Warehouses, XML-Dateien und Flatfiles zugreifen. Spectrum™ Technology Platform unterstützt SQL-Abfragen mit komplexen Zusammenführungen und Aggregationen und bietet ein visuelles Tool zu Abfragenentwicklung. Zudem kann Spectrum™ Technology Platform über REST- und SOAP-Webservices auf Daten zugreifen.

Spectrum™ Technology Platform kann aufgrund des Vorhandenseins einer oder mehrerer Quelldateien in einem angegebenen Ordner Batch-Verarbeitungen auslösen. Dieses Auslösen anhand eines Überwachungsordners ist nützlich, um Hochladen über FTP zu überwachen und nach Abschluss die Datei sofort zu verarbeiten.

Einige dieser Datenintegrationsfunktionen erfordern eine Lizenz für das Enterprise Data Integration-Modul. Wenden Sie sich für weitere Informationen an Ihren Kundenbetreuer.

Schließlich lässt sich Spectrum™ Technology Platform noch in paketisierte Anwendungen wie SAP und Siebel integrieren.

Datenqualität/Datenkontrolle

Prozesse für Datenqualität und Datenkontrolle überprüfen Ihre Daten auf Datensatzdubletten und auf inkonsistente sowie ungenaue Informationen.

Dublettenabgleich identifiziert mögliche Datensatzdubletten oder Beziehungen zwischen Datensätzen anhand von Namen und Adresse oder einem beliebigen anderen Informationstyp. Mit Spectrum™ Technology Platform können Sie über boolesche Vergleichsmethoden, Bewertungsmethoden, Schwellenwerte, Algorithmen und Gewichtungen ein konsistentes Set geschäftlicher Vergleichsregeln spezifizieren und feststellen, ob eine Gruppe von Datensätzen Dubletten enthält. Spectrum™ Technology Platform ist äußerst anpassungsfähig, sodass Sie die Regeln auf die einzigartigen Anforderungen Ihres Unternehmens zuschneiden können.

Sobald Datensatzdubletten identifiziert wurden, können Sie die Datensätze konsolidieren. Mit Spectrum™ Technology Platform sind Sie in der Lage, zu spezifizieren, wie Datensatzdubletten verknüpft oder zusammengeführt werden sollen, damit Sie den genauesten und vollständigsten Datensatz aus den gesammelten Informationen erstellen können. Beispielsweise kann aus allen Datensätzen für einen Haushalt ein einzelner Best-of-Breed-Datensatz erstellt werden. Das Advanced Matching-Modul wird verwendet, um Dubletten zu erkennen und zu beseitigen.

Datenqualitätsprozesse können zudem Ihre Daten standardisieren. Standardisierung ist ein wichtiger Prozess, da standardisierte Datenelemente erforderlich sind, um beim Abgleichen und Identifizieren von Beziehungen zwischen Datensätzen die bestmöglichen Ergebnisse zu erzielen. Während verschiedene Module eine Standardisierung des einen oder anderen Typs durchführen, bietet das Spectrum™ Technology Platform Data Normalization-Modul das umfassendste Set an Standardisierungsfunktionen. Darüber hinaus bietet das Universal Name-Modul bestimmte Datenqualitätsfunktionen zur Verarbeitung von Personennamen und Unternehmensnamen.

Standardisierte Daten sind nicht notwendigerweise genaue Daten. Spectrum™ Technology Platform kann Ihre Daten anhand bekannter, aktueller Referenzdaten auf ihre Richtigkeit hin überprüfen. Die für diesen Prozess verwendeten Quellen können von Behörden wie dem U.S. Postal Service, von

externen Datenanbietern wie Experian oder D&B oder von internen Referenzquellen Ihres Unternehmens wie Buchhaltungsdaten stammen. Spectrum™ Technology Platform ist besonders leistungsstark bei der Adressdatenüberprüfung. Adressen in 250 Ländern und Territorien auf der ganzen Welt können überprüft und standardisiert werden. Es gibt zwei Module zur Adressenüberprüfung: das Address Now-Modul und das Universal Addressing-Modul.

Besprechen Sie Ihre Anforderungen mit Ihrem Kundenbetreuer, um festzustellen, welches Modul für Sie das richtige ist.

Obwohl Spectrum™ Technology Platform automatisch viele Datenqualitätsprobleme beheben kann, gibt es einige Situationen, in denen eine manuelle Überprüfung durch einen Datenverantwortlichen erfolgen sollte. Um dies zu unterstützen, bietet das Business Steward-Modul eine Möglichkeit, die Regeln zu spezifizieren, die eine manuelle Überprüfung auslösen. Zudem bietet das Modul ein webbasiertes Tool zur Überprüfung von Ausnahmedatensätzen. Es beinhaltet integrierten Zugriff auf Drittanbietertools wie Bing-Karten und Experian-Daten, um Datenverantwortlichen im Prozess der Überprüfung und Auflösung zu helfen.

Datenanreicherung

Prozesse zur Datenanreicherung erweitern Ihre Daten um zusätzliche Informationen. Die Anreicherung kann auf Geodaten, Marketingdaten oder Daten aus anderen Quellen basieren, die Sie verwenden möchten, um Ihre Daten um zusätzliche Details zu erweitern. Wenn Sie beispielsweise über eine Datenbank mit Kundenadressen verfügen, könnten Sie die Adressen geocodieren, um Breitengrad/Längengrad-Koordinaten der Adressen zu ermitteln, und diese Koordinaten als Teil der Datensätze speichern. Sie könnten dann Ihre Kundendaten für verschiedene Geodatenberechnungen verwenden, wie das Suchen nach der nächsten Bankfiliale zu einer Kundenadresse. Mit Spectrum™ Technology Platform können Sie Ihre Daten um verschiedenste Informationen erweitern, einschließlich Geocoding (über das Enterprise Geocoding-Modul), Steuerhoheitszuweisungen (über das Enterprise Tax-Modul), Geodatenberechnungen (über das Location Intelligence-Modul) und Richtungsanweisungen zum Fahren oder Gehen zwischen Punkten (über das Enterprise Routing-Modul).

Master Data Management Hub

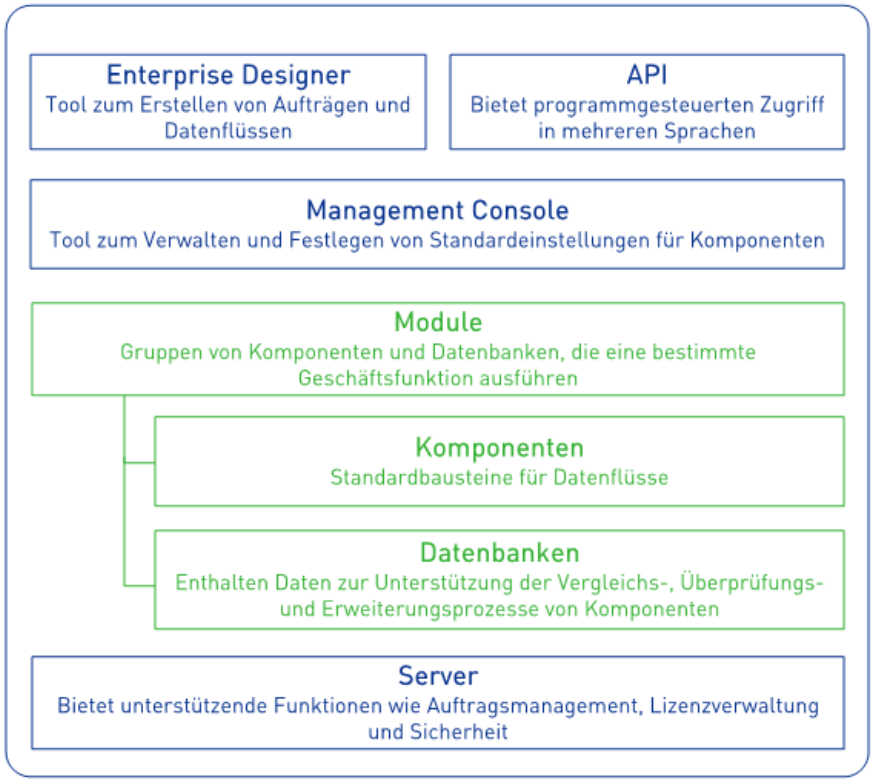
Der Master Data Management (MDM) Hub ermöglicht schnelle Modellierung von Entitäten und ihren komplexen Beziehungen über Rollen, Prozesse und Interaktionen hinweg. Er bietet integrierte Funktionen zur Analyse von soziale Netzwerken, um Ihnen zu helfen, Beeinflusser zu verstehen, Kundenabwanderungen vorherzusagen, nicht offensichtliche Beziehungen und Betrugsmuster zu erkennen und Empfehlungen zu geben.

Spectrum™ Technology Platform unterstützt zwei Ansätze für den MDM Hub. Im Master Hub-Ansatz werden die Daten in einer einzelnen MDM-Datenbank gepflegt und Anwendungen greifen auf Daten in der MDM-Datenbank zu. Im Registrierungsansatz werden die Daten in den einzelnen Unternehmensanwendungen gepflegt und die MDM Hub-Registrierung enthält Schlüssel, die zum Suchen nach zugehörigen Datensätzen verwendet werden. Beispielsweise kann ein Kundendatensatz in einer Auftragserfassungsdatenbank und in einer Kundendienstdatenbank vorhanden sein. Die MDM-Registrierung würde dann einen einzelnen Schlüssel enthalten, der an beiden Speicherorten für den Zugriff auf die Kundendaten verwendet werden kann.

Das Data Hub-Modul bietet MDM-Funktionen.

Architektur von Spectrum™ Technology Platform

Spectrum™ Technology Platform von Pitney Bowes besteht aus einem Server, der eine Anzahl von Modulen ausführt. Diese Module bieten verschiedene Funktionen wie unter anderem Adressenüberprüfung, Geocoding und erweitertes Parsing. Im folgenden Diagramm wird die Architektur von Spectrum™ Technology Platform veranschaulicht.



Server

Die Basis von Spectrum™ Technology Platform ist der Server. Der Server übernimmt die Datenverarbeitung, synchronisiert Datenbankdaten und verwaltet die Kommunikation. Sie bietet Auftragsmanagement- und Sicherheitsfeatures.

Module

Module sind Sets von Features, die eine bestimmte Funktion ausführen. Das Universal Addressing-Modul standardisiert Adressen beispielsweise, damit diese Adressen postalischen Standards entsprechen. Das Enterprise Tax-Modul bestimmt die Steuerhoheit für eine angegebene

Adresse. Module werden zusammen gruppiert, um häufige Geschäftsprobleme zu lösen, und zusammen als Pakete lizenziert.

Komponenten

Module sind aus Komponenten zusammengesetzt, die eine bestimmte Funktion in einem Fluss oder als ein Dienst ausführen. Die Komponente „Geocode US Address“ des Enterprise Geocoding-Moduls gibt beispielsweise den Längen- und Breitengrad für eine Eingabeadresse zurück, die Komponente „Get City State Province“ des Universal Addressing-Moduls gibt den Ort und das Bundesland oder den Kanton einer eingegebenen Postleitzahl zurück.

Welche Komponenten auf Ihrem System verfügbar sind, hängt davon ab, welches Spectrum™ Technology Platform-Bundle Sie lizenziert haben.

Datenbanken

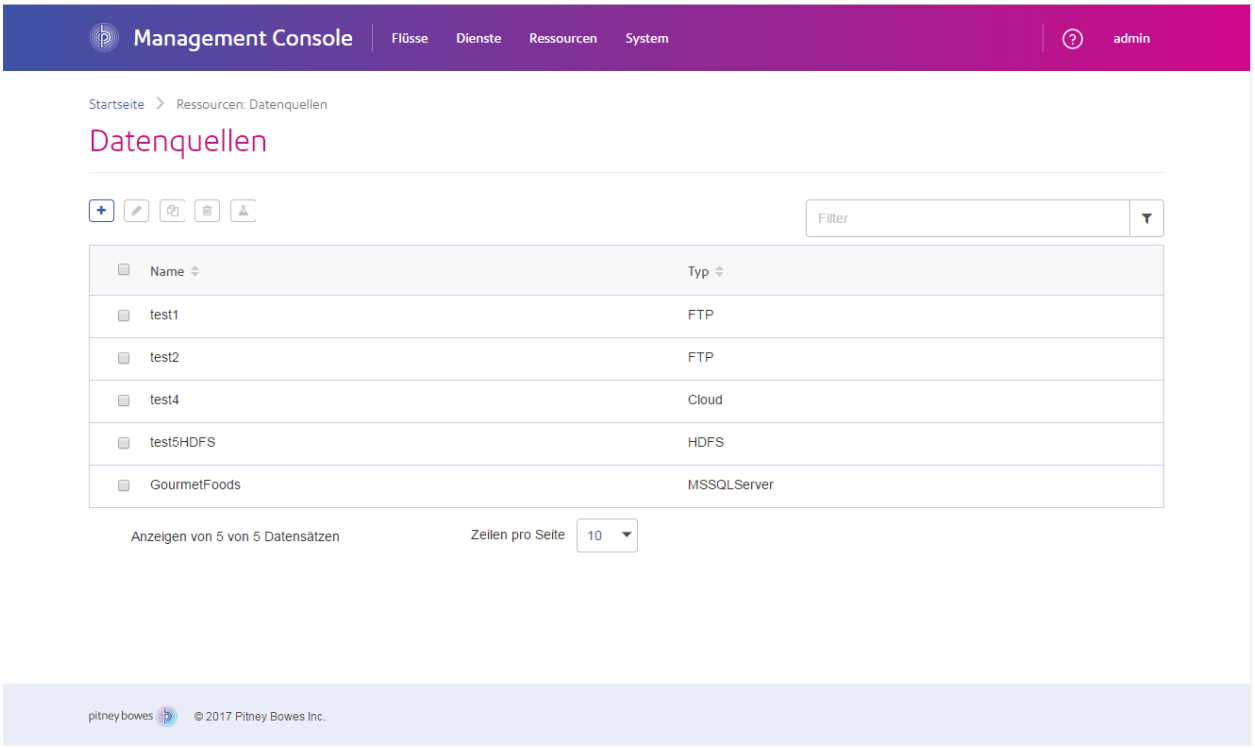
Einige Module sind von Datenbanken abhängig, die Referenzdaten enthalten. Das Universal Addressing-Modul muss beispielsweise Zugriff auf Daten des US-amerikanischen Postdiensts haben, um Adressen in den USA zu überprüfen und zu standardisieren. Datenbanken werden separat installiert und einige von ihnen werden regelmäßig aktualisiert, um Ihnen aktuelle Daten bereitzustellen.

Module verfügen neben erforderlichen Datenbanken über optionale Datenbanken. Optionale Datenbanken enthalten Daten, die für bestimmte Features erforderlich sind, die Ihren Spectrum™ Technology Platform-Prozess erweitern können.

Management Console

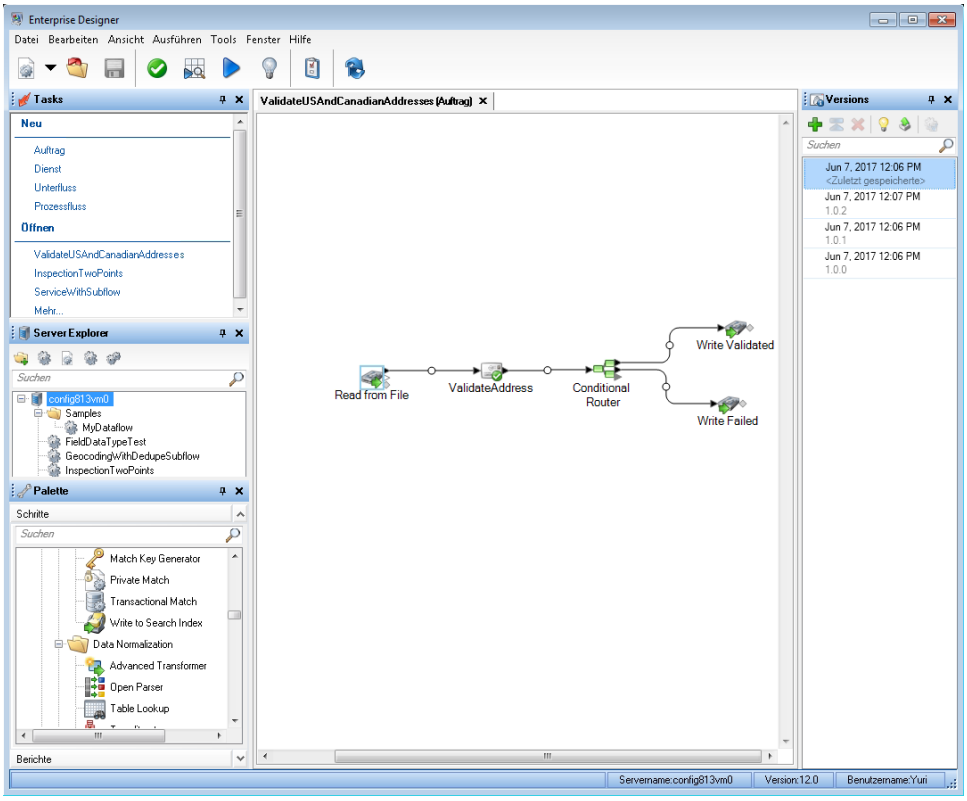
Die Management Console ist ein Administrationstool für Spectrum™ Technology Platform. Sie können die Management Console für folgende Aktionen verwenden:

- Definieren der Verbindungen zwischen Spectrum™ Technology Platform und Ihren Daten
- Spezifizieren der Standardeinstellungen für Dienste und Flüsse
- Verwalten von Benutzerkonten einschließlich Berechtigungen und Kennwörtern
- Anzeigen von Protokollen
- Anzeigen von Lizenzen einschließlich Informationen zum Ablauf



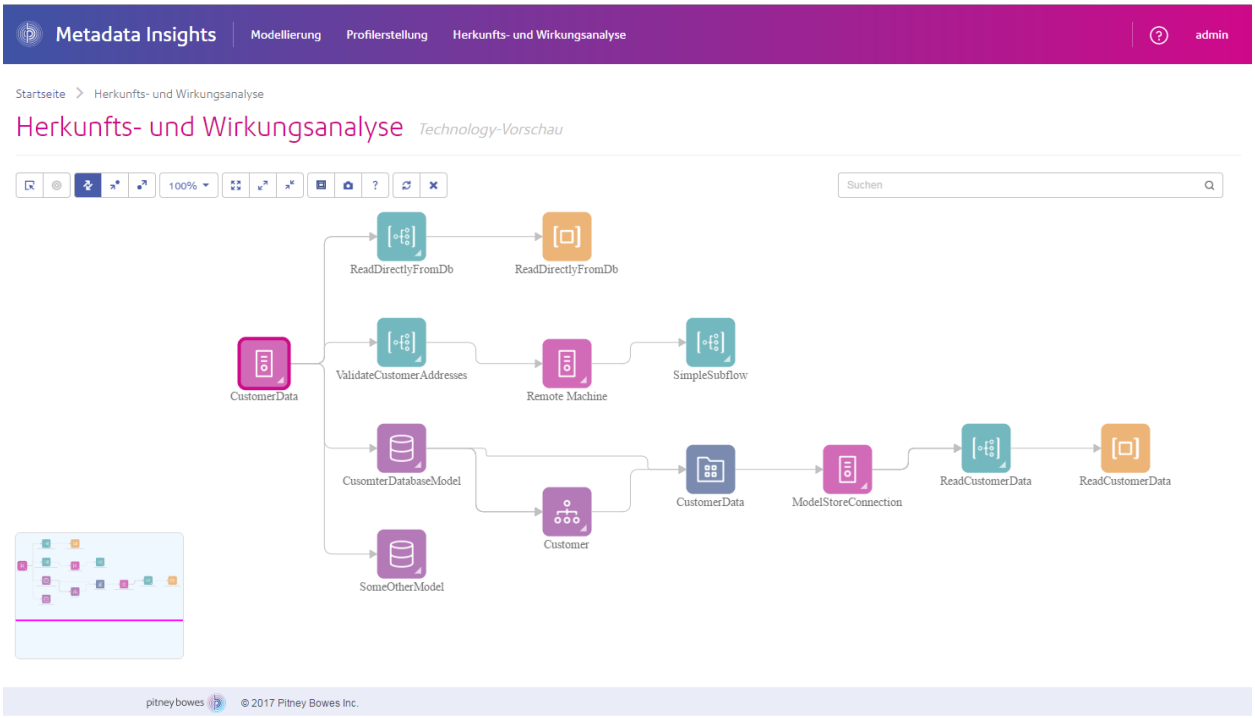
Enterprise Designer

Der Enterprise Designer ist ein Tool zum Erstellen von Aufträgen, Diensten, Unterflüssen und Prozessflüssen für Spectrum™ Technology Platform. Über die einfache Oberfläche können Sie durch Ziehen und Ablegen komplexe Datenflüsse grafisch erstellen.



Metadata Insights

Metadata Insights gibt Ihnen die Kontrolle, die Sie benötigen, um genaue und rechtzeitige datengesteuerte Einblicke in Ihr Geschäft zu geben. Mit Metadata Insights können Sie Datenmodelle entwickeln, den Fluss von Daten von der Quelle bis zur Geschäftsanwendung verfolgen und die Qualität Ihrer Daten durch Profiling bewerten. Mit diesen Einblicken können Sie die Datenressourcen identifizieren, die für die Beantwortung bestimmter Geschäftsfragen, für die Anpassung und Optimierung von Prozessen zur Verbesserung der Nützlichkeit und Konsistenz von Daten in Ihrem gesamten Geschäft und für die Behebung von Datenproblemen verwendet werden.



Webservices und API

Sie können Spectrum™ Technology Platform-Funktionen in Ihre Anwendungen integrieren. Verwenden Sie dazu Webservices und Programmier-APIs. Diese Schnittstellen bieten eine einfache Integration, optimieren die Datenbankverarbeitung und unterstützen die Abwärtskompatibilität für zukünftige Versionen.

Die Spectrum™ Technology Platform-API ist für die folgenden Sprachen verfügbar:

- C
- C++
- COM
- Java
- .NET

Webservices sind über SOAP und REST verfügbar.

Module und Komponenten

Tabelle 10: Module und Komponenten

Moduls	Beschreibung	Komponenten
Address Now-Modul	Es bietet verbesserte Überprüfung und Standardisierung für Adressen außerhalb der USA und weitere Adressenverarbeitungsfunktionen.	Build Global Address Get Global Candidate Addresses Validate Global Address
Advanced Matching-Modul	Vergleicht Datensätze innerhalb und/oder zwischen Eingabedateien.	Best Of Breed Candidate Finder Duplicate Synchronization Filter Interflow Match Intraflow Match Match Key Generator Transactional Match
Advanced Scoring-Modul	Vergleicht Datensätze innerhalb und/oder zwischen Eingabedateien.	PMML Model Scoring Read from Miner Dataset Write to Miner Dataset
Business Steward-Modul	Identifiziert Ausnahmedatensätze und stellt ein browserbasiertes Tool für die manuelle Überprüfung von Ausnahmedatensätzen bereit.	Exception Monitor Read Exceptions Write Exceptions
Country Identifier	Wählt einen Ländernamen oder eine Kombination aus Postleitzahl und Bundesland/Kanton und gibt den zweistelligen ISO-Ländercode, den dreistelligen Universal Postal Union (UPU)-Code und den englischen Ländernamen zurück.	Country Identifier

Moduls	Beschreibung	Komponenten
Data Hub-Modul	Verknüpft und analysiert Daten, identifiziert Beziehungen und Trends.	Write to Hub Read from Hub Query Hub Graph Visualization
Data Integration-Modul	Bietet nützliche Funktionen für Data Warehousing, Datenqualität, Systemintegration und Migration.	Field Selector Generate Time Dimension Query Cache Write to Cache
Data Normalization-Modul	Entfernt Inkonsistenzen in Daten.	Advanced Transformer Open Parser Table Lookup Transliterator
Enterprise Data Integration	Stellt für verschiedene Geschäftsanforderungen, einschließlich Data Warehousing, Datenqualität, Systemintegration und Migration, eine Verbindung zu Daten in mehreren Quellen her.	Call Stored Procedure Field Selector Generate Time Dimension Query Cache Write to Cache
Enterprise Geocoding-Modul	Bestimmt die geografischen Koordinaten einer Adresse. Bestimmt außerdem die Adresse zu gegebenem Längengrad und Breitengrad.	Geocode Address AUS Geocode Address GBR Geocode Address Global Geocode Address World Geocode US Address GNAF PID Location Search Reverse APN Lookup Reverse Geocode Address Global Reverse Geocode US Location

Moduls	Beschreibung	Komponenten
Enterprise Routing-Modul	Ruft Wegbeschreibungen (zum Fahren oder Gehen) ab, berechnet Fahrtzeit und Entfernung und identifiziert Orte innerhalb einer bestimmten Zeitspanne oder Entfernung von einem Startpunkt.	Get Route Data Get Travel Boundary Get Travel Cost Matrix Get Travel Directions Persistent Update
Enterprise Tax-Modul	Bestimmt die Steuerhoheit für einen angegebenen Standort	Assign GeoTAX Info Calculate Distance
GeoConfidence-Modul	Bestimmt die Wahrscheinlichkeit, dass eine Adresse oder Straßenkreuzung innerhalb eines gegebenen Bereichs liegt.	Geo Confidence Surface CreatePointsConvexHull
Global Addressing-Modul	Bietet eine erweiterte Adressenstandardisierung und -prüfung für Adressen außerhalb der USA. Schlägt zudem automatisch Adressen vor, während Sie tippen, und gibt sofort Kandidaten basierend auf Ihrer Eingabe zurück. Teilen Sie postalische Adresszeichenfolgen unter Verwendung von Machine Learning-Techniken in einzelne Adresselemente auf.	Global Address Validation Globaler Typ voraus Globaler Adressen-Parser
Global Geocoding-Modul	Bestimmt die geografischen Koordinaten einer Adresse. Bestimmt außerdem die Adresse zu gegebenem Längengrad und Breitengrad.	Globaler Geocode Globaler Reverse-Geocode
Global Sentry	Versucht, Transaktionen mit von Regierungsbehörden bereitgestellten Suchlisten abzugleichen, die Daten aus verschiedenen Ländern enthalten.	Global Sentry Global Sentry Address Check Global Sentry ID Number Check Global Sentry Name Check Global Sentry Other Data Check

Moduls	Beschreibung	Komponenten
Location Intelligence-Modul	Führt Punkt-in-Polygon-Analysen und radiale Analysen mit verschiedenen Geodaten-Datenbanken durch.	Closest Site Find Nearest Point In Polygon Query Spatial Data Read Spatial Data Spatial Calculator Spatial Union
Machine Learning-Modul	Bietet die Möglichkeit, ein Binning für numerische Daten durchzuführen, überwachte und unüberwachte Machine Learning-Modelle anzupassen und Daten in diesen Modellen zu bewerten.	Binning Binning Lookup Java Model Scoring K-Means Clustering Linear Regression Logistic Regression Principal Component Analysis Random Forest Classification Random Forest Regression
Metadata Insights	Gibt Ihnen die Kontrolle, die Sie benötigen, um genaue und rechtzeitige datengesteuerte Einblicke in Ihr Geschäft zu geben. Entwickelt Datenmodelle, verschafft Ihnen einen Überblick über den Fluss von Daten von der Quelle bis zur Geschäftsanwendung und bewertet die Qualität Ihrer Daten durch Profilerstellung. So können Sie einfacher die Datenressourcen identifizieren, die Sie für die Beantwortung bestimmter Geschäftsfragen und für die Optimierung von Prozessen zur Verbesserung der Nützlichkeit und Konsistenz von Daten in Ihrem gesamten Geschäft verwenden sollten.	Modelle (logische und physische) Modellspeicher Profil Herkunfts- und Wirkungsanalyse

Moduls	Beschreibung	Komponenten
SAP-Modul	Verbindet Spectrum™ Technology Platform über eine Schnittstelle mit Anwendungen des SAP Customer Relationship Management-Moduls.	SAP Generate Match Key SAP Generate Match Score SAP Generate Search Key SAP Generate Search Key Constant SAP Generate Search Key Metaphone SAP Generate Search Key Substring SAP Validate Address With Candidates
Siebel-Modul	Verbindet Spectrum™ Technology Platform über eine Schnittstelle mit Anwendungen von Siebel.	Siebel Generate Match Key Siebel Generate Match Score Siebel Generate Search Key Siebel Business Name Standardization Siebel Standardize Name Siebel Geocode US Address With Candidates Siebel Geocode US Address With No Candidates Siebel Get Global Candidate Addresses Siebel Validate Address With Candidates Siebel Validate Address With No Candidates
Universal Addressing-Modul	Standardisiert und überprüft Adressen gemäß den Standards der Postbehörde.	Get Candidate Addresses Get City State Province Postleitzahlen abrufen Validate Address Validate Address AUS Validate Address Global

Moduls	Beschreibung	Komponenten
Universal Name-Modul	Parst Personennamen, Unternehmensnamen, Adressen sowie viele andere Begriffe und Abkürzungen.	Name Parser (veraltet) Name Variant Finder Open Name Parser

Notices

© 2017 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

© 2017 Pitney Bowes Software Inc.
All rights reserved