

Spectrum™ Technology Platform

Version 12.0 SP1

Enterprise Data Integration-Handbuch



Inhalt

1 - Einführung

Architektur der Unternehmensdatenverwaltung	5
Das „Star Schema“ als Data Warehouse-Design	8

2 - Verbinden mit Datenquellen und Data Warehouses

Datenquellenverbindungen	14
Verbindungstypen	18
Löschen einer Verbindung	71

3 - Ausfüllen des Data Warehouses

Vorbereiten Ihrer Daten	74
Ausfüllen einer Zeitdimensionstabelle	75
Ausfüllen von Dimensionstabellen	76
Ausfüllen einer Faktentabelle	78
Hinzufügen eines Zeitstempels zu Datensätzen in einem Data Warehouse	83

4 - Aktualisieren des Data Warehouses

Definieren eines Aktualisierungsplans für ein Data Warehouse	87
Aktualisieren einer Faktentabelle	88
Globalen Verwenden eines globalen Caches für Abfragen	93
Verwenden eines lokalen Caches für Abfragen	94

5 - Schrittreferenz

Call Stored Procedure	97
DB Change Data Reader	99
DB Loader	103
Field Parser	113
Field Combiner	117
Field Selector	119
Generate Time Dimension	120
Query Cache	127
Query DB	136
Query NoSQL DB	140
Read from DB	143
Read from File	152
Read from Hadoop Sequence File	171
Read from Hive File	175
Read from HL7 File	178
Read from NoSQL DB	192
Read from SAP	197
Read from Spreadsheet	202
Read from Variable Format File	205
Read from XML	220
SQL Command	229
Transposer	239
Unique ID Generator	243
Write to Cache	252
Write to DB	254
Write to File	261
Write to Hadoop Sequence File	280
Write to Hive File	284
Write to NoSQL DB	291
Write to Spreadsheet	295
Write to Variable Format File	298
Write to XML	308
Datums- und Zahlenmuster	318

6 - Konfigurationen

Oracle LogMiner-Konfiguration 326

7 - Optimieren der Leistung

Ermitteln einer idealen Abrufgröße 329

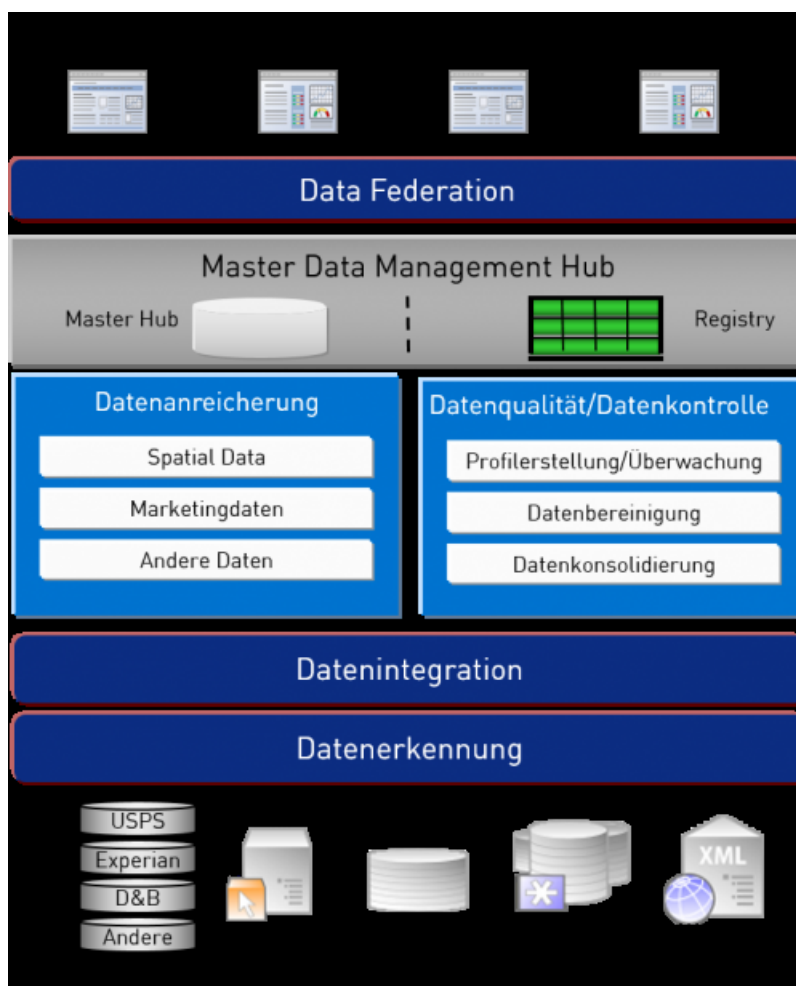
1 - Einführung

In this section

Architektur der Unternehmensdatenverwaltung	5
Das „Star Schema“ als Data Warehouse-Design	8

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.

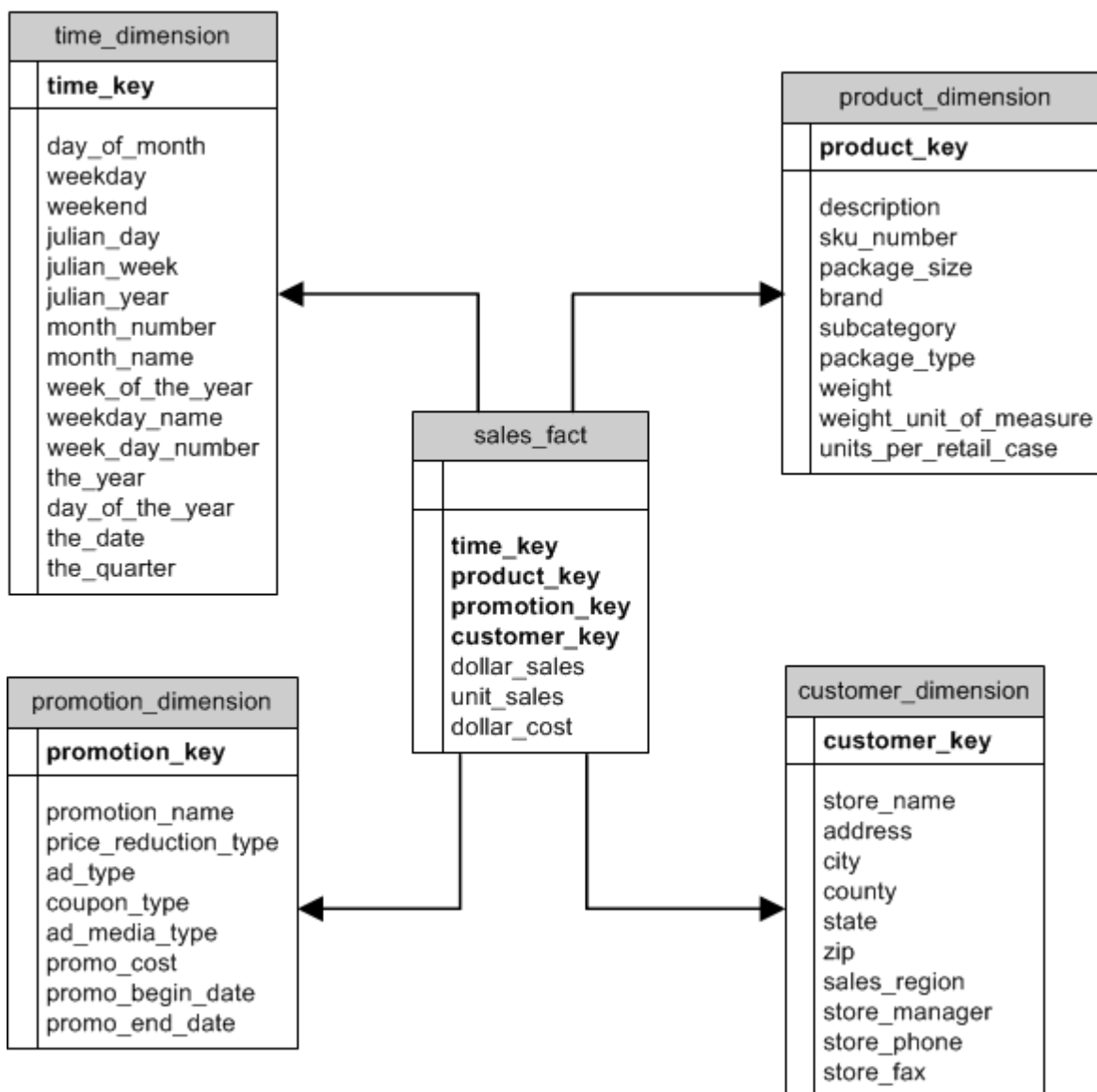
Spectrum™ Technology Platform unterstützt zwei Ansätze für den MDM Hub. Im Master Hub-Ansatz werden die Daten in einer einzelnen MDM-Datenbank gepflegt und Anwendungen greifen auf Daten in der MDM-Datenbank zu. Im Registrierungsansatz werden die Daten in den einzelnen Unternehmensanwendungen gepflegt 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.

Das „Star Schema“ als Data Warehouse-Design

Spectrum™ Technology Platform unterstützt die Erstellung und Wartung von Data Warehouses mit „Star Schema“-Design. In einem „Star Schema“ sind Daten entweder als Fakten – spezifische Beschreibungen eines Ereignisses – gespeichert oder als dimensionale Attribute – Beschreibungen der Fakten in der Faktentabelle. Fakten ändern sich auf regelmäßiger Basis, während sich Dimensionen selten oder nie ändern.

Die folgende Abbildung zeigt das Design eines „Star Schema“:



Die Abbildung zeigt die Hauptmerkmale eines „Star Schema“: eine Faktentabelle, Dimensionstabellen und Verbindungen.

Faktentabelle

Faktentabellen sind die zentralen Tabellen im „Star Schema“ Ihres Data Warehouse. Faktentabellen enthalten normalerweise numerische oder quantitative Informationen (Messwerte genannt), die ein spezifisches Ereignis beschreiben. Wenn Sie beispielsweise über ein Data Warehouse verfügen, das Sie zur Generierung eines Berichts zum Unternehmensumsatz verwenden, haben Sie möglicherweise die Spalten „dollar_sales“ und „dollar_cost“ in Ihrer Faktentabelle, wie in der Abbildung oben dargestellt. Normalerweise sind Fakten durchgehend bewertet und additiv. „Durchgehend bewertet“ bedeutet, dass der Fakt ein numerischer Messwert ist und für jeden Messzeitpunkt einen Wert enthält. „Additiv“ bedeutet, dass der Fakt durch Addition aufsummiert werden kann.

Faktentabellen enthalten außerdem einen Satz Spalten, die einen verketteten oder zusammengesetzten Schlüssel bilden. Jede Spalte des zusammengesetzten Schlüssels ist ein Fremdschlüssel mit Bezug zum Primärschlüssel einer Dimensionstabelle. Die Faktentabelle in der obigen Abbildung enthält beispielsweise eine Spalte „product_key“, die den Fakt zu einem bestimmten Produkt in der Tabelle „product_dimension“ zuordnet.

Die Detailebene in einer Faktentabelle wird als Granularität bezeichnet. Jede Zeile in der Faktentabelle muss in derselben Detailebene aufgezeichnet werden. Im obigen Diagramm sind die Messwerte in der Faktentabelle Tagesgesamtwerte der Verkäufe in Dollar, Verkäufe in Einheiten und Kosten jedes verkauften Produkts in Dollar. Die Granularität ist ein Tag. Jeder Datensatz in der Faktentabelle stellt die gesamten Verkäufe eines bestimmten Produkts in einem Einzelhandelsgeschäft an einem Tag dar. Jede neue Kombination von Produkt, Geschäft oder Tag generiert einen anderen Datensatz in der Faktentabelle.

Faktentabellen werden mit Daten gefüllt, die aus einer Datenquelle extrahiert werden. Die Datenquelle kann entweder ein OLTP-System oder ein Data Warehouse sein. Spectrum™ Technology Platform erstellt nach einem Zeitplan regelmäßig eine Momentaufnahme der Quelldaten und verschiebt die Daten zum Data Warehouse. Dies geschieht üblicherweise zum selben Zeitpunkt an jedem Tag, in jeder Woche oder in jedem Monat.

Ein „Star Schema“ kann mehrere Faktentabellen enthalten. Verwenden Sie ein Schema mit mehreren Faktentabellen, um Sätze mit Messwerten, die eine gemeinsame Teilmenge an Dimensionstabellen verwenden, zu trennen oder um Messwerte mit unterschiedlicher Granularität nachzuverfolgen.

Dimensionstabelle

Dimensionstabellen speichern Daten, die die Informationen in der Faktentabelle beschreiben. Wenn sich beispielsweise „sales_total“ von Monat zu Monat unterscheidet, können Sie über die Dimensionen herausfinden, warum. Dieselbe Dimensionstabelle kann mit unterschiedlichen Faktentabellen verwendet werden.

Dimensionstabellen verfügen über Attribute und einen einteiligen Primärschlüssel, über den eine Dimensionstabelle mit der Faktentabelle verbunden ist. Attribute sind die Spalten in der Dimensionstabelle. Über den einteiligen Primärschlüssel können Sie eine einzelne Dimensionstabelle schnell durchsuchen. Das Durchsuchen einer Dimensionstabelle kann Ihnen helfen, die beste Möglichkeit zur Abfrage der Faktentabelle zu finden.

Zeitdimensionstabellen werden für genaue zeitbasierte Berechnungen benötigt, da die erforderlichen Datumsdaten manchmal nicht einfach aus den Datensätzen extrahiert werden können. Als Beispiel dienen hier die folgenden Datensätze aus einer Umsatzdatenbank. Beachten Sie, dass zwischen den Datensätzen zeitliche Lücken vorhanden sind. Beispielsweise ist kein Datensatz für den 01.04.2012 vorhanden.

Datum	Produkt	Betrag
03.01.2012	Rotes Hemd	10,00 \$

Datum	Produkt	Betrag
05.01.2012	Rotes Hemd	5,00 \$
07.01.2012	Rotes Hemd	15,00 \$

Wenn Sie diese Datensätze abfragen und den durchschnittlichen Tagesumsatz berechnen, wäre das Ergebnis 10,00 \$ (30 \$ / 3 Datensätze). Dieses Ergebnis ist jedoch falsch, da die drei Datensätze eigentlich einen Zeitraum von fünf Tagen abdecken. Wenn Sie über eine Zeitdimensionstabelle mit einem Datensatz für jeden Tag verfügen, können Sie diese Tabelle mit der obigen Tabelle zusammenführen. Sie erhalten dann Folgendes:

Datum	Produkt	Betrag
03.01.2012	Rotes Hemd	10,00 \$
04.01.2012		
05.01.2012	Rotes Hemd	5,00 \$
06.01.2012		
07.01.2012	Rotes Hemd	15,00 \$

Bei der Berechnung des durchschnittlichen Tagesumsatzes erhalten Sie bei diesen Datensätzen das korrekte Ergebnis, 6,00 \$ (30 \$ / 5 Tage).

Darüber hinaus können Sie in Ihrer Berechnung beliebige Zeitattribute berücksichtigen wie Feiertage, Wochenenden und Quartale. Wenn beispielsweise der 06.01.2012 ein Feiertag war und Sie nur am durchschnittlichen Umsatz pro Arbeitstag interessiert sind, ist das Ergebnis 7,50 \$.

Joins

Joins definieren die Beziehungen zwischen einer Faktentabelle und den Dimensionstabellen im „Star Schema“. Der Primärschlüssel in der Dimensionstabelle ist der Fremdschlüssel in der Faktentabelle. Die Faktentabelle muss einen Primärschlüsselwert aus jeder Dimensionstabelle enthalten. Der Verweis vom Fremdschlüssel auf den Primärschlüssel ist der Mechanismus, über den Werte zwischen den beiden Tabellen abgeglichen werden. Join-Beziehungen dieses Typs stellen die referentielle Integrität eines Data Warehouse sicher. Die referentielle Integrität muss erhalten bleiben, um gültige Abfrageergebnisse zu gewährleisten.

Jeder Datensatz in einer Dimensionstabelle kann viele Datensätze in der Faktentabelle beschreiben. Die Verbindungskardinalität von Dimensionstabellen zu Faktentabellen ist daher eine-zu-viele.

In der Abbildung oben ist „product_key“ der Primärschlüssel in der Tabelle „product_dimension“ und der Fremdschlüssel in der Tabelle „sales_fact“. Dieser Join repräsentiert die Beziehung zwischen den Produkten des Unternehmens und den Verkäufen.

Vorteile eines „Star Schema“

Mit einem gut entworfenen Schema können Sie große multidimensionale Datasets schnell verstehen, in ihnen navigieren und sie analysieren. Die größten Vorteile eines „Star Schema“ in einer Entscheidungsfindungsumgebung stellen sich wie folgt dar:

Abfrageleistung

Abfragen an eine „Star Schema“-Datenbank laufen schneller als Abfragen an ein OLTP-System, da das „Star Schema“ über weniger Tabellen und klare Verbindungspfade verfügt. In einem „Star Schema“-Design werden Dimensionen über die zentrale Faktentabelle verknüpft. Dimensionen werden miteinander über einen Verbindungspfad verknüpft, der über die Faktentabelle verläuft. Durch dieses Design-Feature werden genaue und konsistente Abfrageergebnisse erzwungen.

Ladeleistung und Administration

Die Struktur des „Star Schema“ verringert die erforderliche Zeit zum Laden großer Daten-Batches in eine Datenbank. Durch das Definieren von Fakten und Dimensionen und durch ihre Trennung in unterschiedlichen Tabellen werden die Auswirkungen auf Ladevorgänge verringert. Dimensionstabellen können einmal gefüllt und dann gelegentlich aktualisiert werden. Neue Fakten können durch Anfügen von Datensätzen an eine Faktentabelle regelmäßig und selektiv hinzugefügt werden.

Enthaltene referentielle Integrität

Das Design eines „Star Schema“ erzwingt die referentielle Integrität von geladenen Daten. Die referentielle Integrität wird durch die Verwendung von Primär- und Fremdschlüsseln durchgesetzt. Primärschlüssel von Dimensionstabellen werden zu Fremdschlüsseln von Faktentabellen, um jeden Datensatz über Dimensions- und Faktentabellen hinweg zu verknüpfen.

Effiziente Navigation durch Daten

Navigieren durch Daten ist aufgrund der Verbindung der Dimensionen durch Faktentabellen effizient. Diese Verbindungen sind signifikant, da sie die grundlegenden Beziehungen realer Geschäftsprozesse darstellen. Sie können eine einzelne Dimensionstabelle durchsuchen, um Attributwerte auszuwählen und eine effiziente Abfrage zu erstellen.

2 - Verbinden mit Datenquellen und Data Warehouses

In this section

Datenquellenverbindungen	14
Verbindungstypen	18
Löschen einer Verbindung	71

Datenquellenverbindungen

Eine Datenquelle ist eine Datenbank, ein Dateiserver, ein Cloud-Dienst oder eine andere Quelle von Daten, die Sie über Spectrum™ Technology Platform verarbeiten möchten. Spectrum™ Technology Platform kann eine Verbindung zu über 20 Datenquellentypen herstellen.

Um eine Verbindung zwischen Spectrum™ Technology Platform und einer Datenquelle herzustellen, verwenden Sie Management Console, um die Verbindung zu definieren. Wenn Sie beispielsweise Daten aus einer XML-Datei in einen Datenfluss lesen möchten und sich die XML-Datei auf einem Remote-Dateiserver befindet, müssen Sie eine Verbindung zum Dateiserver definieren, bevor Sie die Eingabe-XML-Datei in einem Datenfluss definieren können. Ähnlich müssen Sie zuerst die Datenbank als externe Ressource definieren, wenn Sie die Datenflussausgabe in eine Datenbank schreiben möchten.

Anmerkung: Wenn Sie Daten aus einer Datei lesen oder in eine Datei schreiben möchten, die sich auf dem Spectrum™ Technology Platform-Server selbst befindet, müssen Sie keine Verbindung definieren.

Unterstützung von Komprimierung für Cloud-Dateiserver

Die Dateiserver Amazon S3, Google-Cloud-Speicher und MS Azure Blobstore unterstützen die komprimierten Formate `gzip (.gz)` und `zip (.zip)`.

Spectrum™ Technology Platform übernimmt die Komprimierung und die Dekomprimierung von Daten, die auf die Dateiserver geschrieben und von ihnen gelesen werden.

Anmerkung: Sie können denselben Dateiserver verwenden, um normale Lese- und Schreibvorgänge von Dateien und auch die Komprimierung und die Dekomprimierung von Daten zu übernehmen.

Datei in komprimiertem Format lesen

Beim Lesen einer Datei vom Server wird ihr Komprimierungsformat aus der Metadaten-Schlüsseleigenschaft `Content-Encoding`, die vom Server abgerufen wird, abgeleitet.

Datei in komprimiertem Format schreiben

Beim Schreiben einer Daten auf einen Server wird das erforderliche Komprimierungsformat angegeben: `.gz` oder `.zip`. Die Datei wird basierend auf der angegebenen Komprimierungserweiterung komprimiert.

Die Metadaten-Schlüsseleigenschaft `Content-Encoding` wird auch entsprechend dem ausgewählten Komprimierungsformat festgelegt. Dieser Eigenschaftswert wird beim Schreiben der Datei an den Cloud-Dateiserver übergeben.

Unterstützte Entitäten und Vorgänge

Spectrum™ Technology Platform unterstützt für jeden Verbindungstyp folgende Entitäten und Vorgänge:

Verbindungstyp	Tabelleneigenschaften
----------------	-----------------------

Marketo	<p>Die Entitäten weisen die folgenden Typen auf:</p> <ol style="list-style-type: none"> 1. Entität 2. Entitätsaktualisierung <p>Anmerkung: Entitätsaktualisierung ist eine virtuelle Tabelle, die für eine Aktualisierung in der Lead-Entität verwendet wird. Merge_Leads sollte beispielsweise für das Zusammenführen verschiedener Marketo-Leads verwendet werden.</p>
MS Dynamics CRM	<p>Die Entitäten weisen die folgenden Typen auf:</p> <ol style="list-style-type: none"> 1. Benutzereigen 2. Organisationseigen 3. Geschäftseigen 4. Keine
Netsuite	<p>Die Entitäten weisen die folgenden Typen auf:</p> <ul style="list-style-type: none"> • Standarddatensätze • Benutzerdefinierte Datensätze • Joins • Gespeicherte Suchvorgänge <p>Wenn das Schema der erstellten Baseview angezeigt wird, wird für jeden Entitätstyp das resultierende Schema angezeigt.</p> <p>Bei dem Datensatz eines gespeicherten Suchvorgangs wird beispielsweise das Schema des Suchergebnisses angezeigt. Bei dem Datensatz eines Joins wird das Schema des Join-Ergebnisses angezeigt.</p> <p>Anmerkung: In einer NetSuite-Verbindungstabelle lautet die Primärschlüsselspalte <code>internalId</code>.</p>

Verbindungstyp Tabelleneigenschaften

SAP

- Die Entitätsspalten weisen zwei Typen auf:
 - **Nativ:** Spalten mit nativen Datentypen werden mit den entsprechenden Datentypen angezeigt.
 - **Benutzerdefiniert:** Spalten mit benutzerdefinierten Datentypen werden mit einem leeren Datentyp angezeigt.

Um eine aus einer SAP-Verbindung abgeleitete Virtual Data Source bereitzustellen, dürfen die Metaviews und Baseviews nur solche Entitäten enthalten, deren Spalten native Datentypen sind. Wenn die Baseviews und Metaviews Entitäten mit benutzerdefinierten Datentypen aufweisen, kann die Virtual Data Source nicht bereitgestellt werden.

Siebel Die Geschäftskomponenten werden im Format `Business Object.Business Component` angezeigt.

Splunk

Unterstützte Vorgänge	LIKE, ORDER BY, LIMIT, IN, BETWEEN, !=, <=, >=, <, >, mehrere AND/OR-Operatoren				
Unterstützte Funktionen	<table border="0"> <tr> <td>Zeichenfolgefunktionen</td> <td>upper, lower, length, len, ltrim, rtrim, substring, max, min</td> </tr> <tr> <td>Mathematische Funktionen</td> <td>abs, ceil, exp, floor, sqrt, round</td> </tr> </table>	Zeichenfolgefunktionen	upper, lower, length, len, ltrim, rtrim, substring, max, min	Mathematische Funktionen	abs, ceil, exp, floor, sqrt, round
Zeichenfolgefunktionen	upper, lower, length, len, ltrim, rtrim, substring, max, min				
Mathematische Funktionen	abs, ceil, exp, floor, sqrt, round				

Anmerkung: Verwenden Sie bei allen anderen Abfragevorgängen die Splunk-Spalte `search`, wie unten erläutert.

Spectrum™ Technology Platform stellt in der Splunk-Tabelle die Spalte `search` bereit, über die Sie die erforderlichen Daten in der Splunk-Verbindung suchen können.

Verwenden Sie beim Ausführen einer `select`-Abfrage in der `SplunkTable` in einem der nachfolgenden Szenarien die Spalte `search` in der Klausel `where`:

- Um die Suchkriterien einzuschließen, die mit der ANSI SQL-Syntax nicht angegeben werden können.
- Um die Splunk-spezifischen Suchkriterien einzuschließen, die nicht im Rahmen der SQL-Hauptabfrage eingeschlossen werden können.

Beispiel: In der nachfolgenden Abfrage wird nach dem Wert `_raw` gesucht, der den Schlüssel `opp` mit dem Wert `ACC` enthält.

```
select "_raw" from SplunkTable where "search"='search opp=ACC'
```


Verbindungstyp Tabelleneigenschaften

- SuccessFactors Die Entitäten weisen zwei Typen auf:
1. **Entität:** Stellt eine Tabelle mit einer Geschäftsentität dar.
 2. **Join:** Stellt ein Mapping zwischen zwei Tabellen vom Typ *Entität* dar: einer übergeordneten Tabelle und einer der zugehörigen untergeordneten Tabellen.

Anmerkung: Zwischen Tabellen in dem Baseview-Schema, das über eine SuccessFactors-Verbindung abgeleitet wurde, sind keine Links vorhanden. Grund dafür ist, dass Fremdschlüssel in SuccessFactors-Tabellen nicht vorhanden sind und Joins zwischen Tabellen von Tabellen vom Typ *Join* im Spectrum™ Technology Platform angegeben werden.

Die Features von Join-Tabellen lauten wie folgt:

1. Der Name einer *Join*-Tabelle gibt die zwei *Entität*-Tabellen an, die gemeinsam zugeordnet wurden.
2. Jeder Datensatz einer *Join*-Tabelle enthält den Primärschlüssel der übergeordneten Entität und die Spalten der entsprechenden untergeordneten Entität. Dadurch wird der bestimmte Primärschlüssel der übergeordneten Entität den Details der untergeordneten Entität zugeordnet.

Beispiel: `User#HR` ist eine *Join*-Tabelle, in der `User` die übergeordnete Entität und `Hr` die untergeordnete Entität ist. Dieser Join stellt alle Benutzer und die entsprechenden HR-Mitarbeiter dar. Die Join-Tabelle `User#HR` enthält daher den primären Schlüssel der übergeordneten Tabelle `User`, `UserId`, der den Spalten der untergeordneten Tabelle `HR` zugeordnet ist, z. B. `hr_userId`, `hr_username`, `hr_email` usw.

3. Bei *Join*-Tabellen arbeiten die Funktionen `insert` und `update` wie die Funktion `upsert`. Grund dafür ist, dass es sich bei *Join*-Tabellen nicht um tatsächliche Entitäten in SuccessFactors, sondern um Mappings zwischen Entitäten und den zugehörigen Navigationseigenschaften oder untergeordneten Tabellen handelt.

Zum `insert/update` einer beliebigen *Join*-Tabelle wird die übergeordnete Entität aktualisiert, während in der untergeordneten Tabelle ein neuer Datensatz eingefügt oder der vorhandene Datensatz dem übergeordneten Datensatz entsprechend aktualisiert wird.

Anmerkung: Beim Aktualisieren wird das Mapping zwischen einem übergeordneten und einem untergeordneten Element geändert. Es ist auch möglich, die einzelnen Attribute des untergeordneten Elements nach Bedarf zu ändern.

SugarCRM **Unterstützte Vorgänge:**

`LIKE` (der Vorgang ist darauf beschränkt, dass Optionen beginnend mit dem angegebenen Wert übernommen werden, wie z. B. in der Anweisung `WHERE name LIKE 's%'`, bei der alle Namen übernommen werden, die im Alphabet mit „S“ beginnen), `ISNULL`, `IS NOT NULL`, `IN`, `NOT IN`, `>`, `>=`, `<`, `<=`, `=`, `<>`, `AND`, `OR`

Verbindungstyp Tabelleneigenschaften

Oracle Eloqua

Unterstützte Vorgänge

Es werden drei Entitätstypen unterstützt:

1. **Entität:** Bezeichnet eine Tabelle mit einer Geschäftsentität.
 2. **Aktivität:** Bezeichnet eine Tabelle mit einer Geschäftsentität, wobei die Daten auf Basis einer Aktivität generiert werden.
 3. **Benutzerdefinierte Entität:** Bezeichnet Entitäten, die als Teil von besonderen Vorgängen verwendet werden, die mit dem Connector zur Verfügung gestellt werden.
-


Verbindungstypen

Verbinden mit Amazon

Verbinden mit Amazon DynamoDB

Damit Spectrum™ Technology Platform auf Daten in Amazon DynamoDB zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu Amazon DynamoDB definieren. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten von Amazon DynamoDB lesen und dorthin schreiben können.

Anmerkung: Diese Verbindung wird im Metadata Insights-Modul verwendet.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ Amazon DynamoDB** aus.
6. Geben Sie im Feld **Zugriffsschlüssel-ID** die 20-stellige alphanumerische Sequenz ein, die Ihnen für den Zugriff auf Ihr Amazon AWS-Konto bereitgestellt wurde.
7. Geben Sie den 40-stelligen Schlüssel, der für die Authentifizierung der Verbindung erforderlich ist, in das Feld **Geheimer Zugriffsschlüssel** ein.


8. Wählen Sie die Region des Amazon AWS-Kontos im Feld **Region** aus.
9. Um die Verbindung zu testen, klicken Sie auf **Testen**.
10. Klicken Sie auf **Speichern**.

Einschränkungen von Amazon DynamoDB

1. Hierarchische Datentypen wie Listen, Gruppen und Zuordnungen werden als String-Datentypen interpretiert. Grund dafür ist, dass diese Datentypen nicht unterstützt werden.
2. Nullwerte in einer DynamoDB-Datenquelle werden als leere Spaltenwerte interpretiert.
3. Die Aggregierungsfunktion `count` wird in Abfragen an den Modellspeicher nicht unterstützt.

Verbinden mit Amazon S3

Damit Spectrum™ Technology Platform auf Daten in Amazon S3 zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu Amazon S3 definieren. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten von Amazon S3 lesen und dorthin schreiben können.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ Cloud** aus.
6. Wählen Sie im Feld **Cloud-Dienste** die Option **Amazon S3** aus.
7. Geben Sie im Feld **Bucket-Name** den Bucket-Namen ein, wie er in Ihrem Amazon S3-Cloud-Dienst definiert ist. Dies ist der Bucket, in dem Spectrum™ Technology Platform Dateien liest und schreibt.
8. Geben Sie Ihren Zugriffsschlüssel und Ihren geheimen Schlüssel ein, die Ihnen von Amazon zugewiesen wurden.
9. Wählen Sie im Feld **Speichertyp** die Redundanzstufe aus, die Sie für die Speicherung von Daten zulassen möchten.

Standard Dies ist die Standardstufe von Amazon S3 für die Redundanz.

Reduzierte Redundanz Nicht kritische und einfach reproduzierbare Daten werden mit niedrigerer Redundanz gespeichert. Sie erhalten so angemessen zuverlässigen Speicher zu geringeren Kosten.

10. Wählen Sie im Abschnitt **Verschlüsselung** die Verschlüsselungsmethode für die Daten aus. Sie können serverseitige Verschlüsselung, clientseitige Verschlüsselung oder beides auswählen.

Serverseitiger Schlüssel Die Daten werden auf Serverseite ver- und entschlüsselt. Ihre Daten werden als Klartext an den Amazon-Cloud-Dienst übertragen, wo sie dann verschlüsselt und gespeichert werden. Beim Abruf werden die Daten vom Amazon-Cloud-Dienst entschlüsselt und dann als Klartext an Ihr System übertragen.

Sie haben zwei Möglichkeiten, den Schlüssel anzugeben:

- **Von AWS verwaltet:** Der Schlüssel wird automatisch vom Amazon S3-Cloud-Dienst generiert.
- **Vom Kunden bereitgestellt:** Geben Sie den Schlüssel ein, der vom Amazon S3-Cloud-Dienst verwendet werden soll, um die Daten serverseitig zu ver- und zu entschlüsseln.

Clientseitiger Schlüssel Die Daten werden auf Client-Seite ver- und entschlüsselt. Die Daten werden lokal auf Ihrem Client-System verschlüsselt und dann zum Amazon S3-Cloud-Speicher übertragen. Beim Abruf werden die Daten in verschlüsseltem Format zurück zu Ihrem System übertragen und dann auf dem Client-System entschlüsselt.


Clientseitiger Schlüssel: Geben Sie den Schlüssel ein, der von Ihrem Client-System zum Ver- und Entschlüsseln von Daten verwendet werden soll.

Wenn Sie sowohl **Serverseitiger Schlüssel** als auch **Clientseitiger Schlüssel** auswählen, werden Verschlüsselung und Entschlüsselung auf Server- und auf Client-Seite durchgeführt. Die Daten werden zuerst mit Ihrem clientseitigen Schlüssel verschlüsselt und dann in verschlüsseltem Format zu Amazon übertragen, wo sie noch einmal mit dem serverseitigen Schlüssel verschlüsselt und dann gespeichert werden. Beim Abruf entschlüsselt Amazon zuerst die Daten mit dem serverseitigen Schlüssel, überträgt die Daten in verschlüsseltem Format zu Ihrem System, wo sie dann mit dem clientseitigen Schlüssel endgültig entschlüsselt werden.

Anmerkung: Um das Verschlüsselungsfeature der Amazon S3-Cloud zu verwenden, müssen Sie die Amazon S3-Sicherheits-JAR-Dateien installieren. Weitere Informationen finden Sie unter [Verwenden von Amazon S3-Cloud-Verschlüsselung](#) auf Seite 21.

Weitere Informationen über die Verschlüsselungsfeatures von Amazon S3 finden Sie unter:

docs.aws.amazon.com/AmazonS3/latest/dev/UsingEncryption.html

11. Wenn Sie Zugriffsberechtigungen festlegen möchten, klicken Sie im Abschnitt **Berechtigungen** auf .

Es gibt drei Empfängertypen:

Jeden Jeder, der nicht authentifizierter Benutzer ist oder der Gruppe „Protokollbereitstellung“ angehört.

Authentifizierte Benutzer Benutzer, die bei Amazon angemeldet sind.

Protokollbereitstellung Benutzer, die in einem vom Benutzer angegebenen Bucket Aktivitätsprotokolle schreiben, wenn die Bucket-Protokollierung aktiviert ist.

Wählen Sie für jeden Empfänger die gewünschten Berechtigungen aus:

Öffnen/herunterladen Gestatten Sie dem Benutzer, die Datei herunterzuladen.

Anzeigen Gestatten Sie dem Benutzer, die aktuellen Berechtigungen der Datei anzuzeigen.

Bearbeiten Gestatten Sie dem Benutzer, die Berechtigungen für die Datei zu ändern und festzulegen.

12. Um die Verbindung zu testen, klicken Sie auf **Testen**.

13. Klicken Sie auf **Speichern**.

Verwenden von Amazon S3-Cloud-Verschlüsselung

Um das Sicherheitsfeature Verschlüsselung des Amazon S3-Cloud-Dienstes zu verwenden, müssen Sie die Sicherheits-JAR-Dateien herunterladen und sie auf dem Spectrum™ Technology Platform-Server ablegen. Die Verwendung von Verschlüsselung ist optional.

1. Navigieren Sie zur Download-Seite.

Für Windows- und Linux-Plattformen, die Java 7 verwenden, können die JAR-Dateien heruntergeladen werden von:

<http://www.oracle.com/technetwork/java/javase/downloads/jce-7-download-432124.html>

Für AIX-Plattformen, die Java 7 verwenden, können die JAR-Dateien heruntergeladen werden von:

<https://www14.software.ibm.com/webapp/iwm/web/preLogin.do?source=jcesdk>

2. Laden Sie diese beiden JAR-Dateien herunter:

- local_policy.jar
- US_export_policy.jar

3. Platzieren Sie die JAR-Dateien in folgendem Speicherort:

```
SpectrumFolder\Pitney Bowes\Spectrum\java64\jre\lib\security
```


4. Starten Sie den Server neu.

Verbinden mit Amazon SimpleDB

Damit Spectrum™ Technology Platform auf Daten in Amazon SimpleDB zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu Amazon SimpleDB definieren. Danach

können Sie im Enterprise Designer Flüsse erstellen, die Daten von Amazon SimpleDB lesen und dorthin schreiben können.

Anmerkung: Diese Verbindung wird im Metadata Insights-Modul verwendet.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ Amazon SimpleDB** aus.
6. Geben Sie im Feld **Zugriffsschlüssel-ID** die 20-stellige alphanumerische Sequenz ein, die Ihnen für den Zugriff auf Ihr Amazon AWS-Konto bereitgestellt wurde.
7. Geben Sie den 40-stelligen Schlüssel, der für die Authentifizierung der Verbindung erforderlich ist, in das Feld **Geheimer Zugriffsschlüssel** ein.
8. Um die Verbindung zu testen, klicken Sie auf **Testen**.
9. Klicken Sie auf **Speichern**.

Amazon SimpleDB-Beschränkungen

Schreibbeschränkung

Im „Write to DB“-Schritt ist der Schreibmodus **Aktualisieren** nicht verfügbar, wenn in eine Amazon SimpleDB-Tabelle geschrieben wird. Die Option **Einfügen** befasst sich mit Einfügungs- und Aktualisierungsvorgängen. Sie unterscheidet anhand des eindeutigen Wertes in der Spalte `ItemName`, die in allen Amazon SimpleDB-Tabellen enthalten ist, zwischen einer Einfügung und einer Aktualisierung.

Grund: Eine Aktualisierungsabfrage erfordert einen Primärschlüssel für jeden Datensatz in der zu aktualisierenden Tabelle, der nicht von Amazon SimpleDB-Datenbanken unterstützt wird.

Lesebeschränkung


Die Aggregierungsfunktionen `SUM` und `AVG` werden bei der Ausführung von Abfragen an den Modellspeicher nicht unterstützt.

Verbinden mit Apache Cassandra

Damit Spectrum™ Technology Platform auf Daten in einer Cassandra-Datenbank zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zur Cassandra-Datenbank definieren.

Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten aus der Cassandra-Datenbank lesen und dorthin schreiben können.

Anmerkung: Diese Verbindung wird im Metadata Insights-Modul verwendet.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.


5. Wählen Sie im Feld **Typ Apache Cassandra** aus.
6. Geben Sie den Rechnernamen oder die IP-Adresse, unter der die Apache Cassandra-Datenbank installiert ist, in das Feld **Host** ein.
7. Geben Sie den Namen des Keyspace des Rechenzentrums, auf das Sie zugreifen möchten, in das Feld **Keyspace** ein.
8. Geben Sie den Port, unter der die Apache Cassandra-Datenbank konfiguriert ist, in das Feld **Port** ein.
9. Geben Sie den Benutzernamen und das Kennwort ein, die für die Authentifizierung an der Cassandra-Datenbank verwendet werden.
10. Wählen Sie im Feld **Konsistenzgrad** aus, wie konsistent Datenzeilen über Replikatknoten hinweg sein müssen, um eine Datentransaktion als erfolgreich zu werten. Dabei kann es sich um mindestens einen, alle oder eine Kombination aus verfügbaren Knoten handeln.
11. Geben Sie die Anzahl an Resultset-Zeilen, die Sie bei jeder Read-Transaktion abrufen möchten, unter **Abrufgröße** ein.
12. Um die Verbindung zu testen, klicken Sie auf **Testen**.
13. Klicken Sie auf **Speichern**.

Apache Cassandra-Einschränkung

Die Aggregierungsfunktion `count` wird in Abfragen an den Modellspeicher nicht unterstützt.

Verbinden mit Azure Cloud

Damit Spectrum™ Technology Platform auf Daten in Microsoft Azure, zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu Microsoft Azure definieren. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten von Microsoft Azure lesen und dorthin schreiben können.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.


5. Wählen Sie im Feld **Typ Cloud** aus.
6. Wählen Sie im Feld **Cloud-Dienst** die Option **AzureBlobStorage** aus.
7. Geben Sie im Feld **Protokoll** an, ob die Verbindung zwischen Azure und Spectrum™ Technology Platform HTTP oder HTTPS verwenden soll.
8. Geben Sie im Feld **Kontoname** den Namen Ihres Azure-Speicherkontos ein.
9. Geben Sie im Feld **Zugriffsschlüssel** den Zugriffsschlüssel für Ihr Azure-Konto ein.
10. Klicken Sie zum Prüfen der Cloud-Verbindung auf **Testen**.
11. Klicken Sie auf **Speichern**.

Verbinden mit einer Flatfile

Verbinden mit einem Flatfile mit Begrenzungszeichen

Um eine neue Verbindung für Flatfiles mit Begrenzungszeichen hinzuzufügen, navigieren Sie zu **Verbindungen > Flatfile** und wählen Sie für den **Datensatztyp Mit Begrenzungszeichen** aus. Geben Sie die Zugangsdetails der Datei und die Inhaltstypdetails ein, damit das Data Federation-Modul diese korrekt lesen kann.

Anmerkung: Diese Verbindung wird im Metadata Insights-Modul verwendet.

1. Öffnen Sie **Verbindungen > Flatfile**.
2. Standardmäßig wird der Bildschirm im Erstellungsmodus geöffnet. Klicken Sie anderenfalls auf , um eine neue Flatfile-Verbindung hinzuzufügen.
3. Geben Sie einen **Verbindungsnamen** für die Flatfile-Datenverbindung ein.
4. Geben Sie den **Dateipfad** ein, indem Sie auf **Durchsuchen** klicken und das Dateiverzeichnis auswählen.
5. Wählen Sie die **Zeichencodierung** des Flatfile aus der Dropdown-Liste aus.
6. Wählen Sie für den **Datensatztyp** die Option **Mit Begrenzungszeichen** aus.
7. Wählen Sie das erwartete Trennzeichen zwischen zwei beliebigen Feldern eines Dateidatensatzes unter **Feldtrennzeichen** aus.


8. Wählen Sie bei Bedarf das **Textbegrenzungszeichen (optional)** aus, das die Feldwerte eines Dateidatensatzes einschließt.
9. Unter **Zeilentrennzeichen** ist der Wert `Standard` ausgewählt und gibt an, dass das erwartete Zeilentrennzeichen davon abhängig ist, ob Spectrum™ Technology Platform auf einem Unix- oder Windows-System ausgeführt wird.
10. Um anzugeben, ob es sich bei der ersten Zeile der Datei um eine Kopfzeile handelt, schieben Sie den Schieberegler **Erste Zeile als Kopfzeile** auf **Ja** oder **Nein**.
11. Um anzugeben, ob der Datentyp der verschiedenen Felder in einem beliebigen Datensatz der Datei automatisch erkannt werden soll, schieben Sie den Schieberegler **Datentyp anhand von Datei erkennen** auf **Ja** oder **Nein**.
12. Schieben Sie den Schieberegler **Falsch formatierte Datensätze überspringen** auf **An**.
13. Klicken Sie auf **Testen**.
Eine Meldung bestätigt, dass der Verbindungstest erfolgreich war.
14. Klicken Sie auf **Speichern**.
Eine Meldung bestätigt, dass die Verbindung erfolgreich hergestellt wurde.

Um einen Beispieldatensatz anzuzeigen, der über die erstellte Verbindung für Flatfiles mit Begrenzungszeichen abgerufen wurde, klicken Sie in der Kopfleiste auf **Vorschau**. Der Abruf von Dateidatensätzen und die Sortierung von Feldern erfolgt entsprechend der von Ihnen bereitgestellten Angaben.

Verbinden mit einem Flatfile mit fester Breite

Klicken Sie zum Hinzufügen einer Flatfile-Verbindung mit fester Breite auf **Verbindungen > Flatfile**, und wählen Sie für **Datensatztyp** den Wert **Feste Breite** aus. Geben Sie die Zugangsdetails der Datei und die Inhaltstypdetails ein, damit das Data Federation-Modul diese korrekt lesen kann.

Anmerkung: Diese Verbindung wird im Metadata Insights-Modul verwendet.

1. Öffnen Sie **Verbindungen > Flatfile**.
2. Standardmäßig wird der Bildschirm im Erstellungsmodus geöffnet. Klicken Sie anderenfalls auf , um eine neue Flatfile-Verbindung hinzuzufügen.
3. Geben Sie einen **Verbindungsnamen** für die Flatfile-Datenverbindung ein.
4. Geben Sie den **Dateipfad** ein, indem Sie auf **Durchsuchen** klicken und das Dateiverzeichnis auswählen.
5. Wählen Sie die **Zeichencodierung** des Flatfile aus der Dropdown-Liste aus.
6. Wählen Sie für den **Datensatztyp** den Wert **Feste Breite** aus.
7. Geben Sie in das Feld **Datensatzlänge** die Gesamtzahl der Zeichen in einem Dateidatensatz ein.

Wiederholen Sie Schritt 8 bis 13, um die Details für alle in einem Dateidatensatz erwarteten Felder einzugeben.

8. Klicken Sie auf **Feld hinzufügen**, um eine Zeile für ein Feld in einem Dateidatensatz hinzuzufügen.

9. Geben Sie in die Spalte **Name** die Namen für den Feldwert ein.
10. Wählen Sie aus der Spalte **Typ** den Datentyp des Feldwertes aus.
11. Geben Sie in die Spalte **Startposition** die Position im Dateidatensatz ein, an der der Feldwert beginnt.

Beim ersten Feld in einem Dateidatensatz beginnt die **Startposition** mit dem Wert 1.

12. Geben Sie in das Feld **Länge** die Gesamtzahl der Zeichen ein, die das Feld abdeckt. Das Zeichen an der **Startposition** ist Teil davon.

Die Summe der Werte **Startposition** und **Länge** eines Feldes sollte kleiner oder gleich der **Datensatzlänge** sein.

Wenn der Dateidatensatz wie folgt lautet:

```
01234Rob Smith29PitneyBowes
```

Datensatzlänge = 27

Gilt für das Feld „Name“:

Startposition = 6

Länge = 9

```
Name = Rob Smith
```

13. Aktivieren Sie das Kästchen **Kürzen**, wenn Sie Leerzeichen am Anfang und/oder Ende eines Feldwertes entfernen möchten.
14. Klicken Sie auf **Testen**.
Eine Meldung bestätigt, dass der Verbindungstest erfolgreich war.
15. Klicken Sie auf **Speichern**.
Eine Meldung bestätigt, dass die Verbindung erfolgreich hergestellt wurde.

Klicken Sie in der Header-Leiste auf **Vorschau**, um einen Beispieldatensatz anzuzeigen, der mithilfe der erstellten Flatfile-Verbindung mit fester Breite abgerufen wurde. Der Abruf von Dateidatensätzen und die Sortierung von Feldern erfolgt entsprechend der von Ihnen bereitgestellten Angaben.

Datums-/Uhrzeitformat in einer Dateiverbindung

Beim Lesen von Datums- und Uhrzeitwerten aus Dateien mithilfe einer Dateiverbindung in Spectrum™ Technology Platform müssen die Werte bestimmten Datums-/Uhrzeitformaten entsprechen.

Zulässige Datums-/Uhrzeitformate:

- Datum: "yyyy-mm-dd"
- Datum/Uhrzeit: "yyyy-mm-dd HH:mm:ss"
- Zeit: "HH:mm:ss"

Dies sind standardmäßige Datums-/Uhrzeitschreibweisen.

Dateien mit Begrenzungszeichen

Wenn das Feature **Typ erkennen** während der Konfiguration der Verbindung für Dateien mit Begrenzungszeichen aktiviert ist, werden die Datums- und Uhrzeitwerte in den Dateidatensätzen, die den obigen Formaten entsprechen, automatisch als Typ „Date“ erkannt.

Wenn ein Datums-/Uhrzeitwert nicht einem der zulässigen Formate entspricht, wird der Wert als ein Wert vom Typ „String“ gelesen.

Dateien mit fester Breite

Bei Dateien mit fester Breite werden die Datentypwerte während der Erstellung der Verbindung für Dateien mit fester Breite konfiguriert. Folglich werden diese Werte als Werte vom Typ „Date“ gelesen, ungeachtet davon, ob sie den zulässigen Formaten entsprechen oder nicht.

Wenn der Datums-/Uhrzeitwert in einer Datei mit fester Breite nicht den zulässigen Formaten entspricht, muss er mithilfe von **Transformationen** im Erstellungsschritt des logischen Modells bearbeitet werden, indem die untere Kategoriefunktion *Konvertierung* auf den Wert angewendet wird:

```
parsedate(String date, String format)
```

Darin ist *date* der aus der Datei empfangene Wert, während *format* das Datums-/Uhrzeitformat ist, in dem der Wert aus der Datei empfangen wurde. Dies hilft bei der korrekten Analyse des Datums-/Uhrzeitwertes.

Wenn das Datum z. B. 23-Feb-2008 lautet, ergibt sich das Format dd-*MMM*-yyyy.

Resultierende Werteformate

Während der Datenvorschau in einem Modellspeicher:

- Wenn der Wert als Datum/Uhrzeitwert gelesen wurde, wird er in einem der zulässigen Datums-/Uhrzeitformate in der Vorschau wiedergegeben.
- Wenn der Wert als String-Wert gelesen wurde, wird er, wie er ist, in der Vorschau wiedergegeben.


Verbinden mit einem FTP-Server

Damit Spectrum™ Technology Platform auf Dateien auf einem FTP-Server zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zum FTP-Server definieren. Danach können Sie im Enterprise Designer Datenflüsse erstellen, die Daten aus Dateien auf dem FTP-Server lesen und dorthin schreiben können.

Bevor Sie eine Verbindung zu einem FTP-Server herstellen, überprüfen Sie, ob die Timeout-Einstellungen auf dem FTP-Server für die Aufträge angemessen sind, für die diese Verbindung verwendet wird. Abhängig vom Design eines Auftrags können Zeiträume auftreten, in denen die Verbindung nicht genutzt wird, was ein Timeout auslösen kann. Sie könnten beispielsweise einen Datenfluss mit zwei „Read from File“-Schritten haben, die mit einem „Import to Hub“-Schritt

verbunden sind. Während der „Import to Hub“-Schritt Datensätze aus einem „Read from File“-Schritt liest, ist der andere inaktiv, was dazu führen kann, dass seine Verbindung zum FTP-Server aufgrund eines Timeouts getrennt wird. Ziehen Sie in Betracht, den Timeout-Wert auf Ihrem FTP-Server auf 0 festzulegen, um Timeouts von Verbindungen zu vermeiden.

Anmerkung: Der FTP-Server muss in einem aktiven Verbindungsmodus ausgeführt werden. Der passive Verbindungsmodus wird nicht unterstützt.


1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ FTP** aus.
6. Geben Sie in den Feldern **Benutzername** und **Kennwort** die Anmeldeinformationen zur Authentifizierung am FTP-Server ein. Dies ist nur erforderlich, wenn der FTP-Server dies voraussetzt.
7. Geben Sie im Feld **Host** den Hostnamen oder die IP-Adresse des FTP-Servers ein.
8. Geben Sie im Feld **Port** die Netzwerkportnummer ein, die der Server für FTP verwendet.
9. Klicken Sie auf **Testen**, um zu überprüfen, ob der Spectrum™ Technology Platform-Server eine Verbindung zum FTP-Server herstellen kann.
10. Klicken Sie auf **Speichern**.

Herstellen einer Verbindung zu Google-Cloud-Speicher

Damit Spectrum™ Technology Platform auf Daten in einem Google-Cloud-Speicher zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zum Google-Cloud-Speicher definieren. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten aus dem Google-Cloud-Speicher lesen und dorthin schreiben können.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ Cloud** aus.
6. Wählen Sie im Feld **Cloud-Dienst** die Option **GoogleCloudStorage** aus.
7. Geben Sie im Feld **Bucket-Name** den Bucket-Namen gemäß der Definition in Ihrem Google Cloud-Dienst ein. Dies ist der Bucket, in dem Spectrum™ Technology Platform Dateien liest und schreibt.
8. Geben Sie Ihren Anwendungsnamen, das Dienstkonto und die private Schlüsseldatei an, die Sie von Google erhalten haben.

Anmerkung: Stellen Sie sicher, dass die private Schlüsseldatei auf dem Spectrum™ Technology Platform-Server vorhanden ist.

9. Sie können Zugriffsberechtigungen im Abschnitt **Berechtigungen** festlegen.

Daten und Berechtigung verwalten Hiermit kann der Benutzer Daten und Berechtigungen verwalten.

Daten anzeigen Hiermit kann der Benutzer Daten anzeigen.

Daten verwalten Hiermit kann der Benutzer Daten verwalten.


10. Um die Verbindung zu testen, klicken Sie auf **Testen**.
11. Klicken Sie auf **Speichern**.

Weitere Informationen finden Sie in der Dokumentation [Dienstkonto-Authentifizierung](#) von Google.

Verbinden mit Hadoop

Damit Spectrum™ Technology Platform auf Daten in Hadoop zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu Hadoop definieren. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten von Hadoop lesen und dorthin schreiben können.

Achtung: *Hadoop 2.x* für Kerberos auf Windows-Plattformen wird von Spectrum™ Technology Platform nicht unterstützt.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ HDFS** aus
6. Geben Sie im Feld **Host** den Hostnamen oder die IP-Adresse des Namensknotens im HDFS-Cluster ein.

7. Geben Sie im Feld **Port** die Netzwerkportnummer ein.
8. Wählen Sie unter **Benutzer** eine Option aus:
 - Serverbenutzer** Wählen Sie diese Option aus, wenn in Ihrem HDFS-Cluster Authentifizierung aktiviert ist. Diese Option verwendet die Benutzeranmeldedaten, unter denen der Spectrum™ Technology Platform-Server läuft, um sich gegenüber HDFS zu authentifizieren.
 - Benutzername** Wählen Sie diese Option aus, wenn in Ihrem HDFS-Cluster Authentifizierung deaktiviert ist.
9. Aktivieren Sie **Kerberos**, wenn Sie das Feature Kerberos-Authentifizierung für diese HDFS-Dateiserververbindung aktivieren möchten.
10. Wenn Sie sich für die Aktivierung der **Kerberos**-Authentifizierung entschieden haben, geben Sie den Pfad zur Schlüsseltabellendatei im Feld **Schlüsseltabellen-Dateipfad** ein.

Anmerkung: Vergewissern Sie sich, dass sich die Schlüsseltabellendatei auf dem Spectrum™ Technology Platform-Server befindet.

11. Wählen Sie im Feld **Protokoll** eine der folgenden Optionen aus:
 - WEBHDFS** Wählen Sie diese Option aus, wenn auf dem HDFS-Cluster HDFS 1.0 oder höher läuft. Dieses Protokoll unterstützt sowohl Lese- als auch Schreibvorgänge.
 - HFTP** Wählen Sie diese Option aus, wenn auf dem HDFS-Cluster eine Version läuft, die älter ist als HDFS 1.0, oder wenn Ihre Organisation die Verwendung des WEBHDFS-Protokolls nicht zulässt. Dieses Protokoll unterstützt nur Lesevorgänge.
 - HAR** Wählen Sie diese Option aus, um auf Hadoop-Archivdateien zuzugreifen. Wenn Sie diese Option auswählen, geben Sie im Feld **Pfad** den Pfad zur Archivdatei an. Dieses Protokoll unterstützt nur Lesevorgänge.
12. Erweitern Sie **Erweiterte Optionen**.
13. Wenn Sie das WEBHDFS-Protokoll ausgewählt haben, können Sie folgende erweiterte Optionen nach Bedarf festlegen:

- Replikationsfaktor** Geben Sie hier an, zu wie vielen Datenknoten jeder Block repliziert werden soll. Die Standardeinstellung 3 repliziert beispielsweise jeden Block zu drei verschiedenen Knoten im Cluster. Der maximale Replikationsfaktor ist 1024.
- Blockgröße** Geben Sie hier die Größe eines Blocks an. HDFS teilt eine Datei in Blöcke auf, die die von Ihnen hier angegebene Größe aufweisen. Wenn Sie beispielsweise den Standardwert 64 MB angeben, wird jede Datei in 64-MB-Blöcke aufgeteilt. Jeder Block wird dann zu der Anzahl von Knoten im Cluster repliziert, die im Feld **Replikationsfaktor** angegeben ist.

Dateiberechtigungen Legen Sie hier die Zugriffsstufe auf Dateien fest, die von Spectrum™ Technology Platform in das HDFS-Cluster geschrieben werden. Sie können Lese- und Schreibberechtigungen für jede der folgenden Optionen festlegen:

Anmerkung: Die Berechtigung *Ausführen* ist bei Spectrum™ Technology Platform nicht anwendbar.

- | | |
|-----------------|---|
| Benutzer | Dies ist der oben angegebene Benutzer, entweder Serverbenutzer oder der im Feld Benutzername angegebene Benutzer. |
| Group | Dies bezieht sich auf jede Gruppe, der der Benutzer angehört. Wenn beispielsweise der Benutzer john123 ist, dann gelten die Gruppenberechtigungen für alle Gruppen, denen john123 angehört. |
| Andere | Dies bezieht sich auf alle anderen Benutzer und auf Gruppen, denen der angegebene Benutzer nicht angehört. |

14. Geben Sie im Raster unter der Tabelle **Dateiberechtigungen** die Servereigenschaften für Hadoop ein, um sicherzustellen, dass die Sortier- und Filterfeatures wie gewünscht arbeiten, wenn die Verbindung in einem Schritt oder einer Aktivität verwendet wird. Sie haben mehrere Möglichkeiten, Eigenschaften hinzuzufügen:

- Klicken Sie auf , und fügen Sie die Eigenschaften und die zugehörigen Werte in den Feldern **Eigenschaft** und **Wert** hinzu.
- Klicken Sie auf , und laden Sie Ihre XML-Konfigurationsdatei hoch. Die XML-Datei sollte hdfs-site.xml, yarn-site.xml oder core-site.xml ähneln.

Anmerkung: Die Konfigurationsdatei muss auf dem Server platziert werden.

In der folgenden Tabelle werden die Eigenschaften und ihre Werte abhängig von dem Schritt oder der Aktivität, die die Hadoop-Verbindung verwenden wird, beschrieben. Die Eigenschaften hängen außerdem von der verwendeten Hadoop-Version ab (*Hadoop 1.x* oder *Hadoop 2.x*).

Schritt oder Aktivität mit Verwendung der HDFS-Verbindung	Erforderliche Servereigenschaften
---	-----------------------------------

- Schritt **Read from Sequence File**
- Aktivität **Run Hadoop Pig**

Schritt oder Aktivität mit
Verwendung der
HDFS-Verbindung

Erforderliche Servereigenschaften

Parameter für Hadoop 1.x

fs.default.name

Gibt den Knoten und den Port an, auf dem Hadoop läuft.

Beispiel:

`hdfs://152.144.226.224:9000`

mapred.job.tracker

Gibt den Hostnamen oder die IP-Adresse und den Port an, auf dem MapReduce Job Tracker läuft. Wenn der Hostname als lokal eingegeben wird, werden die Aufträge als einzelne „Map and Reduce“-Aufgaben ausgeführt.

Beispiel: `152.144.226.224:9001`

dfs.namenode.name.dir

Gibt an, wo ein DFS-Namensknoten die Namenstabelle auf dem lokalen Dateisystem speichern soll. Wenn dies eine durch Komma getrennte Liste von Verzeichnissen ist, wird die Namenstabelle aus Redundanzgründen zu allen Verzeichnissen repliziert.

Beispiel:

`file:/home/hduser/Data/namenode`

dfs.datanode.data.dir

Gibt an, wo ein DFS-Datenknoten seine Blöcke auf dem lokalen Dateisystem speichern soll. Wenn dies eine durch Komma getrennte Liste von Verzeichnissen ist, werden Daten in allen aufgeführten Verzeichnissen gespeichert, die sich normalerweise auf unterschiedlichen Geräten befinden. Nicht vorhandene Verzeichnisse werden ignoriert.

Beispiel:

`file:/home/hduser/Data/datanode`

hadoop.tmp.dir

Schritt oder Aktivität mit
Verwendung der
HDFS-Verbindung

Erforderliche Servereigenschaften

Gibt den Basisspeicherort für andere
temporäre Verzeichnisse an.

Beispiel: `/home/hduser/Data/tmp`

Parameter für Hadoop 2.x

Schritt oder Aktivität mit
Verwendung der
HDFS-Verbindung

Erforderliche Servereigenschaften

fs.defaultFS

Gibt den Knoten und den Port an, auf dem Hadoop läuft.

Beispiel:

```
hdfs://152.144.226.224:9000.
```

Anmerkung: In

Spectrum-Version 11.0 und älter muss der Parametername `fs.defaultfs` verwendet werden. Beachten Sie, dass die Groß-/Kleinschreibung sich unterscheidet.

Ab Version 11 SP1 sind beide Namen, `fs.defaultfs` und `fs.defaultFS`, gültig. Wir empfehlen, ab Spectrum™ Technology Platform 11 SP1 den Parameternamen `fs.defaultFS` zu verwenden.

yarn.resourcemanager.resource-tracker.address

Gibt den Hostnamen oder die IP-Adresse des Ressourcenmanagers an.

Beispiel: `152.144.226.224:8025`

yarn.resourcemanager.scheduler.address

Gibt die Adresse der Planerschnittstelle an.

Beispiel: `152.144.226.224:8030`

yarn.resourcemanager.address

Gibt die Adresse der Schnittstelle des Anwendungsmanagers an, der im Ressourcenmanager enthalten ist.

Beispiel: `152.144.226.224:8041`

mapreduce.jobhistory.address

Gibt den Hostnamen oder die IP-Adresse und den Port, auf dem MapReduce Job

Schritt oder Aktivität mit
Verwendung der
HDFS-Verbindung

Erforderliche Servereigenschaften

History Server läuft.

Beispiel: 152.144.226.224:10020

mapreduce.application.classpath

Gibt den CLASSPATH für
MapReduce-Anwendungen an. Dieser
CLASSPATH legt den Speicherort fest, an
dem sich die Klassen bezüglich
MapReduce-Anwendungen befinden.

Anmerkung: Die Einträge
sollten durch Komma getrennt
sein.

Zum Beispiel:

```
$HADOOP_CONF_DIR,  
$HADOOP_COMMON_HOME/share/hadoop/common/*,  
$HADOOP_COMMON_HOME/share/hadoop/common/lib/*,  
$HADOOP_HDFS_HOME/share/hadoop/hdfs/*,  
$HADOOP_HDFS_HOME/share/hadoop/hdfs/lib/*,  
$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*,  
$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/lib/*,  
$HADOOP_YARN_HOME/share/hadoop/yarn/*,  
$HADOOP_YARN_HOME/share/hadoop/yarn/lib/*
```

Schritt oder Aktivität mit
Verwendung der
HDFS-Verbindung

Erforderliche Servereigenschaften

mapreduce.app-submission.cross-platform

Dies verarbeitet verschiedene Plattformprobleme, die auftreten, wenn Ihr Spectrum-Server auf einem Windows-Computer läuft und Sie Cloudera darauf installieren. Wenn Ihr Spectrum-Server und Cloudera auf verschiedenen Betriebssystemen laufen, geben Sie für diesen Parameter den Wert `wahr` ein. Geben Sie anderenfalls `falsch` ein.

Anmerkung: Cloudera unterstützt keine Windows-Clients. Die Konfiguration dieses Parameters umgeht das Problem und stellt keine Lösung für alle daraus resultierenden Plattformprobleme dar.

Wenn Sie oben das Kästchen **Kerberos** aktiviert haben, fügen Sie zusätzlich die unten aufgeführten Kerberos-Konfigurationseigenschaften hinzu:

Schritt oder Aktivität mit
Verwendung der
HDFS-Verbindung

Erforderliche Servereigenschaften

hadoop.security.authentication

Gibt den verwendeten
Authentifizierungssicherheitstyp an. Geben
Sie den Wert `kerberos` ein.

yarn.resourcemanager.principal

Das Kerberos-Prinzip, das für den
Ressourcenmanager für Ihren Hadoop
YARN-Ressourcenvermittler verwendet wird.

Beispiel: `yarn/_HOST@HADOOP.COM`

dfs.namenode.kerberos.principal

Das Kerberos-Prinzip, das für den
Namensknoten Ihres Hadoop Distributed File
System (HDFS) verwendet wird.

Beispiel: `hdfs/_HOST@HADOOP.COM`

dfs.datanode.kerberos.principal

Das Kerberos-Prinzip, das für den
Datenknoten Ihres Hadoop Distributed File
System (HDFS) verwendet wird.

Beispiel: `hdfs/_HOST@HADOOP.COM`

Schritt oder Aktivität mit
Verwendung der
HDFS-Verbindung

Erforderliche Servereigenschaften

- Schritt **Read from File**
- Schritt **Write to File**
- Schritt **Read from Hive ORC File**
- Schritt **Write to Hive ORC File**

Parameter für Hadoop 1.x

fs.default.name

Gibt den Knoten und den Port an, auf dem Hadoop läuft.

Beispiel:

`hdfs://152.144.226.224:9000`

Parameter für Hadoop 2.x

fs.defaultFS

Gibt den Knoten und den Port an, auf dem Hadoop läuft.

Beispiel:

`hdfs://152.144.226.224:9000.`

Anmerkung: In

Spectrum-Version 11.0 und älter muss der Parametername `fs.defaultfs` verwendet werden. Beachten Sie, dass die Groß-/Kleinschreibung sich unterscheidet.

Ab Version 11 SP1 sind beide Namen, `fs.defaultfs` und `fs.defaultFS`, gültig. Wir empfehlen, ab Spectrum™ Technology Platform 11 SP1 den Parameternamen `fs.defaultFS` zu verwenden.

15. Um die Verbindung zu testen, klicken Sie auf **Testen**.

16. Klicken Sie auf **Speichern**.


Nachdem Sie eine Verbindung zu einem HDFS-Cluster definiert haben, ist diese in Quell- und Datenladeschritten im Enterprise Designer wie „Read from File“ und „Write to File“ verfügbar. Sie können das HDFS-Cluster beim Definieren einer Datei in einem Quell- oder Datenladeschritt auswählen, indem Sie auf **Remote-Computer** klicken.

Unterstützung von Komprimierung für Hadoop

Spectrum™ Technology Platform unterstützt auf Hadoop die komprimierten Formate `gzip` (`.gz`) und `bzip2` (`.bz2`). Fügen Sie bei Verwendung der Schritte **Read from File** und **Write to File** mit einer HDFS-Verbindung im Feld **Dateiname** die Erweiterung an, die der erforderlichen Komprimierung entspricht: `.gz` oder `.bz2`. Die Datei wird entsprechend der angegebenen Komprimierungserweiterung komprimiert bzw. dekomprimiert. Spectrum™ Technology Platform übernimmt die Komprimierung und die Dekomprimierung der Dateien.

Verbinden mit einer JDBC-Datenbank

Damit Spectrum™ Technology Platform auf Daten in einer JDBC-Datenbank zugreifen kann, müssen Sie mithilfe der Management Console eine Datenbank definieren. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten aus einer JDBC-Datenbank lesen und dorthin schreiben können.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ** den Datenbanktyp aus, zu dem Sie eine Verbindung herstellen möchten.


Das Spectrum™ Technology Platform Data Integration-Modul enthält JDBC-Treiber für SQL Server-, Oracle- und PostgreSQL-Datenbanken. Wenn Sie eine Verbindung zu einem anderen Datenbanktyp herstellen möchten, müssen Sie die JDBC-Treiber hinzufügen, bevor Sie eine Verbindung definieren.

6. Geben Sie im Feld **URL** die JDBC-Verbindungs-URL ein. Sie können diese URL von Ihrem Datenbankadministrator erhalten.

Um beispielsweise eine Verbindung zu einer MySQL-Datenbank namens „Beispieldatenbank“ auf einem Server namens „MeinServer“ herzustellen müssen Sie Folgendes eingeben:

```
jdbc:mysql://MeinServer/Beispieldatenbank
```

7. Abhängig vom JDBC-Treiber müssen Sie möglicherweise zusätzliche Felder ausfüllen. Die Felder repräsentieren die Eigenschaften in der Verbindungszeichenfolge für den JDBC-Treiber, den Sie im Feld **Typ** ausgewählt haben. Weitere Informationen zu den spezifischen Verbindungseigenschaften und den für den Verbindungstyp erforderlichen Werten finden Sie in der Anbieterdokumentation des JDBC-Treibers oder erhalten Sie von Ihrem Datenbankadministrator.

8. Klicken Sie auf **Speichern**.
9. Testen Sie die Verbindung, indem Sie das Kästchen neben der neuen Verbindung aktivieren und auf die Schaltfläche „Testen“  klicken.

Importieren eines JDBC-Treibers

Spectrum™ Technology Platform kann mithilfe eines JDBC-Treibers auf Daten einer beliebigen Datenbank zugreifen. Im Spectrum™ Technology Platform Data Integration-Modul sind Treiber für SQL, Oracle und PostgreSQL und andere Datenbanktypen enthalten. Wenn im Spectrum™ Technology Platform Data Integration-Modul kein Treiber für den von Ihnen benötigten Datenbanktyp enthalten ist, können Sie einen JDBC-Treiber hinzufügen.

Bei dieser Prozedur importieren Sie einen JDBC-Treiber, indem Sie die Treiberdateien auf den Spectrum™ Technology Platform-Server kopieren. Nach Abschluss dieser Prozedur steht der Treiber beim Definieren einer JDBC-Datenbankverbindung in der Management Console zur Verfügung.

Anmerkung: Die Prozedur funktioniert für JDBC 4.x-Treiber. Wenn der Treiber, den Sie hinzufügen möchten, eine ältere Version von JDBC verwendet, müssen Sie den Treiber manuell in der Management Console hinzufügen. Weitere Informationen finden Sie unter [Manuelles Hinzufügen eines JDBC-Treibers](#) auf Seite 41.

1. Speichern Sie alle JDBC-Treiberdateien für die Datenbank in einem Ordner mit folgendem Namen:

Name.jdbc

Für *Name* können Sie einen beliebigen Namen eingeben. Der Ordnername muss mit *.jdbc* enden.

2. Melden Sie sich auf dem Server an, auf dem Spectrum™ Technology Platform ausgeführt wird.
3. Kopieren Sie den Ordner, der den Treiber enthält, in folgenden Ordner:

Spectrum Location\server\app\drivers

Der Treiber wird automatisch importiert.

4. Um zu überprüfen, ob der Treiber erfolgreich importiert wurde, melden Sie sich in der Management Console an und öffnen Sie **System > Treiber**. Der Treiber sollte in der Liste vorhanden sein.

Wenn der Treiber nicht in der Liste vorhanden ist, öffnen Sie in der Management Console das Systemprotokoll und suchen Sie nach Fehlern bezüglich der Bereitstellung von JDBC-Treibern.


Manuelles Hinzufügen eines JDBC-Treibers

Spectrum™ Technology Platform kann mithilfe eines JDBC-Treibers auf Daten einer beliebigen Datenbank zugreifen. Im Spectrum™ Technology Platform Data Integration-Modul sind Treiber für SQL, Oracle und PostgreSQL und andere Datenbanktypen enthalten. Wenn im Spectrum™

Technology Platform Data Integration-Modul kein Treiber für den von Ihnen benötigten Datenbanktyp enthalten ist, können Sie einen JDBC-Treiber hinzufügen.

Mit dieser Prozedur fügen Sie JDBC-Treiberdateien zum Server hinzu und definieren dann manuell die Verbindungszeichenfolge und die Verbindungseigenschaften. Vergewissern Sie sich zu Anfang, dass Sie das Format der Verbindungszeichenfolge und die Eigenschaften, die für den Treiber erforderlich sind, kennen. Sie müssen diese genau definieren, damit der Treiber funktioniert. Normalerweise finden Sie Informationen zur Verbindungszeichenfolge und zu den Eigenschaften eines Treibers auf der Webseite des Treiberanbieters.

Anmerkung: Wir empfehlen, diese Prozedur nur zu nutzen, wenn Sie einen JDBC-Treiber hinzufügen, der JDBC 1.x, 2.x oder 3.x verwendet. Wenn der Treiber JDBC 4.x verwendet, empfehlen wir, den Treiber über die Importmethode hinzuzufügen. Weitere Informationen finden Sie unter [Importieren eines JDBC-Treibers](#) auf Seite 41.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **System > Treiber**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für den Treiber ein. Sie können den Namen frei wählen.
5. Geben Sie im Feld **Name der JDBC-Treiberklasse** den Java-Klassennamen des Treibers ein. Diesen Klassennamen finden Sie normalerweise in der Dokumentation Ihres JDBC-Treibers.

Zur Verwendung des Microsoft JDBC-Treibers könnten Sie beispielsweise Folgendes eingeben:

```
com.microsoft.sqlserver.jdbc.SQLServerDriver
```

6. Geben Sie im Feld **Verbindungszeichenfolgen-Vorlage** die JDBC-Verbindungs-URL ein, die zur Herstellung einer Verbindung zur Datenbank verwendet wird. Berücksichtigen Sie in der Verbindungszeichenfolge alle Eigenschaften, die Sie festlegen möchten. Verschiedene Datenbankanbieter verwenden unterschiedliche Verbindungszeichenfolgen. Konsultieren Sie daher die Dokumentation Ihrer Datenbank, um weitere Informationen zur Verbindungszeichenfolge zu erhalten.

Wenn der Treiber von mehr als einer Datenbankverbindung verwendet wird, ziehen Sie die Verwendung von Eigenschaftstoken in der Verbindungszeichenfolge in Betracht, anstatt Eigenschaftswerte fest zu kodieren, die sich möglicherweise je nach Verbindung unterscheiden. Wenn Sie beispielsweise möchten, dass für einige Verbindungen Verschlüsselung verwendet wird und für andere nicht, könnten Sie ein Eigenschaftstoken für die Eigenschaft Verschlüsselung definieren.

Um ein Eigenschaftstoken in der Verbindungszeichenfolge zu nutzen, verwenden Sie folgende Syntax:

```
${PropertyToken}
```

Jedes Eigenschaftstoken, das Sie in der Verbindungszeichenfolgen-Vorlage verwenden, wird bei der Definition einer Datenbankverbindung zu einem Pflichtfeld.

Anmerkung: Verwenden Sie den Eigenschaftstokennamen `${password}` für die Eigenschaft, die das Datenbankkennwort enthalten wird. Wenn Sie diesen Tokennamen verwenden, wird das Kennwort im Feld in der Management Console maskiert und in der Datenbank verschlüsselt werden.

Beispielsweise enthält folgende Verbindungszeichenfolge für SQL Eigenschaftstoken für Host, Port, Instanz und Verschlüsselung:

```
jdbc:sqlserver://${host}:${port};databaseName=${instance};encrypt=${encryption};  
TrustServerCertificate=true
```

Diese Token sind bei der Definition einer Datenbankverbindung, die diesen Treiber verwendet, Pflichtfelder:

[Startseite](#) > [Ressourcen: Datenquellen](#) > [Datenquelle hinzufügen](#)

Datenquelle hinzufügen

*Name

ExampleConnection

Verbindung

*Typ

ExampleDriver


*Host

*Port

*Instance

*encryption

Testen

7. Wenn Eigenschaften vorhanden sind, die Sie als optional für Datenbankverbindungen definieren möchten, definieren Sie diese im Abschnitt **Verbindungseigenschaften**.
 - a) Klicken Sie im Abschnitt **Verbindungseigenschaften** auf die Schaltfläche „Hinzufügen“ .
 - b) Geben Sie im Feld **Beschriftung** eine benutzerfreundliche Beschreibung der Eigenschaft ein. Die Beschriftung, die Sie hier eingeben, wird als Feldbeschriftung im Verbindungsfenster verwendet, wenn eine Verbindung über diesen Treiber erstellt wird.

- c) Geben Sie im Feld **Eigenschaftstoken** das Token für die optionale Eigenschaft ein. Schlagen Sie in der Dokumentation des Datenbanktreibers die Eigenschaften nach, die vom Treiber unterstützt werden.

Anmerkung: Verwenden Sie den Eigenschaftstokennamen `password` für die Eigenschaft, die das Datenbankkennwort enthalten wird. Wenn Sie diesen Tokennamen verwenden, wird das Kennwort im Feld in der Management Console maskiert und in der Datenbank verschlüsselt werden.

Wenn Sie beispielsweise für Datenbankverbindungen, die diesen Treiber verwenden, Verschlüsselung optional machen möchten, könnten Sie die Eigenschaft Verschlüsselung wie folgt definieren:

Startseite > System: Treiber > Treiber bearbeiten

Treiber bearbeiten

*Name

ExampleDriver

*Name der JDBC-Treiberklasse ?

com.microsoft.sqlserver.jdbc.SQLServerDriver

*Verbindungszeichenfolgen-Vorlage ?

jdbc:sqlserver://{host}:{port};databaseName=\${insi

Eigenschaften und Treiber

Verbindungseigenschaften ?



Beschriftung	Eigenschaftstoken
<input type="checkbox"/> username	user
<input type="checkbox"/> password	password
<input type="checkbox"/> Use SSL	useSSL

Wenn eine Datenbankverbindung diesen Treiber verwendet, würde die Eigenschaft Verschlüsselung als optionale Eigenschaft in der Datenbankverbindung angezeigt:

Startseite > Ressourcen: Datenquellen > Datenquelle hinzufügen

Datenquelle hinzufügen

*Name

Verbindung

Type

*Host

*Port


*Instance

*encryption

User Name

Password

Use SSL

8. Melden Sie sich am Server an, auf dem Spectrum™ Technology Platform läuft, und platzieren Sie die Datenbanktreiberdatei in einem Ordner auf dem Server. Der Speicherort spielt keine Rolle.
9. Klicken Sie im Abschnitt **Treiberdateien** auf die Schaltfläche „Hinzufügen“ .
10. Geben Sie im Feld **Dateipfad** den Pfad zur Datenbanktreiberdatei auf dem Server an.

11. Klicken Sie auf **Speichern**.

Löschen eines importierten JDBC-Treibers

Ein JDBC-Treiber kann nicht über Management Console gelöscht werden, wenn er in Spectrum™ Technology Platform importiert und nicht manuell über Management Console hinzugefügt wurde. Folgen Sie in dem Fall dieser Prozedur, um den Treiber zu löschen.

Wichtig: Vergewissern Sie sich vor dem Löschen eines Treibers, dass dieser nicht von einer Datenbankverbindung verwendet wird.

1. Stoppen Sie den Spectrum™ Technology Platform-Server.
2. Öffnen Sie den folgenden Ordner:

Spectrum Location\server\app\drivers

3. Löschen Sie im Ordner `drivers` den Ordner, der den Treiber enthält.
4. Starten Sie den Spectrum™ Technology Platform-Server.
5. Um zu überprüfen, ob der Treiber gelöscht wurde, melden Sie sich in der Management Console an, öffnen Sie **System > Treiber**, und vergewissern Sie sich, dass der Treiber nicht mehr aufgeführt wird.

Unterstützte Datenbank-Datentypen

Spectrum™ Technology Platform unterstützt folgende üblicherweise in Datenbanken verwendeten 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.
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 $(2-2^{-52}) \times 2^{1023}$ 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 $(2-2^{-23}) \times 2^{127}$ 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.
time	Ein Datentyp, der die Tageszeit enthält. Beispiel: 21:15:59 oder 9:15:59 PM.
Raw	Ein Oracle-Datentyp zur Speicherung von Binärdaten variabler Länge. Die maximale Größe beträgt 2000 Byte. (Die maximale Länge in Oracle 7 betrug 255 Byte.)

Andere Datenbank-Datentypen werden automatisch einem der unterstützte Datentypen wie folgt zugeordnet:

Datenbank-Datentyp	Unterstützter Datentyp
Typen für Datum/Uhrzeit	
TIMESTAMP	datetime
Zeichenfolgentypen	
CHAR	string
CLOB	string
LONGVARCHAR	string
NCHAR	string
NVARCHAR	string
VARCHAR	string
Numerische Typen	
BIGINT	long
Decimal	double
FLOAT	double
NUMERIC	bigdecimal

Datenbank-Datentyp	Unterstützter Datentyp
REAL	float
SMALLINT	integer
TINYINT	integer
Boolesche Typen	
BIT	boolean

Unterstützte Datenbank-Datentypen des Location Intelligence-Moduls

Diese Datenbank-Datentypen werden automatisch einem der unterstützte Datentypen des Location Intelligence-Moduls wie folgt zugeordnet.

Datenbank-Datentyp	Unterstützter Datentyp
SQL Server	
tinyint	SHORT_INTEGER
smallint	SHORT_INTEGER
int	INTEGER
bigint	LONG_INTEGER
float	DOUBLE
real	DOUBLE
decimal(10, 5)	DOUBLE
numeric(10, 5)	DOUBLE
date	DATE
time	TIME

Datenbank-Datentyp	Unterstützter Datentyp
datetime	DATE_TIME
smalldatetime	DATE_TIME
char(10)	STRING
varchar(10)	STRING
nchar(10)	STRING
nvarchar(10)G	STRING
binary(10)	BINARY
varbinary(10)	BINARY
PostGIS	
smallint	SHORT_INTEGER
integer	INTEGER
bigint	LONG_INTEGER
numeric(10, 5)	DOUBLE
real	DOUBLE
double precision	DOUBLE
serial	INTEGER
bigserial	LONG_INTEGER
bytea	BINARY
date	DATE

Datenbank-Datentyp	Unterstützter Datentyp
time	TIME
timestamp	DATE_TIME
character(10)	STRING
character varying(10)	STRING
nchar(10)	STRING
Oracle	
NUMBER	DOUBLE
CHAR(10)	STRING
VARCHAR(10)	STRING
VARCHAR2(10)	STRING
NCHAR(10)	STRING
NVARCHAR2(10)	STRING
DATE	DATE_TIME
TIMESTAMP	DATE_TIME
BLOB	BINARY
SAP HANA	
tinyint	SHORT_INTEGER
smallint	SHORT_INTEGER


Datenbank-Datentyp	Unterstützter Datentyp
integer	INTEGER
bigint	LONG_INTEGER
smalldecimal	DOUBLE
decimal(10, 5)	DOUBLE
real	DOUBLE
double	DOUBLE
float(30)	DOUBLE
varchar(30)	STRING
nchar(10)	STRING
nvarchar(30)	STRING
alphanum(30)	STRING
date	DATE
time	TIME
seconddate	DATE_TIM
timestamp	DATE_TIM
varbinary(30)	BINARY

Einschränkungen

- MongoDB/Cassandra-Connectors über PrestoDB werden in Metadata Insights nicht unterstützt. Es gibt unterschiedliche Connectors für MongoDB und Cassandra.
- Write to Any DB über Presto wird von Presto DB nicht empfohlen und wird vom Presto JDBC-Connector nicht unterstützt.

Verbindung zu Knox herstellen

Mit einem Apache Knox Gateway können Sie über den Knox-Sicherheitslayer auf einen Hadoop-Dienst zugreifen. Damit Spectrum™ Technology Platform über Knox auf Daten in Hadoop zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu Knox definieren. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten über Knox von Hadoop lesen und dorthin schreiben können.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.


5. Wählen Sie im Feld **Typ** `Gateway` aus.
6. Wählen Sie im Feld **Gatewaytyp** die Option **Knox** aus.
7. Geben Sie im Feld **Host** den Hostnamen oder die IP-Adresse des Knotens im HDFS-Cluster ein, auf dem das Gateway ausgeführt wird.
8. Geben Sie in das Feld **Port** die Portnummer für das Knox-Gateway ein.
9. Geben Sie in das Feld **Benutzername** den Benutzernamen für das Knox-Gateway ein.
10. Geben Sie das Feld **Kennwort** das Kennwort für Ihren Zugriff auf das Knox-Gateway ein.
11. Geben Sie in das Feld **Gatewayname** den Namen des Knox-Gateways ein, auf das Sie zugreifen möchten.
12. Geben Sie in das Feld **Clustername** den Namen des Hadoop-Clusters ein, auf das zugegriffen werden soll.
13. Wählen Sie im Feld **Protokoll** `webhdfs` aus.
14. Geben Sie in das Feld **Dienstname** den Namen des Hadoop-Dienstes ein, auf den zugegriffen werden soll.
15. Um die Verbindung zu testen, klicken Sie auf **Testen**.
16. Klicken Sie auf **Speichern**.

Nachdem Sie eine Knox-Verbindung zu einem HDFS-Cluster definiert haben, können Sie die Verbindung im Enterprise Designer in den Schritten **Read from File** und **Write to File** verwenden. Sie können das HDFS-Cluster beim Definieren einer Datei in einem Quell- oder Datenladeschritt auswählen, indem Sie auf **Remote-Computer** klicken.

Verbindung mit Marketo

Damit Spectrum™ Technology Platform auf Daten in Marketo zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu Marketo definieren. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten von Marketo lesen und dorthin schreiben können.

Anmerkung: Diese Verbindung wird im Metadata Insights-Modul verwendet.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ Marketo** aus.
6. Geben Sie in das Feld **Endpunkt-URL** die Endpunkt-URL Ihres Marketo-Kontos ein.

Um Ihre Endpunkt-URL zu ermitteln, melden Sie sich an Ihrem Marketo-Konto an und navigieren Sie zu **Admin > Integration > Webservices**. Sie finden die Endpunkt-URL unter der Überschrift **REST-API** in folgendem Format:

```
https://AccountID.mktorest.com/rest
```

Kopieren Sie den Teil der URL vor `/rest`. Beispiel: `https://AccountID.mktorest.com`.

7. Geben Sie die Client-ID und den geheimen Schlüssel Ihres Marketo-Kontos ein.

Um Ihre Client-ID und den geheimen Schlüssel zu ermitteln, melden Sie sich an Ihrem Marketo-Konto an und navigieren Sie zu **Admin > Integration > LaunchPoint > API-REST > Details anzeigen**. Die Details werden in einem Popup-Fenster angezeigt.

8. Um die Verbindung zu testen, klicken Sie auf **Testen**.
9. Klicken Sie auf **Speichern**.

Marketo-Beschränkungen

1. Diese Abfrage gilt nur für `List-` und `Activity_type`-Entitäten. Geben Sie für andere Entitäten den Filtertyp an.

```
Select * from Marketo_Table
```

2. Marketo unterstützt keine Joins, außer zwischen Lead- und Lead_List-Entitäten. Die Join-Abfrage zwischen Lead und Lead_List für eine List_Id lautet wie folgt:

```
Select Lead.* from Lead Inner Join Lead_List
On Lead.ID = Lead_List.Lead_ID
And Lead_List.List_ID = <List ID>
```


Verbinden mit Microsoft Dynamics 365

Verbinden mit Microsoft Dynamics 365 Online

Damit Spectrum™ Technology Platform auf Daten in Microsoft Dynamics 365 Online zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu Microsoft Dynamics 365 Online definieren. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten aus Microsoft Dynamics 365 Online lesen und dorthin schreiben können.

Anmerkung: Diese Verbindung wird im Metadata Insights-Modul verwendet.

Führen Sie die angegebenen Schritte aus, um eine Verbindung zu **Microsoft Dynamics 365 Online** zu definieren.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ Microsoft Dynamics 365** aus.
6. Wählen Sie im Feld **Entwicklungstyp** den Eintrag **Online** aus.
7. Geben Sie in das Feld **Benutzername** Ihren Benutzernamen für Microsoft Dynamics ein.
8. Geben Sie in das Feld **Kennwort** Ihr Kennwort für Microsoft Dynamics ein.
9. Geben Sie den eindeutigen Organisationsnamen, der Ihre CRM-Instanz identifiziert, in das Feld **Organisationsname** ein.

Um Ihren eindeutigen Organisationsnamen zu ermitteln, melden Sie sich in Microsoft Dynamics an und navigieren Sie zu **Einstellungen > Anpassung > Anpassungen > Entwicklerressourcen**. Ihr eindeutiger Organisationsname wird angezeigt.

10. Wählen Sie im Feld **Region** die geografische Region Ihres Microsoft Dynamics-Kontos aus.
11. Um die Verbindung zu testen, klicken Sie auf **Testen**.

12. Klicken Sie auf **Speichern**.

Verbinden mit Microsoft Dynamics 365 On Premises

Derzeit unterstützt Spectrum anspruchsbasierte Authentifizierung für Microsoft Dynamics 365 On Premises.

Voraussetzungen:

Importieren Sie das Zertifikat in die Keystore-Datei: Um die Dynamics CRM Server-Zertifikate in den Keystore der Spectrum Java-Distribution zu kopieren, gehen Sie wie folgt vor:


1. Kopieren Sie die Serverzertifikate in einen lokalen Ordner.
2. Navigieren Sie zu folgendem Pfad für die Spectrum JAVA-Distribution:
`<SPECTRUM_HOME>\java\jre\lib\security.`
3. Führen Sie den folgenden Befehl aus, um die Zertifikate zu importieren: `<codeph>keytool -importcert -alias <certificate alias name> -file "<certificate path>\<certificate name>" -keystore keystore.jks</codeph>` unter Windows und `<codeph>keytool -import -alias <certificate alias name> -file "<certificate path>/<certificate name>" -keystore keystore.jks</codeph>` unter UNIX.

Definieren einer Verbindung zu Microsoft Dynamics 365 On Premises

Damit Spectrum™ Technology Platform auf Daten in Microsoft Dynamics 365 On Premises zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu Microsoft Dynamics 365 On Premises definieren. Nachdem Sie die Verbindung definiert haben, können Sie im Enterprise Designer Flüsse erstellen, die Daten aus Microsoft Dynamics 365 On Premises lesen und dorthin schreiben können.

Anmerkung: Diese Verbindung wird im Metadata Insights-Modul verwendet.

Definieren Sie eine Verbindung zu **Microsoft Dynamics 365 On Premises** wie folgt:

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie unter **Typ** die Option **Microsoft Dynamics 365** aus.
6. Wählen Sie unter **Entwicklungstyp** die Option **On Premises** aus.
7. Geben Sie in das Feld **Benutzername** Ihren Microsoft Dynamics-Benutzernamen ein.
8. Geben Sie in das Feld **Kennwort** Ihr Microsoft Dynamics-Kennwort ein.

9. Geben Sie in das Feld **Hostname** den Namen des Hosts ein.
10. Geben Sie in das Feld **Portname** den Namen des Ports ein.
11. Geben Sie in das Feld **STS-URL** die URL des STS ein.
12. Klicken Sie auf **Testen**, um die Verbindung zu testen.
13. Klicken Sie auf **Speichern**.


Einschränkungen

Folgende Einschränkungen bestehen:

1. **Erstellen/Aktualisieren:** Erstellen/Aktualisieren kann fehlgeschlagen, wenn eine Spalte eine Entität mehreren Referenzentitäten zugeordnet ist. Beispielsweise kann in „Kunde“ die Spalte „ParentCustomerId“ den Entitäten „Konto“, „Lead“ und anderen zugeordnet sein. Zur Lösung müssen die Daten dieser Spalten im Format „ReferenceEntityName:GUID“ anstelle von „GUID“ vorliegen.

Herstellen einer Verbindung zu einem Modellspeicher

Stellen Sie eine Verbindung zu einem Modellspeicher her, um aus verschiedenen Quellen wie Datenbanken, Dateiservern und Cloud-Diensten zusammengefasste Daten zu verwenden. Nachdem Sie eine Verbindung definiert haben, können Sie die Daten aus logischen und physischen Modellen eines Modellspeichers in den Schritten **Read from DB** und **Write to DB** des Enterprise Designer verwenden.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ Modellspeicher** aus.
6. Geben Sie in das Feld **Modellspeicher** den Namen des Modellspeichers ein, zu dem Sie eine Verbindung herstellen.

Die Namen der verfügbaren Modellspeicher finden Sie in Metadata Insights unter „Modeling“ auf der Registerkarte **Modellspeicher**.

7. Um die Verbindung zu testen, klicken Sie auf **Testen**.
8. Klicken Sie auf **Speichern**.

Verbinden mit NetSuite


Damit Spectrum™ Technology Platform auf Daten in NetSuite zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu NetSuite definieren. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten von NetSuite lesen und dorthin schreiben können. Beim Lesen aus einer NetSuite-Verbindung und Schreiben in dieser Verbindung werden sowohl der interaktive Modus als auch der Batch-Modus unterstützt.

Anmerkung: Diese Verbindung wird im Metadata Insights-Modul verwendet.

Spectrum™ Technology Platform unterstützt die folgenden NetSuite-Entitätstypen:

- Standarddatensätze
- Benutzerdefinierte Datensätze
- Gespeicherte Suchvorgänge
- Joins zwischen Standarddatensätzen

So stellen Sie eine Verbindung mit NetSuite her:

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ NetSuite** aus.
6. Geben Sie die E-Mail-Adresse, die mit dem für die Verbindung zu verwendenden NetSuite-Konto verknüpft ist, in das Feld **E-Mail** ein.
7. Geben Sie das Kennwort des NetSuite-Kontos in das Feld **Kennwort** ein.
8. Geben Sie den Benutzernamen für das NetSuite-Konto in das Feld **Konto** ein.
9. Wählen Sie die entsprechende Rolle für diese Verbindung aus den Rollen, die einem bestimmten NetSuite-Benutzerkonto zugeordnet sind, im Feld **Rolle** aus.

Das Feld **Rolle** ist optional. Wenn Sie das Feld **Rolle** leer lassen, wird die Standardrolle für die Anmeldung über die Verbindung verwendet.

Achtung: Nur Standardrollen werden unterstützt. Benutzerdefinierte Rollen werden nicht unterstützt.

10. Um die Verbindung zu testen, klicken Sie auf **Testen**.
11. Klicken Sie auf **Speichern**.

Anmerkung: Um einen `INSERT`-Vorgang für einen Datensatz mithilfe einer NetSuite-Verbindung durchzuführen, verwenden Sie eine `UPSERT`-Abfrage, bei der der Primärschlüssel (`internalId`) leer ist.

NetSuite-Beschränkungen

1. Wenn mithilfe von Joins eine Abfrage durchgeführt wird, müssen Sie bestimmte Spalten angeben. Beispielsweise wird die folgende Abfrage nicht unterstützt:

```
select * from CUSTOMER_M
```

2. Gleichzeitige Verbindungen zu NetSuite werden nicht unterstützt, da NetSuite nur eine einzige Anmeldung pro Konto zulässt.
3. Sie können nur Standarddatensätze und benutzerdefinierte Datensätze schreiben.
4. Für `UPDATE`- und `UPSERT`-Abfragen wird ein `UPSERT`-Vorgang ausgeführt.
5. Im „Write to DB“-Schritt beträgt die maximal zulässige Batchgröße bei einem `insert`-Vorgang 200 und bei einem `update`-Vorgang 100.
- 6.


Verbinden mit NoSQL

Damit Spectrum™ Technology Platform auf Daten in einer NoSQL-Datenbank zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zur NoSQL-Datenbank definieren.

Folgende NoSQL-Datenbanktypen werden unterstützt:

1. Couchbase
2. MongoDB

Sobald Sie die gewünschte NoSQL-Verbindung definiert haben, können Sie im Enterprise Designer Flüsse erstellen, die Daten aus dieser Datenbank lesen und dorthin schreiben können.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.


5. Wählen Sie im Feld **Typ** eine der folgenden Optionen aus:
 - Couchbase

- MongoDB
6. Geben Sie **Host**, **Port**, **Datenbank**, **Benutzername** und **Kennwort** zu der spezifischen NoSQL-Datenbank an, auf die Sie zugreifen möchten.
 7. Klicken Sie auf **Testen**, um zu überprüfen, ob die Verbindung zur Datenbank erfolgreich hergestellt wurde.
 8. Klicken Sie auf **OK**.

Verbinden mit Salesforce

Damit Spectrum™ Technology Platform auf Daten in Salesforce zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu Salesforce definieren. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten von Salesforce lesen und dorthin schreiben können.

Anmerkung: Diese Verbindung wird im Metadata Insights-Modul verwendet.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ Salesforce** aus.
6. Geben Sie die auf dem Salesforce-Datenspeicher registrierte E-Mail-ID in das Feld **Benutzername** ein.
7. Geben Sie eine Kombination aus dem Kennwort für das Salesforce-Portal und dem über das Salesforce-Portal generierten Sicherheits-Token in das Feld **Kennwort** ein.

Wenn Ihr Kennwort beispielsweise „Sales@Test“ lautet und der von Salesforce für Sie bereitgestellte Sicherheits-Token 56709367 ist, würde das Kennwort zur Authentifizierung dieser Salesforce-Verbindung „Sales@Test56709367“ lauten.

8. Um die Verbindung zu testen, klicken Sie auf **Testen**.
9. Klicken Sie auf **Speichern**.

Anmerkung: Prüfungsfelder sind standardmäßig auf allen Tabellen aktiviert. Die folgenden Prüffelder sind in Salesforce vorhanden:

- Erstellungsdatum
- Datum der letzten Änderung
- Erstellt von

- Zuletzt geändert von

Achtung: Physische Modelle, die in Spectrum™ Technology Platform Version 10 sowie in früheren Versionen mithilfe von Salesforce-Verbindungen erstellt wurden, müssen geöffnet und erneut gespeichert werden, um Prüfungsfelder auf ihren Tabellen zu aktivieren.


Salesforce-Einschränkung

Die Aggregierungsfunktionen werden bei der Ausführung von Abfragen an den Modellspeicher nicht unterstützt.

Verbinden mit SAP NetWeaver

Durch die Erstellung einer SAP NetWeaver-Verbindung in der Management Console mithilfe der OData Services können Sie Ihre CRM- und ERP-Daten lesen, schreiben und synchronisieren. Während des Lesens aus und Schreibens in eine SAP-Verbindung werden sowohl interaktive als auch Batchmodi unterstützt.

Anmerkung: Diese Verbindung wird im Metadata Insights-Modul verwendet.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ SAP** aus.
6. Geben Sie den Benutzernamen für den Zugriff auf den SAP-Webservice in das Feld **Benutzername** ein.
7. Geben Sie das Kennwort des SAP-Webservices in das Feld **Kennwort** ein.
8. Geben Sie die Adressen des für diese Verbindung zu verwendenden Odata-Webservices in das Feld „OdataURL“ ein.
9. Klicken Sie auf **Testen**.
Eine Meldung bestätigt, dass der Verbindungstest erfolgreich war.
10. Klicken Sie auf **Speichern**.
Eine Meldung bestätigt, dass die Verbindung erfolgreich hergestellt wurde.


Anmerkung: Um Abrufvorgänge durchzuführen, muss ein OData-Dienst die Vorgänge \$skip und \$top unterstützen. Wenn der Dienst diese Vorgänge nicht unterstützt, weisen die abgerufenen Datensätze Inkonsistenzen in der Vorschau für den Modellspeicher auf.

SAP NetWeaver-Beschränkungen

Für UPDATE- und UPSERT-Vorgänge wird ein UPDATE-Vorgang ausgeführt.

Verbinden mit SharePoint

Damit Spectrum™ Technology Platform auf Daten in SharePoint zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu SharePoint definieren. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten von SharePoint lesen und dorthin schreiben können.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ Cloud** aus.
6. Wählen Sie im Feld **Cloud-Dienst** die Option **SharePoint** aus.
7. Wählen Sie im Feld **Version v2010** aus. Spectrum™ Technology Platform unterstützt derzeit SharePoint in der Version 2010.
8. Wählen Sie im Feld **Protokoll** das Protokoll aus, das zur Herstellung einer Verbindung mit SharePoint erforderlich ist.
9. Geben Sie in das Feld **Serveradresse** den Hostnamen oder die IP-Adresse des SharePoint-Servers ein, zu dem Sie eine Verbindung herstellen möchten.
10. Geben Sie den Benutzernamen und das Kennwort ein, die für die Authentifizierung in SharePoint verwendet werden.
11. Geben Sie im Feld **Projekt** das spezifische Projekt ein, auf dessen SharePoint-Speicherort Sie zugreifen möchten.
12. Um die Verbindung zu testen, klicken Sie auf **Testen**.
13. Klicken Sie auf **Speichern**.

Beispiel

Beispiel: Sie möchten eine Verbindung zu folgender SharePoint-URL erstellen:

```
https://sharepoint.example.com/sites/myportal
```


Füllen Sie die Felder **Protokoll**, **Serveradresse** und **Projekt** wie folgt aus:

- **Protokoll:** https
- **Serveradresse:** sharepoint.example.com
- **Projekt:** myportal

Verbinden mit Siebel

Damit Spectrum™ Technology Platform auf Daten in Siebel zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu Siebel definieren. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten von Siebel lesen und dorthin schreiben können.

Anmerkung: Diese Verbindung wird im Metadata Insights-Modul verwendet.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ Siebel** aus.
6. Geben Sie in das Feld **Host** den Hostnamen des Servers ein, auf dem Siebel installiert ist.
7. Geben Sie in das Feld **Port** die Nummer des Überwachungsports für die SCBroker-Komponente ein.

Anmerkung: Verwenden Sie 2321 für den SCB-Port.

8. Geben Sie in das Feld **Server** den Namen Ihres Siebel Enterprise Server ein.
9. Geben Sie in das Feld **Objekt-Manager** den Namen des Objekt-Managers ein.
10. Geben Sie den Benutzernamen und das Kennwort ein, die für die Authentifizierung in Siebel verwendet werden.
11. Wählen Sie im Feld **Gebietsschema** den Typ des Objekt-Managers und Sprachpakets für die Siebel Business-Anwendung aus.

Anmerkung: Für andere Gebietschemas außer **Englisch** müssen Sie die spezifischen JAR-Dateien installieren. Weitere Informationen finden Sie unter **Erforderliche JAR-Dateien**.

12. Um die Verbindung zu testen, klicken Sie auf **Testen**.
13. Klicken Sie auf **Speichern**.

Innerhalb eines physischen Modells werden die Siebel Business Components im Format `<Business Object>.<Business Component>` angezeigt.

Anmerkung: Business Components, die nicht Teil eines Business Object sind, werden in einem physischen Modell nicht angezeigt.

Siebel-Beschränkungen

1. Siebel Business Components-Felder mit dem gleichen Namen, aber unterschiedlicher Schreibweise, werden verarbeitet, indem das erste Vorkommen unverändert bleibt und an alle nachfolgenden Vorkommen `_PB_` und ein ansteigender numerischer Wert angehängt werden.

Felder wie z. B. `DeDup Token`, `Dedup Token` und `DEdup Token` werden in einer Siebel Business Component in `DeDup Token`, `Dedup Token_PB_1` bzw. `DEdup Token_PB_2` umbenannt und verwendet.

2. Wenn ein Feldname in einem Siebel-Schema einen Punkt (.) enthält, werden die Punkte aus dem Feldnamen entfernt, damit er in Spectrum™ Technology Platform verwendet werden kann.
3. Während der Erstellung eines physischen Modells aus einer Siebel-Verbindung werden Links zwischen zwei Siebel Business Components nur angezeigt, wenn der Name eines Links mit dem Business Component-Namen im Siebel-Schema übereinstimmt.


Wenn z. B. in einem Siebel-Schema ein Link zwischen den Business Components `Account` und `Contact` definiert wird, während eine andere Business Component `Contact Custom` die Links von `Custom` verwendet, dann wird in dem für diese Siebel-Verbindung erstellten physischen Modell kein Link zwischen `Account` und `Contact Custom` angezeigt. Links werden nur zwischen `Account` und `Contact` angezeigt, da der ursprüngliche Link zwischen diesen beiden im Siebel-Schema besteht.

4. Felder einer Business Component, die im Siebel-Schema als inaktiv gekennzeichnet sind, werden nicht in physischen Modellen, logischen Modellen und dem Modellspeicher angezeigt.

Verbinden mit Splunk

Damit Spectrum™ Technology Platform auf Daten in Splunk zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu Splunk definieren. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten von Splunk lesen und dorthin schreiben können.

Anmerkung: Diese Verbindung wird im Metadata Insights-Modul verwendet.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ Splunk** aus.
6. Geben Sie in das Feld **Benutzername** den Benutzernamen des Splunk-Kontos ein, um die Splunk-Instanz zu authentifizieren.
7. Geben Sie in das Feld **Kennwort** das Kennwort des Splunk-Kontos ein.
8. Geben Sie in das Feld **Hostname** die Adresse oder den Hostnamen des Servers ein, auf dem die Splunk-Datenquelle gehostet wird.
9. Geben Sie in das Feld **Port** die Portnummer der Splunk-Datenquelle ein.
10. Um die Verbindung zu testen, klicken Sie auf **Testen**.
11. Klicken Sie auf **Speichern**.

Splunk-Beschränkungen


Diese Abfrage wird nicht unterstützt:

```
select count(*) from SplunkTable
```

Verbinden mit SuccessFactors

Damit Spectrum™ Technology Platform auf Daten in SuccessFactors zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu SuccessFactors definieren. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten von SuccessFactors lesen und dorthin schreiben können.

Anmerkung: Diese Verbindung wird im Metadata Insights-Modul verwendet.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ SuccessFactors** aus.
6. Geben Sie in das Feld **Unternehmens-ID** Ihre Unternehmens-ID ein, die die eindeutige Instanz Ihres Unternehmens in einem bestimmten SuccessFactors-Rechenzentrum identifiziert.
7. Geben Sie in das Feld **Dienst-URL** die URL für den SuccessFactors-Server ein, zu dem Sie eine Verbindung herstellen möchten. Diese URL verweist auf das globale Rechenzentrum, dem Ihre Unternehmens-ID zugeordnet ist.
8. Geben Sie Ihren Benutzernamen und Ihr Kennwort für Ihre SuccessFactors-Client-Instanz ein.
9. Um die Verbindung zu testen, klicken Sie auf **Testen**.
10. Klicken Sie auf **Speichern**.


SuccessFactors-Beschränkungen

1. Batchvorgänge können nur mithilfe von `upsert`-Abfragen ausgeführt werden. Folglich werden `insert`- und `update`-Abfragen in Batchvorgängen ebenfalls als `upsert`-Abfragen ausgeführt.
2. Die Tabellen-/Spalteneigenschaften, wie sie im Schema des physischen Modells einer SuccessFactors-Verbindung angezeigt werden, könnten während der Verwendung des entsprechenden Vorgangs nicht wie erwartet funktionieren. Eine Spalte, die beispielsweise als aktualisierbar gekennzeichnet wurde, könnte eine Systemausnahme zurückgeben, wenn Sie versuchen, diese Spalte zu aktualisieren.

Verbinden mit SugarCRM

Damit Spectrum™ Technology Platform auf Daten in SugarCRM zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu SugarCRM definieren. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten von SugarCRM lesen und dorthin schreiben können. Sowohl Online- als auch lokale Versionen von SugarCRM werden unterstützt.

Anmerkung: Diese Verbindung wird im Metadata Insights-Modul verwendet.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ SugarCRM** aus.

6. Geben Sie Ihren Benutzernamen und Ihr Kennwort für SugarCRM ein.
7. Geben Sie die URL des für diese Verbindung zu verwendenden SugarCRM-Kontos in das Feld **URL** ein.
8. Geben Sie die **Client-ID** und den **geheimen Clientschlüssel** Ihres SugarCRM-Kontos ein.
9. Um die Verbindung zu testen, klicken Sie auf **Testen**.
10. Klicken Sie auf **Speichern**.

SugarCRM-Beschränkungen


1. Für `UPDATE`- und `UPSERT`-Abfragen wird ein `UPSERT`-Vorgang ausgeführt.
2. In den Spalten **Nullwerte zulassen** und **Aktualisierbar** der Tabelleneigenschaften, wie sie im **Schema des physischen Modells** der Verbindung angezeigt werden, wird möglicherweise nicht der richtige Vorgang dargestellt. Beispielsweise könnte eine Spalte, die nicht mit „Aktualisierbar“ gekennzeichnet ist, eine Systemausnahme auslösen, wenn Sie versuchen, diese zu aktualisieren. Umgekehrt könnte eine Spalte, die mit „Nullwerte zulassen“ gekennzeichnet ist, eine Aktualisierungsausnahme auslösen.
3. Bei der Abfrage mithilfe von Joins müssen Sie einen Alias verwenden.

Verbinden mit Oracle Eloqua

Damit Spectrum™ Technology Platform auf Daten in Oracle Eloqua zugreifen kann, müssen Sie mithilfe der Management Console eine Verbindung zu Oracle Eloqua definieren.

. Danach können Sie im Enterprise Designer Flüsse erstellen, die Daten von Eloqua lesen und dorthin schreiben können.

Anmerkung: Diese Verbindung wird im Metadata Insights-Modul verwendet.

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.
3. Klicken Sie auf die Schaltfläche „Hinzufügen“ .
4. Geben Sie im Feld **Name** einen Namen für die Verbindung ein. Sie können den Namen frei wählen.

Anmerkung: Sobald Sie eine Verbindung gespeichert haben, können Sie den Namen nicht mehr ändern.

5. Wählen Sie im Feld **Typ Oracle Eloqua** aus.
6. Geben Sie den **Standortnamen** ein, der dem Unternehmensnamen entspricht.
7. Geben Sie in das Feld **Benutzername** den Benutzernamen ein.
8. Geben Sie in das Feld **Kennwort** das Kennwort ein.
9. Klicken Sie auf **Testen**, um die Verbindung zu testen.

10. Klicken Sie auf **Speichern**.

Unterstützende Entitäten

In der folgenden Tabelle werden unterstützende Entitäten und Vorgänge für Oracle Eloqua beschrieben.

Tabelle 1:

Entitätsname	Erstellen	Lesen	Update	Löschen	Batch-Unterstützung	Maximale Batchgröße
Kunden	X	X	X	X	Erfüllen/Aktualisieren*	1000
Kontogruppen		X				
Kampagne		X				
Kontakte	X	X	X	X	Erfüllen/Aktualisieren*	1000
Kontaktliste	X	X	X	X		
Kontaktsegment	X	X	X	X		
E-Mails		X				
E-Mail-Ordner		X				
E-Mail-Gruppen		X				
Microsites		X				
Benutzer		X				
Besucher		X				
Aktivität						
E-Mail öffnen		X				
EmailClickthrough		X				

Entitätsname	Erstellen	Lesen	Update	Löschen	Batch- Einfügung	Maximale Batchgröße
E-Mail senden		X				
Abonnieren		X				
Abmelden		X				
Bounceback		X				
WebVisit		X				
PageView		X				
FormSubmit		X				
Benutzerdefinierte Entitäten						
ContactListMembers	X	X		X	Einfügen/Löschen	1000
ContactSegmentMembers		X				

* Aktualisierungsvorgang wird als Upsert ausgeführt.

Besondere Vorgänge

1. Verwenden Sie die folgende Join-Abfrage, um Kontakte in einer Kontaktliste abzurufen:

```
select * from Contacts inner join ContactListMembers on
Contacts.Eloqua_Contact_ID = ContactListMembers.Contact_Id where
ContactListMembers.ContactList_Id = '<id>'
```

Verwenden Sie die folgende Join-Abfrage, um Kontakte in einem Kontaktsegment abzurufen:

```
select * from Contacts inner join ContactSegmentMembers on
Contacts.Eloqua_Contact_ID = ContactSegmentMembers.Contact_Id where
ContactSegmentMembers.Contactlist_Id = '<id>'
```

2. Verwenden Sie die folgende Anweisung, um Kontakte in eine Kontaktliste einzufügen:

```
insert into ContactListMembers (ContactList_ID,Contact_ID) values
('<contactlist_id>','<contact_id>')
```

3. Verwenden Sie die folgende Anweisung, um Kontakte aus einer Kontaktliste zu löschen:

```
delete from ContactListMembers where ContactList_ID =
'<contactlist_id>' and Contact_ID = '<contact_id>'
```

Einschränkungen

Folgende Einschränkungen bestehen:

1. Erstellen/Aktualisieren:

- a. Einfügen/Upsert schlägt fehl, wenn keine Nicht-Null-Spalten vorhanden sind oder diese leer sind.
- b. Einfügen/Upsert schlägt fehl, wenn Werte in eindeutigen Spalten für einen bestimmten Batch nicht eindeutig sind.
- c. Um eine Rollback-Ausnahme zu vermeiden, lassen Sie den Wert von **Anzahl zu übergebender Batches** auf „1“.

2. Lesen:


- a. Bei benutzerdefinierten Entitäten kann der Vorgang **Auswählen** nur auf Join-Vorgänge mit Entitäten „Kontakt“ angewendet werden.

3. Filter:

- a. Folgende Filter werden unterstützt; =, !=, >, <, >=, <=
- b. Die Bedingungsoperatoren **IN** und **NOT IN** werden nicht unterstützt, wenn mehr als ein Wert angegeben wird.
- c. **Joins** zwischen Entitäten werden nicht unterstützt.
- d. Der Bedingungsoperator **OR** wird nur für die Entitäten „Konten“ und „Kontakte“ unterstützt.
- e. Der Bedingungsoperator **AND** kann nur zwischen zwei Bedingungen verwendet werden.
- f. Der Filter = funktioniert nicht immer bei Felder des Datentyps **Zeitstempel**.

Löschen einer Verbindung

1. Öffnen Sie die Management Console.
2. Öffnen Sie **Ressourcen > Datenquellen**.

3. Aktivieren Sie das Kästchen neben der Verbindung, die Sie löschen möchten, und klicken Sie dann auf die Schaltfläche „Löschen“ .

3 - Ausfüllen des Data Warehouses

In this section

Vorbereiten Ihrer Daten	74
Ausfüllen einer Zeitdimensionstabelle	75
Ausfüllen von Dimensionstabellen	76
Ausfüllen einer Faktentabelle	78
Hinzufügen eines Zeitstempels zu Datensätzen in einem Data Warehouse	83

Vorbereiten Ihrer Daten

Bevor Sie Ihr Data Warehouse füllen, sollten Sie sich vergewissern, dass Ihre Daten eine hohe Qualität aufweisen und alle erforderlichen Attribute vorhanden sind, um geschäftlichen Benutzern aussagekräftige Einsichten zu liefern. Ein üblicher Ansatz besteht darin, zu diesem Zweck einen Betriebsdatenspeicher (Operational Data Store, ODS) zu verwenden. Bei einem ODS handelt es sich um eine Datenbank, in der Sie Vorgänge zur Datenqualität ausführen können, die Ihre Daten zum Laden in das Data Warehouse vorbereiten. Spectrum™ Technology Platform verfügt über verschiedene Features, die in einen ODS implementiert werden können, um die Qualität Ihrer Daten zu verbessern und Ihre Daten auch um geografische, demografische oder andere Daten zu erweitern. Wenn Sie die folgenden Features derzeit nicht lizenziert haben, kontaktieren Sie Pitney Bowes, um weitere Informationen zu erhalten.

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.

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 Verbrauchssteuer
- persönliche Grundsteuer
- Versicherungsbeitragssteuer

Das Enterprise Tax-Modul bietet diese Funktionalität.

Ausfüllen einer Zeitdimensionstabelle

Eine Zeitdimensionstabelle ist eine Tabelle in einer Datenbank, die es ermöglicht, historische Daten ohne komplexe SQL-Berechnungen zu analysieren. Sie können beispielsweise Ihre Daten nach Arbeitstagen und Feiertagen, nach Werktagen und Wochenenden, nach Abrechnungszeiträumen oder nach besonderen Ereignissen analysieren.

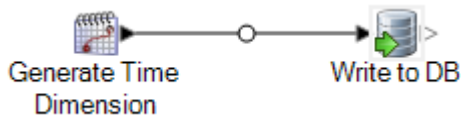
Die folgenden Schritte beschreiben, wie Sie Spectrum™ Technology Platform verwenden, um eine Zeitdimensionstabelle in Ihrem Data Warehouse zu füllen.

Anmerkung: Vorher müssen Sie Verbindungen zu dem Data Warehouse definieren, in dem Sie eine Zeitdimensionstabelle erstellen möchten. Wenn Sie die erforderlichen Verbindungen nicht definiert haben, lesen Sie weiter unter [Datenquellenverbindungen](#) auf Seite 14.

1. Klicken Sie im Enterprise Designer auf **Datei > Neu > Datenfluss > Auftrag**.
2. Ziehen Sie den „Generate Time Dimension“-Schritt auf die Arbeitsfläche.

3. Ziehen Sie einen „Write to DB“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem „Generate Time Dimension“-Schritt.

Der Datenfluss sollte nun folgendermaßen aussehen:



4. Doppelklicken Sie auf den „Generate Time Dimension“-Schritt und konfigurieren Sie ihn, die gewünschten Zeitdimensionen zu erstellen. Weitere Informationen finden Sie unter **Generate Time Dimension** auf Seite 120.

Anmerkung: Wenn die Granularität einen Tag oder mehr beträgt, wird normalerweise der julianische Tag als Schlüsselwert für eine Zeitdimensionstabelle verwendet. Wenn die Granularität geringer als ein Tag ist, können Sie einen separaten Schlüssel generieren, indem Sie einen „Unique ID Generator“-Schritt zum Datenfluss hinzufügen. Wenn Sie den julianischen Tag als Schlüssel verwenden, konfigurieren Sie den „Generate Time Dimension“-Schritt, eine Integer-Spalte für julianische Tageswerte und eine Spalte mit dem Datentyp „Datum“ oder „DateTime“ für Datumswerte zu erstellen.

5. Doppelklicken Sie auf den „Write to DB“-Schritt auf der Arbeitsfläche und konfigurieren Sie ihn, auf die Datenbank und Tabelle zu verweisen, in der Sie die Zeitdimensionstabelle erstellen möchten. Weitere Informationen zum Konfigurieren des „Write to DB“-Schrittes finden Sie unter **Write to DB** auf Seite 254.
6. So erhalten Sie eine Vorschau der Zeitdimensionswerte, bevor sie in die Zeitdimensionstabelle geschrieben werden:
 - a) Klicken Sie mit der rechten Maustaste auf den Kanal, der den „Generate Time Dimension“-Schritt mit dem „Write to DB“-Schritt verbindet, und wählen Sie **Überprüfungspunkt hinzufügen** aus.
 - b) Wählen Sie **Ausführen > Aktuellen Fluss überprüfen** aus.
Unten im Enterprise Designer-Fenster wird der Überprüfungsbereich mit den Daten angezeigt, die in die Zeitdimensionstabelle geschrieben werden. Falls erforderlich können Sie den „Generate Time Dimension“-Schritt anpassen und den Überprüfungsprozess neu starten, um die Auswirkungen Ihrer Änderungen zu begutachten.
7. Wenn Sie mit dem Datenfluss zufrieden sind, wählen Sie **Ausführen > Aktuellen Fluss ausführen** aus, um den Datenfluss auszuführen und die Zeitdimensionstabelle zu füllen.

Ausfüllen von Dimensionstabellen

Dimensionstabellen sind Teil eines „Star Schema“ und enthalten detaillierte Informationen über die Spalten in der Faktentabelle. Dimensionstabellen verfügen über Attribute und einen einteiligen Primärschlüssel, über den eine Dimensionstabelle mit der Faktentabelle verbunden ist. Über den

einteiligen Primärschlüssel können Sie eine einzelne Dimensionstabelle schnell durchsuchen. Das Durchsuchen einer Dimensionstabelle kann Ihnen helfen, die beste Möglichkeit zur Abfrage der Faktentabelle zu finden.

Die folgenden Schritte beschreiben, wie Sie Spectrum™ Technology Platform verwenden, um die Dimensionstabelle in Ihrem Data Warehouse erstmals zu füllen.

Anmerkung: Wenn Sie als Datenquelle eine Datenbank, einen Dateiserver oder einen Webservice verwenden, müssen Sie vor Beginn die Verbindungen zu externen Ressourcen definieren, die Sie als Quelle für die Dimensionstabelle verwenden möchten. Sie müssen außerdem eine Verbindung zu dem Data Warehouse definieren, in dem Sie eine Dimensionstabelle erstellen möchten. Wenn Sie die erforderlichen Verbindungen nicht definiert haben, lesen Sie weiter unter [Datenquellenverbindungen](#) auf Seite 14.

1. Erstellen Sie in Ihrem Data Warehouse die Tabelle, die Sie als Dimensionstabelle verwenden möchten.
2. Erstellen Sie in der Management Console Verbindungen zu Ihrer Datenquelle und zum Data Warehouse.
 - Informationen zum Herstellen einer Verbindung zu einer Datenbank finden Sie unter [Verbinden mit einer JDBC-Datenbank](#) auf Seite 40.
 - Informationen zum Herstellen einer Verbindung zu einem Dateiserver finden Sie unter [Verbinden mit einem FTP-Server](#) auf Seite 27.
3. Klicken Sie im Enterprise Designer auf **Datei > Neu > Datenfluss > Auftrag**.
4. Ziehen Sie den Quellschritt auf die Arbeitsfläche.
 - Um Daten aus einer Datenbank zum Füllen der Tabelle zu verwenden, ziehen Sie den **Read from DB**-Schritt auf die Arbeitsfläche.
 - Um Daten aus einem Flatfile zum Füllen der Tabelle zu verwenden, ziehen Sie den **Read from File**-Schritt auf die Arbeitsfläche.
 - Um Daten aus einer Datei mit variablem Format zum Füllen der Tabelle zu verwenden, ziehen Sie den **Read from Variable Format File**-Schritt auf die Arbeitsfläche.
 - Um Daten aus einer XML-Datei zum Füllen der Tabelle zu verwenden, ziehen Sie den **Read from XML**-Schritt auf die Arbeitsfläche.
5. Doppelklicken Sie auf den Quellschritt, den Sie gerade auf der Arbeitsfläche platziert haben, und konfigurieren Sie ihn, auf die Quelle der Daten zu verweisen, mit denen Sie die Dimensionstabelle füllen möchten.
 - Weitere Informationen zum Konfigurieren des „Read from DB“-Schrittes finden Sie unter [Read from DB](#) auf Seite 143.
 - Weitere Informationen zum Konfigurieren des „Read from File“-Schrittes finden Sie unter [Read from File](#) auf Seite 152.
 - Weitere Informationen zum Konfigurieren des „Read from Variable Format File“-Schrittes finden Sie unter [Read from Variable Format File](#) auf Seite 205.

- Weitere Informationen zum Konfigurieren des „Read from XML“-Schrittes finden Sie unter [Read from XML](#) auf Seite 220.
6. Ziehen Sie einen „Unique ID Generator“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem Quellschritt. Wenn Sie beispielsweise „Read from DB“ als Quellschritt verwenden, müssten Sie den „Read from DB“-Schritt mit dem „Unique ID Generator“-Schritt verbinden.
 7. Doppelklicken Sie auf den „Unique ID Generator“-Schritt auf der Arbeitsfläche und konfigurieren Sie ihn, einen Surrogatschlüssel zu erstellen.

Anmerkung: Normalerweise wird der Schlüssel des Betriebssystems nicht als Primärschlüssel einer Dimension im Warehouse verwendet. So kann die historische Konsistenz besser erhalten werden, da ein Schlüsselwert sich im Betriebssystem möglicherweise ändern kann.
 8. Ziehen Sie einen „Write to DB“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem „Unique ID Generator“-Schritt.
 9. Doppelklicken Sie auf den „Write to DB“-Schritt auf der Arbeitsfläche und konfigurieren Sie ihn, auf die Datenbank und die Dimensionstabelle zu verweisen, die Sie füllen möchten. Weitere Informationen zum Konfigurieren des „Write to DB“-Schrittes finden Sie unter [Write to DB](#) auf Seite 254.
 10. Wählen Sie **Datei > Speichern** aus, und speichern Sie den Datenfluss.
 11. Um den Datenfluss jetzt auszuführen und die Dimensionstabelle zu füllen, wählen Sie **Ausführen > Aktuellen Fluss ausführen** aus.

Ausfüllen einer Faktentabelle

Nachdem Sie die Dimensionstabellen in Ihrem Data Warehouse gefüllt haben, sind Sie nun bereit, die Faktentabelle zu füllen. Sie füllen eine Faktentabelle mit numerischen Messwerten aus Tabellen in der OLTP-Datenbank.

Wichtig: Sie müssen die Dimensionstabellen füllen, bevor Sie die Faktentabelle füllen.

Die folgenden Schritte beschreiben, wie Sie Spectrum™ Technology Platform verwenden, um eine Faktentabelle in Ihrem Data Warehouse zu füllen. Sie werden einen Datenfluss erstellen, der Quelldaten aus einer Tabelle in Ihrer Quelldatenbank einliest, die normalen Schlüssel in den Quelltabellen mit Surrogatschlüsseln aus den Dimensionstabellen ersetzt und dann die aktualisierten Datensätze, die die Surrogatschlüssel und die Faktendaten aus den Quelltabellen enthalten, in die Faktentabelle lädt.

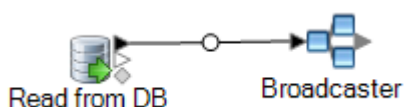
1. Erstellen Sie in der Management Console Verbindungen zu Ihrer Datenquelle und zum Data Warehouse.
 - Informationen zum Herstellen einer Verbindung zu einer Datenbank finden Sie unter [Verbinden mit einer JDBC-Datenbank](#) auf Seite 40.

- Informationen zum Herstellen einer Verbindung zu einem Dateiserver finden Sie unter [Verbinden mit einem FTP-Server](#) auf Seite 27.
2. Klicken Sie im Enterprise Designer auf **Datei > Neu > Datenfluss > Auftrag**.
 3. Ziehen Sie, je nach Quelle der Daten, die Sie in die Faktentabelle schreiben möchten, den entsprechenden Schritt auf die Arbeitsfläche.
 - Um Daten aus einer Datenbank zum Füllen der Tabelle zu verwenden, ziehen Sie den **Read from DB**-Schritt auf die Arbeitsfläche.
 - Um Daten aus einem Flatfile zum Füllen der Tabelle zu verwenden, ziehen Sie den **Read from File**-Schritt auf die Arbeitsfläche.
 - Um Daten aus einer Datei mit variablem Format zum Füllen der Tabelle zu verwenden, ziehen Sie den **Read from Variable Format File**-Schritt auf die Arbeitsfläche.
 - Um Daten aus einer XML-Datei zum Füllen der Tabelle zu verwenden, ziehen Sie den **Read from XML**-Schritt auf die Arbeitsfläche.
 4. Doppelklicken Sie auf den Quellschritt, den Sie gerade auf der Arbeitsfläche platziert haben, und konfigurieren Sie ihn, auf die Quelle der Daten zu verweisen, mit denen Sie die Faktentabelle füllen möchten.
 - Weitere Informationen zum Konfigurieren des „Read from DB“-Schrittes finden Sie unter [Read from DB](#) auf Seite 143.
 - Weitere Informationen zum Konfigurieren des „Read from File“-Schrittes finden Sie unter [Read from File](#) auf Seite 152.
 - Weitere Informationen zum Konfigurieren des „Read from Variable Format File“-Schrittes finden Sie unter [Read from Variable Format File](#) auf Seite 205.
 - Weitere Informationen zum Konfigurieren des „Read from XML“-Schrittes finden Sie unter [Read from XML](#) auf Seite 220.

Anmerkung: Normalerweise liest ein Datenfluss, der eine Faktentabelle füllt, Daten aus einer Datenbank und nicht aus einer Datei. Da dies das gebräuchlichste Szenario ist, wird in den Beispielen der restlichen Schritte „Read from DB“ verwendet.

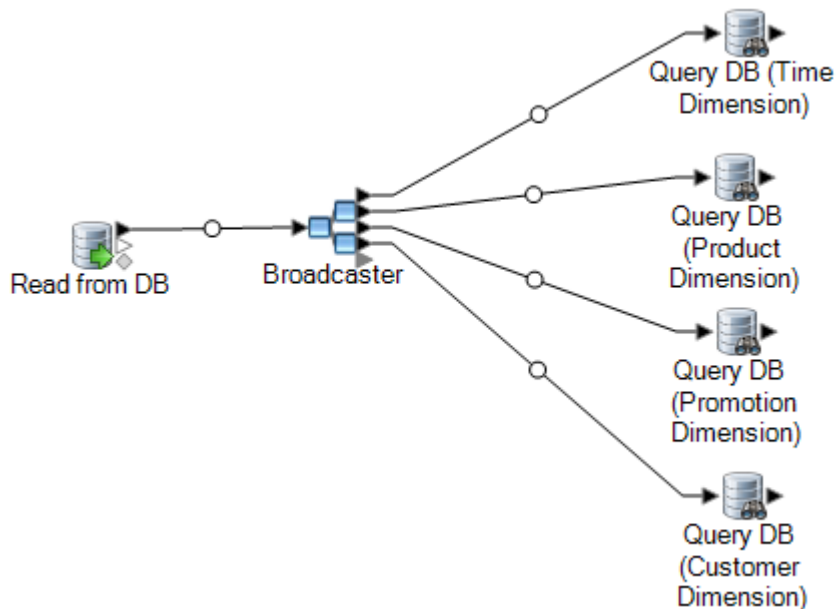
5. Ziehen Sie einen Broadcaster-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem Quellschritt.

Ihr Datenfluss sieht nun folgendermaßen aus:



6. Ziehen Sie für jede Dimensionstabelle in Ihrem Data Warehouse einen „Query DB“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem Broadcaster-Schritt.

Wenn Sie in Ihrem Data Warehouse beispielsweise über vier Dimensionstabellen verfügen, ziehen Sie vier „Query DB“-Schritte auf die Arbeitsfläche. Ihr Datenfluss sieht nun folgendermaßen aus:



Die „Query DB“-Schritte werden verwendet, um den Surrogatschlüssel für jede Dimension über den normalen Schlüssel aus der Datenquelle zu suchen. Der Surrogatschlüssel ersetzt dann den normalen Schlüssel in jedem Datensatz, der in die Faktentabelle geladen wird.

Tipp: Sie können den Namen des Schrittes ändern, um einfacher zu sehen, welche Tabelle von einem Schritt abgefragt wird.

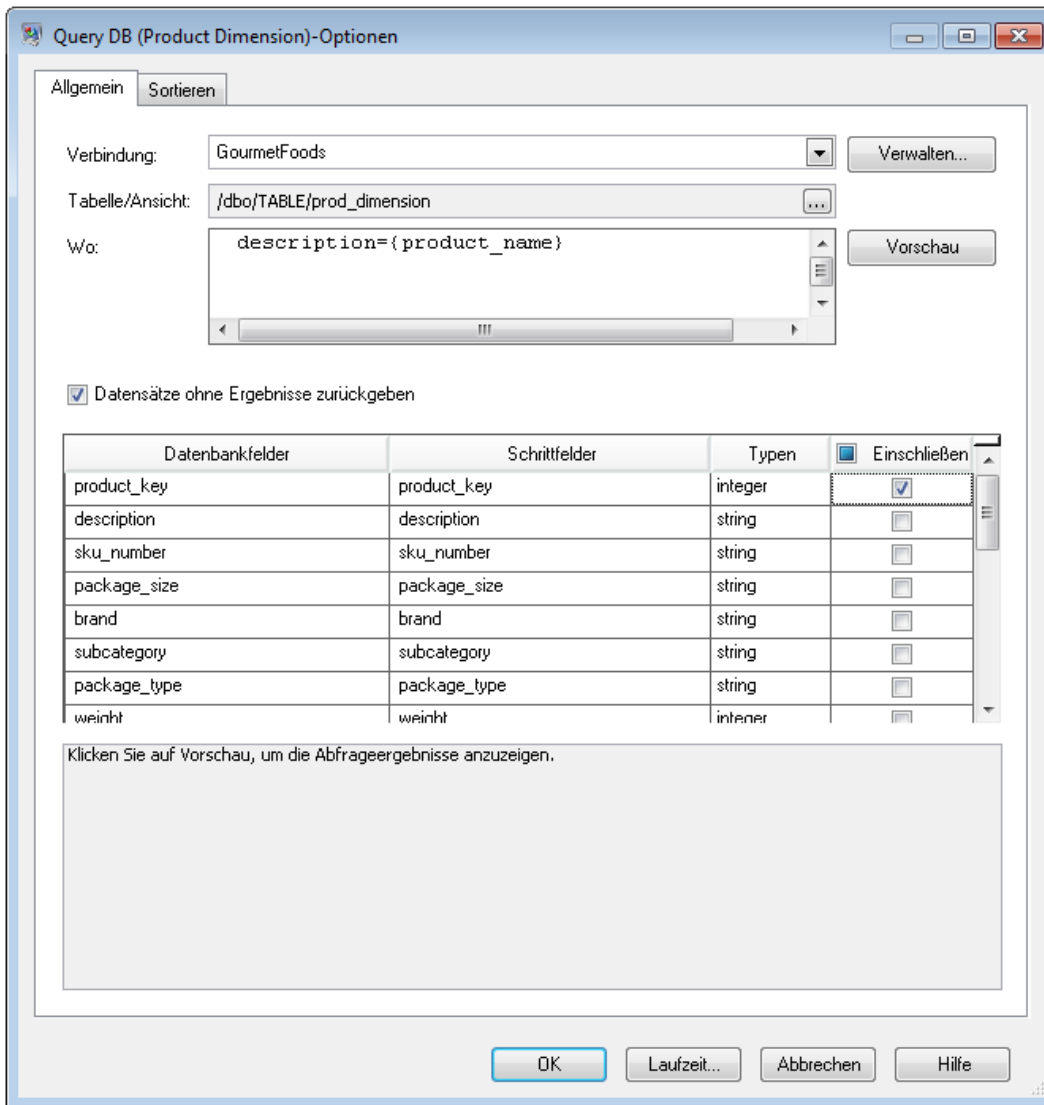
7. Konfigurieren Sie jeden „Query DB“-Schritt, den Surrogatschlüssel für jeden normalen Schlüssel in der Datenquelle zu suchen. Gehen Sie dazu wie folgt vor:
 - a) Geben Sie im Feld **Verbindung** die Verbindung zum Data Warehouse an.
 - b) Wählen Sie im Feld **Tabelle/Ansicht**: die Dimensionstabelle aus, die Sie über diesen Schritt abfragen möchten.
 - c) Geben Sie im Feld **Where** eine `WHERE`-Anweisung ein, die den Surrogatschlüssel basierend auf dem Wert im entsprechenden Datenflussfeld sucht.

Das folgende Beispiel sucht den Surrogatschlüssel für ein Produkt, indem nach dem Datensatz in der Dimensionstabelle gesucht wird, dessen Wert in der Spalte `description` mit dem Wert im Feld `product_name` der Datenquelle übereinstimmt.

```
description=${product_name}
```

- d) Wählen Sie in der Spalte **Einschließen** die Datenbankspalte aus, die den Surrogatschlüssel enthält.

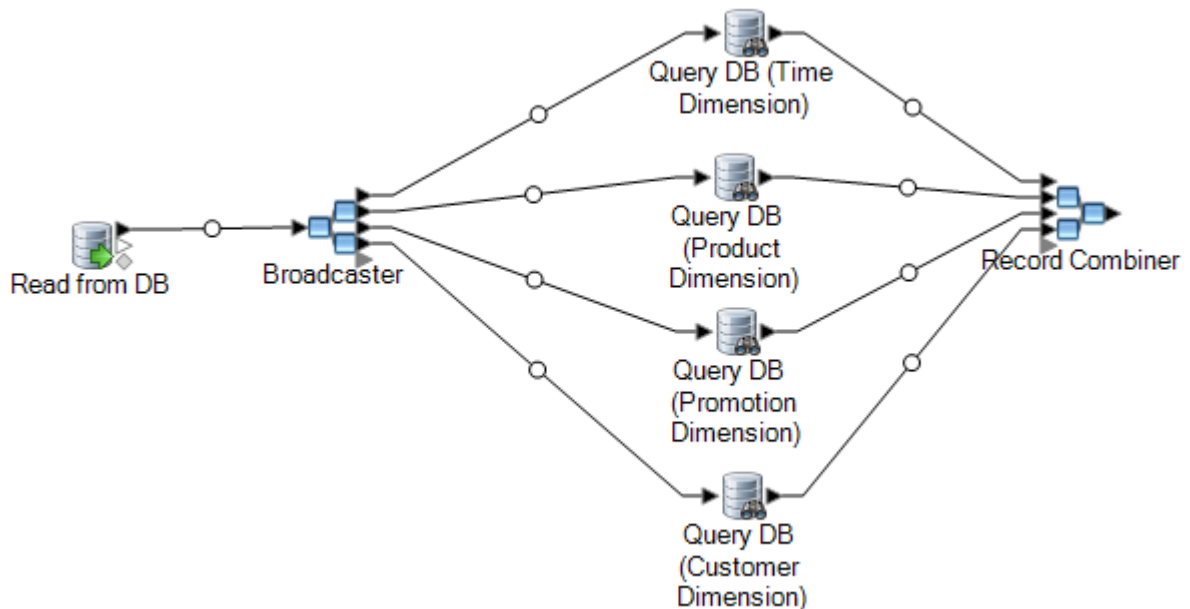
Ein „Query DB“-Schritt, der den Surrogatschlüssel für einen Produktnamen sucht, würde beispielsweise folgendermaßen aussehen:



In diesem Beispiel sucht die Abfrage den Produktschlüssel, indem sie nach dem Datensatz in der Tabelle `prod_dimension` sucht, bei dem der Wert in der Spalte `description` mit dem Wert im Datenflussfeld `product_name` übereinstimmt. Der Schritt gibt das Feld `product_key` aus der Tabelle zurück und fügt es dem Datenfluss hinzu, wie das aktivierte Kästchen in der Spalte **Einschließen** anzeigt.

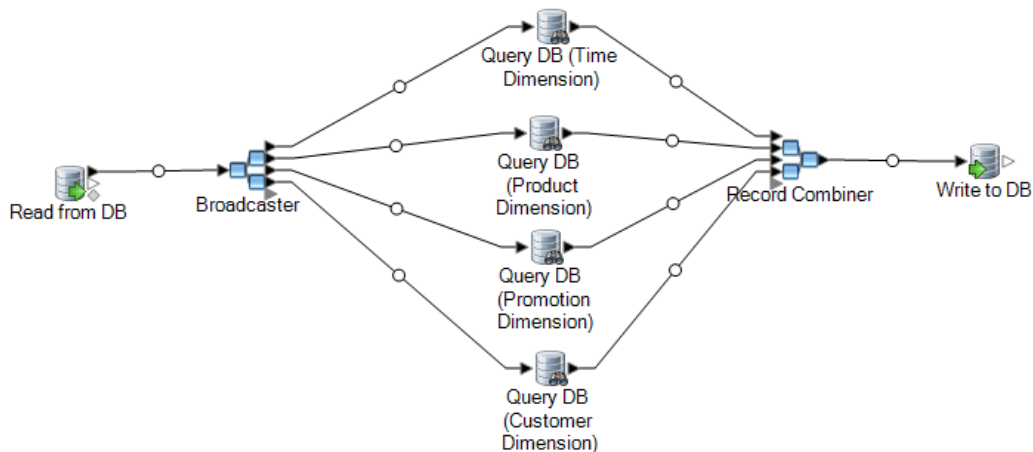
- Ziehen Sie einen „Record Combiner“-Schritt auf die Arbeitsfläche und verbinden Sie alle „Query DB“-Schritte mit diesem.

Ihr Datenfluss sollte nun folgendermaßen aussehen:



9. Ziehen Sie einen „Write to DB“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem „Record Combiner“-Schritt.

Ihr Datenfluss sollte nun folgendermaßen aussehen:



10. Konfigurieren Sie den „Write to DB“-Schritt, die Datensätze in die Faktentabelle zu schreiben. Gehen Sie dazu wie folgt vor:
 - a) Geben Sie im Feld **Verbindung** die Verbindung zum Data Warehouse an.
 - b) Wählen Sie im Feld **Tabelle/Ansicht**: die Faktentabelle aus, die Sie über diesen Schritt abfragen möchten. Wenn die Faktentabelle noch nicht im Data Warehouse vorhanden ist, klicken Sie auf **Tabelle erstellen**, um die Faktentabelle im Data Warehouse zu erstellen.
 - c) Aktivieren Sie für jedes Feld, das Sie in die Faktentabelle schreiben möchten, das Kästchen in der Spalte **Einschließen**.

- d) Beachten Sie, dass auf der Registerkarte **Laufzeit** standardmäßig **Einfügen** als Option für den Schreibmodus ausgewählt ist. Normalerweise wird die Faktentabelle im Einfügemodus gefüllt, Sie können diese Option daher ausgewählt lassen.

11. Speichern Sie Ihren Datenfluss und führen Sie ihn aus.

Beispiel zum Ersetzen von Quelldaten mit Schlüsseln aus der Dimensionstabelle

Betrachten Sie folgenden Datensatz:

```
March 28 2013,Parsley Garlic Pasta,Mile High Gourmet
Market,78.35
```

Dieses Beispiel besteht aus einem Datumsfeld, gefolgt von einem Produktnamen (Parsley Garlic Pasta), einem Kunden (Mile High Gourmet Market) und einer Menge (78.25). Das Data Warehouse verfügt über Dimensionstabellen für das Datum, den Produktnamen und den Kunden. Daher müssen die normalen Schlüssel in diesem Datensatz durch die Surrogatschlüssel aus den Dimensionstabellen ersetzt werden. Dazu muss der Datenfluss drei „Query DB“-Schritte enthalten, einer zur Suche des Surrogatschlüssels für das Datum, einer zur Suche des Surrogatschlüssels für den Produktnamen und einer zur Suche des Surrogatschlüssels für den Kunden.

Jeder „Query DB“-Schritt muss über eine `WHERE`-Anweisung verfügen, die den Surrogatschlüssel sucht.

Als Ergebnis dieser Suchen könnte der Datensatz nach dem Schreiben in die Faktentabelle folgendermaßen aussehen:

```
711,1,15,78.35
```

Beachten Sie, dass die normalen Schlüssel für Datum, Produktname und Kunde durch Surrogatschlüssel ersetzt wurden.

Hinzufügen eines Zeitstempels zu Datensätzen in einem Data Warehouse

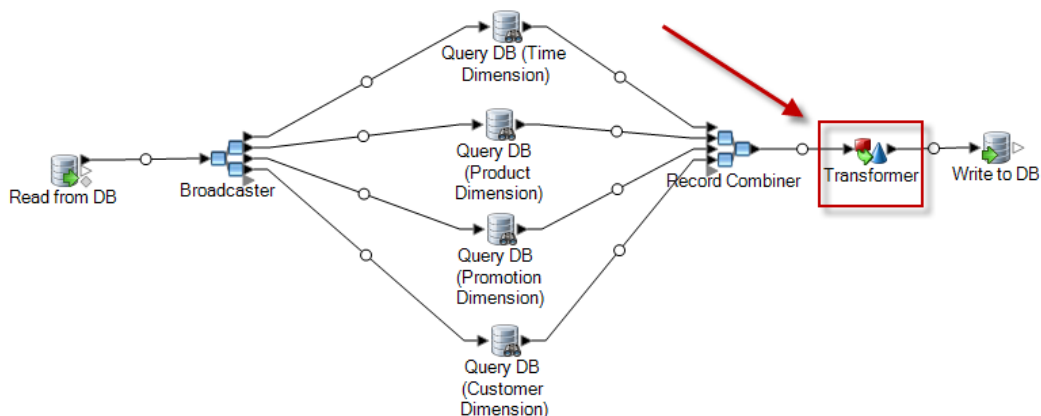
Eine bequeme Möglichkeit zur Sicherstellung der Datenqualität besteht darin, Datensätze im Data Warehouse mit dem Datum zu kennzeichnen, an dem sie geladen wurden. Wenn der Ladeprozess nicht abgeschlossen wird oder Sie Probleme bemerken, nachdem die Daten bereits geladen wurden, hilft eine Zeitstempelspalte, die betroffenen Datensätze zu identifizieren. Sie können dann alle Datensätze löschen, die während eines bestimmten Schrittes verarbeitet wurden, zum Status vor dem Ladevorgang zurückkehren und alle Probleme beheben, bevor Sie erneut versuchen, die Daten

zu laden. Sie können den Ladevorgang mit einem Zeitstempel versehen, indem Sie Ihrer Faktentabelle über den „SQL Command“-Schritt eine zusätzliche Spalte wie „Ladestatus“ hinzufügen.

So lassen Sie den Datenfluss beim Füllen oder Aktualisieren eines Data Warehouse einen Zeitstempel hinzufügen:

1. Öffnen Sie im Enterprise Designer den Datenfluss, der das Data Warehouse füllt oder aktualisiert:
2. Ziehen Sie einen Transformer-Schritt auf die Arbeitsfläche und verbinden Sie ihn kurz vor dem „Write to DB“-Schritt mit dem Datenfluss.

Beispiel:



3. Doppelklicken Sie auf den „Transformer“-Schritt.
4. Klicken Sie auf **Hinzufügen**.
5. Wählen Sie unter **Allgemein Benutzerdefiniert** aus.
6. Geben Sie in das Feld **Name der benutzerdefinierten Transformation** einen Namen für diese Transformation ein. Sie können einen beliebigen Namen auswählen, beispielsweise „Zeitstempel hinzufügen“.
7. Geben Sie im Feld **Benutzerdefiniertes Skript** Folgendes ein:

```
data['<timestamp field>']=currentDateTime()
```

Dabei ist <timestamp field> der Name des Datenflussfeldes, das den Zeitstempel enthalten soll.

Wenn Sie beispielsweise den Zeitstempel in einem Datenflussfeld namens `Timestamp` speichern möchten, sieht Ihr benutzerdefiniertes Skript folgendermaßen aus:

```
data['Timestamp']=currentDateTime()
```

8. Klicken Sie unten im Fenster auf die Schaltfläche **Hinzufügen**.
9. Klicken Sie auf **Schließen**.
10. Klicken Sie auf **OK**, um das Fenster **Transformer-Optionen** zu schließen.

Der Datenfluss fügt nun zu jedem Datensatz die aktuelle Zeit in einem Feld hinzu und erweitert das Data Warehouse damit um einen Zeitstempel, der anzeigt, wann ein Datensatz geladen wurde.

4 - Aktualisieren des Data Warehouses

In this section

Definieren eines Aktualisierungsplans für ein Data Warehouse	87
Aktualisieren einer Faktentabelle	88
Globalen Verwenden eines globalen Caches für Abfragen	93
Verwenden eines lokalen Caches für Abfragen	94

Definieren eines Aktualisierungsplans für ein Data Warehouse

Sie können Datenflüsse von Spectrum™ Technology Platform planen, die aus der normalisierten Struktur in der Datenquelle Daten extrahieren und in die „Star Schema“-Struktur im Data Warehouse transformieren. Das Planen von Datenflüssen ist nützlich, da die meisten Ladeoperationen Systemressourcen benötigen, die während eines Geschäftstags nicht verfügbar sind.

Berücksichtigen Sie bei einem Aktualisierungsplan Folgendes:

- Häufigkeit
- Sequenz
- Abhängigkeiten

Häufigkeit

Sie sollten die Ausführung von Datenflüssen basierend auf der Granularität der detailliertesten Faktentabelle planen. Beispiel:

- Wenn die Granularität der Faktentabelle ein Tag ist, planen Sie, den Datenfluss zum Füllen der Faktentabelle täglich auszuführen.
- Wenn die Granularität ein Monat ist, planen Sie eine monatliche Ausführung des Datenflusses zum Füllen. Planen Sie keine frühere Ausführung, da Benutzer nur mit Daten aus den abgeschlossenen vergangenen Monaten arbeiten.

Die meisten Datenflüsse zum Füllen verarbeiten große Datenmengen, planen Sie daher die Ausführung der Datenflüsse zum Füllen zu Zeiten, wenn der Spectrum™ Technology Platform-Server, die Quelldatenbanken und die Data Warehouse-Datenbanken nur minimal verwendet werden.

Füllen Sie alle Dimensions- und Faktentabellen beim ersten Laden. Aktualisieren Sie die Tabellen nach dem ersten Laden basierend auf hinzugefügten oder geänderten Datensätzen. Im Allgemeinen werden Faktentabellen öfter aktualisiert als Dimensionstabellen. Gründe dafür sind die Folgenden:

- Dimensionstabellen sind normalerweise statisch, es sei denn, ein Attribut wird in der Quelle geändert oder hinzugefügt.
- Faktentabellendaten in einer Datenbank zur Unterstützung der Entscheidungsfindung sind typischerweise historisch und erfordern regelmäßige Aktualisierungen, um aktuell zu bleiben. Das erste Laden und die meisten inkrementellen Ladevorgänge betreffen Faktentabellen.

Sequenz

Es bestehen Abhängigkeiten zwischen Daten in Data Warehouse-Datenbanken. Bestimmen Sie daher die Reihenfolge, in der die Datenflüsse zum Füllen laufen müssen, bevor Sie den Ausführungszeitplan erstellen.

Füllen Sie Dimensionstabellen vor Faktentabellen, da jeder Datensatz und Schlüssel einer Dimensionstabelle vorhanden sein muss, bevor eine diesbezügliche Faktentabelle gefüllt werden kann. Diese Einschränkung beruht auf der Beziehung Primärschlüssel/Fremdschlüssel zwischen Dimensions- und Faktentabellen in einem „Star Schema“.

Aktualisieren Sie Basistabellen, bevor Sie Aggregattabellen in Ihrer Datenbank zur Unterstützung der Entscheidungsfindung füllen. Diese Reihenfolge gewährleistet, dass Basistabellen und Aggregattabellen synchron bleiben.

Die korrekte Reihenfolge zum Ausführen von Datenflüssen zum Füllen ist wie folgt:

1. Basisdimensionstabellen
2. Basisfaktentabellen
3. Aggregatdimensionstabellen
4. Aggregatfaktentabellen

Abhängigkeiten

Sie können Datenflussabhängigkeiten erstellen, wenn mehrere Datenflüsse zum Füllen in einer bestimmten Reihenfolge ausgeführt werden müssen oder wenn die Dauer der Ausführung der Datenflüsse nicht vorhersagbar ist. Ein Datenfluss wird nur ausgeführt, wenn bestimmte Voraussetzungen erfüllt werden. Dazu können der Abschluss oder das Fehlschlagen der vorherigen Datenflussausführung gehören.

Um Datenflussabhängigkeiten zu erstellen, erstellen Sie einen Prozessfluss im Enterprise Designer. Weitere Informationen zu Prozessflüssen finden Sie im *Datenfluss-Designer-Handbuch*.

Aktualisieren einer Faktentabelle

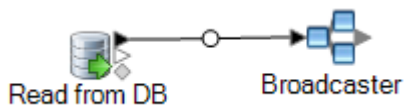
Diese Schritte beschreiben, wie Sie einen Datenfluss erstellen können, der Daten aus einer Quelldatenbank oder einer Datei liest und eine Faktentabelle in Ihrem Data Warehouse mit diesen Daten aktualisiert.

1. Klicken Sie im Enterprise Designer auf **Datei > Neu > Datenfluss > Auftrag**.
2. Ziehen Sie, je nach Quelle der Daten, die Sie in die Faktentabelle schreiben möchten, den entsprechenden Schritt auf die Arbeitsfläche.
 - Um Daten aus einer Datenbank zum Füllen der Tabelle zu verwenden, ziehen Sie den **Read from DB**-Schritt auf die Arbeitsfläche.
 - Um Daten aus einem Flatfile zum Füllen der Tabelle zu verwenden, ziehen Sie den **Read from File**-Schritt auf die Arbeitsfläche.
 - Um Daten aus einer Datei mit variablem Format zum Füllen der Tabelle zu verwenden, ziehen Sie den **Read from Variable Format File**-Schritt auf die Arbeitsfläche.
 - Um Daten aus einer XML-Datei zum Füllen der Tabelle zu verwenden, ziehen Sie den **Read from XML**-Schritt auf die Arbeitsfläche.

Anmerkung: Wenn Sie Daten aus einer Datei und nicht aus einer Datenbank lesen, vergewissern Sie sich, dass die Datei nur die Datensätze enthält, die Sie zur Faktentabelle hinzufügen möchten, und keine Datensätze enthält, die bereits in der Faktentabelle vorhanden sind. Wenn Sie Daten aus einer Datenbank lesen, definieren Sie später in dieser Prozedur eine Abfrage zur Filterung der Datensätze.

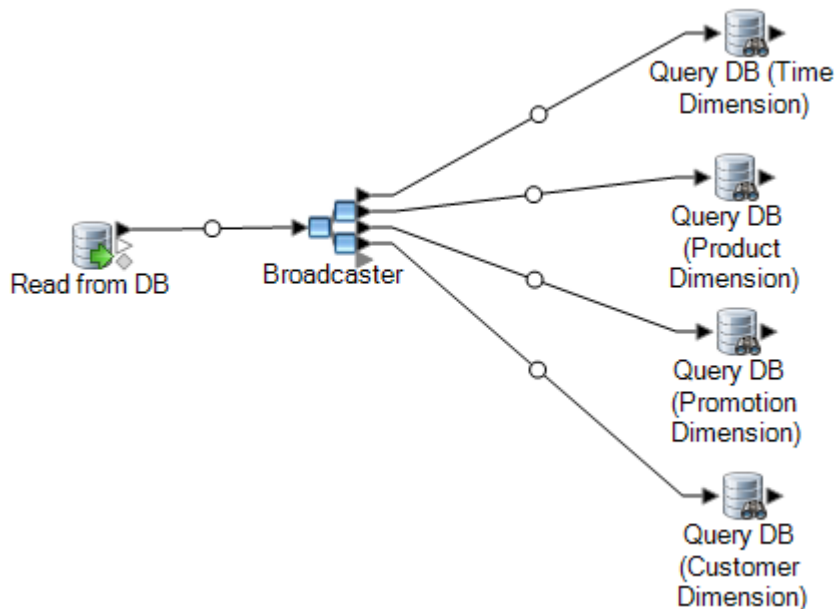
3. Doppelklicken Sie auf den Quellschritt, den Sie gerade auf der Arbeitsfläche platziert haben, und konfigurieren Sie ihn, auf die Quelle der Daten zu verweisen, mit denen Sie die Faktentabelle füllen möchten.
 - Weitere Informationen zum Konfigurieren des „Read from DB“-Schrittes finden Sie unter [Read from DB](#) auf Seite 143.
 - Weitere Informationen zum Konfigurieren des „Read from File“-Schrittes finden Sie unter [Read from File](#) auf Seite 152.
 - Weitere Informationen zum Konfigurieren des „Read from Variable Format File“-Schrittes finden Sie unter [Read from Variable Format File](#) auf Seite 205.
 - Weitere Informationen zum Konfigurieren des „Read from XML“-Schrittes finden Sie unter [Read from XML](#) auf Seite 220.
4. Wenn Sie Daten aus einer Datenbank lesen, filtern Sie die Datensätze, sodass nur neue Datensätze zur Faktentabelle hinzugefügt werden. Sie erreichen dies, indem Sie die SQL SELECT-Anweisung definieren, nur Datensätze einzulesen, die seit der letzten Aktualisierung der Faktentabelle geändert wurden.
5. Ziehen Sie einen Broadcaster-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem Quellschritt.

Ihr Datenfluss sieht nun folgendermaßen aus:



6. Ziehen Sie für jede Dimensionstabelle in Ihrem Data Warehouse einen „Query DB“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem Broadcaster-Schritt.

Wenn Sie in Ihrem Data Warehouse beispielsweise über vier Dimensionstabellen verfügen, ziehen Sie vier „Query DB“-Schritte auf die Arbeitsfläche. Ihr Datenfluss sieht nun folgendermaßen aus:



Die „Query DB“-Schritte werden verwendet, um den Surrogatschlüssel für jede Dimension über den normalen Schlüssel aus der Datenquelle zu suchen. Der Surrogatschlüssel ersetzt dann den normalen Schlüssel in jedem Datensatz, der in die Faktentabelle geladen wird.

Tipp: Sie können den Namen des Schrittes ändern, um einfacher zu sehen, welche Tabelle von einem Schritt abgefragt wird.

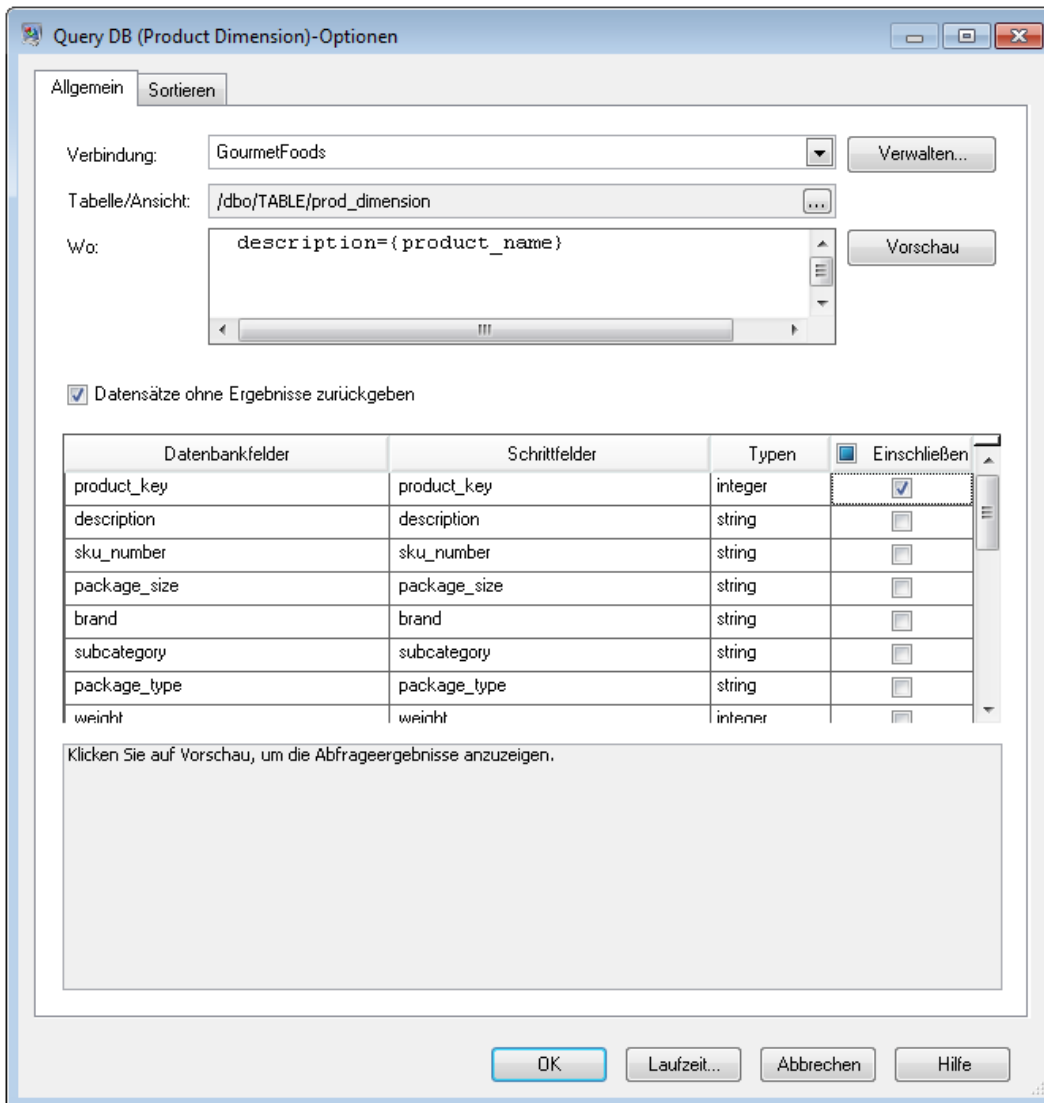
7. Konfigurieren Sie jeden „Query DB“-Schritt, den Surrogatschlüssel für jeden normalen Schlüssel in der Datenquelle zu suchen. Gehen Sie dazu wie folgt vor:
 - a) Geben Sie im Feld **Verbindung** die Verbindung zum Data Warehouse an.
 - b) Wählen Sie im Feld **Tabelle/Ansicht** die Dimensionstabelle aus, die Sie über diesen Schritt abfragen möchten.
 - c) Geben Sie im Feld **Where** eine `WHERE`-Anweisung ein, die den Surrogatschlüssel basierend auf dem Wert im entsprechenden Datenflussfeld sucht.

Das folgende Beispiel sucht den Surrogatschlüssel für ein Produkt, indem nach dem Datensatz in der Dimensionstabelle gesucht wird, dessen Wert in der Spalte `description` mit dem Wert im Feld `product_name` der Datenquelle übereinstimmt.

```
description=${product_name}
```

- d) Wählen Sie in der Spalte **Einschließen** die Datenbankspalte aus, die den Surrogatschlüssel enthält.

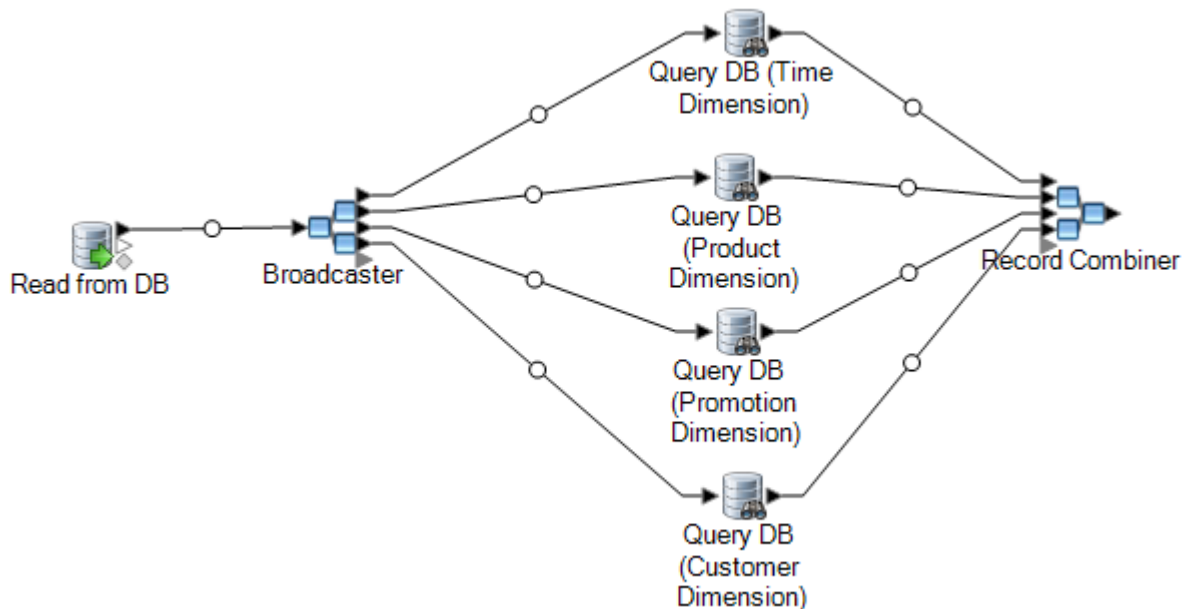
Ein „Query DB“-Schritt, der den Surrogatschlüssel für einen Produktnamen sucht, würde beispielsweise folgendermaßen aussehen:



In diesem Beispiel sucht die Abfrage den Produktschlüssel, indem sie nach dem Datensatz in der Tabelle `prod_dimension` sucht, bei dem der Wert in der Spalte `description` mit dem Wert im Datenflussfeld `product_name` übereinstimmt. Der Schritt gibt das Feld `product_key` aus der Tabelle zurück und fügt es dem Datenfluss hinzu, wie das aktivierte Kästchen in der Spalte **Einschließen** anzeigt.

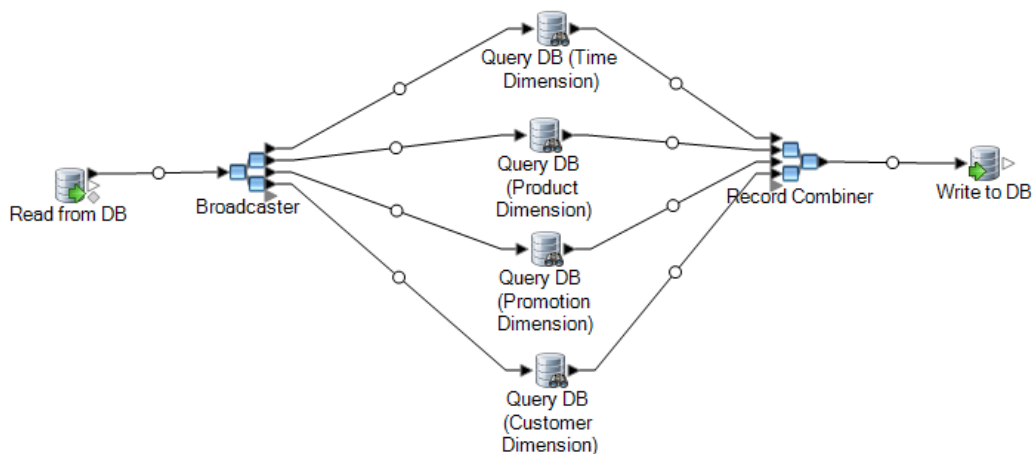
- Ziehen Sie einen „Record Combiner“-Schritt auf die Arbeitsfläche und verbinden Sie alle „Query DB“-Schritte mit diesem.

Ihr Datenfluss sollte nun folgendermaßen aussehen:



9. Ziehen Sie einen „Write to DB“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem „Record Combiner“-Schritt.

Ihr Datenfluss sollte nun folgendermaßen aussehen:



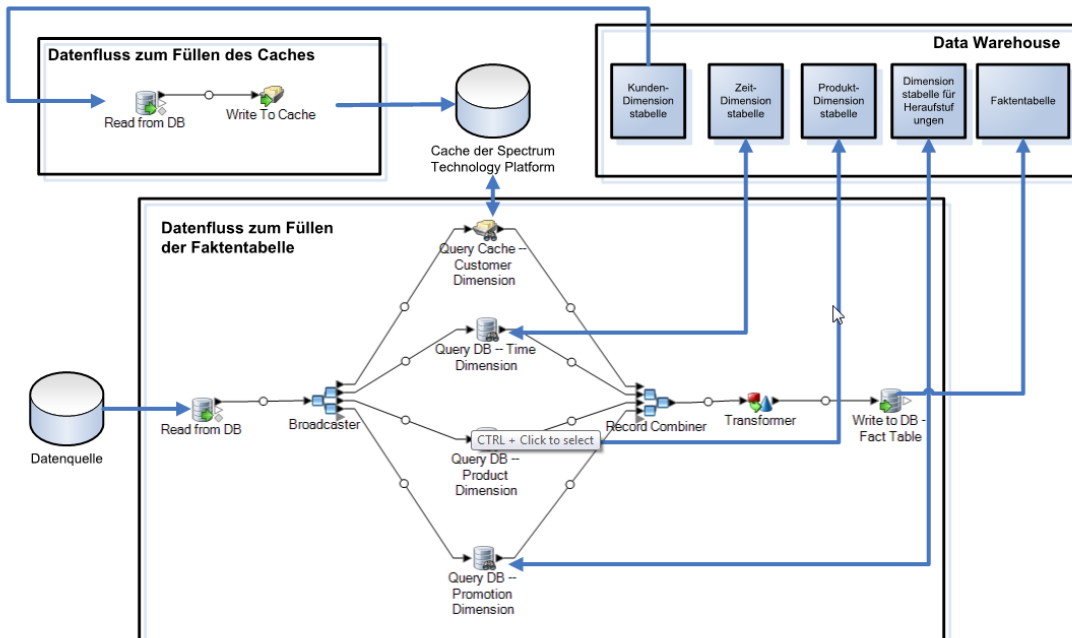
10. Konfigurieren Sie den „Write to DB“-Schritt, die Datensätze in die Faktentabelle zu schreiben. Gehen Sie dazu wie folgt vor:
- Geben Sie im Feld **Verbindung** die Verbindung zum Data Warehouse an.
 - Wählen Sie im Feld **Tabelle/Ansicht**: die Faktentabelle aus, die Sie über diesen Schritt abfragen möchten. Wenn die Faktentabelle noch nicht im Data Warehouse vorhanden ist, klicken Sie auf **Tabelle erstellen**, um die Faktentabelle im Data Warehouse zu erstellen.
 - Aktivieren Sie für jedes Feld, das Sie in die Faktentabelle schreiben möchten, das Kästchen in der Spalte **Einschließen**.

- d) Beachten Sie, dass auf der Registerkarte **Laufzeit** standardmäßig **Einfügen** als Option für den Schreibmodus ausgewählt ist. Normalerweise wird die Faktentabelle im Einfügemodus gefüllt, Sie können diese Option daher ausgewählt lassen.

Globalen Verwenden eines globalen Caches für Abfragen

Wenn Sie über eine große Dimensionstabelle verfügen, können Sie die Daten der Dimensionstabelle in einen Cache laden und den Cache verwenden, um nach Surrogatschlüsseln zu suchen. Die Verwendung eines Cache verbessert die Leistung im Vergleich mit Suchvorgängen, die über Query DB direkt in der Dimensionstabelle durchgeführt werden.

Um einen Cache zu verwenden, müssen Sie zwei Datenflüsse erstellen: einen zum Füllen des Cache mit Dimensionstabellendaten und einen weiteren, der bei der Aktualisierung der Faktentabelle den Cache verwendet. Das folgende Diagramm zeigt, wie die beiden Datenflüsse zusammenarbeiten:



1. Erstellen Sie einen Datenfluss, der den Cache mit Dimensionstabellendaten aus der großen Dimensionstabelle füllt.

Dieser Datenfluss sollte aus zwei Schritten bestehen:

- Ein „Read from DB“-Schritt liest die Daten aus der Dimensionstabelle, die Sie in den Cache laden möchten.
- Ein „Write To Cache“-Schritt füllt den Cache mit den Dimensionstabellendaten.

2. Führen Sie diesen Datenfluss aus, um den Cache zu füllen.
3. Fügen Sie im Datenfluss, der die Faktentabelle füllt, einen „Query Cache“-Schritt hinzu.
4. Konfigurieren Sie den „Query Cache“-Schritt so, dass er den vom „Write To Cache“-Schritt erstellten Cache abfragt.
5. Führen Sie diesen Datenfluss aus, um die Faktentabelle zu füllen.


Wenn Sie sicherstellen möchten, dass der Cache jedes Mal, wenn Sie Ihre Faktentabelle aktualisieren, mit den aktuellsten Daten aus der Dimensionstabelle gefüllt wird, können Sie einen Prozessfluss erstellen, der zuerst den Auftrag zum Füllen der Dimensionstabelle und dann den Auftrag zur Aktualisierung der Faktentabelle ausführt. Sie können dann den Prozessfluss ausführen, um beide Datenflüsse nacheinander ablaufen zu lassen. Weitere Informationen zu Prozessflüssen finden Sie im *Datenfluss-Designer-Handbuch*.

Löschen eines Caches

Ein globaler Cache wird vom „Query Cache“-Schritt zur Datensuche verwendet. Wenn der Cache nicht mehr benötigt wird, können Sie ihn mithilfe folgender Schritte löschen. Das Löschen eines Cache gibt Speicher frei.

Anmerkung: Cache löschen wird nur verwendet, um einen globalen Cache über die Management Console zu löschen.

Warnung: Vergewissern Sie sich vor dem Löschen eines Cache, dass er von keinem Datenfluss mit einem „Query Cache“-Schritt verwendet wird. Wenn Sie einen Cache löschen, der von einem Datenfluss verwendet wird, schlägt dieser Datenfluss fehl.

1. Öffnen Sie die Management Console.
2. Erweitern Sie **Ressourcen > Cacheverwaltung**.
Das Grid zeigt die vorhandenen globalen Caches an.
3. Wählen Sie den zu löschenden Cache aus.
4. Klicken Sie auf .

Verwenden eines lokalen Caches für Abfragen

Wenn Sie über eine große Dimensionstabelle verfügen, können Sie die Daten der Dimensionstabelle in einen Cache laden und den Cache verwenden, um nach Surrogatschlüsseln zu suchen. Die Verwendung eines Cache verbessert die Leistung im Vergleich mit Suchvorgängen, die über Query DB direkt in der Dimensionstabelle durchgeführt werden.

Ein lokaler Cache ist ein temporärer Cache, der nur während der Ausführung des „Query Cache“-Schrittes verwendet wird. Der „Query Cache“-Schritt sucht nach Daten im Cache basierend

auf Schlüsselfeldern und Suchbedingungen und gibt Daten von übereinstimmenden Datensätzen im Cache zurück. Die Daten des Datensatzes im Cache werden dann dem Datensatz im Datenfluss hinzugefügt. Verwenden Sie einen lokalen Cache anstelle eines globalen Cache, wenn dieser nur in einem Datenfluss verwendet wird oder sich die Suchtabelle oft ändert.

So verwenden Sie einen lokalen Cache für Abfragen:

1. Öffnen Sie im Enterprise Designer den Datenfluss, bei dem Sie eine Abfrage über einen Cache ausführen möchten.
2. Ziehen Sie einen „Query Cache“-Schritt auf die Arbeitsfläche und verbinden Sie ihn mit dem Datenfluss.
3. Doppelklicken Sie auf den „Query Cache“-Schritt.
4. Wählen Sie **Lokaler Cache** aus.
5. 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 Datenbankverbindung 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.

6. Geben Sie die Tabelle oder Ansicht in der Datenbank an, die Sie abfragen möchten.
7. Wählen Sie unter der Option **Schlüsselfeld** einen Schlüssel aus.
8. Wählen Sie unter der Option **Eingabefeld** ein Eingabefeld aus. Das **Eingabefeld** enthält die Felder, die aus dem vorherigen Schritt stammen. Wenn der Wert in diesem Feld mit dem Schlüssel im **Schlüsselfeld** in der Datenbank übereinstimmt, gibt die Abfrage die Daten aus diesem Datensatz in der Datenbank zurück.
9. Klicken Sie auf **OK**.
10. Führen Sie den Datenfluss aus.

5 - Schrittreferenz

In this section

Call Stored Procedure	97
DB Change Data Reader	99
DB Loader	103
Field Parser	113
Field Combiner	117
Field Selector	119
Generate Time Dimension	120
Query Cache	127
Query DB	136
Query NoSQL DB	140
Read from DB	143
Read from File	152
Read from Hadoop Sequence File	171
Read from Hive File	175
Read from HL7 File	178
Read from NoSQL DB	192
Read from SAP	197
Read from Spreadsheet	202
Read from Variable Format File	205
Read from XML	220
SQL Command	229
Transposer	239
Unique ID Generator	243
Write to Cache	252
Write to DB	254
Write to File	261
Write to Hadoop Sequence File	280
Write to Hive File	284
Write to NoSQL DB	291
Write to Spreadsheet	295
Write to Variable Format File	298
Write to XML	308
Datums- und Zahlenmuster	318

Call Stored Procedure

„Call Stored Procedure“ ist ein Quellschritt, der eine in einer Datenbank gespeicherte Prozedur ausführt und die Ergebnisse des Aufrufs der gespeicherten Prozedur als Eingabe in den Datenfluss zurückgibt. Verwenden Sie „Call Stored Procedure“, wenn Sie Daten aus einer Datenbank über eine gespeicherte Prozedur der Datenbank abrufen möchten, anstatt eine Abfrage einer Tabelle oder Ansicht zu verwenden.

Anmerkung: Wenn Sie Daten direkt aus einer Tabelle oder Ansicht in einen Datenfluss einlesen möchten, verwenden Sie den „Read from DB“-Schritt.

Wenn Sie Geschäftslogik in einer gespeicherten Prozedur eingebettet haben und diese Logik in Ihrer Spectrum™ Technology Platform-Umgebung verwenden möchten, können Sie „Call Stored Procedure“ verwenden. Viele Betriebssysteme verwenden zum Beispiel keine referentiellen Integritätsüberprüfungen in der Datenbank für große, fortlaufend aktualisierte Tabellen, da die Leistung dadurch signifikant beeinträchtigt werden würde. Um also die referentielle Integrität aufrechtzuerhalten, können Sie gespeicherte Prozeduren erstellen und diese für alle Systemaktualisierungen verwenden.

Sie können gespeicherte Prozeduren auch verwenden, um die Verwaltung der Spectrum™ Technology Platform-Umgebung zu vereinfachen. Wenn Sie beispielsweise Hunderte von ETL-Prozessen eingerichtet haben, die alle dieselben Daten lesen, möchten Sie die Abfrage vielleicht in einer gespeicherten Prozedur gestalten, um sie an einer Stelle verwalten zu können. Die Wartung wird damit einfacher, da Sie nur die eine gespeicherten Prozedur anstatt Hunderte von unterschiedlichen Prozessen ändern müssen.

Optionsname	Beschreibung
Verbindung	<p>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 Datenbankverbindung ändern oder löschen müssen, klicken Sie auf Verwalten.</p> <p>Wenn Sie eine Datenbankverbindung hinzufügen oder ändern, füllen Sie diese Felder aus:</p> <p>Verbindungsname Geben Sie einen Namen für die Verbindung ein. Sie können den Namen frei wählen.</p> <p>Datenbanktreiber Wählen Sie den passenden Datenbanktyp aus.</p> <p>Verbindungsoptionen Geben Sie Host, Port, Instanz, Benutzername und Kennwort an, die für die Verbindung zur Datenbank verwendet werden sollen.</p>

Optionsname	Beschreibung
Schema	Gibt das Schema an, das die gespeicherte Prozedur enthält, die Sie aufrufen möchten.
Prozedur	Gibt die gespeicherte Prozedur an, die Sie aufrufen möchten.
Parameter für gespeicherte Prozedur	<p>In dieser Tabelle werden die Werte für die Parameter der gespeicherten Prozedur angegeben.</p> <p>Parameter In dieser Spalte befinden sich die Parameter, die in der gespeicherten Prozedur definiert werden.</p> <p>Schrittfelder In dieser Spalte wird für die Parameter OUT, INOUT und RETURN der Name des Datenflussfeldes angezeigt, das die Daten enthält, die vom Parameter zurückgegeben werden. Zu Anfang entspricht der Feldname dem Parameternamen. Sie können den Schrittfeldnamen ändern, indem Sie auf den Feldnamen klicken und einen neuen Namen für die Parameter eingeben. Die Spalte wird nicht für IN-Parameter verwendet.</p> <p>Richtung Zur Auswahl stehen:</p> <p>IN Der Parameter ist ein Eingabeparameter. Der Wert, den Sie für diesen Parameter angeben, wird als Eingabe an die gespeicherte Prozedur übergeben.</p> <p>OUT Der Parameter ist ein Ausgabeparameter. Die gespeicherte Prozedur übergibt mit diesem Parameter Daten an den Schritt.</p> <p>INOUT Der Parameter kann sowohl als Eingabeparameter zur Übergabe eines Wertes an die gespeicherte Prozedur als auch als Ausgabeparameter zum Empfang von Daten, die von der gespeicherten Prozedur zurückgegeben werden, verwendet werden.</p> <p>RETURN Der Parameter enthält einen Rückgabecode von der gespeicherten Prozedur.</p> <p>Typen In dieser Spalte wird der Datentyp des Parameterwertes angezeigt. Wenn der Datentyp nicht von Spectrum™ Technology Platform unterstützt wird, wird als Typ „Nicht unterstützt“ angezeigt und die gespeicherte Prozedur kann nicht erfolgreich ausgeführt werden.</p> <p>Wert Geben Sie in dieser Spalte den Wert ein, den Sie für den Parameter festlegen möchten. Die Spalte ist für OUT-Parameter deaktiviert.</p>

Optionsname	Beschreibung
Resultset-Felder	<p>In dieser Tabelle wird angegeben, welche Datenflussfelder für die Daten verwendet werden sollen, die von der gespeicherten Prozedur zurückgegeben werden.</p> <p>Datenbanktabellen In dieser Spalte werden die Tabellen angezeigt, aus denen die gespeicherte Prozedur Daten zurückgibt.</p> <p>Datenbankfelder In dieser Spalte wird das Feld angezeigt, aus dem die gespeicherte Prozedur Daten zurückgibt.</p> <p>Schrittfelder In dieser Spalte wird der Name des Datenflussfeldes angezeigt, das die Daten aus dem Datenbankfeld enthält.</p> <p>Typen In dieser Spalte wird der Datentyp des Feldes angezeigt. Wenn der Datentyp nicht von Spectrum™ Technology Platform unterstützt wird, wird als Typ „Nicht unterstützt“ angezeigt.</p> <p>Einschließen Aktivieren Sie das Kästchen in dieser Spalte, um das Feld in den Datenfluss einzuschließen. Wenn das Kästchen nicht aktiviert ist, wird das Feld nicht im Datenfluss verwendet.</p>
Felder abrufen	Klicken Sie auf diese Schaltfläche, um die Tabelle „Resultset-Felder“ mit dem Resultset-Schema auszufüllen, das von der gespeicherten Prozedur zurückgegeben wird. Hierdurch wird die gespeicherte Prozedur ausgeführt, um das Resultset-Schema abzurufen.
Hinzufügen	Klicken Sie auf diese Schaltfläche, um manuell ein Resultset-Feld hinzuzufügen.
Entfernen	Klicken Sie auf diese Schaltfläche, um ein Resultset-Feld aus der Liste der verfügbaren Felder zu entfernen.

DB Change Data Reader

Mithilfe des Schrittes **DB Change Data Reader** können Sie die Spalten auswählen, die in den aktuellen Auftragsfluss aufgenommen werden sollen, wenn für die Spalten das Feature „Change Data Capture“ aktiviert ist.

In diesem Schritt können Sie eine „Change Data Capture“-Ressource (CDC-Ressource) erstellen, bei der es sich um die erforderliche Datenquellentabelle handelt. Die Spalten der CDC-Ressource, auf denen das Feature „Change Data Capture“ aktiviert ist, werden mithilfe von Kästchen dargestellt.

Change Data Capture

Mithilfe des Feature „Change Data Capture“ können alle an einer Spalte vorgenommenen Änderungen erfasst werden. Für jede aktivierte Spalte werden alle Ergänzungen, Aktualisierungen und Löschungen erfasst.

Unterstützte Datenbanken

Derzeit unterstützt Spectrum™ Technology Platform das Feature „Change Data Capture“ (CDC) lediglich für MS SQL- und Oracle-Datenbanken.

MS SQL Bei MS SQL-Datenquellen kann das Feature „Change Data Capture“ für die Spalten von Tabellen aus dem Back-End aktiviert oder deaktiviert werden. Informationen zu den erforderlichen Schritten finden Sie [hier](#).

Anmerkung: Das Feature „Change Data Capture“ von Spectrum™ wird nicht auf der Express-Edition des SQL-Servers unterstützt.

Oracle Bei Oracle-Datenquellen erfasst Spectrum™ die Änderungen an den Daten in den Tabellenspalten und verwendet dafür das LogMiner-Dienstprogramm von Oracle. CDC kann nicht über Spectrum™ Technology Platform auf Tabellenspalten in einer Oracle-Datenquelle aktiviert oder deaktiviert werden.

Die durch die Abfragen `insert`, `update` und `delete` entstehenden Änderungen an den Daten werden ab einem eingegebenen Startdatum und einer angegebenen Uhrzeit bis zum aktuellen Datum und zur aktuellen Uhrzeit nachverfolgt und erfasst. Diese eingegebene Startzeit gilt während der ersten Abfrageausführung mit aktiviertem CDC-Feature.

In darauffolgenden Ausführungen auf derselben Oracle-Verbindung werden Datenänderungen ab dem Zeitpunkt der letzten Ausführung bis zur aktuellen Uhrzeit erfasst.

Anmerkung: Es handelt sich um einen inkrementellen Vorgang. Bei der ersten Erfassung werden die Änderungen ab dem eingegebenen Startdatum und der eingegebenen Uhrzeit bis zum aktuellen Datum und zur aktuellen Uhrzeit erfasst. Bei den darauffolgenden Erfassungen werden die Änderungen nach Ablauf des Enddatums und der Endzeit der vorangegangenen Erfassung bis zum aktuellen Datum und zur aktuellen Uhrzeit erfasst.

Hinzufügen einer CDC-Ressource

Anmerkung: Um das Feature „Change Data Capture“ von Spectrum™ zu verwenden, müssen Sie sicherstellen, dass der SQL-Server-Agent auf dem MS SQL-Server ausgeführt wird.

1. Öffnen Sie das Pop-up **Change Data Capture-Verwaltung** über einer der folgenden Möglichkeiten:

- Öffnen Sie **Tools > Change Data Capture-Verwaltung**.
 - Fügen Sie einem Auftrag den Schritt **DB Change Data Reader** hinzu, öffnen Sie die Schritteinstellungen, und klicken Sie auf **Verwalten**.
2. Klicken Sie auf **Hinzufügen**.
 3. Geben Sie einen **Namen** für die CDC-Ressource ein.
 4. Wählen Sie im Feld **Verbindung** die gewünschte SQL-Datenbankverbindung aus. Klicken Sie auf **Verwalten**, um eine neue Datenbankverbindung herzustellen. Weitere Informationen zum Erstellen von Datenbankverbindungen finden Sie unter [Datenbankverbindungs-Manager](#).
 5. Geben Sie im Feld **Tabelle/Ansicht** die Tabelle an, deren Spalten in den Auftragsfluss aufgenommen werden sollen. Klicken Sie auf die Schaltfläche „Durchsuchen“ ([...]), um zu Ihrer gewünschten Tabelle oder Ansicht zu navigieren.
Im Raster unten werden alle Spalten der ausgewählten Tabelle zusammen mit den Datentypen jeder Spalte angezeigt.
 6. Wenn Sie eine Oracle-Verbindung im Feld **Verbindung** auswählen, wird das **Startdatum** verfügbar.
Das Feld wird mit dem Standardwert des aktuellen Datums und mit der Uhrzeit 0:00 Uhr aufgefüllt. Sie können ein beliebiges Startdatum und eine beliebige Startzeit eingeben.
Das Enddatum und die Endzeit werden als das aktuelle Datum und die aktuelle Uhrzeit verwendet.
Achtung: Sie sollten über *Ausführungsberechtigungen* im LogMiner-Dienstprogramm von Oracle verfügen, um das CDC-Feature auf einer Oracle-Verbindung verwenden zu können. Weitere Informationen finden Sie unter [Oracle LogMiner-Konfiguration](#) auf Seite 326.
- Anmerkung:** Das Feld **Startdatum** ist bei der Auswahl einer MS SQL-Verbindung im Feld **Verbindung** nicht verfügbar.
7. Klicken Sie auf **OK**.
- Die erstellte CDC-Ressourcentabelle kann nun im **DB Change Data Reader**-Schritt verwendet werden. Ihre Tabellen können in den Auftragsfluss aufgenommen oder davon ausgeschlossen werden.

Der Schritt zeigt die Tabellenspalten an, auf denen das CDC-Feature aktiviert wurde.

Bearbeiten einer CDC-Ressource

1. Öffnen Sie das Pop-up **Change Data Capture-Verwaltung** über einer der folgenden Möglichkeiten:
 - Öffnen Sie **Tools > Change Data Capture-Verwaltung**.
 - Fügen Sie einem Auftrag den Schritt **DB Change Data Reader** hinzu, öffnen Sie die Schritteinstellungen, und klicken Sie auf **Verwalten**.

2. Wählen Sie die zu bearbeitende CDC-Ressource aus.
3. Klicken Sie auf **Bearbeiten**.
4. Ändern Sie die Details der hinzugefügten CDC-Ressource nach Bedarf.
5. Klicken Sie auf **OK**.

Löschen einer CDC-Ressource

1. Öffnen Sie das Pop-up **Change Data Capture-Verwaltung** über einer der folgenden Möglichkeiten:
 - Öffnen Sie **Tools > Change Data Capture-Verwaltung**.
 - Fügen Sie einem Auftrag den Schritt **DB Change Data Reader** hinzu, öffnen Sie die Schritteinstellungen, und klicken Sie auf **Verwalten**.
2. Wählen Sie die zu löschende CDC-Ressource aus.
3. Klicken Sie auf **Löschen**.

Auswählen von „Change Data Reader“-Optionen

„**DB Change Data Reader**“-Optionen stellt die Tabellenspalten der ausgewählten CDC-Ressource dar, für die das CDC-Feature aktiviert wurde.

Sie können auswählen, welche Spalten in den aktuellen Auftragsfluss aufgenommen und welche ausgeschlossen werden sollen.

1. Fügen Sie in einem Auftrag den **DB Change Data Reader**-Schritt hinzu.
2. Öffnen Sie „**DB Change Data Reader**“-Optionen, indem Sie doppelt auf das Symbol des Schrittes klicken.
3. Wählen Sie die gewünschte CDC-Ressource aus der Dropdown-Liste **Ressource auswählen** aus.
 Sie können eine CDC-Ressource hinzufügen oder ändern, indem Sie auf **Verwalten** klicken. Im folgenden Raster werden alle Tabellenspalten mit ihren Datentypen angezeigt. Zudem wird angezeigt, ob eine bestimmte Spalte im Auftragsfluss enthalten ist und ob die Spalte für das Feature „Change Data Capture“ aktiviert ist.
4. Wählen Sie über die Kästchen unter der Rasterspalte **Einschließen** die Tabellenspalten aus, die im Auftragsfluss enthalten sein sollen.
5. Die Kästchen unter der Rasterspalte **CDC aktiviert** stellen die Tabellenspalten dar, auf denen das Feature „Change Data Capture“ aktiviert ist.
 Die schreibgeschützten Kästchen **CDC aktiviert** werden für Spalten aktiviert, auf denen das Feature „Change Data Capture“ aktiviert ist.

Anmerkung: [Hier](#) finden Sie Informationen zu den Schritten, die erforderlich sind, um bei MS SQL-Datenquellen das Feature „Change Data Capture“ auf bestimmten Tabellenspalten aus dem Back-End zu aktivieren oder zu deaktivieren.

6. Klicken Sie auf **OK**.

Die Daten der Tabellenspalten, die für „Change Data Capture“ aktiviert wurden, werden erfasst und gespeichert.

DB Loader

Über den Schritt **DB Loader** können Sie auf Daten in Datenbanken, die im Modul Spectrum™ Data Integration konfiguriert wurden, zugreifen sowie Daten in diese und aus diesen laden. Dieser Schritt stellt eine Schnittstelle zu einem äußerst schnellen Dienstprogramm zum Laden von Daten zur Verfügung. Derzeit unterstützt die Spectrum™ Data Integration-Plattform **Oracle Loader**, **DB2 Loader**, **PostgreSQL Loader** und **Teradata Loader**.

Oracle Loader

Mithilfe von Oracle Loader können Sie Daten in eine beliebige Oracle-Datenbank laden, die in der Spectrum™ Data Integration-Plattform konfiguriert ist.

Anmerkung: Der Oracle-Client muss mit Administrator-Setup installiert sein, bevor Sie den Oracle Loader verwenden können.

Name der Option	Beschreibung
Verbindung	<p>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 Datenbankverbindung ändern oder löschen müssen, klicken Sie auf Verwalten.</p> <p>Wenn Sie eine Datenbankverbindung hinzufügen oder ändern, füllen Sie diese Felder aus:</p> <p>Verbindungsname Geben Sie einen Namen für die Verbindung ein. Sie können den Namen frei wählen.</p> <p>Datenbanktreiber Wählen Sie den passenden Datenbanktyp aus.</p> <p>Verbindungsoptionen Geben Sie Host, Port, Instanz, Benutzername und Kennwort an, die für die Verbindung zur Datenbank verwendet werden sollen.</p>

Name der Option	Beschreibung
Tabelle/Ansicht	<p>Geben Sie nach der Auswahl einer Verbindung die Tabelle oder Ansicht ein, in der geschrieben werden soll. Klicken Sie auf die Schaltfläche zum Durchsuchen ([...]), um zur gewünschten Tabelle oder Ansicht zu navigieren, oder klicken Sie auf Tabelle erstellen, um eine neue Tabelle in der Datenbank zu erstellen.</p> <p>Anmerkung: Wenn Sie in eine SQL-Datenbank schreiben, können Sie nicht in Ansichten schreiben, die mehr als eine Tabelle referenzieren. Das liegt an einer Einschränkung von SQL Server.</p>
Listener	Geben Sie einen Variablennamen an, der die Adress- und Verbindungsdetails enthält, die zur Herstellung einer Verbindung zur Oracle-Datenbank erforderlich sind, zum Beispiel „XE“. Diese Variable ist in <code>tnsnames.ora</code> vorhanden. Dies ist eine Datei, die clientseitige Netzwerkkonfigurationsparameter enthält.
Schrittfelder	In dieser Spalte werden die im Datenfluss verwendeten Feldnamen aufgelistet. Sie können diese Feldnamen nicht ändern.
Typen	In dieser Spalte ist der Datentyp jedes Feldes aufgelistet.

Registerkarte „Laufzeit“

Name der Option	Beschreibung
Lademethode	<p>Gibt an, wie Sie Daten in die Oracle-Datenbank schreiben möchten.</p> <p>Anfügen Fügt die Daten zur Zieltabelle hinzu, ohne die vorhandenen Daten der Tabelle zu löschen.</p> <p>Einfügen Lädt die Daten in die Datenbank. Die Tabelle muss vor dem Laden leer sein. Dies funktioniert nicht für mehrere Laufzeitinstanzen.</p> <p>Abschneiden und einfügen Löscht möglicherweise vorhandene Zeilen und lädt dann die Eingabedaten in die Tabelle. Dies funktioniert nicht für mehrere Laufzeitinstanzen.</p>
Direct Path Loader verwenden	Wählen Sie diese Option aus, um Daten direkt in die Oracle-Datenbank zu laden und einen Großteil der Datenverarbeitung zu vermeiden, die normalerweise erfolgt.
Nicht wiederherstellbar	<p>Dieses Kästchen wird aktiviert, wenn Sie Direct Path Loader verwenden auswählen. Wählen Sie diese Option aus, wenn Sie keine Wiederholungsprotokolle in die Datenbank schreiben möchten. Weitere Informationen zu Wiederholungsprotokollen finden Sie unter http://docs.oracle.com/cd/B28359_01/server.111/b28310/onlineredo001.htm#ADMIN11302</p>

Name der Option	Beschreibung
Protokolldateiordner	Gibt den Pfad zum Ordner an. Klicken Sie auf die Ellipsenschaltfläche (...), um zum gewünschten Ordner zu navigieren. Die Protokolldatei enthält einen Datensatz mit Aktivitäten dieses Schrittes während einer Ladesitzung.
Ordner für ungültige Datei	Gibt den Pfad zum Ordner an. Klicken Sie auf die Ellipsenschaltfläche (...), um zum gewünschten Ordner zu navigieren. Die ungültige Datei enthält eine Liste von Datensätzen, die im Schritt nicht in die Datenbank geladen werden konnten.
Maximal zulässige Anzahl von Fehlern	Geben Sie die maximal zulässige Anzahl von Fehlern an, nach der der Ladevorgang gestoppt werden soll. Um die Ladeoperation beim ersten Fehler zu stoppen, legen Sie den Wert 0 fest. Es sind maximal 32.767 Fehler zulässig.

Anmerkung: Sie können deutliche Leistungsverbesserungen erreichen, indem Sie mehrere Laufzeitinstanzen für diesen Vorgang verwenden. Klicken Sie dazu auf die Schaltfläche **Laufzeit** und geben Sie den erforderlichen Wert im Feld **Laufzeitinstanzen** ein.

DB2 Loader

Mit dem DB2 Loader können Sie Daten in eine beliebige DB2-Datenbank laden, die in der Spectrum™ Data Integration-Plattform konfiguriert ist.

Anmerkung: Der DB2-Laufzeitclient muss mit Administrator-Setup installiert sein, bevor Sie den DB2 Loader verwenden können.

Name der Option	Beschreibung
Verbindung	<p>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 Datenbankverbindung ändern oder löschen müssen, klicken Sie auf Verwalten.</p> <p>Wenn Sie eine Datenbankverbindung hinzufügen oder ändern, füllen Sie diese Felder aus:</p> <p>Verbindungsname Geben Sie einen Namen für die Verbindung ein. Sie können den Namen frei wählen.</p> <p>Datenbanktreiber Wählen Sie den passenden Datenbanktyp aus.</p> <p>Verbindungsoptionen Geben Sie Host, Port, Instanz, Benutzername und Kennwort an, die für die Verbindung zur Datenbank verwendet werden sollen.</p>

Name der Option	Beschreibung
Tabelle/Ansicht	Geben Sie nach der Auswahl einer Verbindung die Tabelle oder Ansicht ein, in der geschrieben werden soll. Klicken Sie auf die Schaltfläche zum Durchsuchen ([...]), um zur gewünschten Tabelle oder Ansicht zu navigieren, oder klicken Sie auf Tabelle erstellen , um eine neue Tabelle in der Datenbank zu erstellen.
Datenbank/Alias	<p>Dies ist die Variable, die den DB2-Server und die Datenbanken katalogisiert.</p> <p>DB2-Server katalogisieren Verwenden Sie die DB2-Befehlszeilenverarbeitung auf dem Spectrum-Server, um folgenden Befehl einzugeben:</p> <pre>CATALOG TCPIP NODE <nodename> REMOTE <hostname> SERVER <port></pre> <p>Dabei gilt Folgendes:</p> <p>nodename: Name der Verbindung</p> <p>hostname: TCP/IP-Name des DB2-Servers</p> <p>port: Serverport</p> <p>Datenbank katalogisieren Verwenden Sie den folgenden Befehl:</p> <pre>CATALOG DATABASE <databasename> AS <local_database_alias> AT NODE <nodename></pre> <p>Dabei gilt Folgendes:</p> <p>databasename: Name der Datenbank auf dem DB2-Server</p> <p>local_database_alias: lokaler Name für die Datenbank, der bei der Verbindungsherstellung zum Server vergeben wird</p> <p>nodename: Name, der im vorherigen Befehl CATALOG TCP/IP verwendet wurde</p>
Schrittfelder	In dieser Spalte werden die im Datenfluss verwendeten Feldnamen aufgelistet. Sie können diese Feldnamen nicht ändern.
Typen	In dieser Spalte ist der Datentyp jedes Feldes aufgelistet.

Registerkarte „Laufzeit“

Name der Option	Beschreibung
Lademethode	<p>Zeigt den Modus an, in dem Daten in eine DB2-Tabelle geschrieben werden.</p> <p>Einfügen Fügt die geladenen Daten in die Tabelle ein, wobei die vorhandenen Tabellendaten nicht geändert werden.</p> <p>Ersetzen Fügt die geladenen Daten in die Tabelle ein, nachdem alle vorhandenen Daten aus ihr gelöscht wurden.</p> <p>Das Tabellenschema und die Indexdefinitionen bleiben unverändert.</p> <p>Neu starten Startet den Datenladevorgang neu, wenn der vorherige Ladeversuch unterbrochen wurde.</p>
Nicht wiederherstellbar	<p>Zeigt an, ob diese Ladetransaktion nicht wiederherstellbar ist.</p> <p>Wenn Sie diese Option auswählen, wird die Ladetransaktion als nicht wiederherstellbar gekennzeichnet. Tabellenbereiche werden nach dem Laden nicht in den Status Backup Pending versetzt. Auch wird während des Ladevorgangs keine Kopie der geladenen Daten angefertigt. Daher kann eine nicht wiederherstellbare Transaktionen nicht wiederhergestellt werden, falls eine Datenladefehler auftritt, auch nicht, wenn später ein <code>rollforward</code> versucht wird.</p> <p>Wenn Sie diese Option auswählen, können Sie den Zustand vor der Transaktionen nicht wiederherstellen, auch nicht, wenn Sie das DB2-Dienstprogramm <code>rollforward</code> verwenden, da das Dienstprogramm solche nicht wiederherstellbaren Transaktionen überspringt und die Tabelle als „ungültig“ kennzeichnet. Außerdem werden auch nachfolgende Transaktionen bezüglich der Tabelle von <code>rollforward</code> ignoriert.</p> <p>Um eine Tabelle wiederherzustellen, die nicht wiederherstellbare Transaktionen enthält, müssen Sie entweder eine Sicherung auf Tabellenbereichebene oder eine vollständige Sicherung an einem Übergabepunkt direkt nach dem nicht wiederherstellbaren Ladevorgang durchführen.</p> <p>Anmerkung: Wählen Sie diese Option nicht aus, wenn die Daten Datalink-Spalten mit vorhandenem Attribut „File Link Control“ enthalten.</p>
CPU	Die Anzahl an parallelen Threads, die das Ladedienstprogramm zum Laden, Parsen und Formatieren von Datensätzen generieren und erhalten kann, während es Tabellenobjekte pro Datenbankpartition erzeugt.
Scheibe	Die Anzahl an parallelen Threads, die das Ladedienstprogramm zum Schreiben von Daten in die Tabellenbereichscontainer generieren und erhalten kann.

Name der Option	Beschreibung
Indizierungsmodus	<p>Zeigt den Verarbeitungsmodus des Ladedienstprogramms für Indizes an.</p> <p>Automatisch auswählen Das Ladedienstprogramm entscheidet aufgrund der Datenmenge und der Tiefe des Indexbaums, ob eine Neuerstellung oder der inkrementelle Modus angewendet wird.</p> <p>Erneut erstellen Alle Indizes werden neu erstellt.</p> <p>Inkrementell Neue Daten werden den vorhandenen Indizes hinzugefügt.</p> <p>Dieser Modus kann nur angewendet werden, wenn das Indexobjekt gültig ist und beim Starten des Ladevorgangs darauf zugegriffen werden kann.</p> <p>Anmerkung: Inkrementelle Indizierung wird nicht unterstützt, wenn ALLE folgenden Bedingungen zutreffen:</p> <ol style="list-style-type: none"> 1. Die Option „Laden/Kopieren“ ist angegeben (<code>logretain</code> oder <code>userexit</code> ist aktiviert). 2. Die Tabelle befindet sich in einem DMS-Tabellenbereich. 3. Das Indexobjekt befindet sich in einem Tabellenbereich, der auch von anderen Tabellenobjekten verwendet wird, die zur zu ladenden Tabelle gehören. <p>Um diese Einschränkung zu umgehen, platzieren Sie Indizes in separaten Tabellenbereichen.</p> <p>Zurückgestellt Das Ladedienstprogramm versucht nicht, einen Index zu erstellen. Vorhandene Indizes werden zur Aktualisierung gekennzeichnet.</p> <p>Anmerkung: Die Erstellung des Index erfordert im Modus „Zurückgestellt“ mehr Zeit als im Modus „Erneut erstellen“. Wenn Sie also mehrere Ladevorgänge durchführen, sollten Sie dem letzten Ladevorgang erlauben, alle Indizes neu zu erstellen, anstatt Indizes beim ersten Nicht-Lade-Zugriff erneut erstellen zu lassen.</p> <p>Anmerkung: Dieser Modus wird nur für Tabellen mit nicht eindeutigen Indizes unterstützt.</p>

Name der Option	Beschreibung
Schnelles Parsen	<p>Zeigt an, ob die Syntaxüberprüfung bei Spaltenwerten übersprungen wird, um die Leistung zu erhöhen.</p> <p>Wenn diese Option aktiviert ist, werden alle Syntaxfehler in den Daten ignoriert, um die Leistung zu optimieren.</p> <p>Wenn beispielsweise ein Zeichenfolgenwert von 12wxvvg56 in einem Feld gefunden wird, das einer Integer-Spalte in einer ASCII-Datei zugeordnet ist, erzeugt das Ladedienstprogramm in der Regel einen Syntaxfehler. Wenn jedoch „Schnelles Parsen“ ausgewählt ist, wird der Syntaxfehler ignoriert und eine willkürliche Zahl in das Integer-Feld geladen.</p> <p>Anmerkung: Achten Sie darauf, diese Option nur bei korrekten und sauberen Daten zu verwenden.</p>
Schemaname	Das Schema, in dem die Ausnahmetabellen gespeichert werden.
Tabellenname	Die Ausnahmetabelle, in die Zeilen kopiert werden, bei denen während des Ladevorgangs ein Fehler auftritt.
Protokolldateiordner	<p>Der Pfad zu dem Verzeichnis, in dem die Protokolldateien gespeichert werden sollen.</p> <p>Eine Protokolldatei enthält eine Liste mit den Datenbankladetransaktionen, die ein DB Loader-Schritt während einer Ladesitzung ausführt.</p> <p>Klicken Sie auf die Ellipsenschaltfläche (...), um das gewünschte Verzeichnis für Protokolldateien anzugeben.</p>
Ordner für ungültige Datei	<p>Der Pfad zu dem Verzeichnis auf dem DB2-Server, in dem die ungültigen Dateien gespeichert werden sollen.</p> <p>Eine ungültige Datei enthält eine Liste der Datensätze, die in einem DB Loader-Schritt nicht in die Datenbank geladen werden konnten.</p> <p>Klicken Sie auf die Ellipsenschaltfläche (...), um das gewünschte Verzeichnis für ungültige Dateien anzugeben.</p>
Maximal zulässige Anzahl von Fehlern	<p>Die maximal zulässige Anzahl von Fehlern, nach der der Ladevorgang abgebrochen wird.</p> <p>Um einen Ladevorgang beim ersten gefundenen Fehler abzubrechen, legen Sie den Wert dieses Felds auf 0 fest.</p> <p>Anmerkung: Es sind maximal 32.767 Fehler zulässig.</p>

- Parallelität** Eine DB2-Datenbank kann in mehrere Partitionen unterteilt werden, indem die Umgebung zu mehreren verschiedenen physischen Knoten geklont wird.
- Separate Datenbankanforderungen zum Abrufen und Aktualisieren von Daten werden automatisch auf die unterschiedlichen Partitionen aufgeteilt und parallel ausgeführt, um die Leistung zu optimieren.
- Ausnahmebehandlung** Bei einer DB2-Datenbank können Sie die Fehler und Ausnahmen, die während der Ausführung von Abfragen und Prozeduren auftreten, aufzeichnen und entsprechend verarbeiten.
- Dazu bietet eine DB2-Datenbank besondere Ausnahmetabellen und -schemas, die die Quelle und auch den Protokollablauf jeder Datenbankausnahme speichern.

PostgreSQL Loader

Mit dem PostgreSQL Loader können Sie Daten in eine beliebige PostgreSQL-Datenbank laden, die in der Spectrum™ Data Integration-Plattform konfiguriert ist.

Name der Option	Beschreibung
Verbindung	<p>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 Datenbankverbindung ändern oder löschen müssen, klicken Sie auf Verwalten.</p> <p>Wenn Sie eine Datenbankverbindung hinzufügen oder ändern, füllen Sie diese Felder aus:</p> <p>Verbindungsname Geben Sie einen Namen für die Verbindung ein. Sie können den Namen frei wählen.</p> <p>Datenbanktreiber Wählen Sie den passenden Datenbanktyp aus.</p> <p>Verbindungsoptionen Geben Sie Host, Port, Instanz, Benutzername und Kennwort an, die für die Verbindung zur Datenbank verwendet werden sollen.</p>
Tabelle/Ansicht	Klicken Sie auf die Schaltfläche „Durchsuchen“ ([...]), um zur Tabelle oder Ansicht zu navigieren, die Sie verwenden möchten.
Datenbankfelder	In dieser Spalte sind die Feldnamen in der Datenbank aufgelistet. Sie können diese Feldnamen nicht ändern.

Name der Option	Beschreibung
Schrittfelder	In dieser Spalte werden die im Datenfluss verwendeten Feldnamen aufgelistet. Sie können diese Feldnamen nicht ändern.
Typen	In dieser Spalte ist der Datentyp jedes Feldes aufgelistet.

Registerkarte „Laufzeit“

Name der Option	Beschreibung
Lademethode	Gibt die Methode an, wie Daten in die PostgreSQL-Datenbanktabellen geschrieben werden. <ul style="list-style-type: none"> Wählen Sie Einfügen aus, um Daten in eine leere Tabelle zu schreiben oder an eine vorhandene Tabelle anzufügen. Wählen Sie Abschneiden und einfügen aus, um die Daten abzuschneiden, bevor sie in die Datenbanktabelle geladen werden.

Anmerkung: Sie können deutliche Leistungsverbesserungen erreichen, indem Sie mehrere Laufzeitinstanzen für diesen Vorgang verwenden. Klicken Sie dazu auf die Schaltfläche **Laufzeit** und geben Sie den erforderlichen Wert im Feld **Laufzeitinstanzen** ein.

Teradata Loader

Mit dem Teradata Loader können Sie Daten in eine beliebige Teradata-Datenbank laden, die in der Spectrum™ Data Integration-Plattform konfiguriert ist.

Anmerkung: Der Loader wird nur auf Windows-Systemen unterstützt.

Name der Option	Beschreibung
Verbindung	<p>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 Datenbankverbindung ändern oder löschen müssen, klicken Sie auf Verwalten.</p> <p>Wenn Sie eine Datenbankverbindung hinzufügen oder ändern, füllen Sie diese Felder aus:</p> <p>Verbindungsname Geben Sie einen Namen für die Verbindung ein. Sie können den Namen frei wählen.</p> <p>Datenbanktreiber Wählen Sie den passenden Datenbanktyp aus.</p> <p>Verbindungsoptionen Geben Sie Host, Port, Instanz, Benutzername und Kennwort an, die für die Verbindung zur Datenbank verwendet werden sollen.</p>
Tabelle/Ansicht	Klicken Sie auf die Schaltfläche „Durchsuchen“ ([...]), um zur Tabelle oder Ansicht zu navigieren, die Sie verwenden möchten.
Datenbankfelder	In dieser Spalte sind die Feldnamen in der Datenbank aufgelistet. Sie können diese Feldnamen nicht ändern.
Schrittfelder	In dieser Spalte werden die im Datenfluss verwendeten Feldnamen aufgelistet. Sie können diese Feldnamen nicht ändern.
Typen	In dieser Spalte ist der Datentyp jedes Feldes aufgelistet.

Registerkarte „Laufzeit“

Name der Option	Beschreibung
Lademethode	<p>Gibt die Methode an, wie Daten in die Teradata-Datenbanktabellen geschrieben werden.</p> <ul style="list-style-type: none"> • Wählen Sie Einfügen aus, um Daten in eine leere Tabelle zu schreiben. • Wählen Sie Anfügen aus, um Daten in eine vorhandene Datentabelle zu schreiben. • Wählen Sie Abschneiden und einfügen aus, um die Daten abzuschneiden, bevor sie in die Datenbanktabelle geschrieben werden.
Protokolldateiordner	Wählen Sie den Speicherort des Ordners aus, in dem die Protokolldateien des Hochladeprozesses gespeichert werden.
Ungültige Datei generieren	Aktivieren Sie dieses Kästchen, um ein Protokoll über die Datensätze zu generieren, bei denen das Hochladen fehlgeschlagen ist.

Name der Option	Beschreibung
Ordner für ungültige Datei	Wählen Sie den Speicherort des Ordners aus, in dem das Protokoll mit fehlgeschlagenen Datensätzen gespeichert wird. Das Protokoll enthält Details wie Fehlercodes und Namen fehlerhafter Felder von fehlgeschlagenen Datensätzen.
Maximal zulässige Anzahl von Fehlern	Geben Sie die Anzahl der maximal zulässigen Fehler für die Hochladesitzung an. Wenn die Fehleranzahl diesen Wert überschreitet, wird der Hochladeprozess pausiert.

Anmerkung: Für Teradata Loader werden mehrere Laufzeitanstzen nicht unterstützt.

Field Parser

Der **Field Parser**-Schritt extrahiert Felder aus XML-Daten und Daten mit Begrenzungszeichen in der angegebenen Eingabespalte. Führen Sie folgende Schritte aus, um **Field Parser**-Optionen zu konfigurieren.

1. Wählen Sie aus dem Feld **Quelle** die Spalte aus, in der sich die zu parsenden XML-Daten oder Daten mit Begrenzungszeichen befinden.

Anmerkung: Das Listenfeld zeigt alle Zeichenfolgeneingabespalten.

2. Wählen Sie je nach Typ der zu parsenden Daten **XML-Format** oder `mit Begrenzungszeichen` aus, und wählen Sie die unten beschriebenen Optionen aus.

Field Parser-Optionen für XML-Daten

Name der Option	Beschreibung
Servername	Gibt an, ob sich die zum Ableiten des Schemas ausgewählte Datei auf dem Computer mit dem Enterprise Designer oder auf dem Server befindet. Wenn Sie eine Datei auf dem lokalen Computer auswählen, lautet der Servername „Arbeitsplatz“. Wenn Sie eine Datei auf dem Server auswählen, lautet der Servername Spectrum™ Technology Platform.

Name der Option	Beschreibung
Schemadatei	<p>Gibt den Pfad zu einer XSD-Schemadatei an. Klicken Sie auf die Ellipsenschaltfläche (...), um zum Speicherort der Datei zu gelangen. Die Schemadatei sich kann auf dem Server oder Ihrem lokalen System befinden.</p> <p>Wahlweise können Sie auch eine XML-Datei statt einer XSD-Datei angeben. Wenn Sie eine XML-Datei angeben, wird das Schema auf Basis der Struktur der XML-Datei abgeleitet. Die Verwendung einer XML-Datei statt einer XSD-Datei hat die folgenden Einschränkungen:</p> <ul style="list-style-type: none"> • Die XML-Datei darf nicht größer als 1 MB sein. Wenn die XML-Datei größer ist, versuchen Sie, einige der Daten zu entfernen und trotzdem die Struktur der XML beizubehalten. • Die Datendatei wird nicht anhand des abgeleiteten Schemas überprüft. <p>Anmerkung: 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.</p>

Ausgabefelder	<p>Dieser Abschnitt zeigt Details des ausgewählten Schemas. Er beinhaltet das Stammelement, gefolgt von den untergeordneten Elementen mit ihren Attributen.</p> <p>Standardmäßig bleiben alle Knoten des Schemas ausgewählt. Sie können jedoch die Kästchen der Knoten deaktivieren, die nicht an den nächsten Schritt übergeben werden sollen.</p> <ul style="list-style-type: none"> • Knoten suchen: Geben Sie den Namen des Knotens ein, zu dem Sie im Schemabaum navigieren möchten. Der eingegebene Knoten wird im Vorschaubereich unter dem Feld markiert. • XPath: Klicken Sie auf eine beliebige Stelle in diesem Feld, um den XML-Pfad (XPath) der Elemente und Attribute des markierten Knotens im Schemabaum anzuzeigen. Klicken Sie auf den Abwärtspfeil am rechten Ende des Feldes, um alle von Ihnen angezeigten vorherigen XPaths zu sehen. <p>Anmerkung: XPath ist eine Sprache, um Informationen in einem XML-Dokument zu finden. Weitere Informationen hierzu finden Sie unter www.w3schools.com/xpath/</p>
---------------	---

Field Parser-Optionen für Daten mit Begrenzungszeichen

Name der Option	Beschreibung
Feldtrennzeichen	<p>Wählen Sie in der Dropdown-Liste das Feldtrennzeichen aus, das in der zu parsenden Spalte mit Begrenzungszeichen verwendet wird.</p> <p>Wenn in der Spalte mit Begrenzungszeichen ein anderes Zeichen als Feldtrennzeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Feldtrennzeichen auszuwählen.</p>

Name der Option	Beschreibung
Textbegrenzungszeichen	<p>Wählen Sie in der Dropdown-Liste das Textbegrenzungszeichen aus, das in der zu parsenden Spalte mit Begrenzungszeichen verwendet wird.</p> <p>Anmerkung: Textbegrenzungszeichen sind Zeichen, die in Daten mit Begrenzungszeichen Textwerte umschließen.</p> <p>Wenn in der Spalte mit Begrenzungszeichen ein anderes Textbegrenzungszeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Textbegrenzungszeichen auszuwählen.</p>
Ausgabebetyp	<p>Legen Sie fest, ob die geparsete Ausgabe in Form einer Liste (hierarchische Anzeige der Werte) oder als Felder erfolgen soll.</p> <p>Anmerkung: Bei einer Liste als Ausgabebetyp können Sie nur ein Ausgabefeld hinzufügen, während Sie bei der Option Felder mehrere Felder hinzufügen können, in denen die Werte während des Parsens aufgeteilt werden.</p>

Name der Option	Beschreibung
-----------------	--------------

Ausgabefelder

In diesem Abschnitt können Sie die verschiedenen Felder, in denen die Details der Daten mit Begrenzungszeichen aufgeteilt werden sollen, hinzufügen/ändern. Sie können außerdem beliebige der hinzugefügten Ausgabefelder löschen.

Um ein neues Feld für die Anzeige der geparsten Ausgabe hinzuzufügen, klicken Sie auf die Schaltfläche **Hinzufügen** und führen Sie folgende Schritte im angezeigten Dialogfeld **Feldeinstellung** aus:

1. Geben Sie den **Namen** des Feldes ein.
2. Wählen Sie aus der Dropdown-Liste **Typ** den Datentyp für die Felder aus, die hinzugefügt werden. Je nach ausgewähltem Typ können mehrere Felder definiert werden. Bei **Datum** können Sie das Format beispielsweise als `M/d/yy`, `MMM d. yyyy` oder `MMMM d. yyyy` definieren. Details zu Datentypen und zum Definieren ihrer Details finden Sie unter [Definieren von Feldern in einer Datei mit Begrenzungszeichen](#) auf Seite 158.

Anmerkung: Wenn Sie als Datentyp **Zeichenfolge** auswählen, werden alle Typen von Daten mit Begrenzungszeichen geparkt. Sie können jedoch je nach Daten, die Sie in dem Feld parsen möchten, auch den spezifischen Typ verwenden.

3. Geben Sie in das Feld **Position** die Position des Datentyps (in der Eingabedatei) an, der in diesem Feld geparkt werden soll. Beispiel: Wenn Sie im folgenden Dateiausschnitt die Datum/Uhrzeit-Werte in dem hinzuzufügenden Feld parsen möchten, geben Sie für **Position** den Wert 3 ein.

```
true;"02/02/2022";"10/2/92 5:05
AM";598985994665542.25634;1;
"Arjun";74785.155;5:05PM,1,Deepak,65152
false;"15/03/1923";"3/23/90 11:55
AM";3425699466554.2563;2;
"sharma";5.1;5:45AM,2,Arjun,365273
```

4. Klicken Sie auf **Feld hinzufügen** und auf **Schließen**.

Das hinzugefügte Feld und seine Details werden im Kästchen angezeigt.

Anmerkung: Wenn Sie überzählige Leerzeichen am Anfang und Ende der Wertezeichenfolge eines Feldes entfernen möchten, aktivieren Sie das Kästchen **Kürzen**.

Ändern: Klicken Sie auf diese Schaltfläche, um Details eines beliebigen hinzugefügten Ausgabefelds zu ändern.

Entfernen: Klicken Sie auf diese Schaltfläche, um ein beliebiges der hinzugefügten Ausgabefelder zu entfernen.

Laufzeit: Verwenden Sie diese Schaltfläche, um mehrere Laufzeitinstanzen des Parsers anzugeben. Dies führt zu einer signifikanten Leistungsverbesserung.

OK: Klicken Sie auf diese Schaltfläche, um alle in diesem Schritt eingegebenen Details zu speichern.

Abbrechen: Klicken Sie auf diese Schaltfläche, um alle vorgenommenen Aktualisierungen abzurechnen.

Hilfe: Klicken Sie auf diese Schaltfläche, um die Hilfedatei für diesen Schritt aufzurufen.

Field Combiner

Der Schritt **Field Combiner** kombiniert Felder, die aus dem vorherigen Schritt in einem Datenfluss stammen, um eine XML-Zeichenfolge zu erstellen.

Field Combiner-Optionen

Name der Option	Beschreibung
Ausgabespaltenname	Gibt den Namen der Spalte an, die die kombinierten Felder als XML-Zeichenfolge enthalten soll.

Name der Option	Beschreibung
Ausgabefelder	<p>Dieser Abschnitt hilft Ihnen, die zu kombinierenden Felder auszuwählen, und verschiedene Aktionen für diese Felder auszuführen.</p> <p>Anmerkung: Die Dropdown-Liste Namespace anwenden sowie die Schaltflächen Element, Attribut und Entfernen werden aktiviert, wenn Sie ein oder mehr Elemente/Attribute im Schemabereich auswählen.</p> <ul style="list-style-type: none"> • Schnell hinzufügen: Klicken Sie auf diese Schaltfläche, um das Popup-Fenster Felder hinzufügen/entfernen zu öffnen. Dieses Fenster zeigt eine Liste aller Felder, die aus dem vorherigen Schritt stammen. Wählen Sie die Felder aus, die Sie kombinieren möchten. <p>Anmerkung: Die ausgewählten Felder werden im Schemabereich unter dem Feld Knoten suchen angezeigt.</p> <ul style="list-style-type: none"> • Knoten suchen: Geben Sie den Namen des Knotens ein, zu dem Sie im Schemabereich navigieren möchten. Der eingegebene Knoten wird hervorgehoben und sein XML-Pfad im Feld XPath unter dem Bereich angezeigt. • Element: Klicken Sie auf diese Schaltfläche, um ein Attribut in ein Element in der XML-Zeichenfolge zu konvertieren. • Attribut: Klicken Sie auf diese Schaltfläche, um ein Element in ein Attribut in der XML-Zeichenfolge zu konvertieren. • Namespace anwenden: Wenn Sie einen XML-Namespace für ein Element oder ein Attribut angeben möchten, wählen Sie ihn hier aus. <p>Anmerkung: Sie können Namespaces erstellen, indem Sie auf die Schaltfläche Namespace oben klicken. Weitere Informationen hierzu finden Sie unter Namespaces definieren.</p> <ul style="list-style-type: none"> • Entfernen: Klicken Sie auf diese Schaltfläche, um die Elemente/Attribute zu entfernen, die nicht in der XML-Zeichenfolge enthalten sein sollen.
XPath	<p>Klicken Sie auf eine beliebige Stelle in diesem Feld, um den XML-Pfad (XPath) des im Schemabereich markierten Knotens anzuzeigen. Klicken Sie auf den Abwärtspfeil am rechten Ende des Feldes, um alle von Ihnen angezeigten vorherigen XPaths zu sehen.</p> <p>Anmerkung: XPath ist eine Sprache, um Informationen in einem XML-Dokument zu finden. Weitere Informationen hierzu finden Sie unter www.w3schools.com/xpath/</p>
	<ul style="list-style-type: none"> • Laufzeit: Verwenden Sie diese Schaltfläche, um mehrere Laufzeitinstanzen von Field Combiner anzugeben. Dies führt zu einer signifikanten Leistungsverbesserung. • OK: Klicken Sie auf diese Schaltfläche, um alle in diesem Schritt eingegebenen Details zu speichern. • Abbrechen: Klicken Sie auf diese Schaltfläche, um alle vorgenommenen Aktualisierungen abzurechnen. • Hilfe: Klicken Sie auf diese Schaltfläche, um die Hilfedatei für diesen Schritt aufzurufen.

Namespace definieren

Mithilfe von Namespaces können Sie Element- und Attributnamen in Ihrer Ausgabe duplizieren, indem Sie jedem Element oder Attribut einen XML-Namespace zuweisen.

So definieren Sie einen oder mehrere Namespaces:

1. Klicken Sie auf dem Bildschirm **Field Combiner-Optionen** auf die Schaltfläche **Namespace**. Das Popup-Fenster **Namespace-Details** wird angezeigt.
2. Geben Sie in die Spalte **Präfix** das Präfix ein, das einem Element oder Attribut zugeordnet werden soll.
3. Geben Sie in der Spalte **Namespace** die URL des Namespace an.
4. Wiederholen Sie diese Schritte, um so viele Namespaces zu definieren, wie Sie verwenden möchten.

Field Selector

Mit „Field Selector“ können Sie auswählen, welche Felder an den nächsten Schritt im Datenfluss weitergegeben werden sollen. Sie können „Field Selector“ verwenden, um unerwünschte Felder aus einem Datenfluss zu entfernen. Wenn Sie beispielsweise ein neues Feld durch Kombinieren der Daten zweier Felder erstellt haben und Sie die beiden Quellfelder nicht mehr benötigen, können Sie „Field Selector“ verwenden, um nur das neue Feld beizubehalten und die beiden Quellfelder aus dem Datenfluss zu entfernen.

Optionen

Option	Beschreibung
Wählen Sie die Felder aus, die an den nächsten Schritt gesendet werden sollen.	Aktivieren Sie das Kästchen neben jedem Feld, das Sie an den nächsten Schritt im Datenfluss senden möchten. Wenn Sie nicht möchten, dass ein Feld an den nächsten Schritt gesendet wird, deaktivieren Sie das Kästchen.
Alle auswählen	Aktivieren Sie dieses Kästchen, um alle Felder im Datenfluss auszuwählen. Deaktivieren Sie dieses Kästchen, um die Auswahl aller Felder aufzuheben.

Generate Time Dimension

Generate Time Dimension erstellt einen Datensatz für jeden Tag des von Ihnen angegebenen Datumsbereichs. Sie können dann diese Datensätze mithilfe des „Write to DB“-Schrittes in eine Zeitdimensionstabelle einer Datenbank schreiben. Sie können anschließend die Zeitdimensionstabelle verwenden, um genaue, auf einem Zeitraum basierende Berechnungen durchzuführen. Beispielsweise handelt es sich bei den Analysen „Verkäufe pro Quartal“, „pro Quartal aufgewendetes Budget“ und „Tagesumsatz“ um Analysen, die eine Zeitdimension erfordern. Über Zeitdimensionstabellen können sie außerdem Geschäftsjahre oder vom Standard abweichende Quartale in Analysen berücksichtigen.

Beispiel der Verwendung einer Zeitdimensionstabelle

Zeitdimensionstabellen werden für genaue zeitbasierte Berechnungen benötigt, da die erforderlichen Datumsdaten manchmal nicht einfach aus den Datensätzen extrahiert werden können. Als Beispiel dienen hier die folgenden Datensätze aus einer Umsatzdatenbank. Beachten Sie, dass zwischen den Datensätzen zeitliche Lücken vorhanden sind. Beispielsweise ist kein Datensatz für den 01.04.2012 vorhanden.

Datum	Produkt	Betrag
03.01.2012	Rotes Hemd	10,00 \$
05.01.2012	Rotes Hemd	5,00 \$
07.01.2012	Rotes Hemd	15,00 \$

Wenn Sie diese Datensätze abfragen und den durchschnittlichen Tagesumsatz berechnen, wäre das Ergebnis 10,00 \$ (30 \$ / 3 Datensätze). Dieses Ergebnis ist jedoch falsch, da die drei Datensätze eigentlich einen Zeitraum von fünf Tagen abdecken. Wenn Sie über eine Zeitdimensionstabelle mit einem Datensatz für jeden Tag verfügen, können Sie diese Tabelle mit der obigen Tabelle zusammenführen. Sie erhalten dann Folgendes:

Datum	Produkt	Betrag
03.01.2012	Rotes Hemd	10,00 \$

Datum	Produkt	Betrag
04.01.2012		
05.01.2012	Rotes Hemd	5,00 \$
06.01.2012		
07.01.2012	Rotes Hemd	15,00 \$

Bei der Berechnung des durchschnittlichen Tagesumsatzes erhalten Sie bei diesen Datensätzen das korrekte Ergebnis, 6,00 \$ (30 \$ / 5 Tage).

Darüber hinaus können Sie in Ihrer Berechnung beliebige Zeitattribute berücksichtigen wie Feiertage, Wochenenden und Quartale. Wenn beispielsweise der 06.01.2012 ein Feiertag war und Sie nur am durchschnittlichen Umsatz pro Arbeitstag interessiert sind, ist das Ergebnis 7,50 \$.

Optionen

Generate Time Dimension verfügt über folgende Optionen.

Option	Beschreibung
Startdatum	Dies ist der erste Tag des Datumsbereichs, für den „Generate Time Dimension“-Datensätze erstellt werden sollen.
Enddatum	Verwenden Sie diese Option, wenn Sie ein bestimmtes Enddatum für die Zeitdimension angeben möchten. Für jeden Tag zwischen dem Startdatum und dem hier angegebenen Datum wird ein Datensatz generiert.
Dauer	Verwenden Sie diese Option, wenn die Zeitdimension eine bestimmte Anzahl von Tagen, Monaten oder Jahren umfassen soll. Für jeden Tag innerhalb der Dauer wird ein Datensatz generiert. Wenn Sie beispielsweise eine Woche angeben, werden sieben Datensätze generiert. Der erste Datensatz ist der Tag, der im Feld Startdatum angegeben wurde.

Option	Beschreibung
Zeitattribut	

Option

Beschreibung

Option	Beschreibung
	Gibt den Typ der Zeitinformation an, die in der Zeitdimension enthalten sein soll. Jedes Attribut erzeugt ein Feld im Datensatz jeden Tags. Zur Auswahl stehen:
Datum	Dies ist das Tagesdatum im Format <Tag> <Monat> <Jahr>. Der Monatsname liegt dabei in der Sprache des eingestellten Gebietsschemas des Servers vor. Wenn der Server beispielsweise mit dem Gebietsschema Englisch läuft, sieht ein Datum folgendermaßen aus: 30 Oktober 2012.
Tag des Monats	Diese Zahl stellt den Tag des Monats dar. 10 bedeutet zum Beispiel, dass der Tag der 10. Tag des Monats ist.
Tag des Jahres	Diese Zahl stellt den Tag des Jahres dar. 304 bedeutet zum Beispiel, dass der Tag der 304. Tag des Jahres ist.
Ist Schaltjahr	Dieser boolesche Wert zeigt an, ob sich der Tag in einem Schaltjahr befindet. In diesem Fall ist der Wert <code>true</code> , andernfalls ist der Wert <code>false</code> .
Ist Werktag	Dieser boolesche Wert zeigt an, ob der Tag ein Werktag (Montag bis Freitag) ist. In diesem Fall ist der Wert <code>true</code> , andernfalls ist der Wert <code>false</code> .
Ist Wochenende	Dieser boolesche Wert zeigt an, ob der Tag ein Wochenendtag (Samstag oder Sonntag) ist. In diesem Fall ist der Wert <code>true</code> , andernfalls ist der Wert <code>false</code> .
Julianischer Tag	Dies ist eine eindeutige Nummer für den Tag. Im julianischen Tagessystem wird der 1. Januar 4713 v. Chr. als Tag 1 gezählt, und die Zahlen für alle folgenden Tage werden entsprechend erhöht. Beispielsweise bedeutet 2456231, dass dies der 2.456.231. Tag seit dem 1. Januar 4713 vor Christus ist.
Julianische Woche	Dies ist eine eindeutige Nummer für die Woche des Tags. Im julianischen System wird die erste Woche des Jahres 4713 v. Chr. als Woche 1 gezählt, und die Zahlen für alle folgenden Wochen werden entsprechend erhöht. Beispielsweise bedeutet 350890, dass der Tag in der 350.890. Woche seit der ersten Woche von 4713 v. Chr. liegt.
Julianisches Jahr	Dies ist eine eindeutige Nummer für das Jahr. Im julianischen System wird das Jahr 4713 v. Chr. als Jahr 1 gezählt, und die Zahlen für alle folgenden Jahre werden entsprechend erhöht. Beispielsweise bedeutet 6725, dass der Tag im Jahr 2012 n. Chr. liegt.
Monatsname	Dies ist der Name des Monats, in dem der Tag liegt, in englischer Sprache.
Monatszahl	Diese Zahl stellt den Monat des Jahres dar. 3 bedeutet zum Beispiel, dass der Tag im dritten Monat des Jahres liegt.
Quartal	Diese Zahl stellt das Quartal des Jahres dar. 1 bedeutet zum Beispiel, dass der Tag im ersten Quartal des Jahres liegt.

Option	Beschreibung
	<p>Woche des Jahres Diese Zahl stellt die Woche des Jahres dar. 43 bedeutet zum Beispiel, dass der Tag in der 43. Woche des Jahres liegt.</p> <p>Wochentagname Dies ist der Name des Tags in deutscher Sprache. Beispiel: Monday.</p> <p>Wochentagnummer Diese Zahl stellt den Tag der Woche dar, wobei Tag 1 Montag ist. Dienstag ist dementsprechend Tag 2, Mittwoch ist Tag 3 usw.</p> <p>Jahr Dies ist das Jahr, in dem der Tag liegt, nach dem gregorianischen Kalender. Beispielsweise bedeutet 2012, dass der Tag im Jahr 2012 liegt.</p>
Feld	Dies ist der Name des Feldes, das im Datenfluss erstellt wird und das Zeitattribut enthält. Hier wird ein Standardfeldname bereitgestellt, Sie können den Feldnamen jedoch ändern.
Typ	Dies ist der Datentyp des Feldes. Generate Time Dimension wählt automatisch den passenden Datentyp für jedes Zeitattribut aus.
Kalender	Gibt an, ob bei der Berechnung des Zeitattributs ein benutzerdefinierter Kalender oder der standardmäßige gregorianische Kalender verwendet werden soll. Um einen benutzerdefinierten Kalender zu definieren, klicken Sie auf Kalender .

KaKaErstellen eines Kalenders

Ein Kalender definiert wichtige Merkmale im Jahr entsprechend des Typs der Analyse, die Sie durchführen möchten. Standardmäßig wird der gregorianische Kalender verwendet. Der standardmäßige gregorianische Kalender eignet sich jedoch nicht für alle Anwendungen. Wenn das Geschäftsjahr Ihres Unternehmens beispielsweise am 1. Juni statt am 1. Januar beginnt, könnten Sie als ersten Monat des Jahres den Monat Juni definieren. Das Attribut „Monat des Jahres“ würde dann Juni als Monat 1, Juli als Monat 2 usw. anstelle von Januar als Monat 1, Februar als Monat 2 usw. bezeichnen.

- Führen Sie eine der folgenden Aktionen aus:
 - Wenn Sie den „Generate Time Dimension“-Schritt im Enterprise Designer konfigurieren, klicken Sie auf **Kalender**.
 - Klicken Sie in der Management Console auf **System > Kalender**.
- Klicken Sie auf **Hinzufügen**.
- Geben Sie dem Kalender im Feld **Kalendernamen** einen für Sie aussagekräftigen Namen.
- Wählen Sie im Feld **Startmonat** den ersten Monat des Jahres aus.

Wenn Sie beispielsweise „Juli“ auswählen, ist Juli Monat 1, August Monat 2 usw.

Anmerkung: Wenn Sie den Startmonat ändern, ändert sich auch die Berechnung der Werte in den folgenden Feldern: DayOfYear, MonthNumber, Quarter und WeekOfYear.

5. Klicken Sie auf **Speichern**.

Ausgabe

Generate Time Dimension erstellt die folgenden Ausgabefelder.

Tabelle 2: Ausgabe von Generate Time Dimension

Feldname	Beschreibung
Date	Dies ist das Tagesdatum im Format <Tag> <Monat> <Jahr>. Der Monatsname liegt dabei in der Sprache des eingestellten Gebietsschemas des Servers vor. Wenn der Server beispielsweise mit dem Gebietsschema Englisch läuft, sieht ein Datum folgendermaßen aus: 30 Oktober 2012.
DayOfMonth	Diese Zahl stellt den Tag des Monats dar. 10 bedeutet zum Beispiel, dass der Tag der 10. Tag des Monats ist.
DayOfYear	Diese Zahl stellt den Tag des Jahres dar. 304 bedeutet zum Beispiel, dass der Tag der 304. Tag des Jahres ist.
IsLeapYear	Dieser boolesche Wert zeigt an, ob sich der Tag in einem Schaltjahr befindet. In diesem Fall ist der Wert <code>true</code> , andernfalls ist der Wert <code>false</code> .
IsWeekday	Dieser boolesche Wert zeigt an, ob der Tag ein Werktag (Montag bis Freitag) ist. In diesem Fall ist der Wert <code>true</code> , andernfalls ist der Wert <code>false</code> .
IsWeekend	Dieser boolesche Wert zeigt an, ob der Tag ein Wochenendtag (Samstag oder Sonntag) ist. In diesem Fall ist der Wert <code>true</code> , andernfalls ist der Wert <code>false</code> .

Feldname	Beschreibung
JulianDay	Dies ist eine eindeutige Nummer für den Tag. Im julianischen Tagessystem wird der 1. Januar 4713 v. Chr. als Tag 1 gezählt, und die Zahlen für alle folgenden Tage werden entsprechend erhöht. Beispielsweise bedeutet 2456231, dass dies der 2.456.231. Tag seit dem 1. Januar 4713 vor Christus ist.
JulianWeek	Dies ist eine eindeutige Nummer für die Woche des Tags. Im julianischen System wird die erste Woche des Jahres 4713 v. Chr. als Woche 1 gezählt, und die Zahlen für alle folgenden Wochen werden entsprechend erhöht. Beispielsweise bedeutet 350890, dass der Tag in der 350.890. Woche seit der ersten Woche von 4713 v. Chr. liegt.
JulianYear	Dies ist eine eindeutige Nummer für das Jahr. Im julianischen System wird das Jahr 4713 v. Chr. als Jahr 1 gezählt, und die Zahlen für alle folgenden Jahre werden entsprechend erhöht. Beispielsweise bedeutet 6725, dass der Tag im Jahr 2012 n. Chr. liegt.
MonthName	Dies ist der Name des Monats, in dem der Tag liegt, in englischer Sprache.
MonthNumber	Diese Zahl stellt den Monat des Jahres dar. 3 bedeutet zum Beispiel, dass der Tag im dritten Monat des Jahres liegt.
Quarter	Diese Zahl stellt das Quartal des Jahres dar. 1 bedeutet zum Beispiel, dass der Tag im ersten Quartal des Jahres liegt.
WeekOfYear	Diese Zahl stellt die Woche des Jahres dar. 43 bedeutet zum Beispiel, dass der Tag in der 43. Woche des Jahres liegt.
WeekdayName	Dies ist der Name des Tags in deutscher Sprache. Beispiel: Monday.
WeekdayNumber	Diese Zahl stellt den Tag der Woche dar, wobei Tag 1 Montag ist. Dienstag ist dementsprechend Tag 2, Mittwoch ist Tag 3 usw.
Year	Dies ist das Jahr, in dem der Tag liegt, nach dem gregorianischen Kalender. Beispielsweise bedeutet 2012, dass der Tag im Jahr 2012 liegt.

Query Cache

Query Cache sucht nach Daten in einem Cache basierend auf Werten in einem oder mehreren Datenflussfeldern und gibt Daten von übereinstimmenden Datensätzen im Cache zurück. Die Daten des Datensatzes im Cache werden dann dem Datensatz im Datenfluss hinzugefügt. Die Datensuche im Cache kann die Leistung im Vergleich zu einer Datensuche in einer Datenbank verbessern.

Es gibt zwei Cachetypen: globaler Cache und lokaler Cache.

Optionen für globalen Cache

Ein globaler Cache ist ein systemweiter, gemeinsam genutzter Cache, der sich im Speicher befindet. Entscheiden Sie sich für einen globalen Cache, wenn der Cache für mehrere Datenflüsse verfügbar sein soll oder wenn die Daten sich nicht oft ändern oder relativ statisch sind und der Speicherplatz nicht begrenzt ist. Ein globaler Cache ist statisch, da Sie ihn nur einmal füllen können. Der Cache kann nach seiner Erstellung nicht mehr aktualisiert werden.

Ein globaler Cache wird mithilfe des „Write To Cache“-Schrittes erstellt. Bevor Sie einen globalen Cache verwenden können, müssen Sie den Cache mit den Daten füllen, die Sie durchsuchen möchten. Erstellen Sie dazu einen Datenfluss, der den **Write to Cache**-Schritt enthält.

Name der Option	Beschreibung
Cachetyp	Wählen Sie die Option „Globaler Cache“ aus.
Cachename	Gibt den abzufragenden Cache an. Um einen Cache zu erstellen, verwenden Sie den „Write To Cache“-Schritt.
Cachefelder	In dieser Spalte sind die Felder im Cache aufgelistet. Sie können diese Feldnamen nicht ändern.
Schrittfelder	In dieser Spalte werden die im Datenfluss verwendeten Feldnamen aufgelistet. Wenn Sie einen Feldnamen ändern möchten, klicken Sie auf den Feldnamen und geben Sie einen neuen Namen ein.
Typ	In dieser Spalte ist der Datentyp jedes Datenflussfeldes aufgelistet.

Name der Option	Beschreibung
Einschließen	Aktivieren Sie das Kästchen in dieser Spalte, damit die Abfrage den Wert des Cache-Feldes zurückgibt. Deaktivieren Sie das Kästchen, wenn die Abfrage das Cache-Feld nicht zurückgeben soll.

Name der Option Beschreibung

Standardfehlerwert

Gibt den im Datenflussfeld anzuzeigenden Wert für den Fall an, dass die Abfrage fehlschlägt. In der Dropdown-Liste werden Werte angezeigt, die für den Datentyp des abgefragten Feldes gültig sind. Bei einem **Integer**-Wert wird beispielsweise **-1** als Option angezeigt.

Sie können auch selbst einen Wert in dieses Feld eingeben. In der folgenden Tabelle finden Sie eine Liste mit gültigen Standardfehlerwerten für verschiedene Datentypen.

Datentyp Gültiger Standardfehlerwert mit Datentyp (in Klammern)

Datentyp	Gültiger Standardfehlerwert mit Datentyp (in Klammern)
NULL	-1
(ganze Zahl)	1899-12-30
(Datum)	1899-12-30
(Uhrzeit)	12:00:00
(Datum/Uhrzeit)	False
	Leer

Date

Ganze Zahl



Lang



Float



Big Dezimalzahl



Double



Zeichenfolge



Zeit



Datum/Uhrzeit



Boolean



Name der Option	Beschreibung
Schlüsselfeld	Gibt das Feld im Cache an, das als Suchschlüssel verwendet wird. Wenn der Wert im Feld in der Spalte Eingabefeld mit dem Wert im Schlüsselfeld im Cache übereinstimmt, gibt die Abfrage die Daten aus diesem Datensatz im Cache zurück.
Eingabefeld	Gibt das Datenflussfeld an, dessen Wert als Schlüssel verwendet wird. Wenn der Wert in diesem Feld mit dem Wert im Schlüsselfeld im Cache übereinstimmt, gibt die Abfrage die Daten aus diesem Datensatz im Cache zurück.

Optionen für lokalen Cache

Ein lokaler Cache ist ein temporärer Cache, der nur während der Ausführung des „Query Cache“-Schrittes verwendet wird. „Query Cache“ erstellt den Cache aus der von Ihnen ausgewählten Datenbanktabelle. Dann sucht der „Query Cache“-Schritt nach Daten im Cache basierend auf Schlüsselfeldern und Suchbedingungen und gibt Daten von übereinstimmenden Datensätzen im Cache zurück. Die Daten des Datensatzes im Cache werden dann dem Datensatz im Datenfluss hinzugefügt.

Ein lokaler Cache ist dynamisch, da er während der Auftragsausführung von Query Cache erstellt wird. Sobald der Datenlesevorgang von Query Cache beendet ist, wird der Cache automatisch aus dem Speicher gelöscht. Ein lokaler Cache wird jedes Mal neu erstellt, wenn der „Query Cache“-Schritt ausgeführt wird. Verwenden Sie einen lokalen Cache, wenn dieser nur in einem Datenfluss verwendet wird oder sich die Suchtabelle oft ändert.

Optionsname	Beschreibung
Cachetyp	Gibt die Option „Lokaler Cache“ an.
Verbindung	<p>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 Datenbankverbindung ändern oder löschen müssen, klicken Sie auf Verwalten.</p> <p>Wenn Sie eine Datenbankverbindung hinzufügen oder ändern, füllen Sie diese Felder aus:</p> <p>Verbindungsname Geben Sie einen Namen für die Verbindung ein. Sie können den Namen frei wählen.</p> <p>Datenbanktreiber Wählen Sie den passenden Datenbanktyp aus.</p> <p>Verbindungsoptionen Geben Sie Host, Port, Instanz, Benutzername und Kennwort an, die für die Verbindung zur Datenbank verwendet werden sollen.</p>
Tabelle/Ansicht	Gibt die Tabelle oder Ansicht in der Datenbank an, die Sie abfragen möchten.

Optionsname	Beschreibung
Datenbankfelder	In dieser Spalte sind die Felder in der Datenbank aufgelistet. Sie können diese Feldnamen nicht ändern.
Schrittfelder	In dieser Spalte werden die im Datenfluss verwendeten Feldnamen aufgelistet. Wenn Sie einen Feldnamen ändern möchten, klicken Sie auf den Feldnamen und geben Sie den neuen Namen ein.
Typ	In dieser Spalte ist der Datentyp jedes Datenflussfeldes aufgelistet.
Einschließen	Aktivieren Sie das Kästchen in dieser Spalte, damit die Abfrage den Wert des Cache-Feldes zurückgibt. Deaktivieren Sie das Kästchen, wenn die Abfrage das Cache-Feld nicht zurückgeben soll.

Optionsname Beschreibung

Standardfehlerwert
 Gibt den im Datenflussfeld anzuzeigenden Wert für den Fall an, dass die Abfrage fehlschlägt. In der Dropdown-Liste werden Werte angezeigt, die für den Datentyp des abgefragten Feldes gültig sind. Bei einem **Integer**-Wert wird beispielsweise **-1** als Option angezeigt. Sie können auch selbst einen Wert in dieses Feld eingeben. In der folgenden Tabelle finden Sie eine Liste mit gültigen Standardfehlerwerten für verschiedene Datentypen.

Datentyp Gültiger Standardfehlerwert mit Datentyp (in Klammern)

Datentyp	Gültiger Standardfehlerwert mit Datentyp (in Klammern)
NULL	-1
(ganze Zahl)	1899-12-30 12:00:00
(Datum)	1899-12-30 12:00:00
(DatumUhrzeit)	12:00:00
False	False
Leer	Leer
Date	
Ganze Zahl	✓
Lang	✓
Float	✓
Big Dezimalzahl	✓
Double	✓
Zeichenfolge	✓ ✓ ✓ ✓ ✓ ✓ ✓
Zeit	✓
DatumUhrzeit	✓
Boolean	✓

Optionsname	Beschreibung
Schlüsselfeld	Gibt das Feld in der Datenbank an, das als Suchschlüssel verwendet wird. Wenn der Wert im Feld in der Spalte Eingabefeld mit dem Wert im Schlüsselfeld in der Datenbank übereinstimmt, gibt die Abfrage die Daten aus diesem Datensatz in der Datenbank zurück.
Eingabefeld	Gibt das Datenflussfeld an, dessen Wert als Schlüssel verwendet wird. Wenn der Wert in diesem Feld mit dem Wert im Schlüsselfeld in der Datenbank übereinstimmt, gibt die Abfrage die Daten aus diesem Datensatz in der Datenbank zurück.

Optionen für erweiterten Cache

Ein erweiterter Cache ist ein temporärer Cache ähnlich dem lokalen Cache. Er wird während der Ausführung des „Query Cache“-Schrittes verwendet. Der Cache wird basierend auf der SQL-Abfrage erstellt und liest die Daten aus den in der Abfrage aufgeführten Tabellen ein. Daten werden dann entsprechend den in der WHERE-Klausel aufgeführten Suchschlüsseln im Cache gesucht. Anschließend werden Daten von übereinstimmenden Datensätzen im Cache zurückgegeben. Die Daten des Datensatzes im Cache werden dann dem Datensatz im Datenfluss hinzugefügt.

Ein erweiterter Cache ist dynamisch, da er während der Auftragsausführung von Query Cache erstellt wird. Sobald der Datenlesevorgang von Query Cache beendet ist, wird der Cache automatisch aus dem Speicher gelöscht. Ein erweiterter Cache wird jedes Mal neu erstellt, wenn der „Query Cache“-Schritt ausgeführt wird. Entscheiden Sie sich für einen erweiterten Cache, wenn Daten aus mehreren Tabellen gelesen werden sollen und eine komplexe Abfrage zur Erstellung des Cache ausgeführt werden muss.

Optionsname	Beschreibung
Cachetyp	Gibt die Option „Erweiterter Cache“ an.
Verbindung	<p>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 Datenbankverbindung ändern oder löschen müssen, klicken Sie auf Verwalten.</p> <p>Wenn Sie eine Datenbankverbindung hinzufügen oder ändern, füllen Sie diese Felder aus:</p> <p>Verbindungsname Geben Sie einen Namen für die Verbindung ein. Sie können den Namen frei wählen.</p> <p>Datenbanktreiber Wählen Sie den passenden Datenbanktyp aus.</p> <p>Verbindungsoptionen Geben Sie Host, Port, Instanz, Benutzername und Kennwort an, die für die Verbindung zur Datenbank verwendet werden sollen.</p>

Optionsname	Beschreibung
Abfrage	Stellt die SQL-Abfrage zur Verfügung, die Daten aus der Datenbank liest. Die Abfrage kann Daten aus mehreren Tabellen lesen.
Wo	Dieser Text wird als WHERE-Klausel zur Suche im Cache verwendet, der auf Basis der Abfrage erstellt wurde. Benutzer können das Eingabefeld in der Abfrage über einen \$-Operator als Präfix angeben. Bei <code>_id = \$_inputId</code> ist beispielsweise „_inputId“ das Eingabefeld und „_id“ die Suchspalte im Cache.
Felder abrufen	Hiermit wird das Grid mit den Felder ausgefüllt, die ausgewählt wurden, durch die SQL-Abfrage in den Cache geschrieben zu werden.
Datenbankfelder	In dieser Spalte sind die Felder aufgelistet, die aus der Datenbank abgerufen werden. Sie können diese Feldnamen nicht ändern.
Schrittfelder	In dieser Spalte werden die im Datenfluss verwendeten Feldnamen aufgelistet. Wenn Sie einen Feldnamen ändern möchten, klicken Sie auf den Feldnamen und geben Sie den neuen Namen ein.
Typ	In dieser Spalte ist der Datentyp jedes Datenflussfeldes aufgelistet.

Optionsname Beschreibung

Standardfehlerwert
 Gibt den im Datenflussfeld anzuzeigenden Wert für den Fall an, dass die Abfrage fehlschlägt. In der Dropdown-Liste werden Werte angezeigt, die für den Datentyp des abgefragten Feldes gültig sind. Bei einem **Integer**-Wert wird beispielsweise **-1** als Option angezeigt. Sie können auch selbst einen Wert in dieses Feld eingeben. In der folgenden Tabelle finden Sie eine Liste mit gültigen Standardfehlerwerten für verschiedene Datentypen.

Datentyp Gültiger Standardfehlerwert mit Datentyp (in Klammern)

Datentyp	Gültiger Standardfehlerwert mit Datentyp (in Klammern)
NULL	-1
(ganze Zahl)	1899-12-30 12:00:00
(Datum)	1899-12-30 12:00:00
(DatumUhrzeit)	12:00:00
False	Leer
Leer	
Date	
Ganze Zahl	✓
Lang	✓
Float	✓
Big Dezimalzahl	✓
Double	✓
Zeichenfolge	✓ ✓ ✓ ✓ ✓ ✓ ✓
Zeit	✓
DatumUhrzeit	✓
Boolean	✓

Registerkarte „Laufzeit“

Die auf der Registerkarte „Laufzeit“ verfügbaren Optionen sind für globale, lokale und erweiterte Caches gleich.

Name der Option	Beschreibung
Vergleichsoptionen	<p>Gibt an, was getan werden soll, wenn mehr als ein Datensatz im Cache als Ergebnis der Abfrage vorhanden ist. Zur Auswahl stehen:</p> <p>Alle Übereinstimmungen zurückgeben Gibt Daten aus allen Datensätzen im Cache zurück, bei denen ein Wert im Schlüsselfeld oder in den Schlüsselfeldern übereinstimmt.</p> <p>Ersten übereinstimmenden Datensatz zurückgeben Gibt nur Daten aus dem ersten Datensatz im Cache zurück, bei dem ein Wert im Schlüsselfeld oder in den Schlüsselfeldern übereinstimmt.</p> <p>Letzten übereinstimmenden Datensatz zurückgeben Gibt nur Daten aus dem letzten Datensatz im Cache zurück, bei dem ein Wert im Schlüsselfeld oder in den Schlüsselfeldern übereinstimmt.</p>
Schrittoptionen	<p>In diesem Abschnitt sind die Datenflussoptionen aufgeführt, die in der SQL-Abfrage dieses Schrittes verwendet werden. Zudem können Sie hier einen Standardwert für diese Optionen angeben. In der Spalte Name sind die Optionen aufgelistet. In der entsprechenden Spalte Wert können Sie die Standardwerte eingeben.</p> <p>Anmerkung: Der hier angegebene Standardwert wird auch im Abschnitt Datenflussoptionen zu Schritten zuordnen des Dialogfeldes Datenflussoptionen angezeigt. Mithilfe der Dialogbox können Sie außerdem den Standardwert ändern. Wenn für Standardwerte einer Option unterschiedliche Werte über Schrittoptionen, Datenflussoptionen und Job Executor angegeben sind, gilt folgende Rangfolge: im Job Executor angegebener Wert > unter Datenflussoptionen definierter Wert > unter Schrittoptionen eingegebener Wert.</p>

Query DB

Mit dem „Query DB“-Schritt können Sie Felder als Parameter in einer Datenbankabfrage verwenden und die Ergebnisse der Abfrage als neue Felder im Datenfluss zurückgeben lassen.

Anmerkung: Wenn Sie eine Geodaten-Datenbanken abfragen möchten, verwenden Sie „Query Spatial Data“ anstelle von „Query DB“.

Registerkarte „Allgemein“

Option	Beschreibung
Verbindung	<p>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 Datenbankverbindung ändern oder löschen müssen, klicken Sie auf Verwalten.</p> <p>Wenn Sie eine Datenbankverbindung hinzufügen oder ändern, füllen Sie diese Felder aus:</p> <p>Verbindungsname Geben Sie einen Namen für die Verbindung ein. Sie können den Namen frei wählen.</p> <p>Datenbanktreiber Wählen Sie den passenden Datenbanktyp aus.</p> <p>Verbindungsoptionen Geben Sie Host, Port, Instanz, Benutzername und Kennwort an, die für die Verbindung zur Datenbank verwendet werden sollen.</p>
Tabelle/Ansicht	Gibt die Tabelle oder Ansicht in der Datenbank an, die Sie abfragen möchten.

Option	Beschreibung
Wo	<p>Wenn Sie eine <code>WHERE</code>-Anweisung verwenden möchten, geben Sie sie hier ein. Beachten Sie, dass Sie nicht das Wort <code>WHERE</code> selbst in der Anweisung verwenden sollten. Der Zweck einer <code>WHERE</code>-Anweisung besteht darin, nur Daten aus Datensätzen zurückzugeben, die die von Ihnen angegebenen Bedingungen erfüllen. Verwenden Sie die folgende Syntax, um einen Wert in einem Datenflussfeld festzulegen:</p> <pre data-bbox="553 562 769 594">\${<field name>}</pre> <p>Dabei steht <code><field name></code> für den Namen eines Feldes im Datenfluss.</p> <p>Beispiel:</p> <pre data-bbox="553 709 1094 741">account_number=\${customer_key}</pre> <p>In diesem Beispiel würde die Abfrage Daten aus Datensätzen zurückgeben, in denen der Wert in der Tabellenspalte <code>account_number</code> mit dem Wert im Datenflussfeld <code>customer_key</code> übereinstimmt.</p> <p>Anmerkung: Wenn Sie eine Datenbank abfragen, bei der zwischen Groß- und Kleinschreibung unterschieden wird, achten Sie darauf, den Feldnamen im selben Format einzugeben, das in der Datenbanktabelle verwendet wird. Schließen Sie also den Feldnamen in Anführungszeichen (") ein, wenn der Feldname beim Erstellen der Tabelle auch in Anführungszeichen gesetzt wurde.</p> <p>Klicken Sie auf Vorschau, um eine Vorschau der Daten (der ersten 50 Datensätze) basierend auf den von Ihnen definierten Kriterien anzuzeigen.</p> <p>Anmerkung: Das Vorschau-Feature in „Query DB“ funktioniert nicht, wenn Sie in der <code>WHERE</code>-Anweisung ein Datenflussfeld verwenden. In dem Fall können Sie eine Vorschau des Ergebnisses über das Tool zur Datenflussüberprüfung im Enterprise Designer anzeigen.</p>
Datensätze ohne Ergebnisse zurückgeben	<p>Aktivieren Sie dieses Kästchen, wenn Sie möchten, dass Datensätze, die bei der Abfrage keine Ergebnisse liefern, trotzdem von „Query DB“ zurückgegeben werden sollen. Wenn Sie dieses Kästchen deaktivieren, wird der Datensatz nicht zurückgegeben. Wir empfehlen, diese Option aktiviert zu lassen.</p>
Einschließen	<p>Wählen Sie in der Feldtabelle die Felder aus, die Sie einschließen möchten, indem Sie das Kästchen Einschließen neben dem Feld aktivieren.</p>

Registerkarte „Sortieren“

Wenn Sie die Datensätze nach dem Wert eines Feldes sortieren möchten, geben Sie die Felder an, nach denen sortiert werden soll.

Parametrisieren von „Query DB“ zur Laufzeit

Sie können den „Query DB“-Schritt konfigurieren, so dass diese Werte in der `WHERE`-Klausel zur Laufzeit angegeben werden. Dies ist in Fällen nützlich, in denen Sie den Spaltennamen in der `WHERE`-Klausel unter Verwendung von Datenflussoptionen konfigurierbar machen möchten.

1. Öffnen Sie den Datenfluss im Enterprise Designer.
2. Konfigurieren Sie die Namensfelder **Verbindung** und **Tabelle/Ansicht**, sodass diese auf die Datenbank verweisen, die Sie abfragen möchten.
3. Geben Sie im Feld **Where** die `WHERE`-Anweisung ein. Verwenden Sie dabei für Werte, die Sie parametrisieren möchten, das folgende Format: `${parameter}`.

Beispiel:

```
${COL}=${EmployeeID}
```

Hier repräsentiert „COL“ eine Datenflussoption, die zur Laufzeit mit dem Spaltennamen für die Tabelle ausgefüllt wird.

4. Schließen Sie das Fenster mit den Optionen für „Query DB“.
5. Klicken Sie in der Symbolleiste auf das Symbol „Datenflussoptionen“ oder klicken Sie auf **Bearbeiten > Datenflussoptionen**. Das Fenster **Datenflussoptionen** wird angezeigt.
6. Klicken Sie auf **Hinzufügen**. Das Fenster **Datenflussoptionen definieren** wird angezeigt.
7. Wählen Sie den „Query DB“-Schritt aus.
8. Geben Sie **Optionsname** und **Optionsbeschriftung** an.

Der Wert im Feld **Optionsname** sollte derselbe Wert sein, der im Format `${parameter}` in der `WHERE`-Klausel eingegeben wurde. Im Feld **Optionsbeschriftung** können Sie eine andere Beschriftung angeben oder den **Optionsnamen** als Beschriftung beibehalten.

Beispiel: COL

9. Geben Sie den **Standardwert** an. Beispiel: EmpID
10. Klicken Sie auf **OK**.

Diese Schritte bilden den tatsächlichen Namen der Datenbankspalte, also „EmpID“, auf den Namen der Laufzeitoption „COL“ ab. Der Name der Datenbankspalte muss mit den korrekten Anführungszeichen der jeweiligen Datenbank versehen sein.

Query NoSQL DB

Der Schritt **Query NoSQL DB** ermöglicht Suchvorgänge nach erforderlichen Daten in einer NoSQL-Datenbank. Der Schritt unterstützt MongoDB-Datenbanken.

Registerkarte „Allgemein“

Feldname	Beschreibung
Verbindung	<p>Wählen Sie die gewünschte Datenbankverbindung aus.</p> <p>Im Enterprise Designer unter Tools > NoSQL-Datenbankverwaltungsverwaltung definierte Verbindungen stehen zur Verfügung.</p> <p>Wenn Sie eine neue Datenbankverbindung herstellen oder eine vorhandene Datenbankverbindung ändern oder löschen möchten, klicken Sie auf Verwalten. Das Fenster NoSQL-Datenbankverwaltungsverwaltung wird mit folgenden Fenstern geöffnet:</p> <p>Verbindungsname Geben Sie einen Namen für die gewünschte Verbindung ein.</p> <p>NoSQL-Datenbank Wählen Sie die NoSQL-Datenbank aus, für die Sie die Verbindung erstellen möchten.</p> <p>Benutzername Geben Sie den Benutzernamen zum Herstellen der Verbindung mit der Datenbank ein.</p> <p>Kennwort Geben Sie das Kennwort zum Herstellen der Verbindung mit der Datenbank ein.</p> <p>Hostname Geben Sie den Hostnamen an, auf dem die Datenbank läuft.</p> <p>Port Geben Sie den Port zum Herstellen einer Verbindung mit der Datenbank an.</p> <p>Datenbank Geben Sie die Datenbank an, aus der Daten abgerufen werden sollen.</p>
Tabelle/Ansicht	<p>Gibt die Sammlung oder Ansicht in der Datenbank an, die Sie abfragen möchten.</p> <p>Anmerkung: In einer MongoDB-Datenbank wird eine Tabelle/Ansicht als <i>Sammlung</i> bezeichnet.</p>

Feldname	Beschreibung
Schemadatei	<p>Klicken Sie auf die Schaltfläche „Durchsuchen“ (...), um eine JSON-Schemadatei auszuwählen. Diese Datei ist optional. Die Felder auf der Registerkarte „Felder“ können entweder über die Schemadatei oder die Datenbanktabelle/-ansicht erneut generiert werden.</p> <p>Um den ausgewählten Dateipfad zu löschen, klicken Sie auf Löschen.</p> <p>Anmerkung: Die Felder werden immer über die Schemadatei generiert, wenn eine ausgewählt ist.</p>
Wo	<p>Geben Sie die <code>where</code>-Klausel ein, um die Kriterien für den Suchvorgang zu definieren.</p> <p>Eine Liste der in der <code>WHERE</code>-Klausel unterstützten Operatoren finden Sie unter http://docs.mongodb.org/manual/reference/operator/query/.</p>
Vorschau	<p>Dies zeigt die Datensätze aus der ausgewählte Tabelle oder Ansicht an.</p> <p>Anmerkung: Wenn Sie auf Vorschau klicken, werden die ersten 50 Datensätze aus der ausgewählten Datenbank abgerufen. Die im Feld Where angegebenen Filterkriterien werden dabei nicht angewendet.</p>
Alle erweitern	Dies erweitert die Elemente im Vorschaubaum.
Alle reduzieren	Dies reduziert die Elemente im Vorschaubaum.

Registerkarte „Felder“

Die Registerkarte „Felder“ ermöglicht Ihnen die Auswahl der Daten, die Sie dem nächsten Schritt übergeben möchten. Weitere Informationen finden Sie unter [Definieren von Feldern – Query NoSQL DB](#) auf Seite 141.

Definieren von Feldern – Query NoSQL DB

Auf der Registerkarte **Felder** werden die Felder und ihre Typen angezeigt, wie sie in der Schemadatei der NoSQL-Datenbank definiert wurden.

1. Klicken Sie auf der Registerkarte **Felder** auf die Option **Erneut generieren**.
Daraufhin werden die aggregierten Daten basierend auf den ersten 50 Datensätzen generiert. Die Daten werden im folgenden Format angezeigt: `Fieldname (datatype)`.

Anmerkung: Beim Durchsuchen der Schemadatei werden die Felder unter Verwendung der Schemadatei generiert. Die Tabelle/Ansicht wird umgangen. Um die Schemadatei zurückzusetzen, klicken Sie auf **Löschen**.

2. Um den Namen und Typ eines Feldes zu ändern, markieren Sie das Feld und klicken Sie auf **Ändern**.
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 „Zeichenfolge“ beibehalten, wenn Sie nicht vorhaben, mathematische Operationen auf die Daten anzuwenden. Wenn Sie allerdings diese 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.

Der Schritt unterstützt die folgenden Datentypen:

double	Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit doppelter Genauigkeit zwischen 2^{-1074} und $(2-2^{-52}) \times 2^{1023}$ 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 $(2-2^{-23}) \times 2^{127}$ 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. Sie können auch zusätzliche Felder hinzufügen, die in der Tabelle oder Schemadatei nicht vorhanden sind. Klicken Sie auf **Hinzufügen**, um ein neues Feld hinzuzufügen. Um ein Feld zu entfernen, klicken Sie auf **Entfernen**.

Anmerkung: Sie können nur ein neues Feld des Typs Liste hinzufügen.

6. Klicken Sie auf **OK**.

Konfigurieren von Datenflussoptionen – Query NoSQL DB

In dieser Prozedur wird beschrieben, wie ein Datenfluss für die Unterstützung von Laufzeitoptionen für **Query NoSQL DB** konfiguriert wird.

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. Erweitern Sie den Schritt **Query NoSQL DB**.
6. Die nachfolgenden Datenflussoptionen werden für die Abfrage einer Mongo DB-Datenbank verfügbar gemacht:
 - Verbindung
 - Tabelle

Der ausgewählte Optionsname **Query NoSQL DB** wird in den Feldern **Optionsname** und **Optionsbeschriftung** angezeigt. Dies ist der Optionsname, der zur Laufzeit angegeben werden muss, um diese Option festzulegen.

7. Geben Sie eine Beschreibung der Option in das Feld **Beschreibung** ein.
8. Wählen Sie im Feld **Ziel** die Option **Ausgewählte(r) Schritt/e** aus.
9. 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.
10. 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](#).

11. Klicken Sie auf **OK**.
12. Fügen Sie auf Wunsch weitere Optionen hinzu.
13. Klicken Sie im Dialogfeld **Datenflussoptionen** auf **OK**, sobald Sie damit fertig sind.
14. Speichern Sie den Datenfluss und machen Sie ihn verfügbar.

Read from DB

Der Schritt **Read from DB** liest Daten aus einer Datenbanktabelle oder -ansicht als Eingabe in einen Datenfluss ein. Der Schritt steht für Aufträge, Dienste und Unterflüsse, jedoch nicht für Prozessflüsse zur Verfügung.

Registerkarte „Allgemein“

Feldname	Beschreibung
Verbindung	<p>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 Datenbankverbindung ändern oder löschen müssen, klicken Sie auf Verwalten.</p> <p>Wenn Sie eine Datenbankverbindung hinzufügen oder ändern, füllen Sie diese Felder aus:</p> <p>Verbindungsname Geben Sie einen Namen für die Verbindung ein. Sie können den Namen frei wählen.</p> <p>Datenbanktreiber Wählen Sie den passenden Datenbanktyp aus.</p> <p>Verbindungsoptionen Geben Sie Host, Port, Instanz, Benutzername und Kennwort an, die für die Verbindung zur Datenbank verwendet werden sollen.</p>
SQL	<p>Geben Sie die SQL-Abfrage ein, um die Datensätze anzugeben, die beim Ausführen des Datenflusses aus der Datenquelle gelesen werden müssen. Sie können die SQL-Abfrage von Hand in dieses Feld eingeben. Alternativ können Sie den Visual Query Builder verwenden, um die Abfrage durch Anklicken von SQL erstellen... zu erstellen.</p> <p>Die SQL-Abfrage kann Variablen anstelle der tatsächlichen Spaltennamen enthalten. Über die Verwendung von Variablen können Sie die Abfrage zur Laufzeit anpassen. Weitere Informationen finden Sie unter Abfragevariablen auf Seite 149.</p>
SQL erstellen	<p>Erstellen Sie eine komplexe Abfrage, indem Sie mehrere Spalten auswählen und Joins sowie verschachtelte Abfragen durch Klicken auf SQL erstellen erstellen. Der Visual Query Builder wird geöffnet. Weitere Informationen finden Sie unter Visual Query Builder auf Seite 146.</p> <p>Anmerkung: Eine mit Visual Query Builder erstellte Abfrage wird im Feld SQL mit vollständig qualifizierten Namen von Spalten und Tabellen angezeigt.</p>
Felder erneut generieren	<p>Zur Anzeige des Schemas der Daten, die von der Abfrage abgerufen werden sollen, klicken Sie auf Felder erneut generieren.</p> <p>Wenn Sie eine vorhandene Abfrage bearbeiten, klicken Sie auf Felder erneut generieren, um das geänderte Schema abzurufen.</p> <p>Anmerkung: Wenn Sie auf Felder erneut generieren klicken, werden die Entitätsnamen in der SQL-Abfrage beibehalten und nicht durch ihre vollständig qualifizierten Namen ersetzt.</p>

Feldname	Beschreibung
Vorschau	Um ein Beispiel der von der SQL-Abfrage abgerufenen Datensätze anzuzeigen, klicken Sie auf Vorschau .

Anmerkung: Über den Schritt **Read from DB** können Sie den Typ eines Eingabefeldes ändern.

Anmerkung: Der **Read from DB**-Schritt liest alle Werte des Datentyps `date` als `String`-Werte ein. Dieses Verhalten beruht auf dem von Spectrum verwendeten Standardtreiber, dem *JTDS-Treiber*. Um alle `date`-Datentypenwerte unverändert zu verarbeiten, verwenden Sie den JDBC-Treiber von Microsoft.

Registerkarte „Laufzeit“

Feldname	Beschreibung
Abrufgröße	<p>Wählen Sie diese Option aus, um die Anzahl der Datensätze anzugeben, die jeweils aus der Datenbanktabelle gelesen werden sollen. Wenn die Abrufgröße beispielsweise 100 beträgt und die Gesamtzahl der zu lesenden Datensätze 1.000 beträgt, würden 10 Abrufe benötigt werden, um alle Datensätze zu lesen.</p> <p>Das Einstellen einer optimalen Abrufgröße kann die Leistung deutlich verbessern.</p> <p>Anmerkung: Sie können eine optimale Abrufgröße für Ihre Umgebung berechnen, in dem Sie die Ausführungszeiten zwischen einem „Read from DB“-Schritt und einem „Write to Null“-Schritt in Tests ermitteln. Weitere Informationen finden Sie unter Ermitteln einer idealen Abrufgröße auf Seite 329.</p>

Schrittoptionen	<p>In diesem Abschnitt sind die Datenflussoptionen aufgeführt, die in der SQL-Abfrage dieses Schrittes verwendet werden. Zudem können Sie hier einen Standardwert für diese Optionen angeben. In der Spalte Name sind die Optionen aufgelistet. In der entsprechenden Spalte Wert können Sie die Standardwerte eingeben.</p> <p>Anmerkung: Der hier angegebene Standardwert wird auch im Abschnitt Datenflussoptionen zu Schritten zuordnen des Dialogfeldes Datenflussoptionen angezeigt. Mithilfe der Dialogbox können Sie außerdem den Standardwert ändern. Wenn für Standardwerte einer Option unterschiedliche Werte über Schrittoptionen, Datenflussoptionen und Job Executor angegeben sind, gilt folgende Rangfolge: im Job Executor angegebener Wert > unter Datenflussoptionen definierter Wert > unter Schrittoptionen eingegebener Wert.</p>
-----------------	---

Visual Query Builder

Visual Query Builder bietet eine Benutzeroberfläche zum Erstellen komplexer SQL-Abfragen im „Read from DB“-Schritt. Um mit Visual Query Builder zu arbeiten, benötigen Sie Grundkenntnisse über SQL-Konzepte.

Um auf Visual Query Builder zuzugreifen, klicken Sie in „Read from DB“ auf die Schaltfläche **SQL erstellen**.

Das Hauptfenster von Query Builder ist in folgende Bereiche unterteilt:

- Der **Bereich zum Erstellen von Abfragen** ist der Hauptbereich, in dem die Abfrage visuell dargestellt wird. In diesem Bereich können Sie zudem Quelldatenbankobjekte definieren, Links zwischen ihnen definieren und Eigenschaften von Tabellen und Links konfigurieren.
- Unter dem Bereich zum Erstellen von Abfragen befindet sich der **Bereich Spalten**. Hier können Sie alle erforderlichen Operationen mit Abfrageausgabespalten und -ausdrücken durchführen. Hier können Sie Aliasse für Felder definieren sowie Kriterien definieren, sortieren und gruppieren.
- Über die Seitensteuerung über dem Bereich zum Erstellen von Abfragen können Sie zwischen der Hauptabfrage und den Unterabfragen wechseln.

Hinzufügen von Objekten zu einer Abfrage

Um ein Objekt zu einer Abfrage hinzuzufügen, verwenden Sie die Struktur **Datenbankobjekte**. Die Objekte innerhalb der Struktur sind nach Datenbank, Schema und Typ gruppiert. Ziehen Sie das hinzuzufügende Objekt zur Abfrage und legen Sie es im Bereich zum Erstellen von Abfragen ab. Alternativ können Sie auf das Objekt auch doppelklicken, um es zur Abfrage hinzuzufügen.

Festlegen von Objektalassen

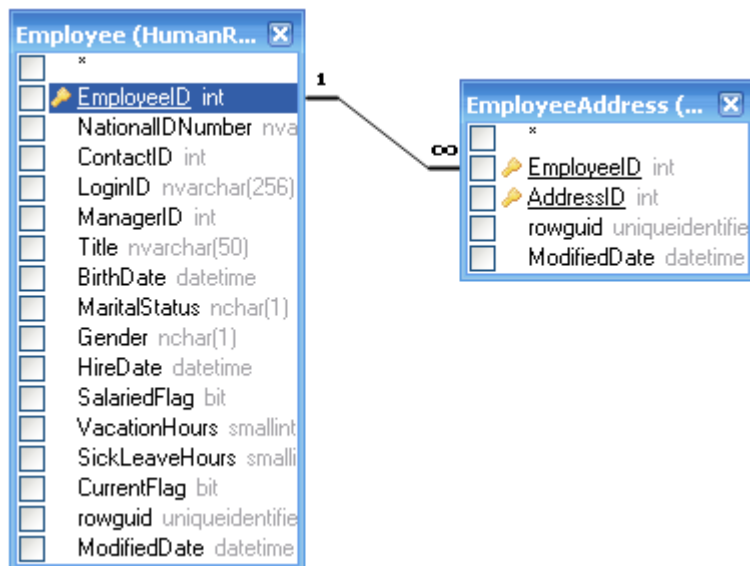
Um für ein Objekt oder eine abgeleitete Tabelle in der Abfrage ein Alias festzulegen, doppelklicken Sie auf das Objekt und wählen Sie **Eigenschaften** aus. Das Dialogfeld **Datenquelleneigenschaften** wird angezeigt. Möglicherweise sind andere serverspezifische Datenquellenoptionen enthalten, die Eigenschaft „Alias“ ist jedoch bei allen Datenbankservern dieselbe.

Zusammenführen von Tabellen

Wenn zwei Objekte, die von einer Fremdschlüsselbeziehung referenziert werden, zur Abfrage hinzugefügt werden, werden sie automatisch über INNER JOIN zusammengeführt. Bei Servern, die die JOIN-Klausel nicht unterstützen, fügt der Query Builder eine Bedingung zum WHERE-Teil der Abfrage hinzu.

Um zwei Objekte manuell zusammenzuführen, wählen Sie das Feld aus, über das Sie das Objekt mit einem anderen verlinken möchten, und ziehen Sie es zu dem entsprechenden Feld des anderen Objekts. Nachdem Sie es hinübergezogen haben, erscheint eine Linie zwischen den verlinkten

Feldern. An den Enden eines Links werden Schlüsselssymbole für die Kardinalität platziert, wenn in der Datenbank eine entsprechende Beziehung vorhanden ist.



Um einen Link zwischen Objekten zu entfernen, doppelklicken Sie auf die Linklinie und wählen Sie **Entfernen** aus.

Um den Join-Typ zu ändern, doppelklicken Sie auf die Linklinie.

Auswählen von Ausgabefeldern

Um ein Feld zu einer Liste von Abfrageausgabefeldern hinzuzufügen, aktivieren Sie im Bereich der **Abfrageerstellung** in der Liste der Datenquellenfelder das Kästchen links vom Feldnamen. Um alle Felder des Objekts einzuschließen, aktivieren Sie das Kästchen links vom Sternchenelement in der Liste der Datenquellenfelder. Sie können alternativ auch die Felder vom Bereich der **Abfrageerstellung** zum Bereich **Spalten** ziehen.

Wenn Sie kein Feld aus den Abfragedatenquellen auswählen, wird ein Sternchenelement zur Auswahlliste der resultierenden Abfrage hinzugefügt („Select * From ...“). Dies geschieht, da eine SELECT-Abfrage ohne Spalten bei den meisten Datenbankservern einen Fehler verursachen würde. Das Sternchenelement wird aus der Abfrage entfernt, wenn Sie ein beliebiges Feld auswählen oder einen beliebigen Ausgabeausdruck zur Abfrage hinzufügen.

Tipp: Eine andere Möglichkeit, ein Feld hinzuzufügen, besteht darin, einen Feldnamen aus der Dropdown-Liste der Spalte „Ausdruck“ im Bereich **Spalten** auszuwählen. Sie können auch einen beliebigen gültigen Ausdruck in der Spalte **Ausdruck** im Bereich **Spalten** eingeben. Um im Bereich **Spalten** eine leere Zeile einzufügen, drücken Sie die Taste Alt+Einfügen.

Um ein Feld aus dem Bereich **Spalten** zu entfernen, deaktivieren Sie das Kästchen links vom Feldnamen im Bereich **Abfrageerstellung** oder drücken Sie die Taste Alt+Entfernen im Bereich **Spalten**.

Um eine Zeile nach oben zu verschieben, drücken Sie die Taste Alt+Nach oben. Um eine Zeile nach unten zu verschieben, drücken Sie die Taste Alt+Nach unten.

Um einen Ausdruck aus der SELECT-Liste der Abfrage zu entfernen, deaktivieren Sie das Kästchen in der Spalte **Ausgabe**.

Um ein Alias für einen Ausdruck festzulegen, geben Sie das Alias in der Spalte **Alias** ein. Aliasse sind im resultierenden Dataset die Spaltenüberschriften.

Sortieren eines Datasets

Um das resultierende Dataset zu sortieren, verwenden Sie die Spalten „Sortierungstyp“ und „Sortierreihenfolge“ im Bereich **Spalten**. Über die Spalte „Sortierungstyp“ können Sie in aufsteigender oder absteigender Reihenfolge sortieren. In der Spalte „Sortierreihenfolge“ können Sie die Reihenfolge festlegen, nach der die Felder sortiert werden, wenn nach mehr als einem Feld sortiert wird.

Um die Sortierung nach einem Feld zu deaktivieren, löschen Sie den Inhalt der Spalte „Sortierungstyp“ für das Feld.

Definieren von Kriterien

Um Kriterien zu definieren, verwenden Sie die Spalte **Kriterien** und die Spalten **Oder** im Bereich **Spalten**. Wenn Sie in diesen Spalten Bedingungen eingeben, lassen Sie den eigentlichen Ausdruck weg. Sie möchten beispielsweise folgende Kriterien in Ihre Abfrage aufnehmen:

```
WHERE (Field1 >= 10) AND (Field1 <= 20)
```

Geben Sie Folgendes in die Zelle „Kriterien“ für den Ausdruck „Field1“ ein:

```
>= 10 AND <= 20
```

In den Spalten **Oder** platzierte Kriterien werden unter Verwendung des AND-Operators nach Spalten gruppiert und dann in der WHERE-Klausel (oder HAVING-Klausel) unter Verwendung des OR-Operators verkettet. Die unten angezeigten Kriterien ergeben beispielsweise die darunter angezeigte SQL-Anweisung. Beachten Sie bitte, dass die Kriterien für „Field1“ sowohl in der Spalte „Kriterien“ als auch in der Spalte „Oder“ platziert sind.

Output	Expression	Aggregate	Alias	Sort Type	Sort Order	Grouping	Criteria	Or...	Or...
<input checked="" type="checkbox"/>	Field1					<input type="checkbox"/>	= 10	= 10	
<input checked="" type="checkbox"/>	Field2					<input type="checkbox"/>	< 0	> 10	
<input type="checkbox"/>						<input type="checkbox"/>			

```
WHERE (Field1= 10) AND ((Field2 < 0) OR (Field2 > 10))
```

Einige Ausdrücke wie beispielsweise die EXISTS-Klausel können vom Typ „boolean“ sein. In diesem Fall sollten Sie bei solch einem Ausdruck in der Spalte „Kriterien“ die Zeichenfolge „= True“ eingeben. Wenn Sie einen NOT-Operator vor den Ausdruck setzen möchten, sollten Sie „= False“ eingeben.

Gruppieren von Ausgabefeldern

Um eine Abfrage mit Gruppierung zu erstellen, markieren Sie die zu gruppierenden Ausdrücke über das Kästchen **Gruppierung**.

Bei einer Abfrage mit Gruppierung sind in der SELECT-Liste nur Gruppierungs- oder Aggregierungsausdrücke zulässig. Daher können Sie bei der Abfrageerstellung das Kästchen „Ausgabe“ nur für Gruppierungs- und Aggregierungsausdrücke aktivieren. Wenn Sie versuchen, dieses Kästchen für eine Spalte ohne eingestellte Gruppierungs- oder Aggregierungsfunktion zu aktivieren, wird automatisch ein Kästchen für Gruppierung aktiviert, um die Gültigkeit der resultierenden SQL-Abfrage zu gewährleisten.

Wenn im Bereich **Spalten** Spalten enthalten sind, bei denen das Kästchen **Gruppierung** aktiviert ist, erscheint eine neue Spalte mit Namen **Kriterien** im Grid. Diese Spalte wendet Kriterien auf Ausdruckgruppen oder ihre Werte an.

Sie verfügen beispielsweise in Ihrer Abfrage über eine Spalte „Menge“ mit der Aggregierungsfunktion „Avg“ und geben in der Spalte „Kriterien“ > 10 ein. Bei eingestelltem Wert „Für Gruppen“ in der Spalte „Kriterien“ enthält die resultierende Abfrage nur Gruppen mit einer mittleren Menge von größer als 10 und Ihre Abfrage verfügt in einer HAVING-Klausel über die Bedingung „Avg(Menge) > 10 “. Bei eingestelltem Wert „Für Werte“ in der Spalte „Kriterien“ berechnet die Abfrage die Aggregierungsfunktion „Average“ nur für Datensätze, deren Wert für „Menge“ größer 10 ist, und Ihre Abfrage verfügt über eine WHERE-Klausel mit der Bedingung „Menge > 10 “.

Definieren von SQL-Abfrageeigenschaften

Sie können Optionen spezifisch für Ihren Datenbankserver definieren, indem Sie das Kontextmenü des Bereichs zur **Abfrageerstellung** verwenden.

Abfragevariablen

Sie können im **Read from DB**-Schritt beim Definieren der auszuführenden Abfrage Variablen anstelle von tatsächlichen Spaltennamen verwenden. Wenn Sie Variablen in der Abfrage verwenden, können Sie die Abfragebedingungen zur Laufzeit (über die **Datenflussoptionen**) oder über den **Job Executor** anpassen.

Sie können aber auch auf der Registerkarte **Laufzeit** Schrittoptionswerte eingeben und unter Verwendung der Schaltfläche **Felder erneut generieren** bzw. **Vorschau** das Schema und Beispieldatensätze, die von der Abfrage abgerufen werden, anzeigen.

Ein Variable wird im Format $\# \{Variable\}$ definiert und entweder in der `select`- oder der `where`-Klausel einer SQL-Abfrage eingefügt.

Anmerkung: Sie können eine mit **Visual Query Builder** generierte Abfrage bearbeiten und Variablen einfügen. Die bearbeitete Abfrage ist dann jedoch nicht mehr von Visual Query Builder lesbar. Die Schaltfläche **SQL erstellen** wird deaktiviert, wenn Sie eine Variable in einer manuell verfassten oder einer generierten SQL-Abfrage einschließen.

Einfügen einer Abfragevariable

1. Öffnen Sie den erforderlichen Auftrag, der den **Read From DB**-Schritt umfasst. Alternativ können Sie den **Read From DB**-Schritt zum Auftrag hinzufügen.
2. Öffnen Sie die „**Read from DB**“-Optionen des **Read From DB**-Schrittes.
3. Erstellen Sie die SQL-Abfrage im Feld **SQL**, entweder manuell oder mit dem Visual Query Builder. Weitere Informationen finden Sie unter **Visual Query Builder** auf Seite 146.
4. Fügen Sie die gewünschten Bedingungen in der Klausel `where` der Abfrage unter Verwendung von Variablen mit der Syntax `#{variable}` hinzu.
Formulieren Sie beispielsweise wie folgt eine SQL-Abfrage in der Tabelle `CUSTOMERS`, die die Spalte `AGE` mit Werten wie 28, 32, 30 usw. und die Spalte `SALARY` mit Werten wie 1000, 1500, 2200 usw. enthält:

```
select * from CUSTOMERS where #{condition1} > 28 and #{condition2} > 1200
```

Anmerkung: Beim Einfügen einer Variable in die Klausel `where` der SQL-Abfrage wird die Schaltfläche **SQL erstellen...** deaktiviert.

5. Geben Sie auf der Registerkarte **Laufzeit** den Wert der Schrittoptionen ein, und klicken Sie anschließend auf die Schaltfläche **Felder erneut generieren** bzw. **Vorschau**, um das Schema und die Beispieldatensätze anzuzeigen, die von der Abfrage abgerufen werden sollen.
6. Klicken Sie auf **OK**.

Die Klausel `where` der SQL-Abfrage kann jetzt zur Laufzeit mit den **Datenflussoptionen** oder während der Ausführung des Auftrags über den Job Executor angepasst werden.

Anmerkung: In der Klausel `select` einer SQL-Abfrage kann auch eine Variable platziert werden. Ein solcher Variablenname sollte jedoch mit dem Namen von einer der Spalten in der Tabelle übereinstimmen, die abgefragt werden.

Konfigurieren einer Abfragevariable als Datenflussoption

1. Öffnen Sie den erforderlichen Auftrag, für den die Abfrage, die die Variablen enthält, in einem **Read from DB**-Schritt definiert wurde.
2. Öffnen Sie **Bearbeiten > Datenflussoptionen...**
3. Klicken Sie auf **Hinzufügen**.
4. Erweitern Sie im Abschnitt **Datenflussoptionen zu Schritten zuordnen** den Eintrag **Read from DB**.
Die in der SQL-Abfrage des **Read from DB**-Schrittes definierten Variablen sind zusammen mit den anderen Attributen des Schrittes aufgelistet.
5. Wählen Sie die anzupassende Variable über das zugehörige Kästchen aus.
6. Geben Sie im Feld **Optionsbeschriftung** einen relevanten Namen für die Variable ein.

- Geben Sie im Feld **Standardwert** den Spaltennamen ein, der anstelle der ausgewählten Variable in der `where`-Klausel der SQL-Abfrage verwendet werden soll. Geben Sie alternativ eine Konstante ein, die anstelle der Variable in der `where`-Klausel verwendet werden soll. Beispielsweise können Sie in der SQL-Abfrage, die im **Read from DB**-Schritt definiert ist,

```
select * from CUSTOMERS where #{condition1} > 28 and #{condition2} > 1200
```

die Spalte `AGE` der Tabelle `CUSTOMERS` als **Standardwert** für die Variable `condition1` und die Spalte `SALARY` als **Standardwert** für die Variable `condition2` auswählen.

Zur Laufzeit wird die Abfrage interpretiert als:

```
select * from CUSTOMERS where AGE > 28 and SALARY > 1200
```

- Wiederholen Sie die Schritte 5 bis 7 für alle Variablen in der SQL-Abfrage des **Read from DB**-Schrittes.
- Klicken Sie auf **OK**.

Beim Ausführen des Datenflusses wird die benutzerdefinierte Abfrage verwendet, um die erforderlichen Daten abzurufen.

Konfigurieren von Abfragevariablen für Job Executor

Anmerkung: Vergewissern Sie sich, dass Sie den Spectrum™ Job Executor auf Ihren Server heruntergeladen haben.

- Erstellen Sie eine Textdatei, die die Standardwerte der Variablen definiert, die in der SQL-Abfrage des **Read from DB**-Schrittes des Auftrags enthalten sind.

Um den Spaltennamen `AGE` zu einer Variable `condition1` als Standardwert zuzuweisen, erstellen Sie eine Textdatei, beispielsweise `variables.txt`, und schließen Sie folgende Zeile in der Datei mit ein:

```
condition1=AGE
```

Um zu einer Variable `condition1` eine Konstante als Standardwert zuzuweisen (z. B. 20), schließen Sie folgende Zeile in der Datei mit ein:

```
condition1=20
```

- Wenn Sie den Auftrag über den Job Executor in der Eingabeaufforderung ausführen, verwenden Sie das Argument `-o` gefolgt vom Pfad zur erstellten Textdatei.

Um beispielsweise den Auftrag `ReadCustomerDataJob`, für dessen Variablen Standardwerte in der Textdatei `variables.txt` definiert wurden, auszuführen, führen Sie den folgenden Befehl in der Eingabeaufforderung aus:

```
java -jar jobexecutor.jar -h "localhost" -u "admin" -p "admin" -s
"8080" -j "ReadCustomerDataJob" -o "variables.txt"
```

Beim Ausführen des Auftrags über den Job Executor wird die benutzerdefinierte Abfrage verwendet, um die erforderlichen Daten abzurufen.

Anmerkung: Informationen zu Anweisungen und zur Syntax in der Befehlszeile finden Sie unter *Auftrag ausführen* im Datenfluss-Designer-Handbuch.

Read from File

Der „Read from File“-Schritt gibt eine Eingabedatei für einen Auftrag oder Unterfluss an. Er steht nicht für Dienste zur Verfügung.

Anmerkung: Wenn Sie eine XML-Datei als Eingabe für Ihren Datenfluss verwenden möchten, verwenden anstelle des „Read from File“-Schrittes den „Read from XML“-Schritt. Wenn Sie eine Datei mit variablem Format als Eingabe verwenden möchten, verwenden Sie den Schritt „Read from Variable Format File“.

Registerkarte „Dateieigenschaften“

Feldname	Beschreibung
Servername	Dies gibt an, ob sich die von Ihnen als Eingabe ausgewählte Datei auf dem Computer mit dem Enterprise Designer oder auf dem Spectrum™ Technology Platform-Server befindet. Wenn Sie eine Datei auf dem lokalen Computer auswählen, lautet der Servername „Arbeitsplatz“. Wenn Sie eine Datei auf dem Server auswählen, lautet der Servername Spectrum™ Technology Platform.

Feldname	Beschreibung
Dateiname	<p>Gibt den Pfad zur Datei an. Klicken Sie auf die Ellipsenschaltfläche (...), um zur gewünschten Datei zu navigieren.</p> <p>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.</p> <p>Beim Lesen einer Datei von einem HDFS-Dateiserver werden die folgenden Komprimierungsformate unterstützt:</p> <ol style="list-style-type: none"> 1. GZIP (.gz) 2. BZIP2 (.bz2) <p>Anmerkung: Die Erweiterung der Datei zeigt das zum Extrahieren zu verwendende Komprimierungsformat an.</p> <p>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.</p>
Datensatztyp	<p>Dies ist das Format der Datensätze in der Datei. Wählen Sie eine Option:</p> <p>Zeilensequenziell 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 über feste Startzeichen- und Endzeichenpositionen verfügt.</p> <p>Feste Breite Dies ist eine Textdatei, in der jeder Datensatz eine bestimmte Anzahl von Zeichen lang ist und jedes Feld über feste Startzeichen- und Endzeichenpositionen verfügt.</p> <p>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.</p>

Feldname	Beschreibung
Zeichencodierung	Dies ist die Codierung der Textdatei. Wählen Sie eine der folgenden Optionen:
	UTF-8 Die Datei unterstützt alle Unicode-Zeichen und ist abwärtskompatibel mit ASCII. Weitere Informationen zu UTF finden Sie unter unicode.org/faq/utf_bom.html .
	UTF-16 Die Datei unterstützt alle Unicode-Zeichen, ist aber nicht abwärtskompatibel mit ASCII. Weitere Informationen zu UTF finden Sie unter unicode.org/faq/utf_bom.html .
	US-ASCII Die Zeichencodierung basiert auf der Reihenfolge des englischen Alphabets.
	UTF-16BE Die Datei ist UTF-16-codiert mit Big-Endian-Reihenfolge (höchstwertiges Byte zuerst).
	UTF-16LE Die Datei ist UTF-16-codiert mit Little-Endian-Reihenfolge (niedrigstwertiges Byte zuerst).
	ISO-8859-1 Dies ist eine ASCII-Zeichencodierung, die normalerweise für westeuropäische Sprachen verwendet wird. Sie ist auch bekannt als Latin-1.
	ISO-8859-3 Dies ist eine ASCII-Zeichencodierung, die normalerweise für südeuropäische Sprachen verwendet wird. Sie ist auch bekannt als Latin-3.
	ISO-8859-9 Dies ist eine ASCII-Zeichencodierung, die normalerweise für Türkisch verwendet wird. Sie ist auch bekannt als Latin-5.
	CP850 Dies ist eine ASCII-Codeseite, die für westeuropäische Sprachen verwendet wird.
	CP500 Dies ist eine EBCDIC-Codeseite, die für westeuropäische Sprachen verwendet wird.
	Shift_JIS Dies ist eine Zeichencodierung für Japanisch.
	MS932 Dies ist eine Shift_JIS-Erweiterung von Microsoft, die besondere NEC-Zeichen, eine NEC-Auswahl aus IBM-Erweiterungen und IBM-Erweiterungen einschließt.
	CP1047 Dies ist eine EBCDIC-Codeseite mit dem vollständigen Latin-1-Zeichensatz.

Feldname	Beschreibung
Feldtrennzeichen	<p>Gibt das Zeichen an, das zum Trennen von Feldern in einer Datei mit Begrenzungszeichen verwendet wird.</p> <p>Im folgenden Beispiel wird ein Pipe-Zeichen () im Datensatz als Feldtrennzeichen verwendet:</p> <pre>7200 13TH ST MIAMI FL 33144</pre> <p>Die folgenden Zeichen stehen zur Definition als Feldtrennzeichen zur Verfügung:</p> <ul style="list-style-type: none">• Leerzeichen• Tabulator• Komma• Punkt• Semikolon• Pipe <p>Wenn in der Datei ein anderes Zeichen als Feldtrennzeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Begrenzungszeichen auszuwählen.</p>
Textbegrenzungszeichen	<p>Dies ist das Zeichen, das in einer Datei mit Begrenzungszeichen Textwerte umschließt.</p> <p>Im folgenden Beispiel werden im Datensatz doppelte Anführungszeichen (") als Textbegrenzungszeichen verwendet.</p> <pre>"7200 13TH ST" "MIAMI" "FL" "33144"</pre> <p>Die folgenden Zeichen stehen zur Definition als Textbegrenzungszeichen zur Verfügung:</p> <ul style="list-style-type: none">• einfaches Anführungszeichen (')• doppeltes Anführungszeichen (") <p>Wenn in der Datei ein anderes Textbegrenzungszeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Textbegrenzungszeichen auszuwählen.</p>

Feldname	Beschreibung
Datensatz-Trennzeichen	<p>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-Zeilende verwenden aktivieren.</p> <p>Die verfügbaren Einstellungen für Datensatz-Trennzeichen sind:</p> <p>Unix (U+000A) Die Datensätze sind durch ein Zeilenvorschubzeichen getrennt. Dies ist bei Unix-Systemen das standardmäßige Datensatz-Trennzeichen.</p> <p>Macintosh (U+000D) Die Datensätze sind durch ein Zeilenumbruchzeichen getrennt. Dies ist bei Macintosh-Systemen das standardmäßige Datensatz-Trennzeichen.</p> <p>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.</p> <p>Wenn in der Datei ein anderes Datensatz-Trennzeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Datensatz-Trennzeichen auszuwählen.</p>
Standard-Zeilende verwenden	<p>Gibt an, dass das standardmäßige Zeilenendzeichen (EOL) des Betriebssystems, auf dem der Spectrum™ Technology Platform-Server läuft, als Datensatz-Trennzeichen der Datei verwendet wird.</p> <p>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-Zeilendzeichen verwendet wird, der Server aber auf Linux läuft, aktivieren Sie diese Option nicht. Wählen Sie stattdessen im Feld Datensatz-Trennzeichen die Option „Windows“ aus.</p>
Datensatzlänge	<p>Geben Sie bei Dateien an, die Datensätze mit fester Breite enthalten, die genaue Anzahl von Zeichen jedes Datensatzes.</p> <p>Geben Sie bei zeilensequenziellen Dateien die Länge des längsten Datensatzes in der Datei in Zeichen an.</p>
Erste Zeile ist Kopfzeilen-Datensatz	<p>Gibt an, ob der erste Datensatz in einer Datei mit Begrenzungszeichen Überschriften und keine Daten enthält.</p> <p>Im folgenden Dateiausschnitt ist der erste Datensatz ein Kopfzeilen-Datensatz.</p> <pre>"AddressLine1" "City" "StateProvince" "PostalCode" "7200 13TH ST" "MIAMI" "FL" "33144" "One Global View" "Troy" "NY" "12180"</pre>

Feldname	Beschreibung
Datensätze mit weniger Feldern als definiert als falsch formatiert behandeln	Datensätze in Dateien mit Begrenzungszeichen, die weniger Felder enthalten, als auf der Registerkarte Felder definiert wurden, werden als falsch formatiert behandelt.
Importieren	Dies importiert die Definition des Dateilayouts, die Codierungseinstellung und Sortieroptionen aus einer Einstellungsdatei. Die Einstellungsdatei wird beim Exportieren von Einstellungen durch einen anderen „Read from File“- oder „Write to File“-Schritt erstellt. Dieser verwendet dieselbe Eingabedatei oder eine Datei, die über dasselbe Layout wie die Datei verfügt, mit der Sie arbeiten.
Exportieren	Dies exportiert die Definition des Dateilayouts, die Codierungseinstellung und Sortieroptionen in eine Einstellungsdatei. Sie können anschließend diese Einstellungen in andere „Read from File“- oder „Write to File“-Schritte importieren, die dieselbe Eingabedatei oder eine Datei mit demselben Layout wie die Datei verwenden, mit der Sie gerade arbeiten. Sie können die Einstellungsdatei auch mit dem Job Executor verwenden, um die Dateieinstellungen zur Laufzeit anzugeben. Informationen zur Einstellungsdatei finden Sie unter Die Dateidefinitions-Einstellungsdatei auf Seite 165.

Registerkarte „Felder“

Die Registerkarte „Felder“ definiert die Namen, Positionen und, bei Dateien vom Typ „Feste Breite“ und „Zeilensequenziell“, die Länge der Felder in der Datei. Weitere Informationen bieten die folgenden Themen:

[Definieren von Feldern in einer Datei mit Begrenzungszeichen](#) auf Seite 158

[Definieren von Feldern in einer Datei vom Typ „Zeilensequenziell“ oder „Feste Breite“](#) auf Seite 161

Registerkarte „Sortierfelder“

Die Registerkarte „Sortierfelder“ definiert Felder, nach denen die Eingabedatensätze sortiert werden sollen, bevor sie in den Datenfluss gesendet werden. Die Sortierung ist optional. Weitere Informationen finden Sie unter [Sortieren von Eingabedatensätzen](#) auf Seite 164.

Registerkarte „Laufzeit“

Feldname	Beschreibung
Dateiname	Zeigt den auf der ersten Registerkarte ausgewählten Dateinamen an.

Feldname	Beschreibung
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, Position, und, für einige Dateitypen, die Längen 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 oder Datum/Uhrzeit-Operationen auf die Daten anzuwenden. Wenn Sie allerdings diese 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.

Spectrum™ Technology Platform 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“.

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 $(2-2^{-52}) \times 2^{1023}$ 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 $(2-2^{-23}) \times 2^{127}$ 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.

5. Wenn Sie den Datentyp „Datum“, „Uhrzeit“ oder „Numerisch“ ausgewählt haben, können Sie das Standardformat für Datum/Uhrzeit oder Zahlen verwenden oder ein anderes Format für dieses bestimmte Feld angeben. Das Standardformat ist entweder das Systemstandardformat, das in den Typkonvertierungsoptionen in der Management Console festgelegt wurde, oder es ist das Standardformat des Datenflusses, das in den Typkonvertierungsoptionen im Enterprise Designer angegeben wurde. Das wirksame Format wird angezeigt. Lassen Sie **Standard** ausgewählt, wenn Sie das Standardformat verwenden möchten. Möchten Sie hingegen ein anderes Format angeben, wählen Sie **Benutzerdefiniert** aus, und gehen Sie wie folgt vor:

Anmerkung: Es ist wichtig, dass Sie ein Datums- und Uhrzeitformat auswählen, das den Daten, die Sie aus der Datei lesen, korrekt entspricht. Wenn die Datei beispielsweise Datumsdaten im Format Monat/Tag/Jahr enthält, Sie aber Tag/Monat/Jahr auswählen, wird keine im Datenfluss ausgeführte Datumsberechnung wie Sortieren nach Datum das korrekte Datum verwenden oder anzeigen. Darüber hinaus kann bei Datensätzen die Typkonvertierung fehlschlagen. In dem Fall kommt das Fehlverhalten zum Tragen, das in den Typkonvertierungsoptionen in der Management Console oder im Enterprise Designer angegeben ist.

- a) Wählen Sie im Feld **Gebietsschema** das Land aus, dessen Formatierungskonventionen Sie verwenden möchten. Durch Ihre Auswahl werden die Standardwerte im Feld **Format** festgelegt. Bei Datumsdaten legt Ihre Auswahl auch die Sprache fest, die verwendet wird, 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 **Format** das Format für die Daten aus. Das Format ist abhängig vom Datentyp des Feldes. Es wird eine Liste der am häufigsten verwendeten Formate für das ausgewählte Gebietsschema bereitgestellt.

Ein Beispiel für das ausgewählte Format wird rechts vom Feld **Format** angezeigt.

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 318 erläuterten Schreibweise in das Feld ein. Geben Sie zur Festlegung eines eigenen Zahlenformats das Format mithilfe der unter **Zahlenmuster** auf Seite 321 erläuterten Schreibweise in das Feld ein.

6. 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.

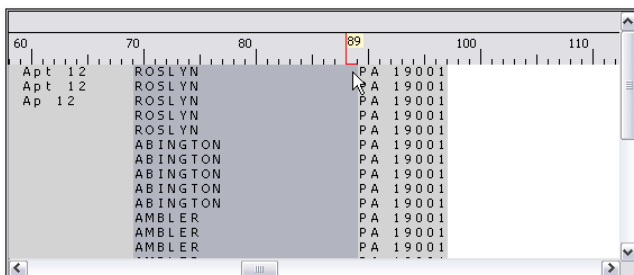
```
"AddressLine1"|"City"|"StateProvince"|"PostalCode"
"7200 13TH ST"|"MIAMI"|"FL"|"33144"
"One Global View"|"Troy"|"NY"|12180
```


7. Wenn Sie überzählige Leerzeichen am Anfang und Ende der Wertezeichenfolge eines Feldes entfernen möchten, aktivieren Sie das Kästchen **Kürzen**.

Definieren von Feldern in einer Datei vom Typ „Zeilensequenziell“ oder „Feste Breite“

Im „Read from File“-Schritt werden auf der Registerkarte **Felder** die Namen, Positionen und bei einigen Dateitypen die Längen der Felder in der Datei definiert. Nach Definieren einer Eingabedatei auf der Registerkarte **Dateieigenschaften** können Sie die Felder definieren.

1. Klicken Sie auf der Registerkarte **Felder** unter **Vorschau** auf den Anfang eines Feldes und ziehen Sie die Maus nach links, sodass das gewünschte Feld markiert ist, wie hier dargestellt:



2. Geben Sie im Feld **Name** das Feld ein, das Sie hinzufügen möchten.
3. Im Feld **Typ** können Sie den Datentyp „Zeichenfolge“ 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 Zeichenfolgendaten aus der Datei in einen Datentyp konvertiert, der die korrekte Manipulation der Daten im Datenfluss ermöglicht.

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 $(2-2^{-52}) \times 2^{1023}$ 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 $(2-2^{-23}) \times 2^{127}$ 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:
	<pre><Names> <Name>John Smith</Name> <Name>Ann Fowler</Name> </Names></pre>
	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.

- Um das Speicherformat `packed decimal` auf das Feld anzuwenden, aktivieren Sie das Kästchen **Packed Decimal**. Beim Typ `packed decimal` werden 4 Bit verwendet. Im Vergleich dazu werden beim Typ `integer` 8 Bit verwendet.

Anmerkung: Dieses Speicherformat ist nur bei der Auswahl der Datentypen `double`, `integer` und `long` beim Lesen von Dateien vom Typ „Zeilensequenziell“ oder „Feste Breite“ verfügbar.

- Wenn Sie den Datentyp „Datum“, „Uhrzeit“ oder „Numerisch“ ausgewählt haben, können Sie das Standardformat für Datum/Uhrzeit oder Zahlen verwenden oder ein anderes Format für

dieses bestimmte Feld angeben. Das Standardformat ist entweder das Systemstandardformat, das in den Typkonvertierungsoptionen in der Management Console festgelegt wurde, oder es ist das Standardformat des Datenflusses, das in den Typkonvertierungsoptionen im Enterprise Designer angegeben wurde. Das wirksame Format wird angezeigt. Lassen Sie **Standard** ausgewählt, wenn Sie das Standardformat verwenden möchten. Möchten Sie hingegen ein anderes Format angeben, wählen Sie **Benutzerdefiniert** aus, und gehen Sie wie folgt vor:

Anmerkung: Es ist wichtig, dass Sie ein Datums- und Uhrzeitformat auswählen, das den Daten, die Sie aus der Datei lesen, korrekt entspricht. Wenn die Datei beispielsweise Datumsdaten im Format Monat/Tag/Jahr enthält, Sie aber Tag/Monat/Jahr auswählen, wird keine im Datenfluss ausgeführte Datumsberechnung wie Sortieren nach Datum das korrekte Datum verwenden oder anzeigen. Darüber hinaus kann bei Datensätzen die Typkonvertierung fehlschlagen. In dem Fall kommt das Fehlerverhalten zum Tragen, das in den Typkonvertierungsoptionen in der Management Console oder im Enterprise Designer angegeben ist.

- a) Wählen Sie im Feld **Gebietsschema** das Land aus, dessen Formatierungskonventionen Sie verwenden möchten. Durch Ihre Auswahl werden die Standardwerte im Feld **Format** festgelegt. Bei Datumsdaten legt Ihre Auswahl auch die Sprache fest, die verwendet wird, 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 **Format** das Format für die Daten aus. Das Format ist abhängig vom Datentyp des Feldes. Es wird eine Liste der am häufigsten verwendeten Formate für das ausgewählte Gebietsschema bereitgestellt.

Ein Beispiel für das ausgewählte Format wird rechts vom Feld **Format** angezeigt.

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 318 erläuterten Schreibweise in das Feld ein. Geben Sie zur Festlegung eines eigenen Zahlenformats das Format mithilfe der unter **Zahlenmuster** auf Seite 321 erläuterten Schreibweise in das Feld ein.

6. Die Felder **Startposition** und **Länge** werden automatisch basierend auf der Auswahl, die Sie in der Dateivorschau getroffen haben, ausgefüllt.
7. Wenn Sie überzählige Leerzeichen am Anfang und Ende der Zeichenfolge eines Feldes entfernen möchten, aktivieren Sie das Kästchen **Leerzeichen abschneiden**.
8. Klicken Sie auf **OK**.

Sortieren von Eingabedatensätzen

Im Schritt „Read from File“ definiert die Registerkarte **Sortierfelder** Felder, nach denen die Eingabedatensätze sortiert werden sollen, bevor sie in den Datenfluss gesendet werden. Die Sortierung ist optional.

1. Klicken Sie auf der Registerkarte **Sortierfelder** 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** die Option Ascending oder Descending aus.
4. 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**.
5. Die Standardoptionen für die Sortierleistung für Ihr System werden in der Management Console festgelegt. Wenn Sie die Standardoptionen für die Sortierleistung Ihres Systems überschreiben möchten, klicken Sie auf **Erweitert**. Das Dialogfeld **Erweiterte Optionen** enthält die folgenden Optionen für die Sortierleistung:

Beschränkung für Datensätze im Arbeitsspeicher Gibt die maximale Anzahl von Datenzeilen an, die ein Sorter im Arbeitsspeicher belässt, bevor er sie auf die Festplatte auslagert. Standardmäßig erfolgt eine Sortierung von bis zu 10.000 Datensätze in einem Arbeitsspeicher und eine Sortierung von mehr als 10.000 Datensätzen auf einem Datenträger. Der maximale Wert umfasst 100.000 Datensätze. In der Regel erfolgt eine Sortierung im Arbeitsspeicher deutlich schneller als auf einem Datenträger. Daher sollte hierfür ein Wert festgelegt werden, der hoch genug ist, damit ein Großteil der Sortierungen im Arbeitsspeicher erfolgen kann und nur umfangreiche Datasets auf einen Datenträger geschrieben werden.

Anmerkung: Vorsicht bei Umgebungen, in denen Aufträge gleichzeitig ausgeführt werden: Durch das Erhöhen der Einstellung **Beschränkung für Datensätze im Arbeitsspeicher** wird die Wahrscheinlichkeit erhöht, dass nicht mehr genügend Arbeitsspeicher verfügbar ist.

Maximale Anzahl temporärer Dateien Gibt die maximale Anzahl temporärer Dateien an, die von einem Sortiervorgang verwendet werden können. Die Verwendung einer größeren Anzahl temporärer Dateien kann zu einer besseren Leistung führen. Die optimale Anzahl hängt jedoch stark von der Konfiguration des Servers ab, auf dem Spectrum™ Technology Platform ausgeführt wird. Sie sollten mit verschiedenen Einstellungen experimentieren und dabei die Auswirkungen auf die Leistung beobachten, wenn mehr oder weniger temporäre Dateien

verwendet werden. Verwenden Sie für die Berechnung der ungefähren Anzahl temporärer Dateien die folgende Gleichung:

$$\frac{(NumberOfRecords \times 2)}{InMemoryRecordLimit} = NumberOfTempFiles$$

Beachten Sie, dass die maximale Anzahl der temporären Dateien 1.000 nicht überschreiten darf.

Komprimierung aktivieren Gibt an, dass temporäre Dateien komprimiert werden, wenn sie auf die Festplatte geschrieben werden.

Anmerkung: Welche Einstellungen für die Sortierleistung optimal sind, hängt von der Hardwarekonfiguration Ihres Servers ab. Dennoch wird mit der folgenden Gleichung im Allgemeinen eine gute Sortierleistung erzielt:

$$\frac{(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2)}{TotalNumberOfRecords} \geq$$

Die Dateidefinitions-Einstellungsdatei

Eine Dateidefinitions-Einstellungsdatei enthält den Dateiaufbau, die Codierung und die Sortieroptionen, die aus einem „Read from File“- oder „Write to File“-Schritt exportiert wurden. Die Dateidefinitions-Einstellungsdatei kann in „Read from File“ oder „Write to File“ importiert werden, um die Optionen des Schrittes schnell festzulegen, anstatt sie manuell anzugeben.

Die einfachste Methode zum Erstellen einer Dateidefinitions-Einstellungsdatei besteht darin, die Dateieinstellungen mit „Read from File“ oder „Write to File“ anzugeben und dann auf die Schaltfläche **Exportieren** zu klicken, um die Dateidefinitions-Einstellungsdatei zu generieren.

Zu Ihrer Information ist jedoch das Schema der Dateidefinitions-Einstellungsdatei unten abgebildet. Jedes Element in der XML-Datei hat einen Typ, und wenn dieser Typ weder „Zeichenfolge“ noch „Ganze Zahl“ ist, werden die zulässigen Werte angezeigt. Diese Werte entsprechen direkt Optionen im Dialogfeld des Schrittes. So entspricht beispielsweise das Element „FileTypeEnum“ dem Feld „Datensatztyp“ auf der Registerkarte „Dateieigenschaften“, und die folgenden drei Werte werden im Schema angezeigt: „LineSequential“, „fixedwidth“ und „Mit Begrenzungszeichen“.

Anmerkung: Wenn Sie „benutzerdefiniert“ für die Felder „LineSeparator“, „FieldSeparator“ oder „TextQualifier“ eingeben, muss auch ein entsprechendes benutzerdefiniertes Element (z. B. „CustomLineSeparator“, „CustomFieldSeparator“ oder „CustomTextQualifier“) mit einer hexadezimalen Zahl, die das Zeichen oder die Zeichenfolge darstellt, angegeben werden.

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema elementFormDefault="qualified"
  xmlns:xs="http://www.w3.org/2001/XMLSchema">
```

```

<xs:element name="FileSchema" nillable="true" type="FileSchema"/>
<xs:complexType name="FileSchema">
  <xs:sequence>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      default="linesequential"
      name="Type"
      type="FileTypeEnum"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      default="UTF-8" name="Encoding" type="xs:string"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="RecordLength"
      type="xs:int"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      default="default"
      name="LineSeparator"
      type="LineSeparatorEnum"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="CustomLineSeparator"
      type="xs:string"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      default="comma"
      name="FieldSeparator"
      type="FieldSeparatorEnum"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="CustomFieldSeparator"
      type="xs:string"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      default="none"
      name="TextQualifier"
      type="TextQualifierEnum"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="CustomTextQualifier"
      type="xs:string"/>
    <xs:element
      minOccurs="0"

```

```

        maxOccurs="1"
        default="false"
        name="HasHeader"
        type="xs:boolean"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  default="true"
  name="EnforceColumnCount"
  type="xs:boolean"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  name="Fields"
  type="ArrayOfFieldSchema"/>
</xs:sequence>
</xs:complexType>
<xs:simpleType name="FileTypeEnum">
  <xs:restriction base="xs:string">
    <xs:enumeration value="linesequential"/>
    <xs:enumeration value="fixedwidth"/>
    <xs:enumeration value="delimited"/>
  </xs:restriction>
</xs:simpleType>
<xs:simpleType name="LineSeparatorEnum">
  <xs:restriction base="xs:string">
    <xs:enumeration value="default"/>
    <xs:enumeration value="windows"/>
    <xs:enumeration value="unix"/>
    <xs:enumeration value="mac"/>
    <xs:enumeration value="custom"/>
  </xs:restriction>
</xs:simpleType>
<xs:simpleType name="FieldSeparatorEnum">
  <xs:restriction base="xs:string">
    <xs:enumeration value="comma"/>
    <xs:enumeration value="tab"/>
    <xs:enumeration value="space"/>
    <xs:enumeration value="semicolon"/>
    <xs:enumeration value="period"/>
    <xs:enumeration value="pipe"/>
    <xs:enumeration value="custom"/>
  </xs:restriction>
</xs:simpleType>
<xs:simpleType name="TextQualifierEnum">
  <xs:restriction base="xs:string">
    <xs:enumeration value="none"/>
    <xs:enumeration value="single"/>
    <xs:enumeration value="double"/>
    <xs:enumeration value="custom"/>
  </xs:restriction>
</xs:simpleType>
<xs:complexType name="ArrayOfFieldSchema">

```

```
<xs:sequence>
  <xs:element
    minOccurs="0"
    maxOccurs="unbounded"
    name="Field"
    nillable="true"
    type="FieldSchema"/>
</xs:sequence>
</xs:complexType>
<xs:complexType name="FieldSchema">
  <xs:sequence>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="Name"
      type="xs:string"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      default="string"
      name="Type"
      type="xs:string"/>
    <xs:element
      minOccurs="1"
      maxOccurs="1"
      name="Position"
      type="xs:int"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="Length"
      type="xs:int"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      default="false"
      name="Trim"
      type="xs:boolean"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="Locale"
      type="Locale"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="Pattern"
      type="xs:string"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      default="none"
      name="Order"
```



```

        type="SortOrderEnum"/>
    </xs:sequence>
</xs:complexType>
<xs:complexType name="Locale">
    <xs:sequence>
        <xs:element
            minOccurs="0"
            maxOccurs="1"
            name="Country"
            type="xs:string"/>
        <xs:element
            minOccurs="0"
            maxOccurs="1"
            name="Language"
            type="xs:string"/>
        <xs:element
            minOccurs="0"
            maxOccurs="1"
            name="Variant"
            type="xs:string"/>
    </xs:sequence>
</xs:complexType>
<xs:simpleType name="SortOrderEnum">
    <xs:restriction base="xs:string">
        <xs:enumeration value="none"/>
        <xs:enumeration value="ascending"/>
        <xs:enumeration value="descending"/>
    </xs:restriction>
</xs:simpleType>
</xs:schema>

```

Definieren von Datenflussoptionen

Die folgenden Schritte beschreiben, wie Sie einen Datenfluss konfigurieren, um Laufzeitoptionen für den „Read from File“-Schritt zu unterstützen.

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. Erweitern Sie den **Read from File**-Schritt.

Die verfügbar gemachten Datenflussoptionen sind:

1. Zeichencodierung

2. Feldtrennzeichen
 3. Textbegrenzungszeichen
 4. Datensatzlänge
 5. Erste Zeile ist Kopfzeilen-Datensatz
 6. Startdatensatz
 7. Max. Anzahl Datensätze
6. Der ausgewählte „Read from File“-Optionsname wird in den Feldern **Optionsname** und **Optionsbeschriftung** angezeigt. Dies ist der Optionsname, der zur Laufzeit angegeben werden muss, um diese Option festzulegen.
 7. Geben Sie eine Beschreibung der Option in das Feld **Beschreibung** ein.
 8. Wählen Sie im Feld **Ziel** die Option **Ausgewählte(r) Schritt/e** aus.
 9. 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.
 10. 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 [Angaben von Standarddienstoptionen](#).

11. Klicken Sie auf **OK**.
12. Fügen Sie auf Wunsch weitere Optionen hinzu.
13. Klicken Sie im Dialogfeld **Datenflussoptionen** auf **OK**, sobald Sie damit fertig sind.
14. Speichern Sie den Datenfluss und machen Sie ihn verfügbar.

Regeln für Datenflussoptionen

1. *Zeichencodierung:* Alle Codierungstypen, die in der zugrunde liegenden JVM gültig sind, sind zulässig. Diese Option darf nicht leer sein.
2. *Feldtrennzeichen:* Jedes aus einem einzelnen Zeichen bestehende Trennzeichen ist zulässig. Derzeit werden keine hexadezimalen Werte und keine Leerzeichen unterstützt.
3. *Textbegrenzungszeichen:* Jedes aus einem einzelnen Zeichen bestehende Begrenzungszeichen ist zulässig. Hexadezimale Werte werden nicht unterstützt.
4. *Datensatzlänge:* Nur ganze Zahlen sind zulässig. Nicht numerische oder leere Werte sind nicht zulässig.
5. *Startdatensatz:* Nur ganze Zahlen sind zulässig. Nicht numerische Werte sind nicht zulässig.
6. *Max. Anzahl Datensätze:* Nur ganze Zahlen sind zulässig. Nicht numerische Werte sind nicht zulässig.
7. *Erste Zeile ist Kopfzeilen-Datensatz:* Nur die booleschen Werte `wahr` und `falsch` sind zulässig. Leere Werte sind nicht zulässig.

Read from Hadoop Sequence File

Der Schritt „Read from Hadoop Sequence File“ liest Daten aus einer Sequenzdatei als Eingabe für einen Datenfluss. Eine Sequenzdatei ist ein Flatfile, das aus binären Schlüssel/Wert-Paaren besteht. Weitere Informationen finden Sie unter wiki.apache.org/hadoop/SequenceFile.

Anmerkung: Der Schritt „Read from Hadoop Sequence File“ unterstützt nur durch Trennzeichen getrennte, unkomprimierte Sequenzdateien im Hadoop Distributed File System (HDFS).

Registerkarte „Dateieigenschaften“

Felder	Beschreibung
Server	Dies gibt an, dass die im Feld Dateiname ausgewählte Datei sich auf dem Hadoop-System befindet. Sie müssen eine Verbindung zum Hadoop-Dateiserver in der Management Console erstellen, bevor Sie ihn im Schritt verwenden. Wenn Sie eine Datei auf dem Hadoop-System auswählen, ist der Servername der Name, den Sie in der Management Console beim Erstellen eines Dateiservers angegeben haben.
Dateiname	Gibt den Pfad zur Datei an. Klicken Sie auf die Ellipsenschaltfläche (...), um die gewünschte Datei zu lokalisieren.
Feldtrennzeichen	<p>Gibt das Zeichen an, das zum Trennen von Feldern in einer Datei mit Begrenzungszeichen verwendet wird.</p> <p>Im folgenden Beispiel wird ein Pipe-Zeichen () im Datensatz als Feldtrennzeichen verwendet:</p> <pre>7200 13TH ST MIAMI FL 33144</pre> <p>Die folgenden Zeichen stehen zur Definition als Feldtrennzeichen zur Verfügung:</p> <ul style="list-style-type: none"> • Leerzeichen • Tabulator • Komma • Punkt • Semikolon • Pipe <p>Wenn in der Datei ein anderes Zeichen als Feldtrennzeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Begrenzungszeichen auszuwählen.</p>

Felder	Beschreibung
Textbegrenzungszeichen	<p>Dies ist das Zeichen, das in einer Datei mit Begrenzungszeichen Textwerte umschließt.</p> <p>Im folgenden Beispiel werden im Datensatz doppelte Anführungszeichen (") als Textbegrenzungszeichen verwendet.</p> <pre>"7200 13TH ST" "MIAMI" "FL" "33144"</pre> <p>Die folgenden Zeichen stehen zur Definition als Textbegrenzungszeichen zur Verfügung:</p> <ul style="list-style-type: none"> • einfaches Anführungszeichen (') • doppeltes Anführungszeichen (") <p>Wenn in der Datei ein anderes Textbegrenzungszeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Textbegrenzungszeichen auszuwählen.</p>

Registerkarte „Felder“

Die Registerkarte „Felder“ definiert die Namen, Positionen und Typen der Felder in der Datei. Weitere Informationen finden Sie unter [Definieren von Feldern in einer Eingabesequenzdatei](#) auf Seite 172.

Registerkarte „Sortierfelder“

Die Registerkarte „Sortierfelder“ definiert Felder, nach denen die Eingabedatensätze sortiert werden sollen, bevor sie in den Datenfluss gesendet werden. Die Sortierung ist optional. Weitere Informationen finden Sie unter [Sortieren von Eingabedatensätzen](#) auf Seite 173.

Registerkarte „Filter“

Die Registerkarte „Filter“ definiert Felder, nach denen die Eingabedatensätze gefiltert werden sollen, bevor sie in den Datenfluss gesendet werden. Weitere Informationen finden Sie unter [Filtern von Eingabedatensätzen](#) auf Seite 174.

Definieren von Feldern in einer Eingabesequenzdatei

Im Schritt „Read from Hadoop Sequence File“ definiert die Registerkarte **Felder** die Namen, Positionen und Längen 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 „Zeichenfolge“ beibehalten, wenn Sie nicht vorhaben, mathematische Operationen auf die Daten anzuwenden. Wenn Sie allerdings diese 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.

Der Schritt unterstützt die folgenden Datentypen:

double	Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit doppelter Genauigkeit zwischen 2^{-1074} und $(2-2^{-52}) \times 2^{1023}$ 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 $(2-2^{-23}) \times 2^{127}$ 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 Eingabedatensätzen

Im Schritt „Read from Hadoop Sequence File“ definiert die Registerkarte **Sortierfelder** Felder, nach denen die Eingabedatensätze sortiert werden sollen, bevor sie in den Datenfluss gesendet werden. Die Sortierung ist optional.

1. Klicken Sie im Schritt „Read from Hadoop Sequence File“ auf die Registerkarte **Sortierfelder**.

2. Klicken Sie auf der Registerkarte **Sortierfelder** auf **Hinzufügen**.
3. 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.
4. Wählen Sie in der Spalte **Sortierung** die Option Ascending oder Descending aus.
5. 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**.

Filtern von Eingabedatensätzen

Im Schritt „Read from Hadoop Sequence File“ definiert die Registerkarte **Filter** Felder, nach denen die Eingabedatensätze gefiltert werden sollen, bevor sie in den Datenfluss gesendet werden. Filterung ist optional.

1. Klicken Sie im Schritt „Read from Hadoop Sequence File“ auf die Registerkarte **Filter**.
2. Wählen Sie im Feld **Ausdrucksmethode kombinieren** **Alle** aus, wenn Sie möchten, dass alle Ausdrücke als wahr (true) ausgewertet werden müssen, damit der Datensatz an diesen Port weitergeleitet wird. Wählen Sie **Beliebige** aus, wenn Sie möchten, dass Datensätze an diesen Port weitergeleitet werden sollen, wenn ein oder mehrere Ausdrücke wahr sind.
3. Klicken Sie auf **Hinzufügen**, geben Sie das zu testende Feld, den Operator und einen Wert an. Die Operatoren finden Sie in folgender Tabelle aufgelistet.

Operator	Beschreibung
Ist gleich	Prüft, ob der Wert im Feld mit dem angegebenen Wert übereinstimmt.
Ist ungleich	Prüft, ob der Wert im Feld nicht mit dem angegebenen Wert übereinstimmt.
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.
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.
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.

Operator	Beschreibung
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.
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.

4. Wählen Sie die Option **Kürzen** aus, falls gewünscht. Bei dieser Option werden zuerst alle Leerzeichen, die möglicherweise vor und hinter dem Wert des Feldes vorhanden sind, abgeschnitten, bevor die Daten im Feld gefiltert werden.
5. Wiederholen Sie diese Schritte, bis Sie alle Eingabefelder hinzugefügt haben, die Sie zum Filtern verwenden möchten.

Read from Hive File

Der Schritt **Read from Hive File** liest Daten aus der ausgewählten Datei, die in einem der folgenden Formate vorliegen kann:

- ORC
- RC
- Parquet
- Avro

Registerkarte *Dateieigenschaften*

Felder	Beschreibung
Server	Dies gibt an, dass die im Feld Dateiname ausgewählte Datei sich auf dem Hadoop-System befindet. Sie müssen eine Verbindung zum Hadoop-Dateiserver in der Management Console erstellen, bevor Sie ihn im Schritt verwenden. Wenn Sie eine Datei auf dem Hadoop-System auswählen, ist der Servername der Name, den Sie in der Management Console beim Erstellen eines Dateiservers angegeben haben.

Felder	Beschreibung
Dateiname	<p>Gibt den Pfad zur Datei an. Klicken Sie auf die Ellipsenschaltfläche (...), um die gewünschte Datei zu lokalisieren.</p> <p>Anmerkung: Das Schema einer Eingabedatei wird importiert, sobald Sie zum korrekten Speicherort navigieren und die Datei auswählen. Dieses importierte Schema kann nicht bearbeitet werden.</p> <p>Sie können die Spalten des Schemas jedoch falls erforderlich umbenennen.</p> <p>Die ersten 50 Datensätze der Datei werden im Raster Vorschau bei Auswahl der Datei abgerufen.</p>
Dateityp	<p>Wählen Sie den Typ der zu lesenden Datei aus:</p> <ul style="list-style-type: none"> • ORC • RC • Parquet • Avro

Anmerkung: Um die Vorschau bei RC-Dateien zu generieren, definieren Sie das Schema auf der Registerkarte **Felder** und klicken Sie dann auf **Vorschau** auf der Registerkarte **Dateieigenschaften**.

Registerkarte **Felder**

Auf der Registerkarte **Felder** werden die Namen, Datentypen, Positionen der Felder definiert, wie sie in der Eingabedatei vorliegen. Zudem werden die benutzerdefinierten Namen der Felder definiert. Weitere Informationen finden Sie unter [Definieren von Feldern zum Lesen aus einer Hive-Datei](#) auf Seite 176.

Definieren von Feldern zum Lesen aus einer Hive-Datei

Auf der Registerkarte **Felder** des Schrittes **Read from Hive File** werden die Schemanamen, Datentypen, Positionen und angegebene Namen der Felder in der Datei aufgelistet.

1. Klicken Sie auf **Erneut generieren**.
Bei ORC-, Avro- und Parquet-Dateien wird dann das Schema basierend auf den Metadaten der vorhandenen Datei generiert. Bei RC-Dateien werden alle Felder, die vor dem Klicken auf **Vorschau** hinzugefügt wurden, geleert.

Im Grid werden die Spalten **Name**, **Typ**, **Schrittfeld** und **Einschließen** angezeigt.

In der Spalte **Name** wird der Feldname angezeigt, wie er aus dem Header-Datensatz der Datei abgeleitet wurde.

In der Spalte **Typ** werden die Datentypen der jeweiligen Felder der Datei aufgelistet.

Der Schritt unterstützt die folgenden Datentypen:

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

Anmerkung: Der Datentyp `datetime` in Spectrum richtet sich nach dem Datentyp `timestamp` von Hive-Dateien.

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

Anmerkung: Bei RC-, Avro- und Parquet Hive-Dateien werden Felder vom Datentyp `decimal` in der Eingabedatei in den Datentyp `bigdecimal` konvertiert.

- 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.

Anmerkung: Der Datentyp `long` in Spectrum richtet sich nach dem Datentyp `bigint` von Hive-Dateien.

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

Anmerkung: Bei RC-Dateien werden der Datentyp `smallint` und komplexe Datentypen nicht unterstützt.

In der Spalte **Position** wird die Startposition des jeweiligen Feldes innerhalb eines Datensatzes angezeigt.

2. Ändern Sie in der Spalte **Schrittfeld** für jedes Feld den vorhandenen Feldnamen in den gewünschten Namen.
Standardmäßig werden in dieser Spalte die Feldnamen angezeigt, die aus der Datei gelesen wurden.
3. Aktivieren Sie in der Spalte **Einschließen** die Kästchen der Felder, die Sie in der Ausgabe des Schrittes einschließen möchten.
Standardmäßig sind alle Felder in dieser Spalte ausgewählt.
4. Bei RC-Dateien können Sie über die folgenden Schaltflächen Felder hinzufügen und entfernen und die Reihenfolge der ausgewählten Spalten in der Ausgabe ändern:

Name der Option	Beschreibung
Hinzufügen	Fügt der Ausgabe ein Feld hinzu.
Ändern	Ändert den Namen und Datentyp des Feldes.
Entfernen	Entfernt das ausgewählte Feld aus der Ausgabe.
Nach oben/Nach unten	Ordnet die Position des ausgewählten Feldes in der Ausgabe neu an.

Anmerkung: Dieses Feature ist nur bei RC-Dateien verfügbar.

5. Klicken Sie auf **OK**.

Read from HL7 File

Der Schritt **Read from HL7 File** liest Health Level Seven-Daten (HL7) aus einer Textdatei als Eingabe in einen Datenfluss. HL7 ist ein Nachrichtenstandard, der im Gesundheitswesen verwendet wird, um Daten zwischen Systemen auszutauschen. Weitere Informationen zu HL7 finden Sie unter www.hl7.org.

HL7-Nachrichtenformat

Daten in einer HL7-Nachricht sind hierarchisch organisiert, wie hier dargestellt ist:

- Nachricht
 - Segment
 - Feld
 - Komponente
 - Unterkomponente

Jede Zeile einer HL7-Nachricht ist ein Segment. Ein *Segment* ist eine logische Gruppierung von Feldern. Die ersten drei Zeichen in einem Segment identifizieren den Segmenttyp. Die Beispielnachricht besteht aus fünf Segmenten: `MSH` (Nachrichten-Header), `PID` (Patienten-ID), zwei `NK1`-Segmente (nächster Verwandter) und `IN1` (Versicherung).

Jedes Segment besteht aus Feldern. Ein *Feld* enthält Informationen bezüglich des Zwecks des Segments wie beispielsweise den Namen der Versicherungsgesellschaft im `IN1`-Segment (Versicherung). Felder werden normalerweise (aber nicht immer) durch ein `|`-Zeichen begrenzt.

Felder können in *Komponenten* unterteilt sein. Komponenten werden normalerweise durch ein `^`-Zeichen angezeigt. In der Beispielnachricht enthält das `PID`-Segment (Patienten-ID) ein Feld für den Patientennamen mit dem Inhalt `LEVERKUHN^ADRIAN^C`, das aus drei Teilen besteht: Nachname (LEVERKUHN), Vorname (ADRIAN) und Initiale (C). Komponenten können in *Unterkomponenten* unterteilt sein. Unterkomponenten werden normalerweise durch ein `&`-Zeichen angezeigt.

Hier folgt ein Beispiel einer HL7-Nachricht:

```
MSH|^~\&|||199908180016||ADT^A04|ADT.1.1698593|P|2.7
PID|1||000395122||LEVERKUHN^ADRIAN^C||19880517180606|M|||6 66TH AVE
NE^^WEIMAR^DL^98052|| (157) 983-3296|||S||12354768|87654321
NK1|1|TALLIS^THOMAS^C|GRANDFATHER|12914 SPEM
ST^^ALIUM^IN^98052| (157) 883-6176
NK1|2|WEBERN^ANTON|SON|12 STRASSE MUSIK^^VIENNA^AUS^11212| (123) 456-7890
IN1|1|PRE2||LIFE PRUDENT BUYER|PO BOX
23523^WELLINGTON^ON^98111|||19601|||||THOMAS^JAMES^M|F|||||||||||||||ZKA535529776
```

Anmerkung: HL7-Datei mit Beispielttext erstellen:

1. Kopieren Sie den Beispielttext und fügen Sie ihn unter Verwendung einer beliebigen Textverarbeitungssoftware (z. B. Notepad++) in ein neues Dokument ein.
2. Nehmen Sie die erforderlichen Änderungen am Inhalt vor.
3. Konfigurieren Sie die Einstellungen, um Zeilenenden (EOL) anzeigen zu lassen. Navigieren Sie in Notepad++ zu **Ansicht > Nicht druckbare Zeichen > Zeilenende anzeigen**.
4. Ändern Sie das Konvertierungsformat für Zeilenende (EOL) auf `CR` (Carriage Return). Navigieren Sie in Notepad++ zu **Bearbeiten > Format Zeilenende > Konvertiere zu Mac (CR)**.

5. Speichern Sie nach dieser Formatänderung die HL7-Datei.

Registerkarte „Dateieigenschaften“

Feldname	Beschreibung
Servername	Dies gibt an, ob sich die von Ihnen als Eingabe ausgewählte Datei auf dem Computer mit dem Enterprise Designer oder auf dem Spectrum™ Technology Platform-Server befindet. Wenn Sie eine Datei auf dem lokalen Computer auswählen, lautet der Servername „Arbeitsplatz“. Wenn Sie eine Datei auf dem Server auswählen, lautet der Servername Spectrum™ Technology Platform.
Dateiname	Gibt den Pfad zur Datei an. Klicken Sie auf die Ellipsenschaltfläche (...), um die gewünschte Datei zu lokalisieren. Anmerkung: 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.
HL7-Version	Dies ist die Version des HL7-Standards in der von Ihnen angegebenen Datei. Beispielsweise bedeutet „2.7“, dass die Datei HL7-Version 2.7 verwendet. Die HL7-Version wird im 12. Feld des MSH-Segments angezeigt.

Feldname	Beschreibung
Zeichencodierung	<p>Dies ist die Codierung der Textdatei. Wählen Sie eine der folgenden Optionen:</p> <p>UTF-8 Die Datei unterstützt alle Unicode-Zeichen und ist abwärtskompatibel mit ASCII. Weitere Informationen zu UTF finden Sie unter unicode.org/faq/utf_bom.html.</p> <p>UTF-16 Die Datei unterstützt alle Unicode-Zeichen, ist aber nicht abwärtskompatibel mit ASCII. Weitere Informationen zu UTF finden Sie unter unicode.org/faq/utf_bom.html.</p> <p>US-ASCII Die Zeichencodierung basiert auf der Reihenfolge des englischen Alphabets.</p> <p>UTF-16BE Die Datei ist UTF-16-codiert mit Big-Endian-Reihenfolge (höchstwertiges Byte zuerst).</p> <p>UTF-16LE Die Datei ist UTF-16-codiert mit Little-Endian-Reihenfolge (niedrigstwertiges Byte zuerst).</p> <p>ISO-8859-1 Dies ist eine ASCII-Zeichencodierung, die normalerweise für westeuropäische Sprachen verwendet wird. Sie ist auch bekannt als Latin-1.</p> <p>ISO-8859-3 Dies ist eine ASCII-Zeichencodierung, die normalerweise für südeuropäische Sprachen verwendet wird. Sie ist auch bekannt als Latin-3.</p> <p>ISO-8859-9 Dies ist eine ASCII-Zeichencodierung, die normalerweise für Türkisch verwendet wird. Sie ist auch bekannt als Latin-5.</p> <p>CP850 Dies ist eine ASCII-Codeseite, die für westeuropäische Sprachen verwendet wird.</p> <p>CP500 Dies ist eine EBCDIC-Codeseite, die für westeuropäische Sprachen verwendet wird.</p> <p>Shift_JIS Dies ist eine Zeichencodierung für Japanisch.</p> <p>MS932 Dies ist eine Shift_JIS-Erweiterung von Microsoft, die besondere NEC-Zeichen, eine NEC-Auswahl aus IBM-Erweiterungen und IBM-Erweiterungen einschließt.</p> <p>CP1047 Dies ist eine EBCDIC-Codeseite mit dem vollständigen Latin-1-Zeichensatz.</p>

Feldname	Beschreibung
„Validate	<p data-bbox="535 367 1429 556">Diese Option gibt an, ob die Datei überprüft werden soll, um sicherzustellen, dass der HL7 2.7-Standard eingehalten wurde. Wenn die Überprüfung bei irgendeiner Nachricht in der Datei einen Fehler ausgibt, wird sie als falsch formatierter Datensatz behandelt. In diesem Fall gelten die Optionen für falsch formatierte Datensätze, die für den Auftrag (im Enterprise Designer unter Bearbeiten > Auftragsoptionen) oder für das System (in der Management Console) angegeben wurden.</p> <p data-bbox="535 567 1429 703">Pflichtfelder Aktivieren Sie dieses Kästchen, wenn Sie sicherstellen möchten, dass jedes Segment, jedes Feld, jede Komponente und Unterkomponente die Elemente enthalten, die nach HL7-2.7-Standard erforderlich sind.</p> <p data-bbox="535 714 1429 840">Länge Aktivieren Sie dieses Kästchen, wenn Sie sicherstellen möchten, dass jedes Element die Anforderungen an Mindest- und Maximallänge des Elements erfüllt, wie sie im HL7-2.7-Standard definiert sind.</p>

Feldname	Beschreibung
----------	--------------

Unerwartete ignorieren	
------------------------	--

Feldname

Beschreibung

Wählen Sie diese Optionen aus, wenn Sie zulassen möchten, dass Nachrichten Segmente, Felder, Komponenten und Unterkomponenten an unerwarteten Positionen enthalten. Die erwarteten Positionen sind im HL7-Standard oder im Falle von benutzerdefinierten Nachrichtentypen im Tool HL7-Schema-Verwaltung im Enterprise Designer definiert.

Hier dient als Beispiel das folgende benutzerdefinierte Nachrichtenschema:

```
MSH
[PID]
{ZSS}
PV1
NK1
{[DG1]}
```

zusammen mit folgenden Daten:

```
MSH|^~\&|Pharm|GenHosp|CIS|GenHosp|198807050000||RAS^O17|RAS1234|P|2.7
ZSS|100|abc
PID|1234||PATID1234^5^M11^ADT1^MR^GOOD HEALTH
HOSPITAL~123456789^^^USSSA^SS|
PV1||O|O/R|||0148^ADDISON, JAMES|0148^ADDISON, JAMES
NK1|Son|John
```

In diesem Fall ist das Segment `PID` an unerwarteter Position, da es sich vor dem Segment `ZSS` befindet.

Nachrichten, die Elemente an unerwarteten Positionen enthalten, werden als falsch formatierte Datensätze behandelt. In diesem Fall gelten die Optionen für falsch formatierte Datensätze, die für den Auftrag (im Enterprise Designer unter **Bearbeiten > Auftragsoptionen**) oder für das System (in der Management Console) angegeben wurden.

Standardmäßig sind alle Optionen unter **Unerwartete ignorieren** aktiviert, damit so viele Datensätze wie möglich erfolgreich verarbeitet werden können.

- | | |
|-------------------------|---|
| Segmente | Aktivieren Sie dieses Kästchen, um zuzulassen, dass Nachrichten Segmente enthalten, die nicht im HL7 2.7-Standard definiert sind. Unerwartete Segmente werden ignoriert, und andere Segmente in der Nachricht werden verarbeitet. |
| Felder | Aktivieren Sie dieses Kästchen, um zuzulassen, dass Segmente Felder enthalten, die nicht im HL7 2.7-Standard definiert sind. Unerwartete Felder werden ignoriert, und andere Felder im Segment werden verarbeitet. |
| Komponenten | Aktivieren Sie dieses Kästchen, um zuzulassen, dass Felder Komponenten enthalten, die nicht im HL7 2.7-Standard definiert sind. Unerwartete Komponenten werden ignoriert, und andere Komponenten im Feld werden verarbeitet. |
| Unterkomponenten | Aktivieren Sie dieses Kästchen, um zuzulassen, dass |

Feldname

Beschreibung

Komponenten Unterkomponenten enthalten, die nicht im HL7 2.7-Standard definiert sind. Unerwartete Unterkomponenten in der Komponente werden ignoriert, und andere Unterkomponenten werden verarbeitet.

Registerkarte „Felder“

Auf der Registerkarte **Felder** werden die Segmente, Felder, Komponenten und Unterkomponenten angezeigt. Verwenden Sie die Registerkarte **Felder**, um die Daten auszuwählen, die Sie in den Datenfluss einlesen möchten.

Segmentgruppen sind Sammlungen von Segmenten, die zusammen verwendet werden und eine Datenkategorie enthalten. Sie werden unter Verwendung eines Nummerierungssystems angezeigt, das deutlich macht, wo die Gruppe im Nachrichtenschema erscheint. Jede Segmentgruppe verfügt über eine Beschriftung „Group_n“, wobei n eine Zahl ist, die der Position der Gruppe im Nachrichtenschema entspricht. Hier ein Beispiel zur Veranschaulichung des Nummerierungssystems:

```

+ MSH
+ SFT
+ UAC
+ NTE
+ RAS_017_Group_1
+ RAS_017_Group_2

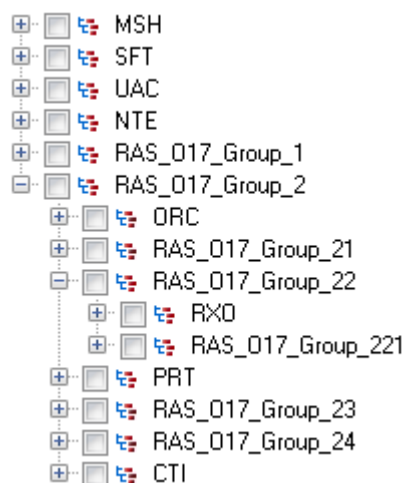
```

In diesem Beispiel wird die Feldliste für die Nachricht „RAS^017“ angezeigt. Diese Nachricht enthält zwei Segmentgruppen: RAS_017_Group_1 und RAS_017_Group_2. Die Segmentgruppe „Group_1“ verweist auf die erste Segmentgruppe im Schema „RAS^017“, und die zweite Gruppe „Group_2“ verweist auf die zweite Gruppe, die im Schema „RAS^017“ aufgelistet ist.

Um festzustellen, welche Segmentgruppe durch „Group_1“ und „Group_2“ repräsentiert wird, suchen Sie nach der Beschreibung der Nachricht „RAS^017“ im Dokument *HL7 Version 2.7 Standard*. Sie können eine Kopie dieses Dokuments unter www.hl7.org herunterladen.

Suchen Sie in der Beschreibung der Nachricht nach der ersten Gruppe. Diese ist bei „RAS^017“ die Gruppe „PATIENT“. Die zweite Gruppe im Schema ist die Gruppe „ORDER“.

Bei Segmentgruppen, die unter einer Segmentgruppe verschachtelt sind, wird eine zusätzliche Nummer an ihre Gruppennummer angehängt. Beispielsweise repräsentiert „Group_21“ die erste verschachtelte Gruppe unter der zweiten Gruppe. Bei weiteren Untergruppen werden zusätzliche Nummern angehängt. Beispielsweise repräsentiert die Gruppe „Group_221“ bei der Nachricht „RAS^017“ die Segmentgruppe „ORDER_DETAIL_SUPPLEMENT“. Im Folgenden finden Sie ein Beispiel für verschachtelte Gruppen:



Die Steuerungselemente auf der Registerkarte **Felder** werden in der folgenden Tabelle beschrieben.

Tabelle 3: Registerkarte „Felder“

Name der Option	Beschreibung
Erneut generieren	<p>Klicken Sie auf diese Schaltfläche, um die Registerkarte Felder mit einer Liste aller Segmente, Felder, Komponenten und Unterkomponenten für den in der Eingabedatei enthaltenen Nachrichtentyp zu füllen. Alle Elemente des Nachrichtentyps werden basierend auf dem HL7-Schema angezeigt, unabhängig davon, ob die Eingabedatei all diese Elemente enthält. Wenn die Datei beispielsweise eine RAS-Nachricht enthält, wird das Schema für den gesamten RAS-Nachrichtentyp angezeigt, unabhängig davon, ob die Eingabedatei tatsächlich Daten für alle möglichen Segmente, Felder, Komponenten und Unterkomponenten enthält.</p> <p>Wenn Sie unter Verwendung des Tools HL7-Schema-Verwaltung im Enterprise Designer irgendwelche benutzerdefinierten Elemente definiert haben, werden auch diese Elemente aufgelistet.</p>
Alle erweitern	<p>Dies erweitert alle Elemente auf der Registerkarte „Felder“, sodass Sie alle Segmente, Felder, Komponenten und Unterkomponenten für den in der Datei enthaltenen Nachrichtentyp sehen können.</p>
Alle reduzieren	<p>Dies schließt alle Knoten in dieser Ansicht, sodass nur die Segmente angezeigt werden. Verwenden Sie diese Option, um einfach einen Überblick über die Segmente für die Nachrichtentypen in der Datei zu erhalten. Sie können dann einzelne Segmente erweitern, um die Felder, Komponenten und Unterkomponenten eines Segments anzuzeigen.</p>

Name der Option	Beschreibung
Alle auswählen	Aktivieren Sie dieses Kästchen, um Datenflussfelder für alle Segmente, Felder, Komponenten und Unterkomponenten für alle in der Datei enthaltenen Nachrichtentypen zu erstellen.

Abflachen von HL7-Daten

HL7-Daten sind hierarchisch organisiert. Da viele Schritte die Daten in einem flachen Format benötigen, müssen Sie die Daten möglicherweise abflachen, damit die Daten in Downstream-Schritten von Aktivitäten wie Adressenüberprüfung oder Geocoding verwendbar sind.

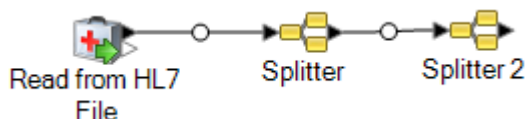
Die folgenden Schritte beschreiben, wie Sie einen Splitter-Schritt verwenden, um HL7-Daten abzuflachen.

1. Fügen Sie Ihrem Datenfluss den Schritt „Read from HL7 File“ hinzu und konfigurieren Sie ihn.
2. Fügen Sie einen Splitter-Schritt hinzu und verbinden Sie ihn mit dem Schritt „Read from HL7 File“.
3. Fügen Sie je nach Bedarf weitere Splitter-Schritte hinzu, sodass Sie über einen Splitter-Schritt für jedes Segment, jedes Feld und jede Komponente verfügen, die Sie abflachen möchten.

Anmerkung: Sie müssen nur die Daten abflachen, die in einem Downstream-Schritt verarbeitet werden sollen. Andere Daten können im hierarchischen Format verbleiben. Wenn Sie beispielsweise nur Adressdaten verarbeiten möchten, müssen Sie auch nur die Adressdaten abflachen.

4. Verbinden Sie alle Splitter-Schritte.

Sie sollten nun einen Datenfluss haben, der wie folgt aussieht:



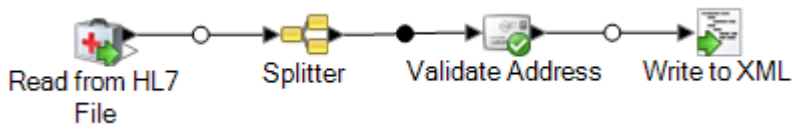
5. Doppelklicken Sie auf den ersten Splitter-Schritt, um die Schrittoptionen zu öffnen.
6. Wählen Sie im Feld **Unterteilen bei** das Segment, das Feld oder die Komponente zum Abflachen aus.
7. Klicken Sie auf **OK**.
8. Konfigurieren Sie jeden zusätzlichen Splitter-Schritt. Wählen Sie im Feld **Unterteilen bei** jedes einzelnen Splitter-Schrittes ein anderes Segment, Feld oder eine andere Komponente aus.
9. Fügen Sie nach Bedarf weitere Schritte nach dem letzten Splitter-Schritt hinzu, um Ihren Datenfluss abzuschließen.

Beispiel

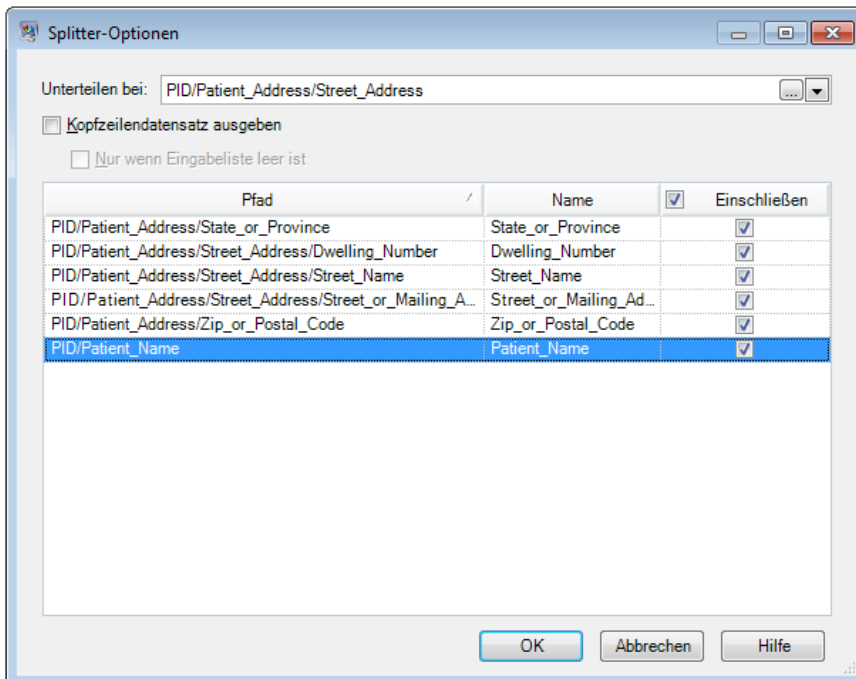
Sie haben folgende HL7-Daten und möchten die Adressen im Segment PID überprüfen.

```
MSH|^~\&||.|||199908180016||RAS^O17|ADT.1.1698594|P|2.7
PID|1||000395122||SMITH^JOHN^D||19880517180606|M|||One Global
View^^Troy^NY^12180||(630)123-4567|||S||12354768|87654321
```

Dazu müssen Sie diese Adressdaten in flache Daten konvertieren, damit sie vom „Validate Address“-Schritt verarbeitet werden können. Sie erstellen also einen Datenfluss, der einen Splitter-Schritt gefolgt von einem „Validate Address“-Schritt enthält, was wie folgt aussehen kann:



Der Splitter-Schritt ist konfiguriert, die Nachricht bei der Komponente „PID/Patient_Address/Street_Address“ zu unterteilen und diese Daten in flache Daten zu konvertieren.



Der Kanal, der den Splitter-Schritt mit dem „Validate Address“-Schritt verbindet, benennt die Felder um, sodass die von „Validate Address“ benötigten Feldnamen verwendet werden: „Street_or_Mailing_Addres“ wird in „AddressLine1“ umbenannt, „State_or_Province“ wird in „StateProvince“ umbenannt und „Zip_or_Postal_Code“ wird in „PostalCode“ umbenannt.

In diesem Beispiel wird die Ausgabe in eine XML-Datei geschrieben, die diese Daten enthält.

```
<?xml version='1.0' encoding='UTF-8'?>
<XmlRoot xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <PatientInformation>
    <Confidence>95</Confidence>
    <RecordType>Normal</RecordType>
    <CountryLevel>A</CountryLevel>
    <ProcessedBy>USA</ProcessedBy>
    <MatchScore>0</MatchScore>
    <AddressLine1>1 Global Vw</AddressLine1>
    <City>Troy</City>
    <StateProvince>NY</StateProvince>
    <PostalCode>12180-8371</PostalCode>
    <PostalCode.Base>12180</PostalCode.Base>
    <PostalCode.AddOn>8371</PostalCode.AddOn>
    <Country>United States Of America</Country>
    <Patient_Name>
      <Family_Name>
        <Surname>SMITH</Surname>
      </Family_Name>
      <Given_Name>JOHN</Given_Name>
    </Patient_Name>
    <Second_and_Further_Given_Names_or_Initials_Thereof>
      D
    </Second_and_Further_Given_Names_or_Initials_Thereof>
  </PatientInformation>
</XmlRoot>
```

Hinzufügen einer benutzerdefinierten HL7-Nachricht

Im Schritt „Read from HL7 File“ werden die Nachrichten anhand des HL7 2.7-Schemas überprüft. Ihre HL7-Daten können jedoch möglicherweise Nachrichten enthalten, die nicht Teil des HL7-Standards sind. Wenn Sie möchten, dass der Schritt „Read from HL7 File“ Ihre benutzerdefinierten HL7-Daten überprüft, müssen Sie ein benutzerdefiniertes HL7-Schema erstellen. In diesem Thema wird beschrieben, wie Sie mit dem Tool HL7-Schema-Verwaltung ein benutzerdefiniertes HL7-Schema erstellen. Weitere Informationen zu HL7 finden Sie unter www.hl7.org.

1. Navigieren Sie im Enterprise Designer zu **Tools > HL7-Schema-Verwaltung**.

Das Fenster „HL7-Schema-Verwaltung“ wird geöffnet, das eine Liste unterstützter Nachrichten enthält. Diese Nachrichten wurden durch HL7 vorab definiert.

2. Klicken Sie im Fenster **HL7-Schema-Verwaltung** auf **Hinzufügen**.

3. Geben Sie im Feld **Nachrichtentyp** einen Typ für eine benutzerdefinierte HL7-Nachricht an.

Der Nachrichtentyp zeigt an, welche gesundheitsbezogenen Informationen in der Nachricht zur Verfügung gestellt werden. Beispielsweise wird ein Nachrichtentyp ADT (Admin Discharge Transfer) verwendet, um den Patientenzustand innerhalb einer Einrichtung des Gesundheitswesens auszutauschen, und ein Nachrichtentyp ORU (Observation Result) wird verwendet, um Beobachtungen und Ergebnisse aus dem LIS (Lab Information System) zum HIS (Hospital Information System) zu übertragen.

4. Geben Sie im Feld **Auslöseereignis** einen Ereigniscode an.

Das Auslöseereignis ist ein Ereignis in der realen Welt, das Kommunikation und das Versenden einer Nachricht einleitet. Sowohl der Nachrichtentyp als auch das Auslöseereignis sind im Feld „MSH-9“ der Nachricht vorhanden. Beispielsweise kann das Feld MSH-9 den Wert „ADT^A01“ enthalten. Das bedeutet, dass der HL7-Nachrichtentyp „ADT“ und das Auslöseereignis „A01“ ist.

5. Geben Sie im Feld **Beschreibung** eine Beschreibung für eine benutzerdefinierte HL7-Nachricht ein.

Dieses Feld soll das Verstehen eines Nachrichtentyps erleichtern. Wenn Sie beispielsweise einen Nachrichtentyp „XYZ“ hinzufügen, können Sie eine Beschreibung angeben, dass dieser Nachrichtentyp dazu verwendet wird, den Patientenzustand innerhalb einer Einrichtung des Gesundheitswesens auszutauschen.

Sie sehen nun unter der **Definition** eine neu erstellte Nachricht. Klicken Sie auf das Pluszeichen, um die Nachricht zu erweitern. Das Segment „MSH“ wird automatisch hinzugefügt.

6. Vorhandenes Segment zu einer Nachricht hinzufügen

- a) Klicken Sie auf **Segment auswählen**.
- b) Wählen Sie die Segmente aus, die Sie zur Nachricht hinzufügen möchten, und klicken Sie auf **OK**.

Im Grid **Segmentschema** wird ein Schema des ausgewählten Segments angezeigt, und die aktivierten Nachrichten werden zum Nachrichtenschema hinzugefügt.

7. Benutzerdefiniertes Segment zu einer Nachricht hinzufügen

- a) Klicken Sie auf **Segment auswählen**.
- b) Klicken Sie auf **Segment hinzufügen**.
- c) Geben Sie im Feld **Name** einen Namen für das Segment ein und klicken Sie auf **OK**.

Das neu hinzugefügte Segment wird unten in der Liste **Unterstützte Segmente** angezeigt.

- d) Wählen Sie das hinzugefügte benutzerdefinierte Segment aus und klicken Sie auf die Schaltfläche **Feld hinzufügen**.

- e) Geben Sie im Feld **Name** einen Feldnamen für das ausgewählte Segment an.
Das Segment „PID (Patient Information)“ enthält beispielsweise die Feldnamen für Patienten-ID, Patientename, Patientenadresse, Ländercode usw.
- f) Wählen Sie im Feld **Typ** einen geeigneten Datentyp aus.
HL7-Datentypen definieren, welche Art von Daten in einem Feld enthalten sein kann, und werden in der gesamten HL7-Nachrichtenstruktur verwendet. Beispielsweise steht ST für Zeichenfolge (String), TX für Text und FT für formatierte Daten (Formatted Data).
- g) Geben Sie im Feld **Normative Länge** die Mindest- und die Maximallänge des Feldes mithilfe des folgenden Formats an: m..n. Sie können auch eine Liste möglicher Werte für die Feldlänge angeben, indem Sie das folgende Format verwenden: x,y,z.
Beispielsweise bedeutet eine Länge von 1..3, dass die Länge des Elements entweder 1, 2 oder 3 sein kann, und eine Länge von 1, 3, 4 bedeutet, dass die Länge des Elements entweder 1, 3 oder 4, aber nicht 2 sein kann. Ein anderer Wert als 1, 3 oder 4 würde als ungültig behandelt.
- h) Geben Sie im Feld **Optionalität** an, ob ein Feld optional oder erforderlich ist.
- O**
Das Feld ist optional.
- R**
Das Feld ist ein Pflichtfeld.
- i) Wenn Sie zulassen möchten, dass das Feld mehr als einmal im Segment enthalten sein darf, aktivieren Sie im Feld **Wiederholung** das Kästchen **Wiederholung** und geben Sie an, wie oft das Feld verwendet werden kann.
Beispielsweise würde der Wert 3 bedeuten, dass das Feld dreimal enthalten sein darf. Wenn Sie hier nichts angeben, ist nur ein Auftreten zulässig, was bedeutet, dass sich dieses Feld nicht wiederholen darf.
8. Klicken Sie auf **OK**.
Sie können auch unter den Segmenteigenschaften die Optionen **Optional** und **Wiederholung** auswählen.
9. Wählen Sie **Optional** aus, um das ausgewählte Segment optional zu machen, und wählen Sie **Wiederholung** aus, um Wiederholungen eines ausgewählten Segments in einer Nachricht zuzulassen.
10. Klicken Sie auf **OK**.
Die neu hinzugefügte Nachricht wird unten in der Liste angezeigt.

Read from NoSQL DB

Der Schritt „Read from NoSQL DB“ liest Daten aus einer Datenbanktabelle als Eingabe in einen Datenfluss ein. Dieser Schritt unterstützt die Datenbanktypen „MongoDB“ und „Couchbase“.

Registerkarte „Allgemein“

Feldname	Beschreibung
Verbindung	<p>Wählen Sie die gewünschte Datenbankverbindung aus. Die Auswahlmöglichkeiten variieren, je nachdem, welche Verbindungen in der NoSQL-Verbindungsverwaltung im Menü Tools des Enterprise Designer definiert sind. Wenn Sie eine neue Datenbankverbindung herstellen oder eine vorhandene Datenbankverbindung ändern oder löschen möchten, klicken Sie auf Verwalten.</p> <p>Verbindungsname Geben Sie einen Namen für die Verbindung ein. Sie können den Namen frei wählen.</p> <p>NoSQL-Datenbank Wählen Sie den passenden Datenbanktyp aus.</p> <p>Benutzername Geben Sie den Benutzernamen zum Herstellen der Verbindung mit der Datenbank ein.</p> <p>Anmerkung: Bei Couchbase ist der Benutzername nicht obligatorisch. Sie können einen beliebigen Benutzernamen eingeben. Die Verbindung wird erfolgreich hergestellt werden, solange Sie das korrekte Kennwort verwenden, das Sie beim Erstellen des Bucket angegeben haben.</p> <p>Kennwort Geben Sie das Kennwort zum Herstellen der Verbindung mit der Datenbank ein.</p> <p>Hostname Geben Sie den Hostnamen an, auf dem die Datenbank läuft.</p> <p>Port Geben Sie den Port zum Herstellen einer Verbindung mit der Datenbank an.</p> <p>Datenbank Geben Sie die Datenbank an, aus der Daten abgerufen werden sollen.</p> <p>Anmerkung: Der auf der Benutzeroberfläche verwendete Begriff Datenbank wird bei Couchbase „Bucket“ genannt.</p>
Tabelle/Ansicht	<p>Gibt die Sammlung oder Ansicht in der Datenbank an, die Sie abfragen möchten.</p> <p>Anmerkung: Der auf der Benutzeroberfläche verwendete Begriff Tabelle/Ansicht wird bei MongoDB <i>Sammlung</i> und bei Couchbase <i>Ansicht</i> genannt.</p>

Feldname	Beschreibung
Schemadatei	<p>Klicken Sie auf die Schaltfläche „Durchsuchen“ (...), um eine JSON-Schemadatei auszuwählen. Diese Datei ist optional. Die Felder auf der Registerkarte „Felder“ können entweder über die Schemadatei oder die Datenbanktabelle/-ansicht erneut generiert werden.</p> <p>Um den ausgewählten Dateipfad zu löschen, klicken Sie auf Löschen.</p> <p>Anmerkung: Die Felder werden immer über die Schemadatei generiert, wenn eine ausgewählt ist.</p>
Wo	<p>Geben Sie die erforderlichen Filterkriterien falls vorhanden unter Verwendung der MongoDB-Syntax ein, um bestimmte Datensätze abzurufen. Lassen Sie das Feld leer, wenn keine Filterkriterien erforderlich sind.</p> <p>Die folgende Syntax gilt für eine Klausel mit einem Operator <i>Gleich</i>:</p> <pre>{ "<column name>" : "<filter value>" }</pre> <p>Sie können mehrere Klauseln mithilfe der erforderlichen Operatoren zusammenführen. Eine Liste der in der <i>where</i>-Klausel unterstützten Operatoren finden Sie unter http://docs.mongodb.org/manual/reference/operator/query/.</p> <p>Um beispielsweise Datensätze abzurufen, in denen der Wert der Spalte <i>customer_name</i> mit dem Wert <i>John</i> übereinstimmt und der Wert der Spalte <i>customer_age</i> größer oder gleich 45 ist, geben Sie Folgendes ein:</p> <pre>{ \$and: [{ "customer_name": "John" }, { \$gte: ["customer_age", "45"] }] }</pre> <p>Achtung: Achten Sie darauf, in diesem Feld nicht das Schlüsselwort <i>where</i> zu verwenden.</p> <p>Anmerkung: Derzeit wird das Feld nur bei Auswahl einer MongoDB-Verbindung angezeigt.</p>
Fehlende Felder ignorieren	<p>Im Schema definierte Felder, die im aktuellen Datensatz nicht vorhanden sind, werden nicht an den nächsten Schritt weitergeleitet, wenn die Option ausgewählt ist.</p> <p>Anmerkung: Wenn Sie diese Option nicht aktivieren, werden Felder, die nicht in der Datenbanktabelle oder -ansicht vorhanden sind, hinzugefügt und mit dem Wert NULL verarbeitet.</p>

Feldname	Beschreibung
Vorschau	<p>Dies zeigt die Datensätze aus der ausgewählte Tabelle an.</p> <p>Anmerkung: Bei MongoDB-Datenquellen werden beim Klicken auf Vorschau die gefilterten Datensätze angezeigt, wenn eine oder mehrere <code>where</code>-Klauseln in das Feld Where eingegeben wurden. Wenn keine <code>WHERE</code>-Klausel eingegeben wurde, werden in der Vorschau alle Datensätze angezeigt.</p> <p>Anmerkung: Bei Couchbase-Datenquellen wird beim Klicken auf Vorschau auch das hinzugefügte Feld <code>_id</code> mit dem enthaltenen Schlüssel angezeigt. Wenn der Datensatz bereits über ein Feld <code>_id</code> verfügt, überschreibt das hinzugefügte Feld <code>_id</code> das vorher vorhandene Feld bei der Vorschau der Felder.</p>
Alle erweitern	Dies erweitert die Elemente im Vorschaubaum.
Alle reduzieren	Dies reduziert die Elemente im Vorschaubaum.

Registerkarte „Felder“

Die Registerkarte „Felder“ ermöglicht Ihnen die Auswahl der Daten, die Sie dem nächsten Schritt übergeben möchten. Weitere Informationen finden Sie unter [Definieren von Feldern in einer NoSQL-Datenbank](#) auf Seite 194.

Definieren von Feldern in einer NoSQL-Datenbank

Auf der Registerkarte **Felder** werden die Felder und ihre Typen angezeigt, wie sie in der Schemadatei der NoSQL-Datenbank definiert wurden.

1. Klicken Sie auf der Registerkarte **Felder** auf die Option **Erneut generieren**.

Daraufhin werden die aggregierten Daten basierend auf den ersten 50 Datensätzen generiert. Die Daten werden im folgenden Format angezeigt: `Fieldname (datatype)`.

Anmerkung: Beim Durchsuchen der Schemadatei werden die Felder unter Verwendung der Schemadatei generiert und die Tabelle oder Ansicht umgangen. Um die Schemadatei zurückzusetzen, klicken Sie auf **Löschen**.

Anmerkung: Beim Lesen der Daten aus einer Couchbase-Datenbank wird auch der Schlüssel jedes Datensatzes gelesen. Dieser Schlüssel wird als Teil des Datensatz gespeichert, indem ein Feld `_id` hinzugefügt wird, wenn die Felder erneut generiert werden. Der Schlüssel ist auch in den Daten enthalten, die zum nächsten Schritt gesendet werden. Wenn der Datensatz bereits über ein Feld `_id` verfügt, wird dies beim erneuten Generieren der Felder mit dem hinzugefügten Feld `_id` überschrieben.

2. Um den Namen und Typ eines Feldes zu ändern, markieren Sie das Feld und klicken Sie auf **Ändern**.
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 „Zeichenfolge“ beibehalten, wenn Sie nicht vorhaben, mathematische Operationen auf die Daten anzuwenden. Wenn Sie allerdings diese 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.

Der Schritt unterstützt die folgenden Datentypen:

double	Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit doppelter Genauigkeit zwischen 2^{-1074} und $(2-2^{-52}) \times 2^{1023}$ 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 $(2-2^{-23}) \times 2^{127}$ 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. Sie können auch zusätzliche Felder hinzufügen, die in der Tabelle oder Schemadatei nicht vorhanden sind. Klicken Sie auf **Hinzufügen**, um ein neues Feld hinzuzufügen. Um ein Feld zu entfernen, klicken Sie auf **Entfernen**.

Anmerkung: Sie können nur ein neues Feld des Typs Liste hinzufügen.

6. Klicken Sie auf **OK**.

Datenflussoptionen für NoSQL-Datenbanken

Diese Schritte beschreiben, wie Sie einen Datenfluss konfigurieren, um Laufzeitoptionen für NoSQL-Datenbanken zu unterstützen.

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. Erweitern Sie den „NoSQLDB“-Schritt.
6. Die Datenflussoptionen werden, wie in folgender Tabelle beschrieben, verfügbar gemacht:

Datenbank	Lesen	Schreiben
Mongo DB	Verbindung	Verbindung
	Tabelle	Tabelle
Couchbase DB	Verbindung	Verbindung
	Anzeigen	
	Name des Designdokuments	

Der ausgewählte „NoSQL DB“-Optionsname wird in den Feldern **Optionsname** und **Optionsbeschriftung** angezeigt. Dies ist der Optionsname, der zur Laufzeit angegeben werden muss, um diese Option festzulegen.

7. Geben Sie eine Beschreibung der Option in das Feld **Beschreibung** ein.
8. Wählen Sie im Feld **Ziel** die Option **Ausgewählte(r) Schritt/e** aus.
9. 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.
10. 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](#).

11. Klicken Sie auf **OK**.
12. Fügen Sie auf Wunsch weitere Optionen hinzu.
13. Klicken Sie im Dialogfeld **Datenflussoptionen** auf **OK**, sobald Sie damit fertig sind.
14. Speichern Sie den Datenfluss und machen Sie ihn verfügbar.

Read from SAP

Der „Read from SAP“-Schritt liest Daten aus einer SAP-Datenbank als Eingabe in einen Datenfluss ein. Er kann Daten aus einer einzelnen Tabelle oder aus mehreren Tabellen lesen. Beim Lesen aus mehreren Tabellen führt der Schritt einen Join-Vorgang durch, um zu bestimmen, welche Datensätze in den Datenfluss einzulesen sind.

Verbinden mit SAP

Um Daten aus SAP in einen Datenfluss mithilfe des „Read from SAP“-Schrittes einzulesen, müssen Sie eine Verbindung zwischen Spectrum™ Technology Platform und Ihrem SAP-System erstellen.

1. Öffnen Sie den SAP-Verbindungs-Manager. Sie können dies im Enterprise Designer unter **Tools > SAP-Verbindungsverwaltung** oder im „Read from SAP“-Schritt durch Klicken auf die Schaltfläche **Verwalten** neben dem Feld **Verbindung** erledigen.
2. Klicken Sie auf **Hinzufügen**.
3. Geben Sie im Feld **Verbindungsname** einen Verbindungsnamen ein.
4. Geben Sie in den anderen Feldern die Informationen zum SAP-Server ein, zu dem Sie eine Verbindung herstellen möchten. Bitten Sie Ihren SAP Basis-Administrator um die erforderlichen Informationen.

Wichtig: Die Benutzer-ID und das Kennwort müssen zu einem SAP-Konto mit Administratorberechtigungen gehören.

5. Klicken Sie auf **Testen**, um die Verbindung zu überprüfen.
6. Klicken Sie auf **OK**.

Sie verfügen jetzt über eine Verbindung, die Sie im „Read from SAP“-Schritt verwenden können, um Daten von SAP in einen Datenfluss zu lesen.

Lesen von Daten aus einer einzelnen SAP-Tabelle

Sie können den „Read from SAP“-Schritt so konfigurieren, dass Daten aus einer einzelnen Tabelle oder aus mehreren Tabellen in der SAP-Datenbank liest. Diese Schritte beschreiben, wie Sie „Read from SAP“ konfigurieren, um Daten aus einer einzelnen Tabelle zu lesen.

1. Ziehen Sie im Enterprise Designer „Read from SAP“ auf die Arbeitsfläche.
2. Doppelklicken Sie auf den „Read from SAP“-Schritt auf der Arbeitsfläche.

3. Wählen Sie im Feld **Verbindung** den SAP-Server aus, der die Daten enthält, die Sie in den Datenfluss einlesen möchten. Wenn für den von Ihnen benötigten SAP-Server keine Verbindung definiert ist, müssen Sie die Verbindung erstellen, indem Sie auf **Verwalten** klicken.
4. Wählen Sie im Feld **Quellentyp Einzeln** aus.
5. Klicken Sie auf **Auswählen**.
6. Wählen Sie die Tabelle aus, die Sie in den Datenfluss einlesen möchten, und klicken Sie dann auf **OK**.

Anmerkung: Es werden nur die ersten 200 Tabellen aufgelistet. Verwenden Sie das Such-Feature, um nach Tabellen zu suchen, die nicht unter den ersten 200 aufgelistet sind. Über das Suchfeld werden nur die Werte in den Spalten **Name** und **Label** durchsucht.

7. Um die Feldnamen anzuzeigen, die im Datenfluss verwendet werden, aktivieren Sie das Kästchen **Technischen Namen anzeigen**.

Felder verfügen in SAP über einen benutzerfreundlichen Namen, der zur Anzeige verwendet wird, und über einen eindeutigen Namen, der möglicherweise schlechter lesbar ist. Beispielsweise kann ein Feld über einen benutzerfreundlichen Namen „Distribution Channel und einen technischen Namen „DIS_CHANNEL“ verfügen. Um sicherzustellen, dass der Feldname im Datenfluss gültig ist, wird der technische Name als Feldname verwendet.

8. Aktivieren Sie das Kästchen in der Spalte **Einschließen** bei jedem Feld, das Sie in den Datenfluss einlesen möchten.
9. Klicken Sie auf **OK**.
10. Wenn Sie nur bestimmte Datensätze einlesen möchten, können Sie auf der Registerkarte **Filter** Filterbedingungen angeben. Damit ein Datensatz in den Datenfluss eingelesen wird, muss er alle von Ihnen definierten Bedingungen erfüllen.
11. Sie können die Leistung verbessern, indem Sie auf der Registerkarte **Laufzeit** eine angemessene Abrufgröße angeben.

Wählen Sie diese Option aus, um die Anzahl der Datensätze anzugeben, die jeweils aus der Datenbanktabelle gelesen werden sollen. Wenn die **Abrufgröße** beispielsweise 100 beträgt und die Gesamtzahl der zu lesenden Datensätze 1.000 beträgt, würden 10 Abrufe benötigt werden, um alle Datensätze zu lesen.

Das Einstellen einer optimalen Abrufgröße kann die Leistung deutlich verbessern.

Anmerkung: Sie können eine optimale Abrufgröße für Ihre Umgebung berechnen, indem Sie die Ausführungszeiten zwischen einem „Read from DB“-Schritt und einem „Write to Null“-Schritt in Tests ermitteln. Weitere Informationen finden Sie unter [Ermitteln einer idealen Abrufgröße](#) auf Seite 329.

Die standardmäßige Abrufgröße für „Read from SAP“ ist 10.000.

12. Klicken Sie auf **OK**.

Der „Read from SAP“-Schritt ist nun so konfiguriert, dass er Daten aus einer einzelnen Tabelle in der SAP-Datenbank in den Datenfluss einliest.

Lesen von Daten aus mehreren SAP-Tabellen

Sie können den „Read from SAP“-Schritt so konfigurieren, dass Daten aus einer einzelnen Tabelle oder aus mehreren Tabellen in der SAP-Datenbank liest. Diese Schritte beschreiben, wie Sie „Read from SAP“ konfigurieren, um Daten aus mehreren Tabellen zu lesen. Um Daten aus mehreren Tabellen zu lesen, müssen Sie eine JOIN-Anweisung definieren, um Daten in einem einzelnen Strom zu kombinieren.

1. Ziehen Sie im Enterprise Designer „Read from SAP“ auf die Arbeitsfläche.
2. Doppelklicken Sie auf den „Read from SAP“-Schritt auf der Arbeitsfläche.
3. Wählen Sie im Feld **Verbindung** den SAP-Server aus, der die Daten enthält, die Sie in den Datenfluss einlesen möchten. Wenn für den von Ihnen benötigten SAP-Server keine Verbindung definiert ist, müssen Sie die Verbindung erstellen, indem Sie auf **Verwalten** klicken.
4. Wählen Sie im Feld **Quellentyp Mehrere** aus.
5. Klicken Sie auf **Hinzufügen**.
6. Wählen Sie die Tabellen aus, die Sie in den Datenfluss einlesen möchten, und klicken Sie dann auf **OK**.

Anmerkung: Es werden nur die ersten 200 Tabellen aufgelistet. Verwenden Sie das Such-Feature, um nach Tabellen zu suchen, die nicht unter den ersten 200 aufgelistet sind. Über das Suchfeld werden nur die Werte in den Spalten **Name** und **Label** durchsucht.

7. Wählen Sie die erste Tabelle in der Liste aus und klicken Sie auf **Beziehung erstellen**. Dies ist die Quelltablelle.
8. Wählen Sie im Feld **Quellschlüssel** die Spalte der Quelltablelle aus, deren Wert verwendet wird, um Datensätze mit Datensätzen aus den anderen Tabellen zu vergleichen.
9. Wählen Sie im Feld **Join-Typ** eine der folgenden Optionen aus:

INNER JOIN	Es werden nur Werte zurückgegeben, bei denen Quell- und Zieltabellen übereinstimmen.
LEFT JOIN	Es werden alle Datensätze der Quelltablelle zurückgegeben, auch wenn keine Übereinstimmung zwischen Quell- und Zieltabellen vorhanden sind. Bei dieser Option werden alle Datensätze der Quelltablelle zuzüglich aller übereinstimmenden Datensätze der Zieltabelle zurückgegeben.
10. Wählen Sie im Feld **Tabelle** die Zieltabelle aus.
11. Wählen Sie im Feld **Tabellenschlüssel** die Spalte der Zieltabelle aus, die die Daten enthält, die Sie mit den Daten im Feld **Quellschlüssel** vergleichen möchten, um zu bestimmen, ob der Datensatz die Join-Bedingung erfüllt.
12. Klicken Sie auf **OK**.
13. Klicken Sie auf **Schema auswählen**.

14. Wählen Sie die Felder aus, die Sie in den Datenfluss einlesen möchten. Um die Feldnamen anzuzeigen, die im Datenfluss verwendet werden, aktivieren Sie das Kästchen **Technischen Namen anzeigen**.

Felder verfügen in SAP über einen benutzerfreundlichen Namen, der zur Anzeige verwendet wird, und über einen eindeutigen Namen, der möglicherweise schlechter lesbar ist. Beispielsweise kann ein Feld über einen benutzerfreundlichen Namen „Distribution Channel und einen technischen Namen „DIS_CHANNEL“ verfügen. Um sicherzustellen, dass der Feldname im Datenfluss gültig ist, wird der technische Name als Feldname verwendet.

15. Klicken Sie auf **OK**.
16. Wenn Sie nur bestimmte Datensätze einlesen möchten, können Sie auf der Registerkarte **Filter** Filterbedingungen angeben. Damit ein Datensatz in den Datenfluss eingelesen wird, muss er alle von Ihnen definierten Bedingungen erfüllen.
17. Sie können die Leistung verbessern, indem Sie auf der Registerkarte **Laufzeit** eine angemessene Abrufgröße angeben.

Wählen Sie diese Option aus, um die Anzahl der Datensätze anzugeben, die jeweils aus der Datenbanktabelle gelesen werden sollen. Wenn die **Abrufgröße** beispielsweise 100 beträgt und die Gesamtzahl der zu lesenden Datensätze 1.000 beträgt, würden 10 Abrufe benötigt werden, um alle Datensätze zu lesen.

Das Einstellen einer optimalen Abrufgröße kann die Leistung deutlich verbessern.

Anmerkung: Sie können eine optimale Abrufgröße für Ihre Umgebung berechnen, in dem Sie die Ausführungszeiten zwischen einem „Read from DB“-Schritt und einem „Write to Null“-Schritt in Tests ermitteln. Weitere Informationen finden Sie unter [Ermitteln einer idealen Abrufgröße](#) auf Seite 329.

Die standardmäßige Abrufgröße für „Read from SAP“ ist 10.000.

Der „Read from SAP“-Schritt ist nun so konfiguriert, dass er Daten aus mehreren Tabellen in der SAP-Datenbank in den Datenfluss einliest.

Filtern von Datensätzen in „Read from SAP“

Die Filtereinstellungen in „Read from SAP“ ermöglichen Ihnen, eine Teilmenge von Datensätzen aus einer SAP-Tabelle zu lesen, anstatt alle Datensätze in der Tabelle zu lesen. Um Datensätze zu filtern, geben Sie die Werte an, die ein Datensatz enthalten muss, damit er in den Datenfluss eingelesen wird. Wenn Sie keine Filterbedingungen angeben, werden alle Datensätze in der Tabelle in den Datenfluss eingelesen. Die Verwendung von Filterbedingungen ist optional.

Anmerkung: Wenn der „Read from SAP“-Schritt konfiguriert wurde, Daten aus mehreren SAP-Tabellen zu lesen, wird der Filter nach Ausführen des JOIN-Vorgangs angewendet.

1. Klicken Sie im Schritt „Read from SAP“ auf die Registerkarte **Filter**.

2. Klicken Sie auf **Hinzufügen**.
3. Wählen Sie im Feld **Tabellename** die Tabelle aus, die die zu filternden Datensätze enthält.
4. Wählen Sie im Feld **Filtern nach** das Feld aus, das die Daten enthält, die Sie als Basis für die Filterung verwenden möchten.
5. Wählen Sie einen der folgenden Operatoren aus:

Anmerkung: Welche Operatoren Ihnen zur Verfügung stehen, hängt vom Datentyp des Feldes ab, das Sie filtern.

Operator	Beschreibung
Enthält	Prüft, ob die Zeichenfolge den angegebenen Wert enthält.
Gleich	Prüft, ob der Wert im Feld mit dem angegebenen Wert übereinstimmt.
Ungleich	Prüft, ob der Wert im Feld nicht mit dem angegebenen Wert übereinstimmt.
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.
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.
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.
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.
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.
Beginnt mit	Prüft, ob das Feld mit dem angegebenen Wert beginnt.
Endet mit	Prüft, ob das Feld mit dem angegebenen Wert endet.

6. Geben Sie den Wert ein, mit der Wert des ausgewählten Feldes verglichen werden soll.
7. Klicken Sie auf **OK**.
8. Fügen Sie zusätzliche Filterbedingungen hinzu, falls erforderlich.

Anmerkung: Wenn Sie mehrere Filterbedingungen angeben, müssen alle Filterbedingungen wahr sein, damit der Datensatz in den Datenfluss eingelesen wird. Wenn eine der Bedingungen nicht wahr ist, wird der Datensatz nicht in den Datenfluss eingelesen.

Read from Spreadsheet

Der „Read from Spreadsheet“-Schritt liest Daten aus einer Excel-Arbeitsmappe als Eingabe in einen Datenfluss ein. Die Dateiformate .xls und .xlsx werden beide unterstützt.

Registerkarte „Dateieigenschaften“

Die Registerkarte **Dateieigenschaften** enthält Optionen, um die Arbeitsmappe und Daten anzugeben, die in einen Datenfluss eingelesen werden sollen.

Feldname	Beschreibung
Servername	Dies gibt an, ob sich die von Ihnen als Eingabe ausgewählte Datei auf dem Computer mit dem Enterprise Designer oder auf dem Spectrum™ Technology Platform-Server befindet. Wenn Sie eine Datei auf dem lokalen Computer auswählen, lautet der Servername „Arbeitsplatz“. Wenn Sie eine Datei auf dem Server auswählen, lautet der Servername Spectrum™ Technology Platform.
Dateiname	Gibt den Pfad zur Datei an. Klicken Sie auf die Ellipsenschaltfläche (...), um die gewünschte Datei zu lokalisieren. Anmerkung: 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.

Feldname	Beschreibung
Datenauswahl	<p>Gibt an, wie Daten in der Arbeitsmappe ausgewählt werden sollen, um in den Datenfluss eingelesen zu werden. Zur Auswahl stehen:</p> <p>Blattdaten Wählen Sie diese Option aus, wenn Sie alle Daten aus einem Blatt in der Arbeitsmappe einlesen möchten.</p> <p>Bereichsdaten Wählen Sie diese Option aus, um eine Teilmenge der Daten in einem Blatt einzulesen, indem Sie einen einzulesenden Zellenbereich angeben.</p> <p>Benannter Bereich Wählen Sie diese Option aus, um eine Teilmenge der Daten in einem Blatt einzulesen, indem Sie einen benannten Bereich der Arbeitsmappe angeben.</p>
Blattauswahl	<p>Wenn Sie im Feld Datenauswahl Blattdaten oder Bereichsdaten ausgewählt haben, verwenden Sie diese Option, um das Blatt auszuwählen, aus dem Daten in den Datenfluss eingelesen werden sollen.</p>
Bereich	<p>Wenn Sie im Feld Datenauswahl Bereichsdaten ausgewählt haben, verwenden Sie diese Option, um die Zellen anzugeben, die Start und Ende des Bereichs festlegen.</p>
Benannter Bereich	<p>Wenn Sie im Feld Datenauswahl Benannter Bereich ausgewählt haben, verwenden Sie diese Option, um den Namen des Bereichs anzugeben, der in den Datenfluss eingelesen werden soll. Bereiche werden in der Arbeitsmappe definiert. Wenn keine Bereiche aufgelistet sind, sind keine Bereiche in der Arbeitsmappe definiert.</p>
Kopfzeile	<p>Aktivieren Sie dieses Kästchen, um eine Zeile anzugeben, die Spaltenüberschriften enthält. Aus den Spaltenüberschriften werden die Feldnamen im Datenfluss generiert. Sie können Feldnamen jedoch auf der Registerkarte Felder ändern. Wenn Sie dieses Kästchen nicht aktivieren, erhalten die Datenflussfelder allgemeine Standardnamen wie Column1, Column2 usw.</p> <p>Die Position des von Ihnen angegebenen Header ist relativ zur Datenauswahl. Wenn Sie beispielsweise Bereichsdaten im Feld Datenauswahl auswählen und der Bereich mit der fünften Zeile beginnt und Sie 1 als Header angeben, wird die fünfte Zeile des Blatts als Header verwendet, da die fünfte Zeile des Blatts die erste Zeile des Bereichs ist.</p>
Versatz der Daten von Kopfzeile	<p>Wenn Sie einen Header angeben, gibt dieses Feld die erste Zeile relativ zum Header an, die Daten enthält. Wenn Sie beispielsweise 1 angeben, ist die erste Zeile unterhalb des Header die erste Zeile, die in den Datenfluss eingelesen wird. Wenn Sie 2 angeben, ist die zweite Zeile unterhalb des Header die erste Zeile, die in den Datenfluss eingelesen wird usw.</p>

Feldname	Beschreibung
Erste Datenzeile	Wenn Sie keinen Header angeben, gibt dieses Feld an, welche Zeile innerhalb der Datenauswahl die erste Zeile mit Daten ist, die in den Datenfluss eingelesen wird. Die Position der von Ihnen angegebenen Zeile ist relativ zur Datenauswahl. Wenn Sie beispielsweise Bereichsdaten im Feld Datenauswahl auswählen und der Bereich mit der fünften Zeile beginnt und Sie 1 als erste Datenzeile angeben, wird die fünfte Zeile als erste Datenzeile in den Datenfluss eingelesen.
Leere Zeilen ignorieren	Wählen Sie diese Option aus, wenn leere Zeilen nicht in den Datenfluss eingelesen werden sollen. Wenn Sie diese Option nicht aktivieren, führen leere Zeilen in der Arbeitsmappe zu leeren Datensätzen im Datenfluss. Anmerkung: Diese Option hat keine Auswirkungen auf die in der Vorschau angezeigten Daten. Leere Zeilen werden in der Vorschau immer angezeigt, auch wenn diese Option aktiviert ist.

Registerkarte „Felder“

Die Registerkarte **Felder** enthält Optionen zum Mapping der Daten aus der Arbeitsmappe zu Feldern im Datenfluss.

Option	Beschreibung
Erneut generieren	Klicken Sie auf diese Schaltfläche, um die Registerkarte „Felder“ mit den Feldern in der auf der Registerkarte Dateieigenschaften definierten Eingabedatei zu füllen.
Typ erkennen	Klicken Sie auf diese Schaltfläche, um den Datentyp aller Felder automatisch erkennen zu lassen. Sie können den Datentyp eines Feldes manuell ändern, indem Sie das Feld auswählen und auf Ändern klicken.
Ändern	Wählen Sie ein Feld aus, und klicken Sie dann auf diese Schaltfläche, um den Feldnamen oder den Datentyp zu ändern.

Read from Variable Format File

Der Schritt „Read from Variable Format File“ liest Daten aus einer Datei mit Datensätzen mit variablem Layout. Jeder Datensatz wird als Listenfeld eingelesen. Sie können das Tag angeben, das den übergeordneten Datensatztyp bezeichnet. Alle anderen Datensatztypen werden dann zu untergeordneten Listenfeldern.

Dateien mit variablem Format haben diese Merkmale:

- Datensätze in der Datei können unterschiedliche Felder und eine unterschiedliche Anzahl von Feldern enthalten.
- Jeder Datensatz muss ein Tag (gewöhnlich eine Zahl) enthalten, das den Typ des Datensatzes angibt.
- Hierarchische Beziehungen werden unterstützt.

Beispiel für eine Datei mit variablem Format

Dieses Beispiel zeigt eine Datei mit variablem Format, die Informationen zur Girokontoaktivität für zwei Kunden, Joe Smith und Anne Johnson, enthält. In diesem Beispiel ist die Datei eine durch Trennzeichen getrennte Datei, in der ein Komma als Feldtrennzeichen verwendet wird.

```
001 Joe,Smith,M,100 Main St,555-234-1290
100 CHK12904567,12/2/2007,6/1/2012,CHK
200 1000567,1/5/2012,Fashion Shoes,323.12
001 Anne,Johnson,F,1202 Lake St,555-222-4932
100 CHK238193875,1/21/2001,4/12/2012,CHK
200 1000232,3/5/2012,Blue Goose Grocery,132.11
200 1000232,3/8/2012,Trailway Bikes,540.00
```

Das erste Feld in jedem Datensatz enthält das Tag, das den Typ des Datensatzes und damit sein Format angibt:

- 001: Kundendatensatz
- 100: Kontodatensatz
- 200: Kontotransaktionsdatensatz

Bei durch Trennzeichen getrennten Dateien ist es häufig so, dass der Tagwert (001, 100, 200) in einer festen Anzahl von Byte am Beginn des Datensatzes steht (siehe das obige Beispiel).

Jeder Datensatz hat ein eigenes Format:

- 001: FirstName,LastName,Gender,Address,PhoneNumber
- 100: AccountID,DateOpened,ExpirationDate,TypeOfAccount
- 200: TransactionID,DateOfTransaction,Vendor,Amount

Das Datensatzformat 100 (Kontodatensatz) ist ein untergeordnetes Element des vorherigen 001-Datensatzes und das Datensatzformat 200 (Kontotransaktionsdatensatz) ein untergeordnetes Element des vorherigen 100-Datensatzes (Kontodatensatz). In der Beispieldatei gab es im Konto von Joe Smith, CHK12904567, am 05.01.2012 eine Transaktion mit dem Betrag 323,12 bei Fashion Shoes. Im Konto von Anne Johnson, CHK238193875, wiederum gab es zwei Transaktionen, eine am 05.03.2012 bei Blue Goose Grocery und eine am 08.03.2012 bei Trailway Bikes.

Registerkarte „Dateieigenschaften“

Name der Option	Beschreibung
Servername	<p>Dies gibt an, ob sich die von Ihnen als Eingabe ausgewählte Datei auf dem Computer mit dem Enterprise Designer oder auf dem Spectrum™ Technology Platform-Server befindet. Wenn Sie eine Datei auf dem lokalen Computer auswählen, lautet der Servername „Arbeitsplatz“. Wenn Sie eine Datei auf dem Server auswählen, lautet der Servername Spectrum™ Technology Platform.</p>
Dateiname	<p>Gibt den Pfad zur Datei an. Klicken Sie auf die Ellipsenschaltfläche (...), um zur gewünschten Datei zu navigieren.</p> <p>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.</p> <p>Beim Lesen einer Datei von einem HDFS-Dateiserver werden die folgenden Komprimierungsformate unterstützt:</p> <ol style="list-style-type: none"> 1. GZIP (.gz) 2. BZIP2 (.bz2) <p>Anmerkung: Die Erweiterung der Datei zeigt das zum Extrahieren zu verwendende Komprimierungsformat an.</p> <p>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.</p>

Name der Option	Beschreibung
Datensatztyp	<p>Dies ist das Format der Datensätze in der Datei. Wählen Sie eine Option:</p> <p>Zeilensequenziell 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 über feste Startzeichen- und Endzeichenpositionen verfügt.</p> <p>Feste Breite Dies ist eine Textdatei, in der jeder Datensatz eine bestimmte Anzahl von Zeichen lang ist und jedes Feld über feste Startzeichen- und Endzeichenpositionen verfügt.</p> <p>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.</p>

Name der Option	Beschreibung
Zeichencodierung	<p>Dies ist die Codierung der Textdatei. Wählen Sie eine der folgenden Optionen:</p> <p>UTF-8 Die Datei unterstützt alle Unicode-Zeichen und ist abwärtskompatibel mit ASCII. Weitere Informationen zu UTF finden Sie unter unicode.org/faq/utf_bom.html.</p> <p>UTF-16 Die Datei unterstützt alle Unicode-Zeichen, ist aber nicht abwärtskompatibel mit ASCII. Weitere Informationen zu UTF finden Sie unter unicode.org/faq/utf_bom.html.</p> <p>US-ASCII Die Zeichencodierung basiert auf der Reihenfolge des englischen Alphabets.</p> <p>UTF-16BE Die Datei ist UTF-16-codiert mit Big-Endian-Reihenfolge (höchstwertiges Byte zuerst).</p> <p>UTF-16LE Die Datei ist UTF-16-codiert mit Little-Endian-Reihenfolge (niedrigstwertiges Byte zuerst).</p> <p>ISO-8859-1 Dies ist eine ASCII-Zeichencodierung, die normalerweise für westeuropäische Sprachen verwendet wird. Sie ist auch bekannt als Latin-1.</p> <p>ISO-8859-3 Dies ist eine ASCII-Zeichencodierung, die normalerweise für südeuropäische Sprachen verwendet wird. Sie ist auch bekannt als Latin-3.</p> <p>ISO-8859-9 Dies ist eine ASCII-Zeichencodierung, die normalerweise für Türkisch verwendet wird. Sie ist auch bekannt als Latin-5.</p> <p>CP850 Dies ist eine ASCII-Codeseite, die für westeuropäische Sprachen verwendet wird.</p> <p>CP500 Dies ist eine EBCDIC-Codeseite, die für westeuropäische Sprachen verwendet wird.</p> <p>Shift_JIS Dies ist eine Zeichencodierung für Japanisch.</p> <p>MS932 Dies ist eine Shift_JIS-Erweiterung von Microsoft, die besondere NEC-Zeichen, eine NEC-Auswahl aus IBM-Erweiterungen und IBM-Erweiterungen einschließt.</p> <p>CP1047 Dies ist eine EBCDIC-Codeseite mit dem vollständigen Latin-1-Zeichensatz.</p>
Datensatzlänge	Geben Sie bei Dateien an, die Datensätze mit fester Breite enthalten, die genaue Anzahl von Zeichen jedes Datensatzes.

Name der Option	Beschreibung
Feldtrennzeichen	<p>Gibt das Zeichen an, das zum Trennen von Feldern in einer Datei mit Begrenzungszeichen verwendet wird.</p> <p>Im folgenden Beispiel wird ein Pipe-Zeichen () im Datensatz als Feldtrennzeichen verwendet:</p> <pre>7200 13TH ST MIAMI FL 33144</pre> <p>Die folgenden Zeichen stehen zur Definition als Feldtrennzeichen zur Verfügung:</p> <ul style="list-style-type: none"> • Leerzeichen • Tabulator • Komma • Punkt • Semikolon • Pipe <p>Wenn in der Datei ein anderes Zeichen als Feldtrennzeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Begrenzungszeichen auszuwählen.</p>
Tag-Trennzeichen	<p>Gibt das Zeichen an, das sich hinter dem Tag-Feld befindet, um in einer Datei mit Begrenzungszeichen das Identifizierungsfeld für jeden Datensatz abzugrenzen. Ein Tag-Trennzeichen muss ein einzelnes Zeichen sein.</p> <p>Standardmäßig stehen die folgenden Zeichen zur Definition als Tag-Trennzeichen zur Verfügung:</p> <ul style="list-style-type: none"> • Leerzeichen • Tabulator • Komma • Punkt • Semikolon • Pipe <p>Wenn in der Datei ein anderes Zeichen als Tag-Trennzeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein benutzerdefiniertes Tag-Trennzeichen hinzuzufügen und auszuwählen.</p> <p>Anmerkung: Standardmäßig ist das Datensatz-Trennzeichen dasselbe Zeichen wie für das ausgewählte Feldtrennzeichen. Um dieses Feld zu aktivieren und ein anderes Zeichen auszuwählen, deaktivieren Sie das Kästchen Wie Feldtrennzeichen.</p>
Wie Feldtrennzeichen	<p>Dies zeigt an, ob das Tag-Trennzeichen dasselbe ist wie das Feldtrennzeichen. Deaktivieren Sie dieses Kästchen, um ein anderes Zeichen als Tag-Trennzeichen auszuwählen.</p> <p>Anmerkung: Standardmäßig ist dieses Kästchen aktiviert und das Feld Tag-Trennzeichen deaktiviert.</p>

Name der Option	Beschreibung
Textbegrenzungszeichen	<p>Dies ist das Zeichen, das in einer Datei mit Begrenzungszeichen Textwerte umschließt.</p> <p>Im folgenden Beispiel werden im Datensatz doppelte Anführungszeichen (") als Textbegrenzungszeichen verwendet.</p> <pre>"7200 13TH ST" "MIAMI" "FL" "33144"</pre> <p>Die folgenden Zeichen stehen zur Definition als Textbegrenzungszeichen zur Verfügung:</p> <ul style="list-style-type: none"> • einfaches Anführungszeichen (') • doppeltes Anführungszeichen (") <p>Wenn in der Datei ein anderes Textbegrenzungszeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Textbegrenzungszeichen auszuwählen.</p>
Datensatz-Trennzeichen	<p>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-Zeilende verwenden aktivieren.</p> <p>Die verfügbaren Einstellungen für Datensatz-Trennzeichen sind:</p> <p>Unix (U+000A) Die Datensätze sind durch ein Zeilenvorschubzeichen getrennt. Dies ist bei Unix-Systemen das standardmäßige Datensatz-Trennzeichen.</p> <p>Macintosh (U+000D) Die Datensätze sind durch ein Zeilenumbruchzeichen getrennt. Dies ist bei Macintosh-Systemen das standardmäßige Datensatz-Trennzeichen.</p> <p>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.</p> <p>Wenn in der Datei ein anderes Datensatz-Trennzeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Datensatz-Trennzeichen auszuwählen.</p>
Root-Tag-Name	<p>Das Tag, das für Datensätze verwendet werden soll, die ein übergeordnetes Element anderer Datensatztypen sind. Wenn Sie beispielsweise die drei Datensatztypen 001, 100 und 200 haben und die Datensatztypen 100 und 200 untergeordnete Elemente des Datensatztyps 001 sind, dann ist 001 das Root-Tag.</p>

Name der Option	Beschreibung
Tags mit fester Breite verwenden	<p>Gibt an, ob ein fester Bereich am Beginn jedes Datensatzes zugeordnet werden soll, an dem das Datensatz-Tag eingefügt wird. Das folgende Beispiel zeigt eine Datei mit den Tags 001, 100, und 200 in einem Feld mit fester Breite:</p> <pre>001 Joe, Smith, M, 100 Main St, 555-234-1290 100 CHK12904567, 12/2/2007, 6/1/2012, CHK 200 1000567, 1/5/2012, Mike's Shoes, 323.12</pre>
Tag-Startposition	<p>Wenn Sie das Kästchen Tags mit fester Breite verwenden aktivieren, gibt diese Option die Position in jedem Datensatz an, an der das Tag beginnt. Wenn das Tag beispielsweise mit dem vierten Zeichen im Datensatz beginnt, müssten Sie 4 angeben.</p>
Tag-Breite	<p>Wenn Sie das Kästchen Tags mit fester Breite verwenden aktivieren, gibt diese Option die Anzahl von Leerzeichen an, die für Tags ab der im Feld Tag-Startposition angegebenen Position zugeordnet werden sollen. Wenn Sie beispielsweise im Feld Tag-Startposition 3 und im Feld Tag-Breite 7 angeben, würden die Positionen 4 bis 10 als Datensatz-Tag betrachtet. Der angegebene Wert muss groß genug sein, um alle Zeichen des längsten Tag-Namens einzuschließen.</p> <p>Der Wert im Feld Tag-Breite wird automatisch erhöht, wenn Sie den Tag-Namen im Feld Root-Tag-Name länger machen.</p> <p>Die maximale Tag-Breite ist 1024.</p>
Standard-Zeilende verwenden	<p>Gibt an, dass das standardmäßige Zeilenendzeichen (EOL) des Betriebssystems, auf dem der Spectrum™ Technology Platform-Server läuft, als Datensatz-Trennzeichen der Datei verwendet wird.</p> <p>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.</p>

Name der Option	Beschreibung
Datensätze mit weniger Feldern als definiert als falsch formatiert behandeln	<p>Wenn Sie diese Option aktivieren, werden untergeordnete Datensätze, die weniger Felder als ein vollständiger Datensatz enthalten, als falsch formatiert betrachtet. Wenn ein falsch formatierter Datensatz gefunden wird, fährt die Verarbeitung mit dem nächsten Root-Tag fort und ignoriert alle dazwischen liegenden untergeordneten Tags. Zudem wird eine Ausnahme im Protokoll festgehalten, die Informationen zum falsch formatierten untergeordneten Datensatz und die Zeilennummer enthält.</p> <p>Anmerkung: Datensätze werden in folgenden Situationen, unabhängig von der Aktivierung dieser Option, immer als falsch formatiert betrachtet.</p> <ul style="list-style-type: none"> • Das Tag ist unbekannt. • Die Zeile ist leer. • Es liegt ein Tag ohne Daten vor. • Ein Datensatz mit einem Tag, das ein untergeordnetes Element zu einem anderen Tag ist, erscheint sofort nach einem Datensatz mit einem Root-Tag.

Registerkarte „Felder“

Auf der Registerkarte **Felder** werden die Merkmale jedes Feldes angegeben, das aus der Datei eingelesen wird.

[Definieren von Feldern in Dateien mit variablem Format und Begrenzungszeichen](#) auf Seite 213

[Definieren von Feldern in einer Datei mit variablem Format vom Typ „Zeilensequenziell“ oder „Feste Breite“](#) auf Seite 216

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.

Feldname	Beschreibung
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 Dateien mit variablem Format und Begrenzungszeichen

Diese Schritte beschreiben, wie Sie im Schritt „Read from Variable Format File“ Felder für Dateien mit Begrenzungszeichen definieren.

1. Klicken Sie im Schritt „Read from Variable Format File“ auf die Registerkarte **Felder**.
2. Klicken Sie auf **Erneut generieren**.

Eine Liste mit allen Felder für jeden Datensatztyp wird angezeigt. Zu jedem Feld werden folgende Informationen angezeigt:

Parent Dies ist das Tag aus der Eingabedatei, das den Datensatztyp anzeigt, in dem das Feld erscheint. Wenn das Tag mit einer Zahl beginnt, erhält das Tag das Präfix „NumericTag_“. Ein Tag mit Namen „100“ würde beispielsweise zu „NumericTag_100“ geändert. Das Präfix ist erforderlich, da Datenflussfeldnamen nicht mit einer Zahl beginnen dürfen.

Feld Dies ist der Name, der im Datenfluss für das Feld verwendet wird. Standardmäßig erhalten Felder Namen im Format <Tag-Name>_<Spalte n>. Das erste Feld des Datensatztyps „Owner“ wäre „Owner_Column1“, der zweite „Owner_Column2“ usw.

Typ Dies ist der Datentyp des Feldes.

Anmerkung: Die ersten 50 Datensätze werden zur Generierung der Feldliste verwendet. Die Eingabedatei muss mindestens zwei Root-Tags enthalten, damit eine Feldliste erstellt werden kann.

3. Wenn Sie die Beziehungen übergeordnetes/untergeordnetes Element zwischen den Tags ändern möchten, gehen Sie wie folgt vor:
 - a) Klicken Sie auf **Tag-Hierarchie ändern**.
 - b) Klicken Sie auf die Tags und ziehen Sie sie, um die gewünschte Tag-Hierarchie zu definieren.
 - c) Klicken Sie auf **OK**.
4. Wenn Sie einen Feldnamen oder Datentyp ändern möchten, wählen Sie das Feld aus und klicken Sie auf **Ändern**.

5. Wählen Sie im Feld **Name** das Feld aus, das Sie hinzufügen möchten, oder geben Sie den Namen in das Feld ein.

Sie werden in der Regel die Standardnamen mit aussagekräftigen Namen ersetzen wollen, die die Daten im Feld repräsentieren. Betrachten Sie als Beispiel folgende Eingabedaten:

```
001 Joe,Smith,M,100 Main St,555-234-1290
```

Dieser Datensatz verfügt über ein übergeordnetes Tag „001“. Folgende Felder würden standardmäßig erstellt werden:

NumericTag_001_Column1: Joe
NumericTag_001_Column2: Smith
NumericTag_001_Column3: M
NumericTag_001_Column4: 100 Main St
NumericTag_001_Column5: 555-234-1290

Wahrscheinlich würden Sie die Felder umbenennen wollen, damit die Namen die Daten beschreiben. Beispiel:

FirstName: Joe
LastName: Smith
Gender: M
AddressLine1: 100 Main St
PhoneNumber: 555-234-1290

Anmerkung: Sie können Listenfelder nicht umbenennen. Listenfelder, die alle Felder eines angegebenen Datensatztyps enthalten, verwenden immer den Tag-Namen aus der Eingabedatei als Feldnamen.

6. Um den Datentyp eines Feldes zu ändern, wählen Sie den gewünschten Datentyp im Feld **Typ** aus.

Die folgenden Datentypen sind verfügbar:

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 $(2-2^{-52}) \times 2^{1023}$ 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 $(2-2^{-23}) \times 2^{127}$ 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: <pre> <Names> <Name>John Smith</Name> <Name>Ann Fowler</Name> </Names> </pre> <p>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.</p>
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.

7. Wenn Sie den Datentyp „Datum“, „Uhrzeit“ oder „Numerisch“ ausgewählt haben, können Sie das Standardformat für Datum/Uhrzeit oder Zahlen verwenden oder ein anderes Format für dieses bestimmte Feld angeben. Das Standardformat ist entweder das Systemstandardformat, das in den Typkonvertierungsoptionen in der Management Console festgelegt wurde, oder es ist das Standardformat des Datenflusses, das in den Typkonvertierungsoptionen im Enterprise Designer angegeben wurde. Das wirksame Format wird angezeigt. Lassen Sie **Standard**

ausgewählt, wenn Sie das Standardformat verwenden möchten. Möchten Sie hingegen ein anderes Format angeben, wählen Sie **Benutzerdefiniert** aus, und gehen Sie wie folgt vor:

Anmerkung: Es ist wichtig, dass Sie ein Datums- und Uhrzeitformat auswählen, das den Daten, die Sie aus der Datei lesen, korrekt entspricht. Wenn die Datei beispielsweise Datumsdaten im Format Monat/Tag/Jahr enthält, Sie aber Tag/Monat/Jahr auswählen, wird keine im Datenfluss ausgeführte Datumsberechnung wie Sortieren nach Datum das korrekte Datum verwenden oder anzeigen. Darüber hinaus kann bei Datensätzen die Typkonvertierung fehlschlagen. In dem Fall kommt das Fehlverhalten zum Tragen, das in den Typkonvertierungsoptionen in der Management Console oder im Enterprise Designer angegeben ist.

- a) Wählen Sie im Feld **Gebietsschema** das Land aus, dessen Formatierungskonventionen Sie verwenden möchten. Durch Ihre Auswahl werden die Standardwerte im Feld **Format** festgelegt. Bei Datumsdaten legt Ihre Auswahl auch die Sprache fest, die verwendet wird, 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 **Format** das Format für die Daten aus. Das Format ist abhängig vom Datentyp des Feldes. Es wird eine Liste der am häufigsten verwendeten Formate für das ausgewählte Gebietsschema bereitgestellt.

Ein Beispiel für das ausgewählte Format wird rechts vom Feld **Format** angezeigt.

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 318 erläuterten Schreibweise in das Feld ein. Geben Sie zur Festlegung eines eigenen Zahlenformats das Format mithilfe der unter **Zahlenmuster** auf Seite 321 erläuterten Schreibweise in das Feld ein.

8. Klicken Sie auf **OK**.

Definieren von Feldern in einer Datei mit variablem Format vom Typ „Zeilensequenziell“ oder „Feste Breite“

Diese Schritte beschreiben, wie Sie im Schritt „Read from Variable Format File“ Felder für Dateien vom Typ „Zeilensequenziell“ oder „Feste Breite“ definieren.

1. Klicken Sie im Schritt „Read from Variable Format File“ auf die Registerkarte **Felder**.
2. Klicken Sie auf **Tags abrufen**.

Eine Liste mit allen Felder für jeden Datensatztyp wird angezeigt. Zu jedem Feld werden folgende Informationen angezeigt:

Parent Dies ist das Tag aus der Eingabedatei, das den Datensatztyp anzeigt, in dem das Feld erscheint. Wenn das Tag mit einer Zahl beginnt, erhält das Tag das Präfix „NumericTag_“. Ein Tag mit Namen „100“ würde beispielsweise zu „NumericTag_100“ geändert. Das Präfix ist erforderlich, da Datenflussfeldnamen nicht mit einer Zahl beginnen dürfen.

Feld Dies ist der Name, der im Datenfluss für das Feld verwendet wird. Standardmäßig erhalten Felder Namen im Format <Tag-Name>_<Spalte n>. Das erste Feld des Datensatztyps „Owner“ wäre „Owner_Column1“, der zweite „Owner_Column2“ usw.

Typ Dies ist der Datentyp des Feldes.

Anmerkung: Die ersten 50 Datensätze werden zur Generierung der Feldliste verwendet. Die Eingabedatei muss mindestens zwei Root-Tags enthalten, damit eine Feldliste erstellt werden kann.

3. Wählen Sie im Feld **Filter** das Tag für den Datensatztyp aus, dessen Felder Sie definieren möchten, und klicken Sie dann auf **Hinzufügen**.

Anmerkung: Der Filter hat keine Auswirkungen darauf, welche Felder in den Datenfluss eingelesen werden. Er filtert nur die Feldliste, damit diese einfacher zu durchsuchen ist.

4. Wählen Sie im Feld **Name** das Feld aus, das Sie hinzufügen möchten, oder geben Sie den Namen in das Feld ein.
5. 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 Zeichenfolgendaten aus der Datei in einen Datentyp konvertiert, der die korrekte Manipulation der Daten im Datenfluss ermöglicht.

Spectrum™ Technology Platform 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“.

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 $(2-2^{-52}) \times 2^{1023}$ 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 $(2-2^{-23}) \times 2^{127}$ 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: <pre><Names> <Name>John Smith</Name> <Name>Ann Fowler</Name> </Names></pre> <p>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.</p>
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.

6. Wenn Sie den Datentyp „Datum“, „Uhrzeit“ oder „Numerisch“ ausgewählt haben, können Sie das Standardformat für Datum/Uhrzeit oder Zahlen verwenden oder ein anderes Format für dieses bestimmte Feld angeben. Das Standardformat ist entweder das Systemstandardformat, das in den Typkonvertierungsoptionen in der Management Console festgelegt wurde, oder es ist das Standardformat des Datenflusses, das in den Typkonvertierungsoptionen im Enterprise Designer angegeben wurde. Das wirksame Format wird angezeigt. Lassen Sie **Standard** ausgewählt, wenn Sie das Standardformat verwenden möchten. Möchten Sie hingegen ein anderes Format angeben, wählen Sie **Benutzerdefiniert** aus, und gehen Sie wie folgt vor:

Anmerkung: Es ist wichtig, dass Sie ein Datums- und Uhrzeitformat auswählen, das den Daten, die Sie aus der Datei lesen, korrekt entspricht. Wenn die Datei beispielsweise

Datumsdaten im Format Monat/Tag/Jahr enthält, Sie aber Tag/Monat/Jahr auswählen, wird keine im Datenfluss ausgeführte Datumsberechnung wie Sortieren nach Datum das korrekte Datum verwenden oder anzeigen. Darüber hinaus kann bei Datensätzen die Typkonvertierung fehlschlagen. In dem Fall kommt das Fehlverhalten zum Tragen, das in den Typkonvertierungsoptionen in der Management Console oder im Enterprise Designer angegeben ist.

- a) Wählen Sie im Feld **Gebietsschema** das Land aus, dessen Formatierungskonventionen Sie verwenden möchten. Durch Ihre Auswahl werden die Standardwerte im Feld **Format** festgelegt. Bei Datumsdaten legt Ihre Auswahl auch die Sprache fest, die verwendet wird, 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 **Format** das Format für die Daten aus. Das Format ist abhängig vom Datentyp des Feldes. Es wird eine Liste der am häufigsten verwendeten Formate für das ausgewählte Gebietsschema bereitgestellt.

Ein Beispiel für das ausgewählte Format wird rechts vom Feld **Format** angezeigt.

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 318 erläuterten Schreibweise in das Feld ein. Geben Sie zur Festlegung eines eigenen Zahlenformats das Format mithilfe der unter **Zahlenmuster** auf Seite 321 erläuterten Schreibweise in das Feld ein.

7. Geben Sie im Feld **Startposition** die Position des ersten Zeichens des Feldes und im Feld **Länge** die Anzahl der Zeichen in dem Feld ein.

Wenn das Feld beispielsweise beim zehnten Zeichen des Datensatzes startet und fünf Zeichen lang ist, müssen Sie eine Startposition von 10 und eine Länge von 5 angeben.

8. Klicken Sie auf **Hinzufügen**.
9. Wiederholen Sie diese Schritte, um weitere Felder zum Datensatztyp hinzuzufügen, oder klicken Sie auf **Schließen**, wenn Sie mit dem Hinzufügen von Feldern fertig sind.

Abflachen von Daten in variablem Format

Daten in variablem Format enthalten oft Datensätze, die über eine hierarchische Beziehung verfügen, bei der ein Datensatztyp das übergeordnete Element zu anderen Datensatztypen darstellt. Da viele Schritte die Daten in einem flachen Format benötigen, müssen Sie die Daten möglicherweise

abflachen, damit die Daten in Downstream-Schritten verwendbar sind. Betrachten Sie als Beispiel folgende Eingabedaten:

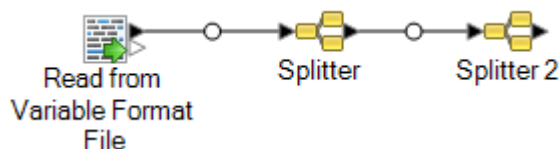
```
001 Joe,Smith,M,100 Main St,555-234-1290
100 CHK12904567,12/2/2007,6/1/2012,CHK
200 1000567,1/5/2012,Fashion Shoes,323.12
001 Anne,Johnson,F,1202 Lake St,555-222-4932
100 CHK238193875,1/21/2001,4/12/2012,CHK
200 1000232,3/5/2012,Blue Goose Grocery,132.11
200 1000232,3/8/2012,Trailway Bikes,540.00
```

Sie möchten möglicherweise die Datensätze abflachen, damit Sie über einen Datensatz pro Transaktionen verfügen. Für das obige Beispiel würde das bedeuten, die Transaktionsdatensätze (Datensätze mit dem Tag „200“) abzuflachen, sodass sie die Informationen zum Kontoinhaber (Datensätze mit dem Tag „001“) und zu den Kontodaten (Datensätze mit dem Tag „100“) enthalten.

Die folgenden Schritte beschreiben, wie Sie Splitter-Schritte verwenden, um Datensätze abzuflachen.

1. Fügen Sie Ihrem Datenfluss den Schritt „Read from Variable Format File“ hinzu und konfigurieren Sie ihn. Weitere Informationen finden Sie unter [Read from Variable Format File](#) auf Seite 205.
2. Fügen Sie einen Splitter-Schritt hinzu und verbinden Sie ihn mit dem Schritt „Read from Variable Format File“.
3. Fügen Sie je nach Bedarf weitere Splitter-Schritte hinzu, sodass Sie über einen Splitter-Schritt für jeden untergeordneten Datensatztyp in Ihren Eingabedaten verfügen.
4. Verbinden Sie alle Splitter-Schritte.

Sie sollten nun einen Datenfluss haben, der wie folgt aussieht:



5. Doppelklicken Sie auf den ersten Splitter-Schritt, um die Schrittoptionen zu öffnen.
6. Wählen Sie im Feld **Unterteilen bei** einen der untergeordneten Datensatztypen aus.
7. Klicken Sie auf **OK**.
8. Konfigurieren Sie jeden zusätzlichen Splitter-Schritt. Wählen Sie im Feld **Unterteilen bei** jedes einzelnen Splitter-Schrittes anderen untergeordneten Datensatztyp aus.

Read from XML

Der „Read from XML“-Schritt liest eine XML-Datei in einen Auftrag oder Unterfluss ein. Sie definiert den Dateipfad und das Datenformat, einschließlich dem XML-Schema und den Details der Datenelemente.

Einfache XML-Elemente werden in flache Felder umgewandelt und an den nächsten Schritt weitergeleitet. Einfache XML-Daten beinhalten Datensätze bestehend aus XML-Elementen, die nur Daten und keine untergeordneten Elemente enthalten. Dies ist beispielsweise eine einfache XML-Datendatei:

```
<customers>
  <customer>
    <name>Sam</name>
    <gender>M</gender>
    <age>43</age>
    <country>United States</country>
  </customer>
  <customer>
    <name>Jeff</name>
    <gender>M</gender>
    <age>32</age>
    <country>Canada</country>
  </customer>
  <customer>
    <name>Mary</name>
    <gender>F</gender>
    <age>61</age>
    <country>Australia</country>
  </customer>
</customers>
```

Beachten Sie, dass in diesem Beispiel jeder Datensatz einfache XML-Elemente wie `<name>`, `<gender>`, `<age>` und `<country>` enthält. Keines der Elemente enthält untergeordnete Elemente.

Der „Read from XML“-Schritt flacht automatisch einfache Daten wie diese ab, da die meisten Schritte Daten in einem flachen Format erfordern. Wenn Sie die hierarchische Struktur beibehalten möchten, verwenden Sie nach „Read from XML“ den Aggregator-Schritt, um die Daten in hierarchische Daten umzuwandeln.

Komplexe XML-Elemente behalten ihr hierarchisches Format bei und werden als Listenfeld weitergeleitet. Da viele Schritte ein flaches Datenformat erfordern, müssen Sie eventuell komplexe XML-Elemente abflachen, um die Daten für die Downstream-Schritte nutzbar zu machen. Weitere Informationen finden Sie unter [Abflachen von komplexen XML-Elementen](#) auf Seite 226.

Anmerkung: „Read from XML“ unterstützt nicht die XML-Typen `xs:anyType` und `xs:anySimpleType`.

Registerkarte „Dateieigenschaften“

Tabelle 4: Registerkarte „Dateieigenschaften“

Name der Option	Beschreibung
Schemadatei	<p>Gibt den Pfad zu einer XSD-Schemadatei an. Klicken Sie auf die Ellipsenschaltfläche (...), um die gewünschte Datei zu lokalisieren. Beachten Sie, dass sich die Schemadatei auf dem Server befinden muss, um die Datendatei im Vergleich zum Schema zu überprüfen. Falls sich die Schemadatei nicht auf dem Server befindet, wird die Überprüfung deaktiviert.</p> <p>Alternativ können Sie anstelle einer XSD-Datei eine XML-Datei angeben. Wenn Sie eine XML-Datei angeben, wird das Schema auf Basis der Struktur der XML-Datei abgeleitet. Die Verwendung einer XML-Datei statt einer XSD-Datei hat die folgenden Einschränkungen:</p> <ul style="list-style-type: none"> • Die XML-Datei darf nicht größer als 1 MB sein. Wenn die XML-Datei größer ist, versuchen Sie, einige der Daten zu entfernen und trotzdem die Struktur der XML beizubehalten. • Die Datendatei wird nicht anhand des abgeleiteten Schemas überprüft. <p>Anmerkung: 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.</p>
Datendatei	<p>Gibt den Pfad zur XML-Datendatei an. Klicken Sie auf die Ellipsenschaltfläche (...), um die gewünschte Datei zu lokalisieren.</p> <p>Anmerkung: 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.</p>
Vorschau	<p>Zeigt eine Vorschau des Schemas oder der XML-Datei an. Wenn Sie eine XSD-Datei angeben, spiegelt die Baumstruktur die ausgewählte XSD-Datei wider. Sobald Sie eine Schemadatei und eine Datendatei angegeben haben, können Sie auf die fett gedruckten Schemaelemente klicken, um eine Vorschau der Daten anzuzeigen, die das Element enthält.</p>

*Registerkarte „Felder“***Tabelle 5: Registerkarte „Felder“**

Name der Option	Beschreibung
Filter	Filtert die Liste der Elemente und Attribute, um das Durchsuchen zu vereinfachen. Der Filter hat keinen Einfluss darauf, welche Felder in die Ausgabe einbezogen werden. Es wird nur die Liste der Elemente und Attribute gefiltert, um das Durchsuchen zu vereinfachen.
XPath	Die XPath-Spalte enthält den XPath-Ausdruck für das Element oder Attribut. Es wird nur zur Information angezeigt. Weitere Informationen zu XPath finden Sie unter www.w3schools.com/xpath/ .
Feld	Der Name, der im Datenfluss für das Element oder Attribut verwendet wird. Doppelklicken Sie zum Ändern des Feldnamens auf den Namen und ändern Sie ihn wie gewünscht.

Name der Option

Beschreibung

Typ

Name der Option

Beschreibung

Der Datentyp für das Feld.

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.

date Ein Datentyp, der einen Monat, einen Tag und ein Jahr enthält. Datumswerte müssen das Format *yyyy-MM-dd* aufweisen. Beispiel: 2012-01-30.

datetime Ein Datentyp, der Tag, Monat, Jahr und Stunden, Minuten und Sekunden enthält. *DateTime* muss das Format *yyyy-MM-dd'THH:mm:ss* aufweisen. Beispiel: 2012-01-30T06:15:30

double Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit doppelter Genauigkeit zwischen 2^{-1074} und $(2-2^{-52}) \times 2^{1023}$ 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 $(2-2^{-23}) \times 2^{127}$ 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 positive als auch negative ganze Zahlen zwischen -2^{63} (-9,223,372,036,854,775,808) und $2^{63}-1$

Name der Option	Beschreibung
	(9,223,372,036,854,775,807) enthält.
string	Eine Folge von Zeichen.
time	Ein Datentyp, der die Tageszeit enthält. „Time“ muss das Format <i>HH:mm:ss</i> aufweisen. Beispiel: 21:15:59.
Einschließen	Gibt an, ob dieses Feld im Datenfluss verfügbar gemacht oder davon ausgeschlossen werden soll.

Beispiel: einfache XML-Datei

In diesem Beispiel möchten Sie die folgende Datei in einen Datenfluss einlesen:

```
<addresses>
  <address>
    <addressline1>One Global View</addressline1>
    <city>Troy</city>
    <state>NY</state>
    <postalcode>12128</postalcode>
  </address>
  <address>
    <addressline1>1825B Kramer Lane</addressline1>
    <city>Austin</city>
    <state>TX</state>
    <postalcode>78758</postalcode>
  </address>
</addresses>
```

In diesem Beispiel können Sie auswählen, `<addressline1>`, `<city>`, `<state>` und `<postalcode>` einzubeziehen. Dies führt zur Erstellung eines Datensatzes für jedes `<address>`-Element, da `<address>` das allgemein übergeordnete Element für `<addressline1>`, `<city>`, `<state>` und `<postalcode>` ist.

Abflachen von komplexen XML-Elementen

Für die meisten Schritte in einem Datenfluss müssen die Daten in einem flachen Format vorliegen. Wenn Sie also hierarchische Daten aus einer XML-Datei in einen Datenfluss einlesen, müssen Sie diese abflachen, wenn die Daten komplexe XML-Elemente enthalten. Ein komplexes XML-Element

ist ein Element, das andere Elemente oder Attribute enthält. In der folgenden Datendatei sind beispielsweise die Elemente `<address>` und `<account>` komplexe XML-Elemente:

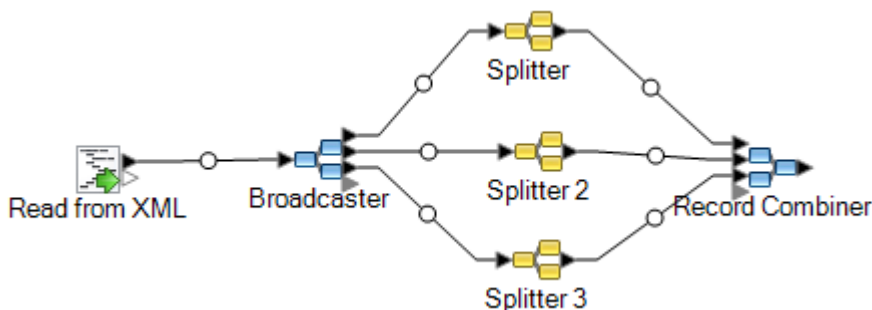
```
<customers>
  <customer>
    <name>Sam</name>
    <gender>M</gender>
    <age>43</age>
    <country>United States</country>
    <address>
      <addressline1>1253 Summer St.</addressline1>
      <city>Boston</city>
      <stateprovince>MA</stateprovince>
      <postalcode>02110</postalcode>
    </address>
    <account>
      <type>Savings</type>
      <number>019922</number>
    </account>
  </customer>
  <customer>
    <name>Jeff</name>
    <gender>M</gender>
    <age>32</age>
    <country>Canada</country>
    <address>
      <addressline1>26 Wellington St.</addressline1>
      <city>Toronto</city>
      <stateprovince>ON</stateprovince>
      <postalcode>M5E 1S2</postalcode>
    </address>
    <account>
      <type>Checking</type>
      <number>238832</number>
    </account>
  </customer>
  <customer>
    <name>Mary</name>
    <gender>F</gender>
    <age>61</age>
    <country>Australia</country>
    <address>
      <addressline1>Level 7, 1 Elizabeth Plaza</addressline1>
      <city>North Sydney</city>
      <stateprovince>NSW</stateprovince>
      <postalcode>2060</postalcode>
    </address>
    <account>
      <type>Savings</type>
      <number>839938</number>
    </account>
  </customer>
</customers>
```

Die folgenden Schritte beschreiben, wie Sie Splitter-Schritte verwenden, um XML-Daten abzuflachen, die mehrere komplexe XML-Elemente enthalten.

Anmerkung: Wenn Ihre Daten ein einzelnes komplexes XML-Element enthalten, können Sie einen einzelnen Splitter-Schritt verwenden, um die Daten abzuflachen, indem Sie einfach den „Read from XML“-Schritt mit dem Splitter-Schritt verbinden. Bei Datendateien, die ein einzelnes komplexes XML-Element enthalten, müssen Sie die Schritte „Broadcaster“ und „Record Combiner“ abweichend von der hier beschriebenen Prozedur nicht verwenden.

1. Fügen Sie Ihrem Datenfluss einen „Read from XML“-Schritt hinzu und konfigurieren Sie ihn. Weitere Informationen finden Sie unter [Read from XML](#) auf Seite 220.
2. Fügen Sie einen „Broadcaster“-Schritt hinzu und verbinden Sie ihn mit dem „Read from XML“-Schritt.
3. Fügen Sie für jedes komplexe XML-Element in Ihren Daten einen Splitter-Schritt hinzu.
4. Verbinden Sie den „Broadcaster“-Schritt mit jedem Splitter-Schritt.
5. Fügen Sie einen „Record Combiner“-Schritt hinzu und verbinden Sie ihn mit jedem Splitter-Schritt.

Sie sollten nun einen Datenfluss haben, der wie folgt aussieht:



6. Doppelklicken Sie auf den ersten Splitter-Schritt, um die Schrittoptionen zu öffnen.
7. Wählen Sie im Feld **Unterteilen bei** eins der komplexen Felder aus. Im obigen Beispiel könnte das das Adressenfeld sein.
8. Klicken Sie auf **OK**.
9. Konfigurieren Sie jeden zusätzlichen Splitter-Schritt. Wählen Sie im Feld **Unterteilen bei** jedes einzelnen Splitter-Schrittes ein anderes komplexes XML-Element aus.

Der Datenfluss ist nun konfiguriert, eine XML-Eingabe zu akzeptieren, die Datensätze mit komplexen XML-Elementen enthält, und die Daten abzuflachen. Die aus dem „Record Combiner“-Schritt resultierenden Datensätze können zu jedem Schritt gesendet werden, der flache Daten erfordert. Sie können beispielsweise den „Record Combiner“-Schritt zur Adressenüberprüfung mit einem „Validate Address“-Schritt verbinden.

SQL Command

SQL Command führt einen oder mehrere SQL-Befehle für jeden Datensatz im Datenfluss aus. Sie können SQL Command für Folgendes verwenden:

- Ausführen von komplexen INSERT/UPDATE-Anweisungen, wie Anweisungen, die Unteranfragen enthalten oder Verbindungen mit anderen Tabellen erstellen.
- Aktualisieren von Tabellen nach dem Einfügen/Aktualisieren von Daten, um die referenzielle Integrität zu wahren.
- Aktualisieren oder Löschen eines Datensatzes in einer Datenbank, bevor ein Ersatzdatensatz geladen wird.
- Aktualisieren von mehreren Tabellen in einer einzigen Transaktion.

Sie können vor und nach den SQL-Hauptbefehlen zusätzliche SQL-Befehle ausführen, und Sie können gespeicherte Prozeduren aufrufen.

Anmerkung: Gespeicherte Prozeduren, die von SQL Command aufgerufen werden, dürfen keine OUT-Parameter verwenden.

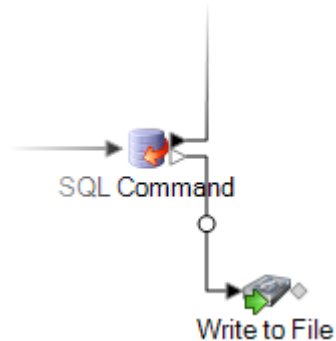
Anmerkung: Es können bestimmte Leistungsverbesserungen erzielt werden, indem Sie mehrere Laufzeitinstanzen von SQL Command verwenden. Klicken Sie auf die Schaltfläche **Laufzeit**, um mehrere Laufzeitinstanzen anzugeben.

Allgemein

Auf der Registerkarte **Allgemein** können Sie dynamische SQL-Anweisungen festlegen, die Sie einmal für jeden Datensatz ausführen möchten. Die folgende Tabelle enthält die Optionen der Registerkarte **Allgemein**.

Option	Beschreibung
Verbindung	<p>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 Datenbankverbindung ändern oder löschen müssen, klicken Sie auf Verwalten.</p> <p>Wenn Sie eine Datenbankverbindung hinzufügen oder ändern, füllen Sie diese Felder aus:</p> <p>Verbindungsname Geben Sie einen Namen für die Verbindung ein. Sie können den Namen frei wählen.</p> <p>Datenbanktreiber Wählen Sie den passenden Datenbanktyp aus.</p> <p>Verbindungsoptionen Geben Sie Host, Port, Instanz, Benutzername und Kennwort an, die für die Verbindung zur Datenbank verwendet werden sollen.</p>
SQL-Anweisungen	<p>Geben Sie die SQL-Anweisungen ein, die Sie für jeden Datensatz im Datenfluss ausführen möchten. Mit Beginn der Eingabe erscheint ein Fenster, das automatisch die gültigen SQL-Befehle anzeigt. Trennen Sie mehrere SQL-Anweisungen mit einem Semikolon (;).</p> <p>Verwenden Sie die folgende Syntax, um einen Wert in einem Datenflussfeld festzulegen:</p> <pre> \${<field name>} </pre> <p>Dabei steht <field name> für den Namen eines Feldes im Datenfluss.</p> <p>Beispiel:</p> <pre> UPDATE MyDatabase.dbo.customer SET name=\${Name} WHERE id=\${ID}; </pre> <p>In diesem Beispiel ersetzt <code>\${Name}</code> den Wert aus dem Datenflussfeld „Name“ und <code>\${ID}</code> den Wert aus dem ID-Feld des Datenflusses.</p> <p>Anmerkung: Anforderungen müssen den vollständig qualifizierten Namen verwenden. Beispiel: MeineDatenbank.dbo.customer.</p>
Transaktionsverarbeitung	<p>Gibt an, ob Datensätze in Batches oder alle auf einmal verarbeitet werden können. Zur Auswahl stehen:</p> <p>Batchgröße Gruppirt Datensätze in Batches der von Ihnen angegebenen Größe und verarbeitet einen Batch nach dem anderen.</p> <p>Gesamte Ausführung Erstellt ein großes Batch für alle Datensätze und verarbeitet alle Transaktionen gleichzeitig.</p>

Option	Beschreibung
Fehlerverarbeitung	<p>Gibt an, was zu tun ist, wenn während der Ausführung der SQL-Befehle ein Fehler auftritt. Zur Auswahl stehen:</p> <p>Datenfluss bei Fehler nicht beenden Der Datenfluss wird weiter ausgeführt, wenn die Datenbank bei der Ausführung der SQL-Befehle einen Fehler zurückgibt.</p> <p>Datenfluss nach Auftreten dieser Anzahl von Fehlern beenden Der Datenfluss wird gestoppt, nachdem die Datenbank die festgelegte Anzahl von Fehlern zurückgibt.</p> <p>Anmerkung: Wenn in SQL ein Syntaxfehler auftritt, wird der Datenfluss ungeachtet der von Ihnen hier ausgewählten Einstellung immer beendet.</p> <p>Zusätzlich können Sie optional Fehlerdatensätze in ein Laden von Daten schreiben, indem Sie den SQL Command-Fehlerport mit dem gewünschten Datenladungstyp verbinden. Der Fehlerport ist das weiße Dreieck auf der rechten Seite des Schrittsymbols im Datenfluss. Zum Schreiben von Fehlerdatensätzen in eine Flatfile müssen Sie z. B. den SQL Command-Fehlerport mit dem „Write to File“-Schritt verbinden, wie hier dargestellt:</p>



Pre-/Post-SQL

Auf der Registerkarte **Pre/Post SQL** legen Sie SQL-Anweisungen fest, die einmal pro Datenflussausführung ausgeführt werden sollen. Dies ist der Gegensatz zur einmaligen Ausführung pro Datensatz wie auf der Registerkarte **Allgemein**. Die folgende Tabelle enthält die Optionen der Registerkarte **Pre-/Post-SQL**.

Option	Beschreibung
Pre-SQL	<p>Geben Sie eine oder mehrere SQL-Anweisungen ein, die Sie ausführen möchten, bevor die in den Schritt eintretenden Datensätze verarbeitet werden. Die von Ihnen eingegebenen SQL-Anweisungen werden einmal pro Ausführung ausgeführt, nachdem der Datenfluss gestartet ist, aber bevor der „SQL Command“-Schritt die ersten Datensätze verarbeitet.</p> <p>Ein Beispiel für die Verwendung von Pre-SQL ist die Erstellung einer Tabelle für die zu verarbeitenden Datensätze.</p>
Automatischer Commit von Pre-SQL	<p>Aktivieren Sie dieses Kästchen, um die Pre-SQL-Anweisungen zu übergeben, bevor die SQL-Anweisungen auf der Registerkarte Allgemein ausgeführt werden.</p> <p>Wenn Sie dieses Kästchen nicht aktivieren, werden die Pre-SQL-Anweisungen in der gleichen Transaktion wie die SQL-Anweisungen auf der Registerkarte Allgemein übergeben.</p> <p>Anmerkung: Wenn Sie weder das Kästchen Automatischer Commit von Pre-SQL noch das Kästchen Automatischer Commit von Post-SQL aktivieren, werden alle SQL-Anweisungen für den Schritt in einer Transaktion übergeben.</p>
Post-SQL	<p>Geben Sie eine oder mehrere SQL-Anweisungen ein, die Sie ausführen möchten, nachdem alle Datensätze verarbeitet wurden. Die von Ihnen hier eingegebenen SQL-Anweisungen werden einmal pro Ausführung ausgeführt, nachdem der „SQL Command“-Schritt beendet ist, aber bevor der Datenfluss beendet wird.</p> <p>Eine Beispielanwendung von Post-SQL wäre der Aufbau eines Index nach der Verarbeitung der Datensätze.</p>
Automatischer Commit von Post-SQL	<p>Aktivieren Sie dieses Kästchen, um die Post-SQL-Anweisungen in ihrer eigenen Transaktion zu übergeben, nachdem die SQL-Befehle der Registerkarte Allgemein übergeben wurden.</p> <p>Wenn Sie dieses Kästchen nicht aktivieren, werden die Post-SQL-Anweisungen in der gleichen Transaktion wie die SQL-Anweisungen auf der Registerkarte Allgemein übergeben.</p> <p>Anmerkung: Wenn Sie weder das Kästchen Automatischer Commit von Pre-SQL noch das Kästchen Automatischer Commit von Post-SQL aktivieren, werden alle SQL-Anweisungen für den Schritt in einer Transaktion übergeben.</p>

Registerkarte „Laufzeit“

Die Registerkarte **Laufzeit** enthält **Schrittoptionen** und gibt Ihnen die Möglichkeit, Standardwertwerte für die Schrittoptionen zu definieren.

Feldname	Beschreibung
Schrittoptionen	<p>In diesem Abschnitt sind die Datenflussoptionen aufgeführt, die in der SQL-Abfrage dieses Schrittes verwendet werden. Zudem können Sie hier einen Standardwert für diese Optionen angeben. In der Spalte Name sind die Optionen aufgelistet. In der entsprechenden Spalte Wert können Sie die Standardwerte eingeben.</p> <p>Anmerkung: Der hier angegebene Standardwert wird auch im Abschnitt Datenflussoptionen zu Schritten zuordnen des Dialogfeldes Datenflussoptionen angezeigt. Mithilfe der Dialogbox können Sie außerdem den Standardwert ändern. Wenn für Standardwerte einer Option unterschiedliche Werte über Schrittoptionen, Datenflussoptionen und Job Executor angegeben sind, gilt folgende Rangfolge: im Job Executor angegebener Wert > unter Datenflussoptionen definierter Wert > unter Schrittoptionen eingegebener Wert.</p>

Festlegen von „SQL Command“ zur Laufzeit

Diese Prozedur beschreibt, wie Sie einen Datenfluss konfigurieren, damit dieser Laufzeitoptionen für „SQL Command“ unterstützt, und wie Sie die Job Executor-Argumente dafür festlegen.

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. Erweitern Sie den „SQL Command“-Schritt.
6. Wählen Sie eine „SQL Command“-Option aus. Dies kann **PreSqlCommand**, **SqlCommand** oder **PostSqlCommand** sein.

PreSqlCommand SQL-Anweisungen, die Sie ausführen möchten, bevor die in den Schritt eintretenden Datensätze verarbeitet werden. Diese SQL-Anweisungen werden einmal pro Ausführung ausgeführt, nachdem der Datenfluss gestartet ist, aber bevor der „SQL Command“-Schritt den ersten Datensatz verarbeitet.

Ein Beispiel für die Verwendung von Pre-SQL ist die Erstellung einer Tabelle für die zu verarbeitenden Datensätze.

SqlCommand SQL-Anweisungen, die Sie für jeden Datensatz im Datenfluss ausführen möchten.

PostSqlCommand SQL-Anweisungen, die Sie ausführen möchten, nachdem alle Datensätze verarbeitet wurden. Diese SQL-Anweisungen werden einmal pro Ausführung ausgeführt, nachdem der „SQL Command“-Schritt beendet ist, aber bevor der Datenfluss beendet wird.

Eine Beispielanwendung von Post-SQL wäre der Aufbau eines Index nach der Verarbeitung der Datensätze.

Der ausgewählte „SQL Command“-Optionsname wird in den Feldern **Optionsname** und **Optionsbeschriftung** angezeigt. Dies ist der Optionsname, der zur Laufzeit angegeben werden muss, um diese Option festzulegen.

7. Geben Sie eine Beschreibung der Option in das Feld **Beschreibung** ein.
8. Wählen Sie im Feld **Ziel** die Option **Ausgewählte(r) Schritt/e** aus.
9. 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.
10. 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 [Angaben von Standarddienstoptionen](#).

11. Klicken Sie auf **OK**.
12. Fügen Sie auf Wunsch weitere Optionen hinzu.
13. Klicken Sie im Dialogfeld **Datenflussoptionen** auf **OK**, sobald Sie damit fertig sind.
14. Speichern Sie den Datenfluss und machen Sie ihn verfügbar.
15. Erstellen Sie eine Textdatei, die die SQL-Anweisung enthält, die Sie zur Laufzeit verwenden möchten.

Die Textdatei könnte wie folgt aussehen:

```
SqlCommand = UPDATE CustomersSET
ContactName='Alfred Schmidt'
City='Hamburg'
WHERE CustomerName='Alfreds Futterkiste';
```

In diesem Beispiel ist „SqlCommand“ einer der Optionsnamen des „SQL Command“-Schrittes.

16. Verwenden Sie das Argument `-o`, wenn Sie einen Job Executor aus der Befehlszeile ausführen.

```
java -jar jobexecutor.jar -h "noipa019sh-11" -u "admin" -p "admin" -s
"8080" -o "options.txt" -j "FetchOracleData" -w
```

Der Dateiname (`options.txt`) gibt den Namen der Textdatei an, die Sie in Schritt 14 erstellt haben.

Weitere Informationen finden Sie unter [Ausführen eines Auftrags über die Befehlszeile](#) auf Seite 235.

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 <i>delimiter</i>	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 <i>property file</i>	Gibt einen Pfad zu einer Auftragseigenschaftsdatei an. Eine Auftragseigenschaftsdatei enthält Job Executor-Argumente. Weitere Informationen zu Auftragseigenschaftsdateien finden Sie unter Verwenden einer Auftragseigenschaftsdatei .
Nein	-h <i>host name</i>	Gibt Namen oder die IP-Adresse des Spectrum™ Technology Platform-Servers an.
Nein	-i <i>poll interval</i>	Gibt an, wie häufig auf abgeschlossene Aufträge geprüft werden sollen (in Sekunden). Gilt nur im synchronen Modus.
Ja	-j <i>job name</i>	Eine durch Kommas getrennte Liste von Aufträgen, die ausgeführt werden sollen. Bei Auftragsnamen wird die Groß-/Kleinschreibung

Erforderlich		Argument	Beschreibung
			beachtet. Aufträge werden in der aufgelisteten Reihenfolge gestartet.
Nein	-n	<i>email list</i>	Gibt eine durch Kommas getrennte Liste zusätzlicher E-Mail-Adressen für konfigurierte Auftragsbenachrichtigungen an.
Nein	-o	<i>property file</i>	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 . So kann etwa eine Datenflussoptions-Eigenschaftsdatei für einen Datenfluss, der einen „Assign GeoTAX Info“-Schritt enthält, wie folgt aussehen: <pre>OutputCasing=U UseStreetLevelMatching=N TaxKey=T Database.GTX=gsl</pre>
Ja	-p	<i>password</i>	Das Kennwort des Benutzers.
Nein	-r		Geben Sie dieses Argument an, um einen detaillierten Bericht über den Auftrag zurückzugeben. Diese Option funktioniert nur, wenn Sie auch <code>-w</code> angeben. Der Bericht enthält die folgenden Informationen: <ul style="list-style-type: none"> • 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

Erforderlich	Argument	Beschreibung
		<ul style="list-style-type: none"> • Position 7 – Anzahl der fehlgeschlagenen Datensätze • Position 8 – Anzahl der falsch formatierten Datensätze • Position 9 – derzeit nicht verwendet <p>Beispiel:</p> <pre>MySimpleJob 4 succeeded 04/09/2010 14:50:47 04/09/2010 14:50:47 100 0 0 </pre> <p>Die Informationen sind das durch Trennzeichen getrennt, das im Argument <code>-d</code> angegeben ist.</p>
Nein	<code>-s port</code>	Der Socket (Port), an dem der Spectrum™ Technology Platform-Server ausgeführt wird. Der Standardwert ist 8080.
Nein	<code>-t timeout</code>	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	<code>-u user name</code>	Der Anmelde-name des Benutzers.
Nein	<code>-v</code>	Gibt eine ausführliche Ausgabe zurück.
Nein	<code>-w</code>	<p>Führt Job Executor im synchronen Modus aus. Das bedeutet, dass Job Executor weiter ausgeführt wird, bis der Auftrag abgeschlossen ist.</p> <p>Wenn Sie <code>-w</code> 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.</p>
Nein	<code>StageName=Protocol:FileName</code>	Überschreibt die Eingabe- oder Ausgabedatei, die in „Read from File“ oder „Write to File“ angegeben ist. Weitere Informationen finden Sie unter Überschreiben von Auftragsdateispeicherorten .

Erforderlich	Argument	Beschreibung
Nein	<i>StageName:schema=Protocol:SchemaFile</i>	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 .

Anwendungsbeispiel für Job Executor

Das folgende Beispiel zeigt Befehlszeilenaufwurf 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“.

Ausführen von SQL-Befehlen vor oder nach einem Datenfluss

Die Aktivität **SQL ausführen** führt während eines Prozessflusses Vorgänge in der Datenbank aus. Mithilfe dieser Aktivität können Sie die SQL-Anweisungen sowohl vor als auch nach der Ausführung des Spectrum™ Technology Platform-Datenflusses oder eines externen Programms ausführen. Die Aktivität **SQL ausführen** kann z. B. zum Löschen von Indizes vor der Ausführung eines Spectrum™ Technology Platform-Datenflusses und zur Erstellung von Indizes nach der Ausführung des Datenflusses verwendet werden. Zur Ausführung von SQL-Anweisungen mithilfe der Aktivität **SQL ausführen** müssen Sie einen Prozessfluss erstellen.

Anmerkung: Anweisungen, wie ein Prozessfluss erstellt und geplant wird, finden Sie im *Datenfluss-Designer-Handbuch*.

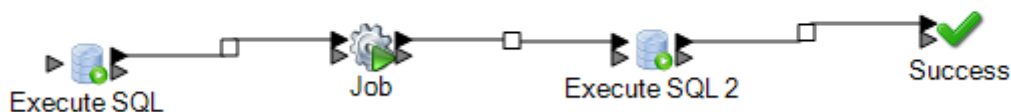
1. Ziehen Sie die Aktivität **SQL ausführen** auf die Arbeitsfläche.
2. Doppelklicken Sie auf die Aktivität **SQL ausführen**.
3. Wählen Sie eine Datenbankverbindung aus, die Sie verwenden möchten.

Wenn Sie eine neue Datenbankverbindung herstellen oder eine vorhandene Datenbankverbindung 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.

- Schreiben Sie die SQL-Anweisung in das Feld **SQL-Anweisung(en)**.
Die Option **Prozessfluss bei Fehler beenden** ist standardmäßig aktiviert. Das bedeutet, dass der Prozessfluss bei einer Ausnahme beendet wird. Wenn die Option **Prozessfluss bei Fehler beenden** deaktiviert ist und eine Ausnahme eintritt, wird der Prozessfluss nicht gestoppt und die Ausnahme wird in die Serverprotokolle eingetragen.
- Fügen Sie die Aktion hinzu, die ein Prozessfluss ausführen soll.
Sie können einen Auftrag hinzufügen, indem Sie das Auftragsymbol auf die Arbeitsfläche ziehen, oder ein externes Programm hinzufügen, indem Sie das Symbol „Programm ausführen“ auf die Arbeitsfläche ziehen.
- Verbinden Sie die zwei Aktivitäten.
- Fügen Sie bei Bedarf eine weitere **SQL ausführen**-Aktivität hinzu.
Informationen zur Ausführung von Aktionen für **SQL ausführen** finden Sie in Schritt 2 bis 5.
- Wenn Sie alle Aufträge hinzugefügt haben (die Aktivitäten „Programm ausführen“ und „SQL ausführen“), die Sie im Prozessfluss ausführen möchten, ziehen Sie eine Erfolg-Aktivität auf die Arbeitsfläche und verbinden Sie sie mit der letzten Aktivität im



~~Poste~~

- Führen Sie den Prozessfluss aus.

Transposer

Transposer konvertiert Spalten in Zeilen. Transponieren von Daten ist das Gegenteil zum Pivotieren von Daten mithilfe des „Group Statistics“-Schrittes, der Zeilendaten in Spalten transformiert.

Betrachten Sie folgendes Beispiel, um Transposer zu verstehen. Eine Tabelle enthält Verkaufsdaten aus vier Quartalen. Sie möchten die generierten Gesamtumsätze hinzufügen und das in den ersten

drei Quartalen erzielte Wachstum analysieren. Verwenden Sie dazu Transposer, um eine Spalte zu erstellen, die den Gesamtumsatz aus den drei transponierten Quartalen enthält. Wenn Sie Transposer verwenden, um den Gesamtumsatz, der in unterschiedlichen Spalten generiert wurde, als eine Spalte hinzuzufügen, ist die Leistung möglicherweise besser, als wenn Sie ihn in verschiedenen Spalten hinzufügen.

In der folgenden Tabelle sind die Optionen im Dialogfeld von Transposer erklärt.

Option	Beschreibung
Kopfzeile transponierter Felder	Geben Sie einen Kopfzeilennamen für die Spalte ein, die die zu transponierenden Spalten enthalten soll. Diese neue Spalte wird automatisch zum Datenfluss hinzugefügt.
Kopfzeile transponierter Werte	Geben Sie einen Kopfzeilennamen für die Spalte ein, die die transponierten Spaltenwerte enthalten soll. Diese neue Spalte wird automatisch zum Datenfluss hinzugefügt.
Transponierte Felder beibehalten	Aktivieren Sie diese Option, um alle transponierten Felder als Spalten in der Ausgabe beizubehalten.
Feldname	Zeigt alle Spaltenkopfzeilen der Eingabedatei an.

Option Beschreibung

Typ Zeigt den Datentyp der jeweiligen Felder (Spaltenkopfzeilen) an.
 Die zu transponierenden Spalten sollten in der Eingabequellendatei über einen kompatiblen Datentyp verfügen. Unten finden Sie die Kompatibilitätsmatrix. Die mit Häkchen markierten Felder zeigen kompatible Datentypen an.

	Integer	Long	String	Date/Time	Double	Big Decimal	Time	Date
Integer	✓	✓	✓		✓	✓		
Long	✓	✓	✓		✓	✓		
String	✓	✓	✓	✓	✓	✓	✓	✓
Datetime			✓	✓				
Double	✓	✓	✓		✓			
Bigdecimal	✓	✓	✓			✓		
Time			✓				✓	
Date			✓					✓

Transponiert Aktivieren Sie das Kästchen neben jedem Feld, das Sie in eine Spalte konvertieren möchten. Deaktivieren Sie das Kästchen, wenn eine Spalte nicht transponiert und in der Ausgabe beibehalten werden soll.

Beispiel der Verwendung von Transposer

Die folgenden Eingabedaten enthalten Verkäufe aus vier Quartalen nach Geschäft. Beachten Sie, dass Q1, Q2, Q3 und Q4 für Verkäufe in vier Quartalen (in Millionen) stehen.

Geschäft (USA)	Q1	Q2	Q3	Q4
New York	100,00	200,10	300,00	400,00

Geschäft (USA)	Q1	Q2	Q3	Q4
Kalifornien	250,10	450,00	550,00	650,00
Illinois	150,00	250,10	350,00	450,00

Die unten aufgeführten Fälle zeigen das Verhalten von Transposer mit den im Schritt angegebenen Optionen. Beachten Sie, dass der Spaltenname für die Kopfzeile transponierter Felder „Quartal“ und der Spaltenname für die Werte transponierter Felder „Umsatz“ ist.

Fall 1

Nehmen Sie an, Sie möchten, dass die Spalten Q1, Q2 und Q3 transponiert werden und Q4 in der Ausgabe beibehalten wird. Aktivieren Sie dazu unter der Kopfzeile **Transponiert** das Kästchen neben jeder Spalte, die transponiert werden soll. Sie sehen nun Q1, Q2 und Q3 als Zeilen, während Q4 als Spalte in der Ausgabe erhalten bleibt.

Geschäft (USA)	Quartal	Umsatz	Q4
New York	Q1	100,00	400,00
New York	Q2	200,10	400,00
New York	Q3	300,00	400,00
Kalifornien	Q1	250,10	650,00
Kalifornien	Q2	450,00	650,00
Kalifornien	Q3	550,00	650,00
Illinois	Q1	150,00	450,00
Illinois	Q2	250,10	450,00
Illinois	Q3	350,00	450,00

Fall 2

Nehmen Sie an, Sie möchten, dass die Spalten Q1 und Q2 transponiert und Q3 und Q4 in der Ausgabe beibehalten werden. Außerdem möchten Sie alle transponierte Felder (Q1 und Q2) als Spalten in der Ausgabe beibehalten. Aktivieren Sie dazu die Option **Transponierte Felder beibehalten** und unter der Kopfzeile **Transponiert** das Kästchen neben jeder Spalte, die transponiert werden soll. Sie sehen nun Q1 und Q2 als Zeilen. Q3 und Q4 werden neben Q1 und Q2 als Spalten in der Ausgabe beibehalten.

Geschäft (USA)	Quartal	Umsatz	Q1	Q2	Q3	Q4
New York	Q1	100,00	100,00	200,10	300,00	400,00
New York	Q2	200,10	100,00	200,10	300,00	400,00
Kalifornien	Q1	250,10	250,10	450,00	550,00	650,00
Kalifornien	Q2	450,00	250,10	450,00	550,00	650,00
Illinois	Q1	150,00	150,00	250,10	350,00	450,00
Illinois	Q2	250,10	150,00	250,10	350,00	450,00

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 $(2-2^{-52}) \times 2^{1023}$ 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 $(2-2^{-23}) \times 2^{127}$ 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.

- 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.

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

- Metaphone (Spanisch)** Gibt einen per Metaphone codierten Schlüssel ausgewählter Felder für die spanische Sprache zurück. Dieser Metaphone-Algorithmus codiert Wörter anhand ihrer spanischen Aussprache.
- Metaphone 3** Übertrifft die Algorithmen „Metaphone“ und „Double Metaphone“ mit genaueren Einstellungen für Konsonanten und inneren Vokalen, was Ihnen ermöglicht, Wörter oder Namen auf phonetischer Basis mehr oder weniger eng mit Suchbegriffen abgeglichen zu erzeugen. Metaphone 3 erhöht die Genauigkeit der phonetischen Codierung auf 98 %. Diese Option wurde entwickelt, um auf Einschränkungen von Soundex zu reagieren.
- NYSIIS** Phonetischer Codealgorithmus, der eine ungefähr übereinstimmende Aussprache mit der exakten Schreibweise vergleicht und Wörter indiziert, deren Aussprache ähnlich ist. Bestandteil des New York State Identification and Intelligence System. Angenommen, Sie suchen die Daten einer Person in einer Personendatenbank. Sie meinen, dass der Name der Person wie „John Smith“ klingt, er wird aber „Jon Smyth“ geschrieben. Würden Sie eine Suche durchführen, in der nach einer genauen Übereinstimmung mit „John Smith“ gesucht wird, würden keine Ergebnisse zurückgegeben. Wenn Sie jedoch die Datenbank mit dem Algorithmus NYSIIS indizieren und eine 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.

Write to Cache

Write to Cache lädt die Ausgabe eines Datenflusses in einen globalen Cache und macht die Daten so für Suchen über den „Query Cache“-Schritt verfügbar. Die Verwendung eines globalen Cache für Datensuchen verbessert die Leistung im Vergleich zu Suchen in Datenbanken.

Ein globaler Cache ist ein systemweiter, gemeinsam genutzter Cache, der sich im Speicher befindet. Entscheiden Sie sich für einen globalen Cache, wenn der Cache für mehrere Datenflüsse verfügbar sein soll oder wenn die Daten sich nicht oft ändern oder relativ statisch sind und der Speicherplatz nicht begrenzt ist. Ein globaler Cache ist statisch, da Sie ihn nur einmal füllen können. Der Cache kann nach seiner Erstellung nicht mehr aktualisiert werden.

Anmerkung: Write to Cache überschreibt den Cache jedes Mal, wenn der Datenfluss ausgeführt wird.

Allgemein

Name der Option	Beschreibung
Cachename	Gibt den Namen des Cache an. Wenn im System bereits Caches vorhanden sind, werden diese aufgelistet und Sie können einen auswählen, wenn Sie den vorhandenen Cache mit neuen Daten füllen möchten. Um einen neuen Cache zu erstellen, geben Sie den gewünschten Namen für den neuen Cache ein. Der Name muss mit einem Buchstaben beginnen. Er kann einen Unterstrich, jedoch keine anderen Sonderzeichen enthalten. Der Name kann numerische Werte enthalten.
Cachefelder	In dieser Spalte werden die im Cache verwendeten Feldnamen aufgelistet. Wenn Sie einen Feldnamen ändern möchten, klicken Sie auf den Feldnamen und geben Sie einen neuen Namen ein.
Schrittfelder	In dieser Spalte werden die im Datenfluss verwendeten Feldnamen aufgelistet. Sie können diese Feldnamen nicht ändern.
Typ	In dieser Spalte ist der Datentyp jedes Feldes aufgelistet.
Einschließen	Aktivieren Sie das Kästchen in dieser Spalte, um das Feld in den Cache einzuschließen. Deaktivieren Sie das Kästchen, wenn das Feld nicht in den Cache geschrieben werden soll.
Schlüsselfeld	Aktivieren Sie das Kästchen in dieser Spalte, wenn das Feld im „Query Cache“-Schritt als Schlüssel verwendet werden soll. Wenn Sie beispielsweise über ein Datenflussfeld namens „Kontonummer“ verfügen und der „Query Cache“-Schritt Daten über eine Abfrage nach passenden Werten im Feld „Kontonummer“ suchen soll, müssen Sie in der Schlüsselfeldspalte das Kästchen für das Feld „Kontonummer“ aktivieren. Die Felder, die Sie als Schlüsselfelder angeben, stehen im „Query Cache“-Schritt als Schlüsselfelder zur Verfügung.

Leeren eines globalen Caches

Um einen globalen Cache zu leeren, müssen Sie einen Prozessfluss erstellen und ausführen. Der Prozessfluss muss eine „Cache löschen“-Aktivität enthalten. Die „Cache löschen“-Aktivität leert den

globalen Cache, löscht ihn aber nicht. Sie können den Cache auch leeren, indem Sie einen Prozessfluss planen.

Anmerkung: Anweisungen, wie ein Prozessfluss erstellt und geplant wird, finden Sie im *Datenfluss-Designer-Handbuch*.



Um den globalen Cache manuell zu leeren, gehen Sie wie folgt vor:

1. Ziehen Sie die Aktivität **Cache löschen** auf die Arbeitsfläche.
2. Ziehen Sie die Aktivität **Erfolg** auf die Arbeitsfläche.
3. Verbinden Sie die zwei Aktivitäten.
4. Doppelklicken Sie auf die Aktivität **Cache löschen**.
5. Wählen Sie den Cache aus. Sie können auch mehrere Caches auswählen, um die zugehörigen Daten zu löschen.
Die Caches, die Sie im „Write To Cache“-Schritt erstellen, werden in der „Cache löschen“-Aktivität aufgelistet.
6. Führen Sie den Prozessfluss aus.

Write to DB

Der „Write to DB“-Schritt schreibt die Ausgabe eines Datenflusses in eine Datenbank.

Anmerkung: Es lassen sich erhebliche Leistungsverbesserungen erzielen, wenn mehrere Laufzeitinstanzen von „Write to DB“ verwendet werden. Klicken Sie auf die Schaltfläche **Laufzeit**, um mehrere Laufzeitinstanzen anzugeben.

Registerkarte „Allgemein“

Name der Option	Beschreibung
Verbindung	<p>Wählen Sie die Verbindung für die Datenbank aus, die Sie im Feld „Verbindung“ verwenden möchten. Klicken Sie auf Verwalten, um eine neue Datenbankverbindung herzustellen. Weitere Informationen zum Erstellen von Datenbankverbindungen finden Sie unter Datenbankverbindungs-Manager.</p>
Tabelle/Ansicht	<p>Geben Sie nach der Auswahl einer Verbindung die Tabelle oder Ansicht ein, in der geschrieben werden soll. Klicken Sie auf die Schaltfläche zum Durchsuchen ([...]), um zur gewünschten Tabelle oder Ansicht zu navigieren, oder klicken Sie auf Tabelle erstellen, um eine neue Tabelle in der Datenbank zu erstellen.</p> <p>Anmerkung: Wenn Sie in eine SQL-Datenbank schreiben, können Sie nicht in Ansichten schreiben, die mehr als eine Tabelle referenzieren. Das liegt an einer Einschränkung von SQL Server.</p>

Name der Option	Beschreibung
Tabelle erstellen	<p>Erstellt eine neue Tabelle in der ausgewählten Datenbank. Wählen Sie im Feld Tabellenbesitzer den Besitzer der Tabelle aus und geben Sie im Feld Tabellenname den Namen für die neue Tabelle an. Bei Tabellennamen wird die Groß-/Kleinschreibung beachtet. Geben Sie einen Primärschlüssel an. Aktivieren Sie dazu ein Kästchen in der Spalte Primärschlüssel. Geben Sie außerdem die Felder an, die Sie in die neue Tabelle schreiben möchten, indem Sie jeweils das zugehörige Kästchen in der Spalte Einschließen aktivieren. Die Spalte Breite gibt die Länge des Feldes für den Datentyp Zeichenfolge an. Der Standardwert ist 512. Falls die Spalte NULL zulassen aktiviert ist und die Eingabefelder einen Nullwert enthalten, schreibt der Datenfluss den Nullwert in die Datenbank. Beachten Sie, dass Sie den Spaltennamen bearbeiten können, indem Sie den Wert in der Spalte „Ausgabefelder“ ändern.</p> <p>Die Schaltfläche Tabelle erstellen unterstützt die Tabellenerstellung in den folgenden Datenbanken:</p> <ul style="list-style-type: none"> • Axion • DB2 • Derby/Cloudscape • Firebird • HSQLDB • Interbase • MaxDB/SapDB • McKoi • MySQL • Oracle • PostgreSQL • SQL Server • Sybase <p>Anmerkung: Für DB2-Datenbanken gilt: Wenn Sie versuchen, eine Tabelle zu erstellen und die Seitengröße kleiner als die Gesamtlänge aller Zeichenfolgenspalten ist, erhalten Sie die Fehlermeldung „Fehler beim Erstellen von Text aus Inhalt. Serialisierbare Klasse nicht für Broker verfügbar.“</p>
Schrittfelder	<p>In der Spalte „Schrittfelder“ können Sie das Feld angeben, das Sie in das in der Spalte „Datenbankfeld“ angezeigte Datenbankfeld schreiben möchten.</p>
Einschließen	<p>Die Spalte „Einschließen“ ermöglicht Ihnen ebenfalls, die Felder auszuwählen, in die Sie schreiben möchten.</p> <p>Anmerkung: Um schlechte Leistung zu verhindern, sollte ein sortierter Index oder Schlüssel in der Datenbanktabelle vorhanden sein.</p>

Anmerkung: Der **Write to DB**-Schritt schreibt alle Werte des Datentyps `date` als `String`-Werte. Dieses Verhalten beruht auf dem von Spectrum verwendeten Standardtreiber, dem *jTDS-Treiber*. Um alle `date`-Datentypenwerte unverändert zu verarbeiten, verwenden Sie den JDBC-Treiber von Microsoft.

Datenbankverbindungs-Manager

Der Datenbankverbindungs-Manager ermöglicht Ihnen, registrierte Datenbankverbindungen zu verwalten. So können Sie Verbindungen hinzufügen, ändern, löschen und testen:

1. Klicken Sie im Dialogfeld „**Write to DB**“-Optionen auf **Verwalten**.
2. Klicken Sie auf **Hinzufügen**, **Ändern** oder **Löschen**.
3. Wenn Sie eine Datenbankverbindung hinzufügen oder ändern, füllen Sie diese Felder aus:
 - Verbindungsname: Geben Sie den Namen der neuen Verbindung ein.
 - Datenbanktreiber: Wählen Sie den passenden Datenbanktyp aus.
 - Verbindungsoptionen: Geben Sie alle Optionen an, normalerweise Host, Port, Instanz, Benutzername und Kennwort.

Anmerkung: Sie können die Verbindung per Klick auf **Testen** testen.

4. Wenn Sie eine Datenbankverbindung löschen möchten, wählen Sie die gewünschte Verbindung aus und klicken Sie auf „Löschen“.

Registerkarte „Laufzeit“

Name der Option	Beschreibung
Schreibmodus	<p>Gibt den Typ von Aktionen an, die beim Schreiben in die Datenbank ausgeführt werden sollen. Zur Auswahl stehen:</p> <p>Einfügen Fügt neue Datensätze in die Datenbank ein, aktualisiert aber keine vorhandenen Datensätze. Dies ist die Standardeinstellung.</p> <p>Update Aktualisiert vorhandene Datensätze in der Datenbank, fügt aber keine neuen Datensätze ein.</p> <p>Anmerkung: Wenn Sie Aktualisieren auswählen, muss der Name der Primärschlüsselspalte in der Eingabetabelle dem Namen der Primärschlüsselspalte in der Ausgabetablelle entsprechen. Wenn Sie versuchen, eine Tabelle zu aktualisieren, bei der der Name der Primärschlüsselspalte nicht der Eingabe entspricht oder bei der die Primärschlüsselspalte nicht definiert ist, funktioniert die Aktualisierung nicht.</p> <p>Einfügen, wenn Aktualisieren nicht möglich Fügt neue Datensätze in die Datenbank ein, wenn der Datensatz nicht vorhanden ist, und aktualisiert andernfalls den vorhandenen Datensatz.</p>
Übergabe als Batch	<p>Wählen Sie diese Option aus, um Änderungen an die Datenbank zu übergeben, nachdem eine angegebene Anzahl von Datensätzen verarbeitet wurde. Standardmäßig ist diese Option nicht ausgewählt, was bedeutet, dass Änderungen sofort übergeben werden, nachdem ein jeder Datensatz verarbeitet wurde. Durch Auswahl dieser Option kann sich die Leistung des „Write to DB“-Schrittes erheblich verbessern.</p>

Name der Option	Beschreibung
Batchgröße	<p>Wenn Sie die Option Übergabe als Batch aktivieren, geben Sie damit die Anzahl von Datensätzen an, die in jedem Batch an die Datenbank übergeben werden sollen. Der Standardwert ist 1.000. Bei mit Spectrum™ Technology Platform 7.0 und älter erstellten Datenflüssen ist der Standardwert 100.</p> <p>Eine größere Batchgröße bietet nicht immer eine bessere Leistung beim Laden. Berücksichtigen Sie die folgenden Faktoren, wenn Sie eine Batchgröße wählen:</p> <ul style="list-style-type: none"> • 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.
Am Ende übergeben	<p>Wählen Sie diese Option aus, um sicherzustellen, dass die Übergabeoperation für die Datenbank ausgeführt wird, nachdem alle Datensätze an die Datenbank übertragen wurden.</p>
Anzahl zu übergebender Batches	<p>Geben Sie einen Wert an, bei dessen Erreichen die Datensätze übergeben werden sollen. Datensätze werden immer dann an die Datenbank übergeben, wenn (Anzahl zu übergebender Batches * Batchgröße) Datensätze an die Datenbank übertragen wurden. Beispiel: Wenn Batchgröße auf 1.000 und Anzahl zu übergebender Batches auf 3 gesetzt ist, erfolgt die Übergabe immer, nachdem 3.000 Datensätze an die Datenbank übertragen wurden.</p>
Tabelle vor dem Einfügen von Daten abschneiden	<p>Wählen Sie diese Option aus, wenn Sie vor dem Schreiben in die Datenbank alle Daten aus der Tabelle löschen möchten.</p>

Name der Option	Beschreibung
Tabelle löschen und neu erstellen, wenn sie bereits vorhanden ist	<p>Wählen Sie diese Option aus, um die Tabelle vor dem Schreiben der Ausgabe des Datenflusses in die Tabelle zu löschen und erneut zu erstellen. Diese Option ist nützlich, wenn Sie möchten, dass das Schema der Tabelle mit den Feldern aus dem Datenfluss übereinstimmt und keine fremden Schemadaten enthält.</p> <p>Die Tabelle, die gelöscht und erneut erstellt wird, ist die Tabelle, die im Feld Tabelle/Ansicht auf der Registerkarte Allgemein angegeben ist. Beispiel: Wenn Sie die Tabelle „Customers“ im Feld Tabelle/Ansicht angeben und Tabelle löschen und neu erstellen, wenn sie bereits vorhanden ist auswählen, wird die Tabelle „Customers“ aus der Datenbank gelöscht und eine neue Tabelle namens „Customers“ erstellt – mit einem Schema, das genau den in die Tabelle geschriebenen Feldern entspricht.</p>

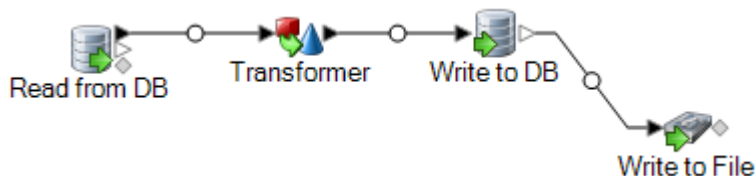
Konfigurieren der Fehlerbehandlung in „Write to DB“

Der „Write to DB“-Schritt hat einen Fehlerport, der Ihnen ermöglicht, Datensätze herauszufiltern, die Datenbankfehler verursachen, wenn der Datensatz in eine Datenbank geschrieben wird, etwa ein Verstoß bei einer Primärschlüsseinschränkung oder bei einer eindeutigen Einschränkung. Diese Datensätze können dann in einem anderen Pfad im Datenfluss weitergeleitet werden, während andere Datensätze erfolgreich übergeben werden. Beispiel: Wenn Sie 100 Datensätze verarbeiten und die Datensätze 4, 23 und 56 einen Datenbankfehler verursachen, würden diese drei Datensätze über den Fehlerport weitergeleitet, während die anderen 97 Datensätze an die Datenbank übergeben würden.

Anmerkung: Sie können den Fehlerport verwenden, müssen es aber nicht. Wenn Sie ihn nicht verwenden, schlägt der Auftrag fehl, wenn ein beliebiger Datensatz einen Fehler verursacht.

- Wählen Sie in der Palette den Typ des Schrittes zur Verarbeitung von Datensätzen mit Fehlern aus (z. B. „Write to File“) und ziehen Sie den Schritt auf die Arbeitsfläche. Sie haben mehrere Optionen zum Auswählen eines Schrittes:
 - Möchten Sie fehlgeschlagene Datensätze in eine Datei schreiben, ziehen Sie einen der folgenden Schritte auf die Arbeitsfläche: „Write to File“, „Write to XML“ oder „Write to Variable Format File“.
 - Um fehlgeschlagene Datensätze einfach zu löschen, ziehen Sie „Write to Null“ auf die Arbeitsfläche.
- Verbinden Sie den Fehlerport bei „Write to DB“ mit dem Schritt, der fehlgeschlagene Datensätze verarbeiten soll.

Das folgende Beispiel zeigt den Fehlerport bei „Write to DB“ verbunden mit einem „Write to File“-Schritt. In diesem Beispiel werden Datensätze, die beim Schreiben in die Datenbank einen Fehler verursachen, stattdessen in die im „Write to File“-Schritt angegebene Datei geschrieben.



Wenn Sie den Datenfluss ausführen, werden Datensätze, die einen Fehler verursachen, über den Fehlerport weitergeleitet. Die Datensätze vom Fehlerport enthalten die in „Write to DB“ angegebenen Felder sowie zusätzlich die folgenden Felder:

Error.code	Dieses Feld enthält den numerischen von der Datenbank zurückgegebenen Fehlercode. Beispiel: Beim Fehler <code>ORA-00001: unique constraint ANKUSH.SYS_C0010018) violated</code> wäre der Wert im Feld „Error.code“ 1. Eine Liste der Fehlercodes finden Sie in der Dokumentation Ihrer Datenbanksoftware.
Error.Message	Dieses Feld enthält die von der Datenbank zurückgegebene Fehlermeldung. Beispiel: <code>ORA-01034 ORACLE not available</code> . In diesem Fall wäre <code>ORACLE not available</code> der Wert im Feld „Error.Message“. Eine Liste der Fehlermeldungen finden Sie in der Dokumentation Ihrer Datenbanksoftware.
Error.SQLState	Dieses Feld enthält den SQLSTATE-Code, der ausführliche Informationen über die Ursache des Fehlers bereitstellt. Eine Liste der SQLSTATE-Codes finden Sie in der Dokumentation Ihrer Datenbanksoftware.
Zeitstempel	Das Datum und die Uhrzeit auf dem Spectrum™ Technology Platform-Server, als der Fehler auftrat.
Benutzername	Der Name des Spectrum™ Technology Platform-Benutzers, der den Datenfluss ausgeführt hat.

Write to File

„Write to File“ schreibt die Datenflussausgabe in ein Flatfile. Die Datensätze enthalten alle dieselben Felder. Wenn Sie Datensätze unterschiedlichen Formats schreiben möchten, lesen Sie [Write to Variable Format File](#) auf Seite 298. Wenn Sie Datensätze in eine XML-Datei schreiben möchten, lesen Sie [Write to XML](#) auf Seite 308.

Tip: Sie können Ihre Quelle kopieren und als Datenladung in den Datenfluss einfügen, um die Datei schnell zu konfigurieren und dieselben Felder zu verwenden, die Sie in der Quelle definiert haben.

Registerkarte „Dateieigenschaften“

Feldname	Beschreibung
Servername	<p>Dies gibt an, ob sich die von Ihnen als Eingabe ausgewählte Datei auf dem Computer mit dem Enterprise Designer oder auf dem Spectrum™ Technology Platform-Server befindet. Wenn Sie eine Datei auf dem lokalen Computer auswählen, lautet der Servername „Arbeitsplatz“. Wenn Sie eine Datei auf dem Server auswählen, lautet der Servername Spectrum™ Technology Platform.</p>
Dateiname	<p>Gibt den Pfad zur Datei an. Klicken Sie auf die Ellipsenschaltfläche (...), um die gewünschte Datei zu lokalisieren.</p> <p>Beim Schreiben einer Datei auf einen HDFS-Dateiserver werden die folgenden Komprimierungsformate unterstützt:</p> <ol style="list-style-type: none"> 1. GZIP (.gz) 2. BZIP2 (.bz2) <p>Anmerkung: Geben Sie beim Dateinamen die passende Erweiterung an, um das gewünschte Komprimierungsformat anzugeben, das beim Schreiben in die Datei verwendet werden soll.</p> <p>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.</p>
Datensatztyp	<p>Dies ist das Format der Datensätze in der Datei. Wählen Sie eine Option:</p> <p>Zeilensequenziell 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 über feste Startzeichen- und Endzeichenpositionen verfügt.</p> <p>Feste Breite Dies ist eine Textdatei, in der jeder Datensatz eine bestimmte Anzahl von Zeichen lang ist und jedes Feld über feste Startzeichen- und Endzeichenpositionen verfügt.</p> <p>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.</p>

Feldname	Beschreibung
Zeichencodierung	<p>Dies ist die Codierung der Textdatei. Wählen Sie eine der folgenden Optionen:</p> <p>UTF-8 Die Datei unterstützt alle Unicode-Zeichen und ist abwärtskompatibel mit ASCII. Weitere Informationen zu UTF finden Sie unter unicode.org/faq/utf_bom.html.</p> <p>UTF-16 Die Datei unterstützt alle Unicode-Zeichen, ist aber nicht abwärtskompatibel mit ASCII. Weitere Informationen zu UTF finden Sie unter unicode.org/faq/utf_bom.html.</p> <p>US-ASCII Die Zeichencodierung basiert auf der Reihenfolge des englischen Alphabets.</p> <p>UTF-16BE Die Datei ist UTF-16-codiert mit Big-Endian-Reihenfolge (höchstwertiges Byte zuerst).</p> <p>UTF-16LE Die Datei ist UTF-16-codiert mit Little-Endian-Reihenfolge (niedrigstwertiges Byte zuerst).</p> <p>ISO-8859-1 Dies ist eine ASCII-Zeichencodierung, die normalerweise für westeuropäische Sprachen verwendet wird. Sie ist auch bekannt als Latin-1.</p> <p>ISO-8859-3 Dies ist eine ASCII-Zeichencodierung, die normalerweise für südeuropäische Sprachen verwendet wird. Sie ist auch bekannt als Latin-3.</p> <p>ISO-8859-9 Dies ist eine ASCII-Zeichencodierung, die normalerweise für Türkisch verwendet wird. Sie ist auch bekannt als Latin-5.</p> <p>CP850 Dies ist eine ASCII-Codeseite, die für westeuropäische Sprachen verwendet wird.</p> <p>CP500 Dies ist eine EBCDIC-Codeseite, die für westeuropäische Sprachen verwendet wird.</p> <p>Shift_JIS Dies ist eine Zeichencodierung für Japanisch.</p> <p>MS932 Dies ist eine Shift_JIS-Erweiterung von Microsoft, die besondere NEC-Zeichen, eine NEC-Auswahl aus IBM-Erweiterungen und IBM-Erweiterungen einschließt.</p> <p>CP1047 Dies ist eine EBCDIC-Codeseite mit dem vollständigen Latin-1-Zeichensatz.</p>

Feldname	Beschreibung
Feldtrennzeichen	<p>Gibt das Zeichen an, das zum Trennen von Feldern in einer Datei mit Begrenzungszeichen verwendet wird.</p> <p>Im folgenden Beispiel wird ein Pipe-Zeichen () im Datensatz als Feldtrennzeichen verwendet:</p> <pre data-bbox="552 525 1412 588">7200 13TH ST MIAMI FL 33144</pre> <p>Die folgenden Zeichen stehen zur Definition als Feldtrennzeichen zur Verfügung:</p> <ul data-bbox="552 651 714 861" style="list-style-type: none">• Leerzeichen• Tabulator• Komma• Punkt• Semikolon• Pipe <p>Wenn in der Datei ein anderes Zeichen als Feldtrennzeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Begrenzungszeichen auszuwählen.</p>
Textbegrenzungszeichen	<p>Dies ist das Zeichen, das in einer Datei mit Begrenzungszeichen Textwerte umschließt.</p> <p>Im folgenden Beispiel werden im Datensatz doppelte Anführungszeichen (") als Textbegrenzungszeichen verwendet.</p> <pre data-bbox="552 1197 1412 1260">"7200 13TH ST" "MIAMI" "FL" "33144"</pre> <p>Die folgenden Zeichen stehen zur Definition als Textbegrenzungszeichen zur Verfügung:</p> <ul data-bbox="552 1344 941 1428" style="list-style-type: none">• einfaches Anführungszeichen (')• doppeltes Anführungszeichen (") <p>Wenn in der Datei ein anderes Textbegrenzungszeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Textbegrenzungszeichen auszuwählen.</p>

Feldname	Beschreibung
Datensatz-Trennzeichen	<p>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-Zeileneende verwenden aktivieren.</p> <p>Die verfügbaren Einstellungen für Datensatz-Trennzeichen sind:</p> <p>Unix (U+000A) Die Datensätze sind durch ein Zeilenvorschubzeichen getrennt. Dies ist bei Unix-Systemen das standardmäßige Datensatz-Trennzeichen.</p> <p>Macintosh (U+000D) Die Datensätze sind durch ein Zeilenumbruchzeichen getrennt. Dies ist bei Macintosh-Systemen das standardmäßige Datensatz-Trennzeichen.</p> <p>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.</p> <p>Wenn in der Datei ein anderes Datensatz-Trennzeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Datensatz-Trennzeichen auszuwählen.</p>
Standard-Zeileneende verwenden	<p>Gibt an, dass das standardmäßige Zeilenendzeichen (EOL) des Betriebssystems, auf dem der Spectrum™ Technology Platform-Server läuft, als Datensatz-Trennzeichen der Datei verwendet wird.</p> <p>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-Zeileneendezeichen verwendet wird, der Server aber auf Linux läuft, aktivieren Sie diese Option nicht. Wählen Sie stattdessen im Feld Datensatz-Trennzeichen die Option „Windows“ aus.</p>
Datensatzlänge	<p>Geben Sie bei Dateien an, die Datensätze mit fester Breite enthalten, die genaue Anzahl von Zeichen jedes Datensatzes.</p> <p>Geben Sie bei zeilensequenziellen Dateien die Länge des längsten Datensatzes in der Datei in Zeichen an.</p>
Erste Zeile ist Kopfzeilen-Datensatz	<p>Gibt an, ob der erste Datensatz in einer Datei mit Begrenzungszeichen Überschriften und keine Daten enthält.</p> <p>Im folgenden Dateiausschnitt ist der erste Datensatz ein Kopfzeilen-Datensatz.</p> <pre data-bbox="560 1764 1421 1869">"AddressLine1" "City" "StateProvince" "PostalCode" "7200 13TH ST" "MIAMI" "FL" "33144" "One Global View" "Troy" "NY" 12180</pre>

Feldname	Beschreibung
Datensätze mit weniger Feldern als definiert als falsch formatiert behandeln	Datensätze in Dateien mit Begrenzungszeichen, die weniger Felder enthalten, als auf der Registerkarte Felder definiert wurden, werden als falsch formatiert behandelt.
Importieren	Dies importiert die Definition des Dateilayouts, die Codierungseinstellung und Sortieroptionen aus einer Einstellungsdatei. Die Einstellungsdatei wird beim Exportieren von Einstellungen durch einen anderen „Read from File“- oder „Write to File“-Schritt erstellt. Dieser verwendet dieselbe Eingabedatei oder eine Datei, die über dasselbe Layout wie die Datei verfügt, mit der Sie arbeiten.
Exportieren	Dies exportiert die Definition des Dateilayouts, die Codierungseinstellung und Sortieroptionen in eine Einstellungsdatei. Sie können anschließend diese Einstellungen in andere „Read from File“- oder „Write to File“-Schritte importieren, die dieselbe Eingabedatei oder eine Datei mit demselben Layout wie die Datei verwenden, mit der Sie gerade arbeiten. Sie können die Einstellungsdatei auch mit dem Job Executor verwenden, um die Dateieinstellungen zur Laufzeit anzugeben. Informationen zur Einstellungsdatei finden Sie unter Die Dateidefinitions-Einstellungsdatei auf Seite 165.

Registerkarte „Felder“

Die Registerkarte „Felder“ definiert die Namen, Positionen und, bei Dateien vom Typ „Feste Breite“ und „Zeilensequenziell“, die Länge der Felder in der Datei. Weitere Informationen bieten die folgenden Themen:

- [Definieren von Feldern in einer durch Trennzeichen getrennten Ausgabedatei](#) auf Seite 267
- [Definieren von Feldern in einer Datei vom Typ „Zeilensequenziell“ oder „Feste Breite“](#) auf Seite 270

Registerkarte „Sortierfelder“

Die Registerkarte „Sortierfelder“ definiert Felder, nach denen die Ausgabesätze sortiert werden sollen, bevor sie in die Ausgabedatei geschrieben werden. Die Sortierung ist optional. Weitere Informationen finden Sie unter [Sortieren von Ausgabedatensätzen](#) auf Seite 273.

Registerkarte „Laufzeit“

Name der Option	Beschreibung
Dateiname	Zeigt die auf der Registerkarte Dateieigenschaften definierte Datei an.
Mehrere Dateien generieren	<p>Wählen Sie diese Option aus, um Datensätze in unterschiedliche Dateien zu schreiben (statt alle Datensätze in eine Datei). Die Datei, in die jeder Datensatz geschrieben wird, ist im Datensatz selbst angegeben. Jeder Datensatz muss ein Feld enthalten, das entweder den Namen oder den vollständigen Pfad zu der Datei angibt, in die der Datensatz geschrieben werden soll. Wenn Sie beispielsweise die Aktienkurse verschiedener Unternehmen (verschiedener Gruppen) allen Kunden gesondert senden möchten, schreibt dieses Feature die Aktienkurse der verschiedenen Unternehmen in separate Dateien, die Sie auf Wunsch jedem der Kunden senden können. Wenn Sie die Option Mehrere Dateien generieren aktivieren, müssen Sie eine Ausgabedatei entweder auf dem Spectrum™ Technology Platform-Server oder einem FTP-Server angeben. Wenn Sie Daten in eine Datei auf einem FTP-Server schreiben möchten, müssen Sie mit Management Console eine Verbindung zum Dateiserver definieren.</p> <p>Anmerkung: Die Datensätze in der im Dateipfadfeld ausgewählten Spalte müssen sortiert vorliegen. Verwenden Sie dieses Feature, wenn der Datensatz entweder den Namen oder den vollständigen Pfad der Datei enthält.</p>
Dateipfadfeld	Wählt das Feld aus, das den Pfad (entweder ein Dateiname oder der vollständige Dateipfad) der Datei enthält, in die Sie den Datensatz schreiben möchten. Dieses Feld ist nur aktiviert, wenn Sie Mehrere Dateien generieren auswählen.
Schreibmodus	<p>Gibt an, ob die Ausgabe des Datenflusses am Ende der Datei hinzugefügt oder ob die in der Datei vorhandenen Daten vor dem Schreiben der Ausgabe gelöscht werden sollen. Zur Auswahl stehen:</p> <p>Überschreiben Ersetzt die vorhandenen Daten in der Ausgabedatei jedes Mal, wenn der Datenfluss ausgeführt wird.</p> <p>Anfügen Fügt die Ausgabe des Datenflusses am Ende der Datei hinzu, ohne die vorhandenen Daten der Datei zu löschen.</p>

Definieren von Feldern in einer durch Trennzeichen getrennten Ausgabedatei

Im „Write to File“-Schritt definiert die Registerkarte **Felder** die Namen, Position, und, für einige Dateitypen, die Längen der Felder in der Datei. Nach Definieren einer Ausgabedatei auf der Registerkarte **Dateieigenschaften** können Sie die Felder definieren.

Wenn die Ausgabedatei einen Kopfzeilendatensatz enthält, können Sie die Felder schnell definieren, indem Sie auf **Erneut generieren** klicken.

Klicken Sie zum Definieren von Feldern ohne Standardwerte für Position, Länge und Datentyp auf **Schnell hinzufügen** und wählen Sie die hinzuzufügenden Felder aus.

Wenn die Eingabedatei keinen Kopfzeilendatensatz enthält oder wenn Sie die Felder manuell definieren möchten, gehen Sie wie folgt vor:

1. Klicken Sie auf **Hinzufügen**.
2. Wählen Sie im Feld **Name** das Feld, das Sie hinzufügen möchten.
3. Wählen Sie im Feld **Typ** den Datentyp des Feldes aus dem Datenfluss aus.

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 $(2-2^{-52}) \times 2^{1023}$ 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 $(2-2^{-23}) \times 2^{127}$ 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.

4. Wenn Sie den Datentyp „Datum“, „Uhrzeit“ oder „Numerisch“ ausgewählt haben, können Sie das Standardformat für Datum/Uhrzeit oder Zahlen verwenden oder ein anderes Format für dieses bestimmte Feld angeben. Das Standardformat ist entweder das Systemstandardformat, das in den Typkonvertierungsoptionen in der Management Console festgelegt wurde, oder es ist das Standardformat des Datenflusses, das in den Typkonvertierungsoptionen im Enterprise Designer angegeben wurde. Das wirksame Format wird angezeigt. Lassen Sie **Standard** ausgewählt, wenn Sie das Standardformat verwenden möchten. Möchten Sie hingegen ein anderes Format angeben, wählen Sie **Benutzerdefiniert** aus, und gehen Sie wie folgt vor:
- a) Wählen Sie im Feld **Gebietsschema** das Land aus, dessen Formatierungskonventionen Sie verwenden möchten. Durch Ihre Auswahl werden die Standardwerte im Feld **Format** festgelegt. Bei Datumsdaten legt Ihre Auswahl auch die Sprache fest, die verwendet wird, 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 **Format** das Format für die Daten aus. Das Format ist abhängig vom Datentyp des Feldes. Es wird eine Liste der am häufigsten verwendeten Formate für das ausgewählte Gebietsschema bereitgestellt.

Ein Beispiel für das ausgewählte Format wird rechts vom Feld **Format** angezeigt.

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 318 erläuterten Schreibweise in das Feld ein. Geben Sie zur Festlegung eines eigenen Zahlenformats das Format mithilfe der unter **Zahlenmuster** auf Seite 321 erläuterten Schreibweise in das Feld ein.

5. Klicken Sie auf **Hinzufügen**.

Nach Definieren der Felder in der Ausgabedatei können Sie ihren Inhalt und ihr Layout bearbeiten.

Name der Option	Beschreibung
Hinzufügen	Fügt der Ausgabe ein Feld hinzu. Sie können ein Feld am Ende des vorhandenen Layouts anfügen oder ein Feld an einer vorhandenen Position einfügen, wobei die Position der restlichen Felder dementsprechend angepasst wird.
Ändern	Ändert den Namen und Typ des Feldes.
Entfernen	Entfernt das ausgewählte Feld aus der Ausgabe.
Nach oben/Nach unten	Ordnet das ausgewählte Feld neu an.

Definieren von Feldern in einer Datei vom Typ „Zeilensequenziell“ oder „Feste Breite“

Im „Write to File“-Schritt definiert die Registerkarte **Felder** die Namen, Position, und, für einige Dateitypen, die Längen der Felder in der Datei. Nach Definieren einer Ausgabedatei auf der Registerkarte **Dateieigenschaften** können Sie die Felder definieren.

Klicken Sie zum Definieren von Feldern ohne Standardwerte für Position, Länge und Datentyp auf **Schnell hinzufügen** und wählen Sie die hinzuzufügenden Felder aus.

Um Felder manuell aus einer Liste von im Datenfluss verwendeten Feldern hinzuzufügen, gehen Sie wie folgt vor:

1. Klicken Sie auf **Hinzufügen**.
2. Wählen Sie im Feld **Name** das Feld, das Sie hinzufügen möchten.
3. Wählen Sie im Feld **Typ** den Datentyp des Feldes aus dem Datenfluss aus.

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 $(2-2^{-52}) \times 2^{1023}$ 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 $(2-2^{-23}) \times 2^{127}$ 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: <pre><Names> <Name>John Smith</Name> <Name>Ann Fowler</Name> </Names></pre> <p>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.</p>
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.

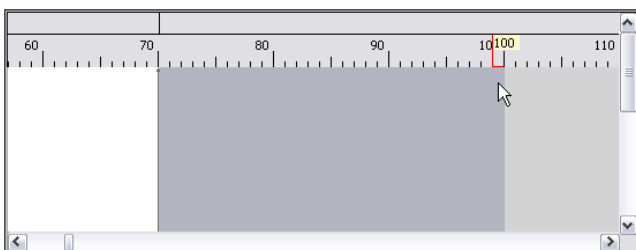
4. Wenn Sie den Datentyp „Datum“, „Uhrzeit“ oder „Numerisch“ ausgewählt haben, können Sie das Standardformat für Datum/Uhrzeit oder Zahlen verwenden oder ein anderes Format für dieses bestimmte Feld angeben. Das Standardformat ist entweder das Systemstandardformat, das in den Typkonvertierungsoptionen in der Management Console festgelegt wurde, oder es ist das Standardformat des Datenflusses, das in den Typkonvertierungsoptionen im Enterprise Designer angegeben wurde. Das wirksame Format wird angezeigt. Lassen Sie **Standard** ausgewählt, wenn Sie das Standardformat verwenden möchten. Möchten Sie hingegen ein anderes Format angeben, wählen Sie **Benutzerdefiniert** aus, und gehen Sie wie folgt vor:
- Wählen Sie im Feld **Gebietsschema** das Land aus, dessen Formatierungskonventionen Sie verwenden möchten. Durch Ihre Auswahl werden die Standardwerte im Feld **Format** festgelegt. Bei Datumsdaten legt Ihre Auswahl auch die Sprache fest, die verwendet wird, 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“.
 - Wählen Sie im Feld **Format** das Format für die Daten aus. Das Format ist abhängig vom Datentyp des Feldes. Es wird eine Liste der am häufigsten verwendeten Formate für das ausgewählte Gebietsschema bereitgestellt.

Ein Beispiel für das ausgewählte Format wird rechts vom Feld **Format** angezeigt.

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 318 erläuterten Schreibweise in das Feld ein. Geben Sie zur Festlegung eines eigenen Zahlenformats das Format mithilfe der unter **Zahlenmuster** auf Seite 321 erläuterten Schreibweise in das Feld ein.

5. Die Felder **Startposition** und **Länge** werden auf Basis der Daten im Datenfluss und der Anzahl von Feldern, die Sie bereits hinzugefügt haben, automatisch ausgefüllt.
6. Klicken Sie auf **Hinzufügen**.

Wahlweise können Sie ein Feld auch hinzufügen, indem Sie zuerst die Startposition und Länge des Feldes definieren. Klicken Sie dazu unter **Beispieldatei** auf die Position, an der das Feld beginnen soll, und ziehen Sie den Zeiger nach links, sodass das gewünschte Feld markiert wird (siehe Abbildung):



Nach Definieren der Felder in der Ausgabedatei können Sie ihren Inhalt und ihr Layout bearbeiten. Die Option **Startposition neu berechnen** weist den „Write to File“-Schritt an, die Positionen der Felder neu zu berechnen, wenn Sie ein Feld in der Ausgabedatei ändern, verschieben oder löschen. Deaktivieren Sie dieses Kästchen, wenn Sie die Positionen nicht neu berechnen und stattdessen

möchten, dass die Felder an ihrer vorhandenen Position bleiben, nachdem Sie die Ausgabedatei bearbeitet haben.

Name der Option	Beschreibung
Hinzufügen	Fügt der Ausgabe ein Feld hinzu.
Ändern	Ändert den Namen, Typ, die Startposition und Länge des Feldes.
Entfernen	Entfernt das ausgewählte Feld aus der Ausgabe.
Nach oben/Nach unten	Ordnet das ausgewählte Feld neu an.

Sortieren von Ausgabedatensätzen

Im Schritt „Write to File“ definiert die Registerkarte **Sortierfelder** Felder, nach denen die Ausgabesätze sortiert werden sollen, bevor sie in die Ausgabedatei geschrieben werden. Die Sortierung ist optional.

1. Klicken Sie in „Write to File“ auf die Registerkarte **Sortierfelder**.
2. Klicken Sie auf **Hinzufügen**.
3. 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 im Datenfluss ab.
4. Wählen Sie in der Spalte **Sortierung** die Option Ascending oder Descending aus.
5. Wiederholen Sie diese Schritte, bis Sie alle Ausgabefelder 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**.
6. Die Standardoptionen für die Sortierleistung für Ihr System werden in der Management Console festgelegt. Wenn Sie die Standardoptionen für die Sortierleistung Ihres Systems überschreiben möchten, klicken Sie auf **Erweitert**. Das Dialogfeld **Erweiterte Optionen** enthält die folgenden Optionen für die Sortierleistung:

Beschränkung für Datensätze im Arbeitsspeicher Gibt die maximale Anzahl von Datenzeilen an, die ein Sorter im Arbeitsspeicher belässt, bevor er sie auf die Festplatte auslagert. Standardmäßig erfolgt eine Sortierung von bis zu 10.000 Datensätze in einem Arbeitsspeicher und eine Sortierung von mehr als 10.000 Datensätzen auf einem Datenträger. Der maximale Wert umfasst 100.000

Datensätze. In der Regel erfolgt eine Sortierung im Arbeitsspeicher deutlich schneller als auf einem Datenträger. Daher sollte hierfür ein Wert festgelegt werden, der hoch genug ist, damit ein Großteil der Sortierungen im Arbeitsspeicher erfolgen kann und nur umfangreiche Datasets auf einen Datenträger geschrieben werden.

Anmerkung: Vorsicht bei Umgebungen, in denen Aufträge gleichzeitig ausgeführt werden: Durch das Erhöhen der Einstellung **Beschränkung für Datensätze im Arbeitsspeicher** wird die Wahrscheinlichkeit erhöht, dass nicht mehr genügend Arbeitsspeicher verfügbar ist.

Maximale Anzahl temporärer Dateien Gibt die maximale Anzahl temporärer Dateien an, die von einem Sortiervorgang verwendet werden können. Die Verwendung einer größeren Anzahl temporärer Dateien kann zu einer besseren Leistung führen. Die optimale Anzahl hängt jedoch stark von der Konfiguration des Servers ab, auf dem Spectrum™ Technology Platform ausgeführt wird. Sie sollten mit verschiedenen Einstellungen experimentieren und dabei die Auswirkungen auf die Leistung beobachten, wenn mehr oder weniger temporäre Dateien verwendet werden. Verwenden Sie für die Berechnung der ungefähren Anzahl temporärer Dateien die folgende Gleichung:

$$\frac{(NumberOfRecords \times 2)}{InMemoryRecordLimit} = NumberOfTempFiles$$

Beachten Sie, dass die maximale Anzahl der temporären Dateien 1.000 nicht überschreiten darf.

Komprimierung aktivieren Gibt an, dass temporäre Dateien komprimiert werden, wenn sie auf die Festplatte geschrieben werden.

Anmerkung: Welche Einstellungen für die Sortierleistung optimal sind, hängt von der Hardwarekonfiguration Ihres Servers ab. Dennoch wird mit der folgenden Gleichung im Allgemeinen eine gute Sortierleistung erzielt:

$$\frac{(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2)}{TotalNumberOfRecords} \geq$$

Die Dateidefinitions-Einstellungsdatei

Eine Dateidefinitions-Einstellungsdatei enthält den Dateiaufbau, die Codierung und die Sortieroptionen, die aus einem „Read from File“- oder „Write to File“-Schritt exportiert wurden. Die Dateidefinitions-Einstellungsdatei kann in „Read from File“ oder „Write to File“ importiert werden, um die Optionen des Schrittes schnell festzulegen, anstatt sie manuell anzugeben.

Die einfachste Methode zum Erstellen einer Dateidefinitions-Einstellungsdatei besteht darin, die Dateieinstellungen mit „Read from File“ oder „Write to File“ anzugeben und dann auf die Schaltfläche **Exportieren** zu klicken, um die Dateidefinitions-Einstellungsdatei zu generieren.

Zu Ihrer Information ist jedoch das Schema der Dateidefinitions-Einstellungsdatei unten abgebildet. Jedes Element in der XML-Datei hat einen Typ, und wenn dieser Typ weder „Zeichenfolge“ noch „Ganze Zahl“ ist, werden die zulässigen Werte angezeigt. Diese Werte entsprechen direkt Optionen im Dialogfeld des Schrittes. So entspricht beispielsweise das Element „FileTypeEnum“ dem Feld „Datensatztyp“ auf der Registerkarte „Dateieigenschaften“, und die folgenden drei Werte werden im Schema angezeigt: „LineSequential“, „fixedwidth“ und „Mit Begrenzungszeichen“.

Anmerkung: Wenn Sie „benutzerdefiniert“ für die Felder „LineSeparator“, „FieldSeparator“ oder „TextQualifier“ eingeben, muss auch ein entsprechendes benutzerdefiniertes Element (z. B. „CustomLineSeparator“, „CustomFieldSeparator“ oder „CustomTextQualifier“) mit einer hexadezimalen Zahl, die das Zeichen oder die Zeichenfolge darstellt, angegeben werden.

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema elementFormDefault="qualified"
  xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="FileSchema" nillable="true" type="FileSchema"/>
  <xs:complexType name="FileSchema">
    <xs:sequence>
      <xs:element
        minOccurs="0"
        maxOccurs="1"
        default="linesequential"
        name="Type"
        type="FileTypeEnum"/>
      <xs:element
        minOccurs="0"
        maxOccurs="1"
        default="UTF-8" name="Encoding" type="xs:string"/>
      <xs:element
        minOccurs="0"
        maxOccurs="1"
        name="RecordLength"
        type="xs:int"/>
      <xs:element
        minOccurs="0"
        maxOccurs="1"
        default="default"
        name="LineSeparator"
        type="LineSeparatorEnum"/>
      <xs:element
        minOccurs="0"
        maxOccurs="1"
        name="CustomLineSeparator"
        type="xs:string"/>
      <xs:element
        minOccurs="0"
        maxOccurs="1"
```

```

        default="comma"
        name="FieldSeparator"
        type="FieldSeparatorEnum"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  name="CustomFieldSeparator"
  type="xs:string"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  default="none"
  name="TextQualifier"
  type="TextQualifierEnum"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  name="CustomTextQualifier"
  type="xs:string"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  default="false"
  name="HasHeader"
  type="xs:boolean"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  default="true"
  name="EnforceColumnCount"
  type="xs:boolean"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  name="Fields"
  type="ArrayOfFieldSchema"/>
</xs:sequence>
</xs:complexType>
<xs:simpleType name="FileTypeEnum">
  <xs:restriction base="xs:string">
    <xs:enumeration value="linesequential"/>
    <xs:enumeration value="fixedwidth"/>
    <xs:enumeration value="delimited"/>
  </xs:restriction>
</xs:simpleType>
<xs:simpleType name="LineSeparatorEnum">
  <xs:restriction base="xs:string">
    <xs:enumeration value="default"/>
    <xs:enumeration value="windows"/>
    <xs:enumeration value="unix"/>
    <xs:enumeration value="mac"/>
    <xs:enumeration value="custom"/>
  </xs:restriction>

```

```

</xs:simpleType>
<xs:simpleType name="FieldSeparatorEnum">
  <xs:restriction base="xs:string">
    <xs:enumeration value="comma"/>
    <xs:enumeration value="tab"/>
    <xs:enumeration value="space"/>
    <xs:enumeration value="semicolon"/>
    <xs:enumeration value="period"/>
    <xs:enumeration value="pipe"/>
    <xs:enumeration value="custom"/>
  </xs:restriction>
</xs:simpleType>
<xs:simpleType name="TextQualifierEnum">
  <xs:restriction base="xs:string">
    <xs:enumeration value="none"/>
    <xs:enumeration value="single"/>
    <xs:enumeration value="double"/>
    <xs:enumeration value="custom"/>
  </xs:restriction>
</xs:simpleType>
<xs:complexType name="ArrayOfFieldSchema">
  <xs:sequence>
    <xs:element
      minOccurs="0"
      maxOccurs="unbounded"
      name="Field"
      nillable="true"
      type="FieldSchema"/>
  </xs:sequence>
</xs:complexType>
<xs:complexType name="FieldSchema">
  <xs:sequence>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="Name"
      type="xs:string"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      default="string"
      name="Type"
      type="xs:string"/>
    <xs:element
      minOccurs="1"
      maxOccurs="1"
      name="Position"
      type="xs:int"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="Length"
      type="xs:int"/>
  </xs:sequence>
</xs:complexType>

```

```

<xs:element
  minOccurs="0"
  maxOccurs="1"
  default="false"
  name="Trim"
  type="xs:boolean"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  name="Locale"
  type="Locale"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  name="Pattern"
  type="xs:string"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  default="none"
  name="Order"
  type="SortOrderEnum"/>
</xs:sequence>
</xs:complexType>
<xs:complexType name="Locale">
  <xs:sequence>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="Country"
      type="xs:string"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="Language"
      type="xs:string"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="Variant"
      type="xs:string"/>
  </xs:sequence>
</xs:complexType>
<xs:simpleType name="SortOrderEnum">
  <xs:restriction base="xs:string">
    <xs:enumeration value="none"/>
    <xs:enumeration value="ascending"/>
    <xs:enumeration value="descending"/>
  </xs:restriction>
</xs:simpleType>
</xs:schema>

```

Definieren von Datenflussoptionen

Diese Schritte beschreiben, wie Sie einen Datenfluss konfigurieren, um Laufzeioptionen für den „Write to File“-Schritt zu unterstützen.

1. Öffnen Sie den Datenfluss im Enterprise Designer.
2. Wenn Sie Laufzeioptionen 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. Erweitern Sie den **Write to File**-Schritt.
Die verfügbar gemachten Datenflussoptionen sind:
 1. Zeichencodierung
 2. Feldtrennzeichen
 3. Textbegrenzungszeichen
 4. Datensatzlänge
 5. Erste Zeile ist Kopfzeilen-Datensatz
6. Der ausgewählte „Write to File“-Optionsname wird in den Feldern **Optionsname** und **Optionsbeschriftung** angezeigt. Dies ist der Optionsname, der zur Laufzeit angegeben werden muss, um diese Option festzulegen.
7. Geben Sie eine Beschreibung der Option in das Feld **Beschreibung** ein.
8. Wählen Sie im Feld **Ziel** die Option **Ausgewählte(r) Schritt/e** aus.
9. 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.
10. 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 [Angaben von Standarddienstoptionen](#).
11. Klicken Sie auf **OK**.
12. Fügen Sie auf Wunsch weitere Optionen hinzu.
13. Klicken Sie im Dialogfeld **Datenflussoptionen** auf **OK**, sobald Sie damit fertig sind.
14. Speichern Sie den Datenfluss und machen Sie ihn verfügbar.

Regeln für Datenflussoptionen

1. *Zeichencodierung*: Alle Codierungstypen, die in der zugrunde liegenden JVM gültig sind, sind zulässig. Diese Option darf nicht leer sein.
2. *Feldtrennzeichen*: Jedes aus einem einzelnen Zeichen bestehende Trennzeichen ist zulässig. Derzeit werden keine hexadezimalen Werte und keine Leerzeichen unterstützt.
3. *Textbegrenzungszeichen*: Jedes aus einem einzelnen Zeichen bestehende Begrenzungszeichen ist zulässig. Hexadezimale Werte werden nicht unterstützt.
4. *Datensatzlänge*: Nur ganze Zahlen sind zulässig. Nicht numerische oder leere Werte sind nicht zulässig.
5. *Startdatensatz*: Nur ganze Zahlen sind zulässig. Nicht numerische Werte sind nicht zulässig.
6. *Max. Anzahl Datensätze*: Nur ganze Zahlen sind zulässig. Nicht numerische Werte sind nicht zulässig.
7. *Erste Zeile ist Kopfzeilen-Datensatz*: Nur die booleschen Werte `wahr` und `falsch` sind zulässig. Leere Werte sind nicht zulässig.

Write to Hadoop Sequence File

Der Schritt „Write to Hadoop Sequence File“ schreibt Daten so in eine Sequenzdatei, wie sie von einem Datenfluss ausgegeben werden. Eine Sequenzdatei ist ein Flatfile, das aus binären Schlüssel/Wert-Paaren besteht. Weitere Informationen finden Sie unter wiki.apache.org/hadoop/SequenceFile.

Anmerkung: Der Schritt „Write to Hadoop Sequence File“ unterstützt nur durch Trennzeichen getrennte, unkomprimierte Sequenzdateien im Hadoop Distributed File System (HDFS).

Registerkarte „Dateieigenschaften“

Felder	Beschreibung
Server	Dies gibt an, dass die im Feld Dateiname ausgewählte Datei sich auf dem Hadoop-System befindet. Sie müssen eine Verbindung zum Hadoop-Dateiserver in der Management Console erstellen, bevor Sie ihn im Schritt verwenden. Wenn Sie eine Datei auf dem Hadoop-System auswählen, ist der Servername der Name, den Sie in der Management Console beim Erstellen eines Dateiservers angegeben haben.
Dateiname	Gibt den Pfad zur Datei an. Klicken Sie auf die Ellipsenschaltfläche (...), um die gewünschte Datei zu lokalisieren.

Felder	Beschreibung
Feldtrennzeichen	<p>Gibt das Zeichen an, das zum Trennen von Feldern in einer Datei mit Begrenzungszeichen verwendet wird.</p> <p>Im folgenden Beispiel wird ein Pipe-Zeichen () im Datensatz als Feldtrennzeichen verwendet:</p> <pre>7200 13TH ST MIAMI FL 33144</pre> <p>Die folgenden Zeichen stehen zur Definition als Feldtrennzeichen zur Verfügung:</p> <ul style="list-style-type: none">• Leerzeichen• Tabulator• Komma• Punkt• Semikolon• Pipe <p>Wenn in der Datei ein anderes Zeichen als Feldtrennzeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Begrenzungszeichen auszuwählen.</p>
Textbegrenzungszeichen	<p>Dies ist das Zeichen, das in einer Datei mit Begrenzungszeichen Textwerte umschließt.</p> <p>Im folgenden Beispiel werden im Datensatz doppelte Anführungszeichen (") als Textbegrenzungszeichen verwendet.</p> <pre>"7200 13TH ST" "MIAMI" "FL" "33144"</pre> <p>Die folgenden Zeichen stehen zur Definition als Textbegrenzungszeichen zur Verfügung:</p> <ul style="list-style-type: none">• einfaches Anführungszeichen (')• doppeltes Anführungszeichen (") <p>Wenn in der Datei ein anderes Textbegrenzungszeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Textbegrenzungszeichen auszuwählen.</p>

Registerkarte „Felder“

Die Registerkarte „Felder“ definiert die Namen, Positionen und Typen der Felder in der Datei. Weitere Informationen finden Sie unter [Definieren von Feldern in einer Ausgabesequenzdatei](#) auf Seite 282.

Definieren von Feldern in einer Ausgabesequenzdatei

Im Schritt „Write to Hadoop Sequence File“ definiert die Registerkarte **Felder** die Namen, Positionen und Längen der Felder in der Datei. Nach Definieren einer Eingabedatei auf der Registerkarte **Dateieigenschaften** können Sie die Felder definieren.

1. Um die gewünschten Felder der Eingabedaten oder eine vorhandene Datei auszuwählen, klicken Sie auf **Schnell hinzufügen**.
 - a) Wählen Sie die jeweiligen Felder der Eingabedaten aus.
 - b) Klicken Sie auf **OK**.
2. Klicken Sie zum Hinzufügen neuer Felder auf **Hinzufügen**.
 - a) Geben Sie den **Namen** des Feldes ein.
 - b) Geben Sie den **Typ** des Feldes ein. Der Schritt unterstützt die folgenden Datentypen:

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

Anmerkung: In Parquet-Dateien werden die Datentypen `datetime` und `time as string` zugeordnet. In RC-Dateien wird der Datentyp `datetime as timestamp` zugeordnet.

double	Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit doppelter Genauigkeit zwischen 2^{-1074} und $(2-2^{-52}) \times 2^{1023}$ 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 $(2-2^{-23}) \times 2^{127}$ 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.
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“.

Anmerkung: Für RC-, Avro- und Parquet-Hive-Dateien wird der Datentyp `bigdecimal` in den Datentyp `decimal` mit Genauigkeit 38 und Dezimalstellenzahl 10 konvertiert.

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.

Anmerkung: In RC-Dateien wird der Datentyp `long` als Datentyp `bigint` zugeordnet.

string Eine Folge von Zeichen.

c) 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"
```

3. Wenn Sie eine vorhandene Datei überschreiben, klicken Sie auf **Erneut generieren**, um das Schema aus der vorhandenen Datei zu übernehmen, und ändern Sie es dann. Dadurch wird das Schema auf Basis der ersten 50 Datensätze in den Eingabedaten für diesen Schritt generiert.
4. Wenn Sie überzählige Leerzeichen am Anfang und Ende der Zeichenfolge eines Feldes entfernen möchten, aktivieren Sie das Kästchen **Leerzeichen abschneiden**.
5. Geben Sie eine der folgenden Optionen an, um den Schlüssel zu generieren:

Automatisch generieren Das Hadoop-Cluster generiert den Schlüssel automatisch. Für „Automatisch generieren“ werden alle Felder im Grid als Wertfelder betrachtet. Der Datentyp des Schlüssels ist „Lang“.

Benutzerdefiniert Standardmäßig wird das erste Feld im Grid als Schlüsselfeld ausgewählt. Ein Symbol wird angezeigt, um anzugeben, dass das betreffende Feld das Schlüsselfeld ist. Sie können auch ein beliebiges anderes Feld als Schlüsselfeld auswählen.

Nach Definieren der Felder in der Ausgabedatei können Sie ihren Inhalt und ihr Layout bearbeiten.

Name der Option	Beschreibung
Hinzufügen	Fügt der Ausgabe ein Feld hinzu. Sie können ein Feld am Ende des vorhandenen Layouts anfügen oder ein Feld an einer vorhandenen Position einfügen, wobei die Position der restlichen Felder dementsprechend angepasst wird.
Ändern	Ändert den Namen und Typ des Feldes.
Entfernen	Entfernt das ausgewählte Feld aus der Ausgabe.
Nach oben/Nach unten	Ordnet das ausgewählte Feld neu an.

Write to Hive File

Der Schritt **Write to Hive File** schreibt die Datenflusseingabe in die angegebene Hive-Ausgabedatei. Sie können beliebige der unterstützten Hive-Dateiformate unten für die Ausgabedatei auswählen:

- ORC
- RC
- Parquet
- Avro

Registerkarte „Dateieigenschaften“

Tabelle 6: Allgemeine Dateieigenschaften

Felder	Beschreibung
Servername	Gibt an, dass die im Feld Dateiname ausgewählte Datei sich auf dem Hadoop-System befindet. Nachdem Sie eine Datei auf einem Hadoop-System ausgewählt haben, entspricht der Servername dem Namen des entsprechenden Dateiservers, wie in der Management Console angegeben.

Felder	Beschreibung
Dateiname	<p>Klicken Sie auf die Ellipsenschaltfläche (...), um die auf dem definierten Hadoop-Dateiserver zu erstellende Hive-Ausgabedatei zu lokalisieren. Die Ausgabedaten dieses Schrittes werden in die ausgewählte Datei geschrieben.</p> <p>Anmerkung: Sie müssen eine Verbindung zum Hadoop-Dateiserver in der Management Console erstellen, bevor Sie ihn im Schritt verwenden.</p>
Dateityp	<p>Wählen Sie eines der vier unterstützten Hive-Dateiformate aus:</p> <ul style="list-style-type: none"> • ORC • RC • Parquet • Avro

Tabelle 7: ORC-Dateieigenschaften

Felder	Beschreibung
Puffergröße	<p>Definiert die Puffergröße, die beim Schreiben in eine ORC-Datei zugeordnet werden soll. Dieser Wert wird in Kilobyte angegeben.</p> <p>Anmerkung: Die Standardpuffergröße ist 256 KB.</p>
Stripe-Größe	<p>Definiert die Größe der beim Schreiben in eine ORC-Datei zu erstellenden Stripes. Dieser Wert wird in Megabyte angegeben.</p> <p>Anmerkung: Die Standard-Stripe-Größe ist 64 MB.</p>
Zeilenindex-Schrittgröße	<p>Definiert die Anzahl der zwischen zwei fortlaufende Zeilenindizes zu schreibenden Zeilen.</p> <p>Anmerkung: Der Standardwert für die Zeilenindex-Schrittgröße ist 10.000 Zeilen.</p>
Komprimierungstyp	<p>Definiert den beim Schreiben in eine ORC-Datei zu verwendenden Komprimierungstyp. Die verfügbaren Komprimierungstypen sind <i>ZLIB</i> und <i>SNAPPY</i>.</p> <p>Anmerkung: Der Standardkomprimierungstyp ist <i>ZLIB</i>.</p>

Felder	Beschreibung
Textabstand	Gibt an, ob beim Schreiben in eine ORC-Datei die Stripes aufgefüllt werden, um Stripes zu minimieren, die HDFS-Blockgrenzen überschreiten. Anmerkung: Standardmäßig ist das Kästchen Padding aktiviert.
Vorschau	Die ersten 50 Datensätze der geschriebenen Datei werden abgerufen und im Grid Vorschau angezeigt, nachdem der Datenfluss mindestens einmal ausgeführt wurde und die Daten in die ausgewählte Datei geschrieben wurden.

Tabelle 8: RC-Dateieigenschaften

Felder	Beschreibung
Puffergröße	Definiert die Puffergröße, die beim Schreiben in eine RC-Datei zugeordnet werden soll. Dieser Wert wird in Kilobyte angegeben. Anmerkung: Die Standardpuffergröße ist 256 KB.
Blockgröße	Definiert die Größe der beim Schreiben in eine RC-Datei zu erstellenden Blöcke. Dieser Wert wird in Megabyte angegeben. Anmerkung: Die Standardblockgröße ist 64 MB.
Komprimierungstyp	Definiert den beim Schreiben in eine RC-Datei zu verwendenden Komprimierungstyp. Die verfügbaren Komprimierungstypen sind NONE und DEFLATE. Anmerkung: Der Standardkomprimierungstyp ist NONE.
Vorschau	Die ersten 50 Datensätze der geschriebenen Datei werden abgerufen und im Grid Vorschau angezeigt, nachdem der Datenfluss mindestens einmal ausgeführt wurde und die Daten in die ausgewählte Datei geschrieben wurden. Die Registerkarte Felder dient dazu, die Reihenfolge und den Datentyp der erforderlichen Felder zu definieren. Anmerkung: Beim Dateityp RC müssen Sie die Metadaten der Ausgabedatei definieren, bevor Sie auf Vorschau klicken, um das Grid Vorschau zu laden.

Tabelle 9: Parquet-Dateieigenschaften

Felder	Beschreibung
Komprimierungstyp	<p>Definiert den beim Schreiben in eine PARQUET-Datei zu verwendenden Komprimierungstyp. Die verfügbaren Komprimierungstypen sind UNCOMPRESSED, GZIP und SNAPPY.</p> <p>Anmerkung: Der Standardkomprimierungstyp ist UNCOMPRESSED.</p>
Blockgröße	<p>Definiert die Größe der beim Schreiben in eine PARQUET-Datei zu erstellenden Blöcke. Dieser Wert wird in Megabyte angegeben.</p> <p>Anmerkung: Die Standardblockgröße ist 128 MB.</p>
Seitengröße	<p>Die Seitengröße ist für die Komprimierung relevant. Beim Lesen kann jede Seite unabhängig dekomprimiert werden. Dieser Wert wird in Kilobyte angegeben.</p> <p>Anmerkung: Die Standardseitengröße ist 1.024 KB.</p>
Wörterbuch aktivieren	<p>Zum Aktivieren/Deaktivieren der Wörterbuchcodierung.</p> <p>Achtung: Das Wörterbuch muss aktiviert sein, damit sich die Wörterbuchseitengröße aktivieren lässt.</p> <p>Anmerkung: Der Standardwert ist <code>wahr</code>.</p>
Wörterbuchseitengröße	<p>Es gibt eine Wörterbuchseite pro Spalte pro Zeilengruppe, wenn die Wörterbuchcodierung verwendet wird. Die Wörterbuchseitengröße funktioniert wie die Seitengröße. Dieser Wert wird in Kilobyte angegeben.</p> <p>Anmerkung: Die Standard-Wörterbuchseitengröße ist 1.024 KB.</p>
Writer-Version	<p>Parquet unterstützt zwei Writer-API-Versionen: <code>PARQUET_1_0</code> und <code>PARQUET_2_0</code>.</p> <p>Anmerkung: Der Standardwert ist <code>PARQUET_1_0</code>.</p>
Vorschau	<p>Die ersten 50 Datensätze der geschriebenen Datei werden abgerufen und im Grid Vorschau angezeigt, nachdem der Datenfluss mindestens einmal ausgeführt wurde und die Daten in die ausgewählte Datei geschrieben wurden.</p>

Tabelle 10: Avro-Dateieigenschaften

Felder	Beschreibung
Synchronisierungsintervall (in Bytes)	<p>Gibt die ungefähre Anzahl von unkomprimierten in jedem Block zu schreibenden Byte an. Die gültigen Werte reichen von 32 bis 2^{30}. Es empfiehlt sich jedoch, ein Synchronisierungsintervall im Bereich von 2 TSD bis 2 Mio zu verwenden.</p> <p>Anmerkung: Das Standardsynchronisierungsintervall ist 16.000.</p>
Komprimierung	<p>Definiert den beim Schreiben in eine Avro-Datei zu verwendenden Komprimierungstyp. Die verfügbaren Komprimierungstypen sind NONE, SNAPPY und DEFLATE. Wenn Sie die Komprimierung DEFLATE wählen, erhalten Sie eine zusätzliche Option, um die Komprimierungsstufe (unten beschrieben) auszuwählen.</p> <p>Anmerkung: Der Standardkomprimierungstyp ist NONE.</p>
Komprimierungsstufe	<p>Dieses Feld wird angezeigt, wenn Sie die Option DEFLATE im obigen Feld Komprimierung auswählen.</p> <p>Es kann Werte von 0 bis 9 aufweisen, wobei 0 für keine Komprimierung steht. Die Komprimierungsstufe nimmt von 1 bis 9 zu, wobei gleichzeitig auch die Zeit zunimmt, die zum Komprimieren der Daten erforderlich ist.</p> <p>Anmerkung: Die Standardkomprimierungsstufe ist 1.</p>
Vorschau	<p>Die ersten 50 Datensätze der geschriebenen Datei werden abgerufen und in diesem Grid angezeigt, nachdem der Datenfluss mindestens einmal ausgeführt wurde und die Daten in die ausgewählte Datei geschrieben wurden.</p>

Registerkarte „Felder“

Die Registerkarte **Felder** definiert die Namen und Typen der Felder, die in der Quellendatei dieses Schrittes vorhanden sind und zum Schreiben in die Ausgabedatei ausgewählt werden sollen.

Weitere Informationen finden Sie unter [Definieren von Feldern zum Schreiben in eine Hive-Datei](#) auf Seite 289.

Registerkarte „Laufzeit“

Die Registerkarte **Laufzeit** bietet die Möglichkeit, eine vorhandene Datei mit demselben Namen auf dem konfigurierten Hadoop-Dateiserver zu **überschreiben**. Wenn Sie das Kästchen **Überschreiben** aktivieren, überschreibt beim Ausführen des Datenflusses die neue Hive-Ausgabedatei eine beliebige vorhandene Datei desselben Namens auf demselben Hadoop-Dateiserver.

Standardmäßig ist das Kästchen **Überschreiben** aktiviert.

Anmerkung: Wenn Sie **Überschreiben** nicht aktivieren, wird beim Ausführen des Datenflusses eine Ausnahmebedingung ausgelöst, wenn die zu schreibende Datei denselben Namen wie eine vorhandene Datei auf demselben Hadoop-Dateiserver hat.

Definieren von Feldern zum Schreiben in eine Hive-Datei

Auf der Registerkarte **Felder** des Schrittes **Write to Hive File** werden die Schemanamen und Datentypen der Felder in den Eingabedaten des Schrittes aufgelistet.

1. Um die gewünschten Felder der Eingabedaten oder eine vorhandene Datei auszuwählen, klicken Sie auf **Schnell hinzufügen**.
 - a) Wählen Sie die jeweiligen Felder der Eingabedaten aus.
 - b) Klicken Sie auf **OK**.
2. Klicken Sie zum Hinzufügen neuer Felder auf **Hinzufügen**.
 - a) Geben Sie den **Namen** des Feldes ein.
 - b) Geben Sie den **Typ** des Feldes ein. Der Schritt unterstützt die folgenden Datentypen:

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

Anmerkung: In Parquet-Dateien werden die Datentypen `datetime` und `time` als `String` zugeordnet. In RC-Dateien wird der Datentyp `datetime` als `timestamp` zugeordnet.

double	Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit doppelter Genauigkeit zwischen 2^{-1074} und $(2-2^{-52}) \times 2^{1023}$ 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 $(2-2^{-23}) \times 2^{127}$ 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.
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“.

Anmerkung: Für RC-, Avro- und Parquet-Hive-Dateien wird der Datentyp `bigdecimal` in den Datentyp `decimal` mit Genauigkeit 38 und Dezimalstellenzahl 10 konvertiert.

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.

Anmerkung: In RC-Dateien wird der Datentyp `long` als Datentyp `bigint` zugeordnet.

string Eine Folge von Zeichen.

- c) 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"
```

3. Wenn Sie eine vorhandene Datei überschreiben, klicken Sie auf **Erneut generieren**, um das Schema aus der vorhandenen Datei zu übernehmen, und ändern Sie es dann. Dadurch wird bei ORC- und Parquet-Ausgabedateien das Schema auf Basis der Metadaten der vorhandenen Datei generiert. Bei RC-Ausgabedateien müssen Sie die Felder explizit hinzufügen, um vorhandene Felder zu überschreiben.

In der Spalte **Name** werden die Namen der verschiedenen Spalten der Eingabedaten aufgelistet. In der Spalte **Typ** werden die Datentypen der jeweiligen Felder der Eingabedaten aufgelistet.

Anmerkung: Beim Dateityp *Parquet* gibt eine weitere Spalte, **NULL zulassen**, an, ob im Feld Nullwerte zugelassen sind oder nicht. Sie können dieses Kästchen für ein bestimmtes Feld aktivieren, damit dieses Nullwerte enthalten darf, oder es andernfalls deaktivieren.

4. Sie können die Namen, Datentypen und die Reihenfolge der ausgewählten Spalten in der Ausgabe mit den Schaltflächen unten ändern:

Name der Option	Beschreibung
Hinzufügen	Fügt der Ausgabe ein Feld hinzu.

Name der Option	Beschreibung
Ändern	Ändert den Namen und Datentyp des Feldes.
Entfernen	Entfernt das ausgewählte Feld aus der Ausgabe.
Nach oben/Nach unten	Ordnet die Position des ausgewählten Feldes in der Ausgabe neu an.

5. Klicken Sie auf **OK**.

Write to NoSQL DB

„Write to NoSQL DB“ schreibt die Ausgabe eines Datenflusses in eine NoSQL-Datenbank. Dieser Schritt unterstützt die Datenbanktypen „MongoDB“ und „Couchbase“.

Anmerkung: Es lassen sich erhebliche Leistungsverbesserungen erzielen, wenn mehrere Laufzeitinstanzen von „Write to NoSQL DB“ verwendet werden. Klicken Sie auf die Schaltfläche **Laufzeit**, um mehrere Laufzeitinstanzen anzugeben.

Registerkarte „Allgemein“

Felder	Beschreibung
Verbindung	<p>Wählen Sie die gewünschte Datenbankverbindung aus. Die Auswahlmöglichkeiten variieren, je nachdem, welche Verbindungen in der NoSQL-Verbindungsverwaltung im Menü Tools des Enterprise Designer definiert sind. Wenn Sie eine neue Datenbankverbindung herstellen oder eine vorhandene Datenbankverbindung ändern oder löschen möchten, klicken Sie auf Verwalten.</p> <p>Verbindungsname Geben Sie einen Namen für die Verbindung ein. Sie können den Namen frei wählen.</p> <p>NoSQL-Datenbank Wählen Sie den passenden Datenbanktyp aus.</p> <p>Benutzername Geben Sie den Benutzernamen zum Herstellen der Verbindung mit der Datenbank ein.</p> <p>Anmerkung: Bei Couchbase ist der Benutzername nicht obligatorisch. Sie können einen beliebigen Benutzernamen eingeben. Die Verbindung wird erfolgreich hergestellt werden, solange Sie das korrekte Kennwort verwenden, das Sie beim Erstellen des Bucket angegeben haben.</p> <p>Kennwort Geben Sie das Kennwort zum Herstellen der Verbindung mit der Datenbank ein.</p> <p>Hostname Geben Sie den Hostnamen an, auf dem die Datenbank läuft.</p> <p>Port Geben Sie den Port zum Herstellen einer Verbindung mit der Datenbank an.</p> <p>Datenbank Geben Sie die Datenbank an, aus der Daten abgerufen werden sollen.</p> <p>Anmerkung: Der auf der Benutzeroberfläche verwendete Begriff Datenbank wird bei Couchbase „Bucket“ genannt.</p>
Tabelle/Ansicht	<p>Geben Sie den Namen der Sammlung an, in die Sie schreiben möchten.</p> <p>Sie können eine neue Sammlung in der NoSQL-Datenbank erstellen, indem Sie im Auswahlfeld „Tabelle/Ansicht“ einen Sammlungsnamen eingeben und auf Tabelle erstellen klicken.</p> <p>Anmerkung: Bei Couchbase sind das Auswahlfeld „Tabelle/Ansicht“ und die Schaltfläche „Tabelle erstellen“ deaktiviert, weil wir statt in eine Ansicht in ein Bucket schreiben. Zusätzlich ist auch die Schaltfläche „Vorschau“ deaktiviert.</p>
NULL-Werte ignorieren	<p>Wenn diese Option aktiviert ist, wird jedes Feld mit Nullwert ignoriert.</p> <p>Anmerkung: Wenn Sie diese Option nicht aktivieren, wird jedes Feld mit Nullwert auch in die Datenbank geschrieben.</p>

Felder	Beschreibung
Vorschau	<p>Dies zeigt die Datensätze aus der ausgewählte Tabelle an.</p> <p>Anmerkung: Bei MongoDB-Datenquellen werden beim Klicken auf Vorschau die gefilterten Datensätze angezeigt, wenn eine oder mehrere <i>where</i>-Klauseln in das Feld Where eingegeben wurden. Wenn keine WHERE-Klausel eingegeben wurde, werden in der Vorschau alle Datensätze angezeigt.</p> <p>Anmerkung: Bei Couchbase-Datenquellen wird beim Klicken auf Vorschau auch das hinzugefügte Feld <code>_id</code> mit dem enthaltenen Schlüssel angezeigt. Wenn der Datensatz bereits über ein Feld <code>_id</code> verfügt, überschreibt das hinzugefügte Feld <code>_id</code> das vorher vorhandene Feld bei der Vorschau der Felder.</p>
Alle erweitern	Dies erweitert die Elemente im Vorschaubaum.
Alle reduzieren	Dies reduziert die Elemente im Vorschaubaum.

Registerkarte „Felder“

Die Registerkarte „Felder“ ermöglicht Ihnen die Auswahl der Daten, die Sie in die Datenbank schreiben möchten. Weitere Informationen finden Sie unter [Definieren von Feldern in einer NoSQL-Datenbank](#) auf Seite 293.

Definieren von Feldern in einer NoSQL-Datenbank

Im Schritt „Write to NoSQL DB“ definiert die Registerkarte **Felder** die Namen und Typen der aus dem vorherigen Schritt abgerufenen Felder.

1. Klicken Sie auf der Registerkarte **Felder** auf die Option **Erneut generieren**.

Dadurch werden die Felder aus dem vorherigen Schritt angezeigt.

Die Daten werden im folgenden Format angezeigt: `Fieldname (datatype)`.

Anmerkung: Bei Couchbase gilt: Wenn der Datensatz ein Feld `_id` enthält, wird dieses Feld als Schlüssel verwendet, um den Datensatz in die Datenbank zu schreiben. Andernfalls wird der Schlüssel für den Datensatz automatisch generiert und in die Datenbank geschrieben.

2. Um den Namen und Typ eines Feldes zu ändern, markieren Sie das Feld und klicken Sie auf **Ändern**.
3. Wählen Sie im Feld **Name** das Feld aus, das Sie hinzufügen möchten, oder geben Sie den Namen in das Feld ein.

- Im Feld **Typ** können Sie den Datentyp „Zeichenfolge“ beibehalten, wenn Sie nicht vorhaben, mathematische Operationen auf die Daten anzuwenden. Wenn Sie allerdings diese 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.

Der Schritt unterstützt die folgenden Datentypen:

double	Ein numerischer Datentyp, der sowohl negative als auch positive Zahlen mit doppelter Genauigkeit zwischen 2^{-1074} und $(2-2^{-52}) \times 2^{1023}$ 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 $(2-2^{-23}) \times 2^{127}$ 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.

- Auf der Registerkarte **Felder** können Sie entweder jedes in die Datenbank zu schreibende Feld einzeln auswählen oder per Klick auf **Alle auswählen** alle Felder auswählen.
- Wahlweise können Sie auf der Registerkarte **Laufzeit** die Batchgröße festlegen. Die Batchgröße gibt die Anzahl von Datensätzen an, die auf einmal in die Datenbank geschrieben werden sollen.
- Klicken Sie auf **OK**.

Datenflussoptionen für NoSQL-Datenbanken

Diese Schritte beschreiben, wie Sie einen Datenfluss konfigurieren, um Laufzeitoptionen für NoSQL-Datenbanken zu unterstützen.

- Öffnen Sie den Datenfluss im Enterprise Designer.
- Wenn Sie Laufzeitoptionen für einen Schritt in einem eingebetteten Datenfluss konfigurieren möchten, öffnen Sie den eingebetteten Datenfluss.
- Klicken Sie in der Symbolleiste auf das Symbol „Datenflussoptionen“ oder klicken Sie auf **Bearbeiten > Datenflussoptionen**. Das Dialogfeld **Datenflussoptionen** wird angezeigt.
- Klicken Sie auf **Add** (Hinzufügen). Das Dialogfeld **Datenflussoptionen definieren** wird angezeigt.
- Erweitern Sie den „NoSQLDB“-Schritt.
- Die Datenflussoptionen werden, wie in folgender Tabelle beschrieben, verfügbar gemacht:

Datenbank	Lesen	Schreiben
Mongo DB	Verbindung	Verbindung
	Tabelle	Tabelle
Couchbase DB	Verbindung	Verbindung
	Anzeigen	
	Name des Designdokuments	

Der ausgewählte „NoSQL DB“-Optionsname wird in den Feldern **Optionsname** und **Optionsbeschriftung** angezeigt. Dies ist der Optionsname, der zur Laufzeit angegeben werden muss, um diese Option festzulegen.

7. Geben Sie eine Beschreibung der Option in das Feld **Beschreibung** ein.
8. Wählen Sie im Feld **Ziel** die Option **Ausgewählte(r) Schritt/e** aus.
9. 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.
10. 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](#).

11. Klicken Sie auf **OK**.
12. Fügen Sie auf Wunsch weitere Optionen hinzu.
13. Klicken Sie im Dialogfeld **Datenflussoptionen** auf **OK**, sobald Sie damit fertig sind.
14. Speichern Sie den Datenfluss und machen Sie ihn verfügbar.

Write to Spreadsheet

„Write to Spreadsheet“ schreibt Daten so in ein Excel-Arbeitsblatt, wie sie von einem Datenfluss ausgegeben werden.

Registerkarte „Dateieigenschaften“

Die Registerkarte **Dateieigenschaften** enthält Optionen, um das Arbeitsblatt und Daten anzugeben, die aus einem Datenfluss geschrieben werden sollen.

Feldname	Beschreibung
Servername	Gibt an, dass die im Feld Dateiname ausgewählte Datei sich auf dem Spectrum™ Technology Platform-Server befindet.
Dateiname	<p>Gibt den Pfad zur Datei an. Klicken Sie auf die Ellipsenschaltfläche (...), um die gewünschte Datei zu lokalisieren.</p> <p>Beim Schreiben einer Datei auf einen HDFS-Dateiserver werden die folgenden Komprimierungsformate unterstützt:</p> <ol style="list-style-type: none"> 1. GZIP (.gz) 2. BZIP2 (.bz2) <p>Anmerkung: Geben Sie beim Dateinamen die passende Erweiterung an, um das gewünschte Komprimierungsformat anzugeben, das beim Schreiben in die Datei verwendet werden soll.</p> <p>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.</p>
Schreibmodus	<p>Gibt an, wie Daten im Arbeitsblatt geschrieben werden sollen, um in den Datenfluss zu schreiben. Sie können mit beliebigen der folgenden Optionen ein Arbeitsblatt zur Laufzeit erstellen. Die Optionen sind:</p> <p>Erstellen oder überschreiben Erstellt eine neue Datei und ersetzt jedes Mal, wenn der Datenfluss ausgeführt wird, die vorhandenen Daten in der Ausgabedatei.</p> <p>Einfügen Fügt die Ausgabe des Datenflusses dem zugeordneten Bereich hinzu und verschiebt die Daten nach unten, falls sie dort bereits vorhanden sind.</p> <p>Anfügen Fügt die Ausgabe des Datenflusses am Ende der Datei hinzu, ohne die vorhandenen Daten der Datei zu löschen.</p>
Blattname	Gibt einen Blattnamen im Arbeitsblatt an, in das Sie Daten in den Datenfluss schreiben möchten.
Schreiben beginnend ab	Gibt entweder eine Zeilen-/Spaltenkombination (A1 oder B2 ...) oder eine Spalte an, ab der die Daten geschrieben werden sollen. Für die Option Einfügen müssen Sie sowohl die Zeile als auch die Spalte angeben. Für die Option Anfügen können Sie nur die Spalte angeben, da der Zeilenwert ignoriert wird.
Erste Zeile als Spaltenüberschrift	Gibt an, dass die erste Zeile in einer Datei Überschriften und keine Daten enthält.

Registerkarte „Felder“

Die Registerkarte „Felder“ definiert Spalten, Positionen, Typen und zulässige Nullwerte für die Felder in der Datei. Weitere Informationen bietet das folgende Thema:

Definieren von Feldern in einer Ausgabedatei auf Seite 297

Definieren von Feldern in einer Ausgabedatei

Im Schritt „Write to Spreadsheet“ definiert die Registerkarte **Felder** die Namen, Position, und, für einige Dateitypen, die Längen der Felder in der Datei. Nach Definieren einer Ausgabedatei auf der Registerkarte **Dateieigenschaften** können Sie die Felder definieren. Falls die Spalte **NULL zulassen** aktiviert ist und das Feld **Name** einen Nullwert enthält, schreibt der Datenfluss den Nullwert in das Arbeitsblatt.

Wenn die Ausgabedatei einen Kopfzeilendatensatz enthält, können Sie die Felder schnell definieren, indem Sie auf **Erneut generieren** klicken.

Klicken Sie zum Definieren von Feldern ohne Standardwerte für Position, Länge und Datentyp auf **Schnell hinzufügen** und wählen Sie die hinzuzufügenden Felder aus.

Wenn die Eingabedatei keinen Kopfzeilendatensatz enthält oder wenn Sie die Felder manuell definieren möchten, gehen Sie wie folgt vor:

1. Klicken Sie auf **Hinzufügen**.
2. Wählen Sie im Feld **Name** das Feld, das Sie hinzufügen möchten.
3. Wählen Sie im Feld **Typ** den Datentyp des Feldes aus dem Datenfluss aus.

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 $(2-2^{-52}) \times 2^{1023}$ 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 $(2-2^{-23}) \times 2^{127}$ 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.
time	Ein Datentyp, der die Tageszeit enthält. Beispiel: 21:15:59 oder 9:15:59 PM.

4. Klicken Sie auf **Hinzufügen**.

Nach Definieren der Felder in der Ausgabedatei können Sie ihren Inhalt und ihr Layout bearbeiten.

Name der Option	Beschreibung
Hinzufügen	Fügt der Ausgabe ein Feld hinzu. Sie können ein Feld am Ende des vorhandenen Layouts anfügen oder ein Feld an einer vorhandenen Position einfügen, wobei „Write to Spreadsheet“ die restlichen Felder dementsprechend anpasst.
Ändern	Ändert den Namen und Typ des Feldes.
Entfernen	Entfernt das ausgewählte Feld aus der Ausgabe.
Nach oben/Nach unten	Ordnet das ausgewählte Feld neu an.

Write to Variable Format File

„Write to Variable Format File“ schreibt Datensätze mit variablem Layout in eine Datei.

Dateien mit variablem Format haben diese Merkmale:

- Datensätze in der Datei können unterschiedliche Felder und eine unterschiedliche Anzahl von Feldern enthalten.
- Jeder Datensatz muss ein Tag (gewöhnlich eine Zahl) enthalten, das den Typ des Datensatzes angibt.
- Hierarchische Beziehungen werden unterstützt.

Beispiel für eine Datei mit variablem Format

Dieses Beispiel zeigt eine Datei mit variablem Format, die Informationen zur Girokontoaktivität für zwei Kunden, Joe Smith und Anne Johnson, enthält. In diesem Beispiel ist die Datei eine durch Trennzeichen getrennte Datei, in der ein Komma als Feldtrennzeichen verwendet wird.

```
001   Joe,Smith,M,100 Main St,555-234-1290
100   CHK12904567,12/2/2007,6/1/2012,CHK
200   1000567,1/5/2012,Fashion Shoes,323.12
001   Anne,Johnson,F,1202 Lake St,555-222-4932
100   CHK238193875,1/21/2001,4/12/2012,CHK
200   1000232,3/5/2012,Blue Goose Grocery,132.11
200   1000232,3/8/2012,Trailway Bikes,540.00
```

Das erste Feld in jedem Datensatz enthält das Tag, das den Typ des Datensatzes und damit sein Format angibt:

- 001: Kundendatensatz
- 100: Kontodatensatz
- 200: Kontotransaktionsdatensatz

Bei durch Trennzeichen getrennten Dateien ist es häufig so, dass der Tagwert (001, 100, 200) in einer festen Anzahl von Byte am Beginn des Datensatzes steht (siehe das obige Beispiel).

Jeder Datensatz hat ein eigenes Format:

- 001: FirstName,LastName,Gender,Address,PhoneNumber
- 100: AccountID,DateOpened,ExpirationDate,TypeOfAccount
- 200: TransactionID,DateOfTransaction,Vendor,Amount

Das Datensatzformat 100 (Kontodatensatz) ist ein untergeordnetes Element des vorherigen 001-Datensatzes und das Datensatzformat 200 (Kontotransaktionsdatensatz) ein untergeordnetes Element des vorherigen 100-Datensatzes (Kontodatensatz). In der Beispieldatei gab es im Konto von Joe Smith, CHK12904567, am 05.01.2012 eine Transaktion mit dem Betrag 323,12 bei Fashion Shoes. Im Konto von Anne Johnson, CHK238193875, wiederum gab es zwei Transaktionen, eine am 05.03.2012 bei Blue Goose Grocery und eine am 08.03.2012 bei Trailway Bikes.

Registerkarte „Dateieigenschaften“

Name der Option	Beschreibung
Servername	Dies gibt an, ob sich die von Ihnen als Eingabe ausgewählte Datei auf dem Computer mit dem Enterprise Designer oder auf dem Spectrum™ Technology Platform-Server befindet. Wenn Sie eine Datei auf dem lokalen Computer auswählen, lautet der Servername „Arbeitsplatz“. Wenn Sie eine Datei auf dem Server auswählen, lautet der Servername Spectrum™ Technology Platform.
Dateiname	Gibt den Pfad zur Datei an. Klicken Sie auf die Ellipsenschaltfläche (...), um die gewünschte Datei zu lokalisieren. Anmerkung: 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.
Root-Tag-Name	Das Tag, das für Datensätze verwendet werden soll, die ein übergeordnetes Element anderer Datentypen sind. Wenn Sie beispielsweise die drei Datentypen 001, 100 und 200 haben und die Datentypen 100 und 200 untergeordnete Elemente des Datentyps 001 sind, dann ist 001 das Root-Tag.
Tags mit fester Breite verwenden	Gibt an, ob ein fester Bereich am Beginn jedes Datensatzes zugeordnet werden soll, an dem das Datensatz-Tag eingefügt wird. Das folgende Beispiel zeigt eine Datei mit den Tags 001, 100, und 200 in einem Feld mit fester Breite: <pre>001 Joe, Smith, M, 100 Main St, 555-234-1290 100 CHK12904567, 12/2/2007, 6/1/2012, CHK 200 1000567, 1/5/2012, Mike's Shoes, 323.12</pre>
Tag-Breite	Wenn Sie das Kästchen Tags mit fester Breite verwenden aktivieren, gibt diese Option die Anzahl von Leerzeichen an, die für Tags am Beginn jedes Datensatzes zugeordnet werden sollen. Wenn Sie beispielsweise 7 angeben, dann werden die ersten sieben Stellen in jedem Datensatz für das Tag vorgesehen. Der angegebene Wert muss größer oder gleich groß wie der längste Tag-Name in Zeichen sein. Informationen zu Tag-Namen finden Sie unter Tag-Namen in Dateien mit variablem Format auf Seite 307. Der Wert im Feld Tag-Breite wird automatisch erhöht, wenn Sie auf der Registerkarte Felder Felder hinzufügen, die einen Namen haben, der länger als der angegebene Wert ist. Die maximale Tag-Breite ist 1024.

Name der Option	Beschreibung
Numerisches Tag-Präfix entfernen	Löscht den Teil „NumericTag_“ des Feldnamens, bevor das Tag in die Datei geschrieben wird. Das Präfix „NumericTag_“ wird Tag-Namen, die mit einer Zahl beginnen, durch den Schritt „Read from Variable Format File“ hinzugefügt. Dies deshalb, weil der Tag-Name als Name eines Listendatenflussfeldes verwendet wird, das die Daten im Datensatz enthält, und Datenflussfeldnamen nicht mit einer Zahl beginnen dürfen. So würde beispielsweise das Tag 100 in das Listenfeld namens „NumericTag_100“ geändert. Wenn Sie diese Option aktivieren, würde dieses Feld als Datensatz mit einem Tag „100“ statt „NumericTag_100“ in die Ausgabedatei geschrieben.
Zeichencodierung	<p>Dies ist die Codierung der Textdatei. Wählen Sie eine der folgenden Optionen:</p> <p>UTF-8 Die Datei unterstützt alle Unicode-Zeichen und ist abwärtskompatibel mit ASCII. Weitere Informationen zu UTF finden Sie unter unicode.org/faq/utf_bom.html.</p> <p>UTF-16 Die Datei unterstützt alle Unicode-Zeichen, ist aber nicht abwärtskompatibel mit ASCII. Weitere Informationen zu UTF finden Sie unter unicode.org/faq/utf_bom.html.</p> <p>US-ASCII Die Zeichencodierung basiert auf der Reihenfolge des englischen Alphabets.</p> <p>UTF-16BE Die Datei ist UTF-16-codiert mit Big-Endian-Reihenfolge (höchstwertiges Byte zuerst).</p> <p>UTF-16LE Die Datei ist UTF-16-codiert mit Little-Endian-Reihenfolge (niedrigstwertiges Byte zuerst).</p> <p>ISO-8859-1 Dies ist eine ASCII-Zeichencodierung, die normalerweise für westeuropäische Sprachen verwendet wird. Sie ist auch bekannt als Latin-1.</p> <p>ISO-8859-3 Dies ist eine ASCII-Zeichencodierung, die normalerweise für südeuropäische Sprachen verwendet wird. Sie ist auch bekannt als Latin-3.</p> <p>ISO-8859-9 Dies ist eine ASCII-Zeichencodierung, die normalerweise für Türkisch verwendet wird. Sie ist auch bekannt als Latin-5.</p> <p>CP850 Dies ist eine ASCII-Codeseite, die für westeuropäische Sprachen verwendet wird.</p> <p>CP500 Dies ist eine EBCDIC-Codeseite, die für westeuropäische Sprachen verwendet wird.</p> <p>Shift_JIS Dies ist eine Zeichencodierung für Japanisch.</p> <p>MS932 Dies ist eine Shift_JIS-Erweiterung von Microsoft, die besondere NEC-Zeichen, eine NEC-Auswahl aus IBM-Erweiterungen und IBM-Erweiterungen einschließt.</p> <p>CP1047 Dies ist eine EBCDIC-Codeseite mit dem vollständigen Latin-1-Zeichensatz.</p>

Name der Option	Beschreibung
Feldtrennzeichen	<p>Gibt das Zeichen an, das zum Trennen von Feldern in einer Datei mit Begrenzungszeichen verwendet wird.</p> <p>Im folgenden Beispiel wird ein Pipe-Zeichen () im Datensatz als Feldtrennzeichen verwendet:</p> <pre>7200 13TH ST MIAMI FL 33144</pre> <p>Die folgenden Zeichen stehen zur Definition als Feldtrennzeichen zur Verfügung:</p> <ul style="list-style-type: none"> • Leerzeichen • Tabulator • Komma • Punkt • Semikolon • Pipe <p>Wenn in der Datei ein anderes Zeichen als Feldtrennzeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Begrenzungszeichen auszuwählen.</p>
Tag-Trennzeichen	<p>Gibt das Zeichen an, das sich hinter dem Tag-Feld befindet, um in einer Datei mit Begrenzungszeichen das Identifizierungsfeld für jeden Datensatz abzugrenzen. Ein Tag-Trennzeichen muss ein einzelnes Zeichen sein.</p> <p>Standardmäßig stehen die folgenden Zeichen zur Definition als Tag-Trennzeichen zur Verfügung:</p> <ul style="list-style-type: none"> • Leerzeichen • Tabulator • Komma • Punkt • Semikolon • Pipe <p>Wenn in der Datei ein anderes Zeichen als Tag-Trennzeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein benutzerdefiniertes Tag-Trennzeichen hinzuzufügen und auszuwählen.</p> <p>Anmerkung: Standardmäßig ist das Datensatz-Trennzeichen dasselbe Zeichen wie für das ausgewählte Feldtrennzeichen. Um dieses Feld zu aktivieren und ein anderes Zeichen auszuwählen, deaktivieren Sie das Kästchen Wie Feldtrennzeichen.</p>
Wie Feldtrennzeichen	<p>Dies zeigt an, ob das Tag-Trennzeichen dasselbe ist wie das Feldtrennzeichen. Deaktivieren Sie dieses Kästchen, um ein anderes Zeichen als Tag-Trennzeichen auszuwählen.</p> <p>Anmerkung: Standardmäßig ist dieses Kästchen aktiviert und das Feld Tag-Trennzeichen deaktiviert.</p>

Name der Option	Beschreibung
Textbegrenzungszeichen	<p>Dies ist das Zeichen, das in einer Datei mit Begrenzungszeichen Textwerte umschließt.</p> <p>Im folgenden Beispiel werden im Datensatz doppelte Anführungszeichen (") als Textbegrenzungszeichen verwendet.</p> <pre>"7200 13TH ST" "MIAMI" "FL" "33144"</pre> <p>Die folgenden Zeichen stehen zur Definition als Textbegrenzungszeichen zur Verfügung:</p> <ul style="list-style-type: none"> • einfaches Anführungszeichen (') • doppeltes Anführungszeichen (") <p>Wenn in der Datei ein anderes Textbegrenzungszeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Textbegrenzungszeichen auszuwählen.</p>
Datensatz-Trennzeichen	<p>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-Zeilende verwenden aktivieren.</p> <p>Die verfügbaren Einstellungen für Datensatz-Trennzeichen sind:</p> <p>Unix (U+000A) Die Datensätze sind durch ein Zeilenvorschubzeichen getrennt. Dies ist bei Unix-Systemen das standardmäßige Datensatz-Trennzeichen.</p> <p>Macintosh (U+000D) Die Datensätze sind durch ein Zeilenumbruchzeichen getrennt. Dies ist bei Macintosh-Systemen das standardmäßige Datensatz-Trennzeichen.</p> <p>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.</p> <p>Wenn in der Datei ein anderes Datensatz-Trennzeichen verwendet wird, klicken Sie auf die Ellipsenschaltfläche, um ein anderes Zeichen als Datensatz-Trennzeichen auszuwählen.</p>
Standard-Zeilende verwenden	<p>Gibt an, dass das standardmäßige Zeilenendzeichen (EOL) des Betriebssystems, auf dem der Spectrum™ Technology Platform-Server läuft, als Datensatz-Trennzeichen der Datei verwendet wird.</p> <p>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-Zeilendzeichen verwendet wird, der Server aber auf Linux läuft, aktivieren Sie diese Option nicht. Wählen Sie stattdessen im Feld Datensatz-Trennzeichen die Option „Windows“ aus.</p>

Registerkarte „Felder“

Die Registerkarte **Felder** steuert, welche Felder vom Datenfluss in die Ausgabedatei eingeschlossen werden.

Name der Option	Beschreibung
Hinzufügen	Klicken Sie darauf, um der Ausgabe ein Feld hinzuzufügen. Informationen zum Erstellen von Datenflussfeldern für „Write to Variable Format File“ finden Sie unter Schreiben von flachen Daten in eine Datei mit variablem Format auf Seite 305.
Ändern	Klicken Sie darauf, um den Namen des Tags zu ändern. Diese Schaltfläche ist nur aktiviert, wenn ein Tag ausgewählt ist. Falls die Option Tags mit fester Breite verwenden auf der Registerkarte Dateieigenschaften aktiviert ist, wird die Tag-Breite automatisch angepasst, wenn Sie einen längeren Tag-Namen eingeben. Anmerkung: Wenn Sie mit dieser Schaltfläche den Root-Tag-Namen ändern, hat das denselben Effekt wie eine Änderung des Wertes im Feld Root-Tag-Name auf der Registerkarte Dateieigenschaften .
Entfernen	Entfernt das ausgewählte Feld aus der Ausgabe. Wenn Sie ein Listenfeld entfernen, werden alle untergeordneten Felder ebenfalls entfernt. Wenn Sie ein untergeordnetes Feld entfernen, wird nur das ausgewählte untergeordnete Feld aus dem Listenfeld entfernt.
xx Alle entfernen	Entfernt alle Felder aus der Ausgabe.
Nach oben/Nach unten	Ordnet das ausgewählte Feld neu an.

Registerkarte „Laufzeit“

Name der Option	Beschreibung
Dateiname	Zeigt die auf der Registerkarte Dateieigenschaften definierte Datei an.

Name der Option	Beschreibung
Mehrere Dateien generieren	<p>Wählen Sie diese Option aus, um Datensätze in unterschiedliche Dateien zu schreiben (statt alle Datensätze in eine Datei). Die Datei, in die jeder Datensatz geschrieben wird, ist im Datensatz selbst angegeben. Jeder Datensatz muss ein Feld enthalten, das entweder den Namen oder den vollständigen Pfad zu der Datei angibt, in die der Datensatz geschrieben werden soll. Wenn Sie beispielsweise die Aktienkurse verschiedener Unternehmen (verschiedener Gruppen) allen Kunden gesondert senden möchten, schreibt dieses Feature die Aktienkurse der verschiedenen Unternehmen in separate Dateien, die Sie auf Wunsch jedem der Kunden senden können. Wenn Sie die Option Mehrere Dateien generieren aktivieren, müssen Sie eine Ausgabedatei entweder auf dem Spectrum™ Technology Platform-Server oder einem FTP-Server angeben. Wenn Sie Daten in eine Datei auf einem FTP-Server schreiben möchten, müssen Sie mit Management Console eine Verbindung zum Dateiserver definieren.</p> <p>Anmerkung: Die Datensätze in der im Dateipfadfeld ausgewählten Spalte müssen sortiert vorliegen. Verwenden Sie dieses Feature, wenn der Datensatz entweder den Namen oder den vollständigen Pfad der Datei enthält.</p>
Dateipfadfeld	<p>Wählt das Feld aus, das den Pfad (entweder ein Dateiname oder der vollständige Dateipfad) der Datei enthält, in die Sie den Datensatz schreiben möchten. Beachten Sie, dass nur die direkt einem Root-Tag zugeordneten Elemente einfachen Typs im Dateipfadfeld aufgelistet werden. Dieses Feld ist nur aktiviert, wenn Sie die Option Mehrere Dateien generieren auswählen.</p>
Schreibmodus	<p>Gibt an, ob die Ausgabe des Datenflusses am Ende der Datei hinzugefügt oder ob die in der Datei vorhandenen Daten vor dem Schreiben der Ausgabe gelöscht werden sollen. Zur Auswahl stehen:</p> <p>Überschreiben Ersetzt die vorhandenen Daten in der Ausgabedatei jedes Mal, wenn der Datenfluss ausgeführt wird.</p> <p>Anfügen Fügt die Ausgabe des Datenflusses am Ende der Datei hinzu, ohne die vorhandenen Daten der Datei zu löschen.</p>

Schreiben von flachen Daten in eine Datei mit variablem Format

In einem Spectrum™ Technology Platform-Datenfluss weist jeder Datensatz dieselben Felder auf. In einer Datei mit variablem Format hingegen enthalten nicht alle Datensätze dieselben Felder. Um flache Daten aus einem Datenfluss in eine Datei mit variablem Format zu schreiben, müssen Sie jeden Datensatz im Datenfluss aufspalten, die Felder aus jedem Datensatz in Listenfeldern gruppieren, die den Datentypen entsprechen, die Sie in der Datei mit variablem Format verwenden möchten. Ein Listenfeld ist eine Sammlung von Feldern. So könnten beispielsweise die Felder „FirstName“, „LastName“, „Gender“, „Address“ und „Phone“ zusammen in einem Listenfeld namens „AccountOwner“ gruppiert werden.

Um flache Daten in eine Datei mit variablem Format zu schreiben, verwenden Sie einen Aggregatorschritt, um Felder in Listenfeldern zu gruppieren, die den Datensatztypen entsprechen, die Sie in die Datei mit variablem Format schreiben möchten. Gehen Sie dazu wie folgt vor:

1. Platzieren Sie aus den Schritt „Write To Variable Format File“ einen Aggregatorschritt an einer beliebigen Stelle weiter oben in Ihrem Datenfluss.
2. Doppelklicken Sie auf den Aggregatorschritt, um das zugehörige Optionsfenster zu öffnen.
3. Wählen Sie **Gruppieren nach** aus, und klicken Sie dann auf **Hinzufügen**.
4. Wählen Sie im Feld **Gruppieren nach** das Feld aus, das einen eindeutigen Bezeichner enthält, mit dessen Hilfe sich zugehörige Daten identifizieren lassen. Der Wert dieses Feldes sollte für alle Datensätze in den flachen Daten eindeutig sein. Beispiele: eine Kontonummer, eine Sozialversicherungsnummer oder eine Telefonnummer.

Anmerkung: Das ausgewählte Feld sollte sortiert werden. Falls dies nicht so ist, verwenden Sie einen Sorter-Schritt, um die Datensätze nach dem Feld zu sortieren.

5. Klicken Sie auf **OK**.
6. Wählen Sie **Ausgabelisten** aus, und klicken Sie dann auf **Hinzufügen**.
Jede Ausgabeliste stellt einen Datensatztyp in der Datei mit variablem Format dar.
7. Wählen Sie **Neuer Datentyp** aus und geben Sie im Feld **Typname** den Typ der Informationen an, die in diesem Datentyp enthalten sein sollen. Daraus wird dann ein Datensatztyp in der Datei mit variablem Format. Beispiel: Dieser Datentyp soll Datensätze bezüglich Kontotransaktionen enthalten, daher könnten Sie den Typ „AccountTransaction“ nennen.
8. Geben Sie im Feld **Name** den Namen ein, den Sie diesem Feld geben möchten. Dies kann derselbe Name sein, den Sie im Feld **Typname** angeben.
9. Klicken Sie auf **OK**.
10. Wählen Sie den gerade erstellten Datentyp aus und klicken Sie auf **Hinzufügen**.
11. Lassen Sie die Option **Vorhandenes Feld** aktiviert und wählen Sie eines der Felder aus, die Sie in diesen Datentyp einschließen möchten. Klicken Sie dann auf **OK**. Beachten Sie, dass daraus dann ein Datensatztyp in der Datei mit variablem Format wird. Wiederholen Sie diese Schritte, um zusätzliche Felder zu diesem Datensatztyp hinzuzufügen.
12. Erstellen Sie zusätzliche Ausgabelisten für jede Datensatztyp, den Sie in der Datei mit variablem Format benötigen. Klicken Sie zum Abschluss auf **OK**, um die Aggregatoroptionen zu schließen.

Die Felder, die aus dem Aggregatorschritt stammen, werden jetzt in Listenfeldern gruppiert, die den Datensatztypen entsprechen, die Sie in die Ausgabe der Datei mit variablem Format einschließen möchten.

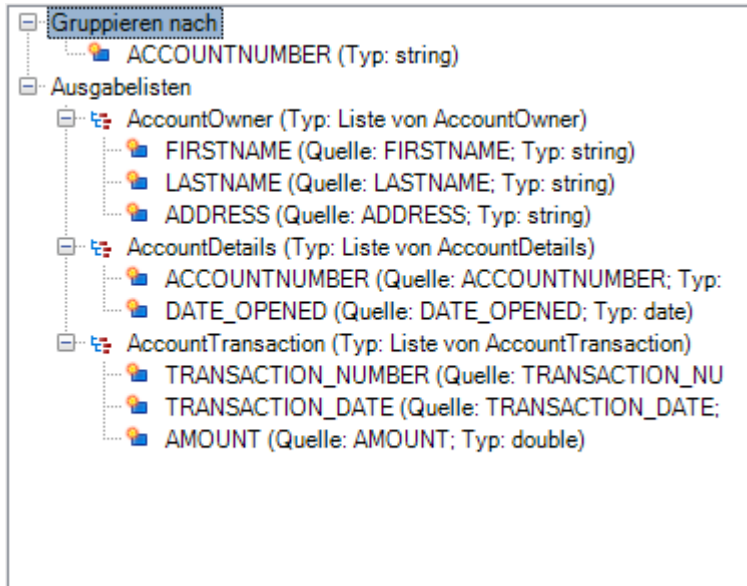
Betrachten wir beispielsweise diese flachen Daten:

```
FIRSTNAME, LASTNAME, ADDRESS, ACCOUNTNUMBER, DATE_OPENED, TRANSACTION_NUMBER, TRANSACTION_DATE, AMOUNT
Joe, Smith, 100 Main St, CHK12904567, 12/2/2007, 1000567, 1/5/2012, 323.12
```

Sie könnten sie in etwas wie das Folgende in der Datei mit variablem Format konvertieren:

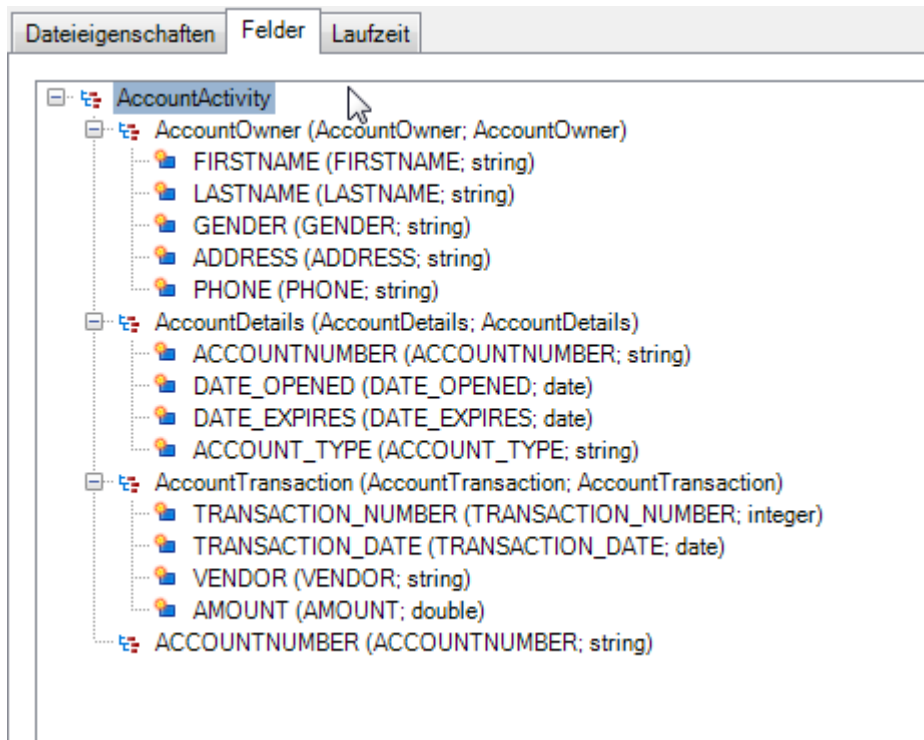
```
AccountOwner      Joe,Smith,100 Main St
AccountInformation CHK12904567,12/2/2007
Transaction       1000567,1/5/2012,323.12
```

Dazu würden Sie einen Aggregatorschritt erstellen, die wie folgt konfiguriert wird:



Tag-Namen in Dateien mit variablem Format

In einer Datei mit variablem Format enthält jeder Datensatz in der Ausgabedatei ein Tag, das den Datensatztyp angibt. In „Write To Variable Format File“ wird der Feldname als Tag-Name in der Ausgabedatei verwendet. Nehmen Sie diese Felder als Beispiel:



Diese Felder würden wie folgt in die Datei geschrieben. Beachten Sie, dass in diesem Beispiel das Konto zwei „AccountTransaction“-Datensätze aufweist.

```
AccountOwner      Anne,Johnson,F,1202 Lake St,555-222-4932
AccountDetails    CHK238193875,1/21/2001,4/12/2012,CHK
AccountTransaction 1000232,3/5/2012,Blue Goose Grocery,132.11
AccountTransaction 1000232,3/8/2012,Trailway Bikes,540.00
```

Anmerkung: Nur Listenfelder, die einfache Felder wie Zeichenfolgen enthalten, werden in die Ausgabedatei geschrieben. Wenn ein Listenfeld nur aus anderen Listenfeldern besteht, wird es nicht in die Ausgabedatei geschrieben. Im obigen Beispiel würde kein Datensatz mit einem „AccountActivity“-Tag in die Ausgabedatei geschrieben, weil AccountActivity nur aus anderen Listenfeldern (AccountOwner, AccountDetails und AccountTransaction) besteht.

Write to XML

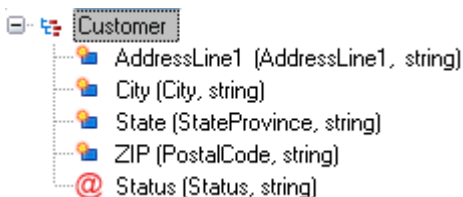
Der „Write to XML“-Schritt schreibt die Ausgabe eines Auftrags oder Unterflusses in eine XML-Datei.

Registerkarte „Dateieigenschaften“

Feldname	Beschreibung
Datendatei	<p>Gibt den Pfad zur XML-Ausgabedatei an. Klicken Sie auf die Ellipsenschaltfläche (...), um die gewünschte Datei zu lokalisieren.</p> <p>Anmerkung: 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.</p>
Tatsächliche Datei	<p>Zeigt die auf der Registerkarte Felder angegebene Struktur an. Wenn Sie auf ein Element klicken und die im Feld Datendatei angegebene Datei das Element enthält, wird eine Vorschau der Daten angezeigt. Beachten Sie, dass nur Daten aus einfachen Elementen in der Vorschau angezeigt werden können.</p>
Schema exportieren	<p>Klicken Sie auf diese Schaltfläche, um eine XSD-Datei zu speichern, die das in der Ansicht „Tatsächliche Datei“ angezeigte Schema darstellt. Die Schemadatei wird sofort an dem von Ihnen angegebenen Speicherort gespeichert.</p>

Registerkarte „Felder“

Die Registerkarte **Felder** definiert die Felder, die Sie in die XML-Ausgabedatei einbeziehen möchten. Beim Hinzufügen von Feldern werden diese in einer Baumstruktur angezeigt. Die Baumstruktur zeigt den Namen des Elements oder Attributs an, das in die XML-Datei geschrieben wird. In Klammern hinter dem Element-/Attributnamen steht der Name des Datenflussfeldes gefolgt vom Datentyp. Beispiel:



Dies gibt an, dass vier Elemente und ein Attribut in die XML-Datei geschrieben werden. Das Attribut wird durch ein rotes @-Zeichen angegeben.

Beachten Sie, dass das Element „Bundesland/Kanton“ die Daten aus dem Feld „StateProvince“ enthält, bei denen es sich um Zeichenfolgendaten handelt. Gleichermäßen enthält die Postleitzahl

des Elements die Daten aus dem Feld „PostalCode“, bei denen es sich ebenfalls um Zeichenfolgendaten handelt. Die XML-Datei könnte wie folgt aussehen:

```
<XmlRoot>
  <Customer Status="0">
    <AddressLine1>7713 Mullen Dr</AddressLine1>
    <City>Austin</City>
    <State>TX</State>
    <ZIP>78757-1346</ZIP>
  </Customer>
  <Customer Status="0">
    <AddressLine1>1825B Kramer Ln</AddressLine1>
    <City>Austin</City>
    <State>TX</State>
    <ZIP>78758-4260</ZIP>
  </Customer>
</XmlRoot>
```

Anmerkung: Der Stammelementname (in diesem Beispiel `<XmlRoot>`) wird auf der Registerkarte **Dateieigenschaften** angegeben.

Die folgende Tabelle enthält die Optionen der Registerkarte **Felder**.

Name der Option	Beschreibung
Hinzufügen	Fügt der Ausgabe ein Feld hinzu.

Name der Option

Beschreibung

Ändern

Name der Option

Beschreibung

Ändert die Art, wie das Feld in die XML-Datei geschrieben wird. Sie können Folgendes angeben:

Ausgabetyt

Diese Option ist verfügbar, wenn Sie ein einfaches Feld ändern. Sie gibt an, ob das Datenflussfeld in ein XML-Element oder -Attribut geschrieben werden soll.

Element Wählen Sie diese Option aus, um die Felddaten in ein XML-Element zu schreiben. Geben Sie den zu verwendenden Elementnamen in das Feld **Elementname** ein.

Attribut Schreibt die Felddaten in ein Attribut des übergeordneten Elements. Geben Sie den zu verwendenden Attributnamen in das Feld **Attributname** ein.

Elementname/Attributname Gibt den Namen des Elements oder Attributs an, das in die XML-Datei geschrieben werden soll. Der Standardname ist der Name des Datenflussfeldes.

Alle untergeordneten Elemente ändern in

Diese Option ist verfügbar, wenn Sie ein komplexes Element ändern. Sie gibt den XML-Typ an, den das komplexe Element enthalten soll. Zur Auswahl stehen:

Keine Änderung Die untergeordneten Typen bleiben entsprechend ihrer aktuellen Definition unverändert: entweder Element oder Attribut. Sie können den Typ für jedes Feld einzeln angeben, indem Sie das Feld auswählen und auf **Ändern** klicken.

Elemente Alle einfachen Felder unter dem Element werden als XML-Elemente geschrieben.

Attribute Alle einfachen Felder unter dem Element werden als XML-Attribute geschrieben.

Namespace

Wenn Sie einen XML-Namespace angeben möchten, der für das Element oder Attribut verwendet werden soll, wählen Sie ihn hier aus. Sie können Namespaces auf der Registerkarte **Felder** im „Write to XML“-Schritt erstellen.

Leere Felder einschließen Aktivieren Sie dieses Kästchen für eine Einbeziehung in die XML-Elemente der Ausgabedatei, die einen

Name der Option	Beschreibung
	<p>Nullwert oder keine Daten enthalten. Wenn Sie dieses Kästchen nicht aktivieren, werden leere Elemente nicht in die Ausgabe einbezogen.</p> <p>Wenn Sie beispielsweise ein Element mit dem Namen <code><City></code> definieren, es jedoch einen Datensatz gibt, der keine Daten im Feld „City“ enthält, enthält die XML-Ausgabe das Folgende, wenn Sie das Kästchen Leere Felder einschließen aktivieren:</p> <pre data-bbox="873 632 1263 657"><City xs:nil="true"></City></pre> <p>Wenn Sie dieses Kästchen nicht aktivieren, wird das Element <code><City></code> nicht in die Ausgabedatei geschrieben.</p> <p>Anmerkung: Das Datenflussfeld zeigt das Feld an, dessen Daten in das Element oder Attribut geschrieben werden. Dies wird angezeigt, so dass Sie, wenn Sie den Namen des Elements oder Attributs ändern, dennoch sehen können, welche Felddaten im Element oder Attribut enthalten sind.</p>
Entfernen	Entfernt das ausgewählte Feld aus der Ausgabe. Wenn Sie ein Listefeld entfernen, werden alle untergeordneten Felder ebenfalls entfernt. Wenn Sie ein untergeordnetes Feld entfernen, wird nur das ausgewählte untergeordnete Feld aus dem Listefeld entfernt.
Alle entfernen	Entfernt alle Felder aus der Ausgabe.
Nach oben/Nach unten	<p>Ordnet das ausgewählte Feld neu an.</p> <p>Beachten Sie, dass Sie einfache Elemente nicht in komplexe Elemente verschieben können. Wenn Sie die Elemente in einem komplexen Element ändern möchten, müssen Sie den Aggregatorschritt des Datenflusses ändern, um die gewünschten Datenflussfelder in das komplexe Element einzubeziehen. Weitere Informationen finden Sie unter Erstellen von komplexem XML aus flachen Daten auf Seite 315.</p>
Erneut generieren	Ersetzt die Felder, die aktuell mit den Feldern aus dem Upstream-Kanal im „Write to XML“-Schritt definiert sind.

Registerkarte „Laufzeit“

Name der Option	Beschreibung
Mehrere Dateien generieren	<p>Wählen Sie diese Option aus, um Datensätze in unterschiedliche Dateien zu schreiben (statt alle Datensätze in eine Datei). Die Datei, in die jeder Datensatz geschrieben wird, ist im Datensatz selbst angegeben. Jeder Datensatz muss ein Feld enthalten, das entweder den Namen oder den vollständigen Pfad zu der Datei angibt, in die der Datensatz geschrieben werden soll. Wenn Sie beispielsweise die Aktienkurse verschiedener Unternehmen (verschiedener Gruppen) allen Kunden gesondert senden möchten, schreibt dieses Feature die Aktienkurse der verschiedenen Unternehmen in separate Dateien, die Sie auf Wunsch jedem der Kunden senden können. Wenn Sie die Option Mehrere Dateien generieren aktivieren, müssen Sie eine Ausgabedatei entweder auf dem Spectrum™ Technology Platform-Server oder einem FTP-Server angeben. Wenn Sie Daten in eine Datei auf einem FTP-Server schreiben möchten, müssen Sie mit Management Console eine Verbindung zum Dateiserver definieren.</p> <p>Anmerkung: Die Datensätze in der im Dateipfadfeld ausgewählten Spalte müssen sortiert vorliegen. Verwenden Sie dieses Feature, wenn der Datensatz entweder den Namen oder den vollständigen Pfad der Datei enthält.</p>
Dateipfadfeld	<p>Wählt das Feld aus, das den Pfad (entweder ein Dateiname oder der vollständige Dateipfad) der Datei enthält, in die Sie den Datensatz schreiben möchten. Beachten Sie, dass nur die einfachen Elemente, die direkt einem Stamm zugewiesen sind, im Dateipfadfeld aufgelistet werden. Dieses Feld ist nur aktiviert, wenn Sie die Option Mehrere Dateien generieren auswählen.</p>
Schema zur Laufzeit generieren	<p>Wählen Sie diese Option aus, um ein XSD bei Laufzeit zu generieren und die Referenz <code>noNamespaceSchemaLocation</code> in das Schema der XML-Datei einzufügen. Der Wert des Attributs <code>noNamespaceSchemaLocation</code> ist der XSD-Dateiname in der XML-Datei. Wenn Sie das Schema während der Bearbeitung eines Datenflusses exportieren, gibt es keine Referenz zum XSD in der XML-Ausgabedatei, und der Benutzer muss sie manuell in die Referenz zum XSD hinzufügen.</p>
Schemapfad	<p>Gibt den Pfad zur Speicherung der XSD-Datei an, der das Schema der XML-Ausgabedatei enthält. Klicken Sie auf die Ellipsenschaltfläche (...), um die gewünschte Datei zu lokalisieren. Die Schemadatei wird an dem Speicherort gespeichert, den Sie bei der Ausführung des Datenflusses angeben.</p>

Verwenden von Namespaces in einer XML-Ausgabedatei

Mithilfe von Namespaces können Sie Element- und Attributnamen in Ihrer Ausgabe duplizieren, indem Sie jedem Element oder Attribut einen XML-Namespace zuweisen.

1. Öffnen Sie im Enterprise Designer den Datenfluss.
2. Doppelklicken Sie in der Arbeitsfläche auf den „Write to XML“-Schritt.
3. Klicken Sie auf die Registerkarte **Felder**.

4. Definieren Sie einen oder mehrere Namespaces:
 - a) Geben Sie in die Spalte **Präfix** das Präfix ein, das Sie für die Zuweisung eines Elements oder Attributs zum Namespace verwenden möchten.
 - b) Geben Sie in der Spalte **Namespace** die URL des Namespace an.
 - c) Wiederholen Sie diese Schritte, um so viele Namespaces wie gewünscht für die XML-Ausgabedatei zu erstellen.
5. Weisen Sie ein oder mehrere Elemente oder Attribute zum Namespace zu.
 - a) Wählen Sie auf der Registerkarte **Felder** das Element oder Attribut aus, das Sie einem Namespace zuweisen möchten, und klicken Sie auf **Ändern** oder erstellen Sie ein neues Element oder Attribut durch Klicken auf **Hinzufügen**.
 - b) Wählen Sie im Feld **Namespace** das Namespace-Präfix aus, das Sie dem Element oder Attribut zuweisen möchten.
 - c) Klicken Sie auf **OK**.

Erstellen von komplexem XML aus flachen Daten

Datenflüsse erzeugen häufig Datensätze, die flache Felder enthalten, die als einfache XML-Elemente in XML geschrieben werden. Wenn Sie flache Felder in komplexen XML-Elemente anordnen möchten, um hierarchische Daten zu erzeugen, können Sie dazu eine oder mehr Aggregatorschritte verwenden.

Bei diesen flachen Daten, bei denen die erste Zeile ein Kopfzeilendatensatz ist, gilt beispielsweise:

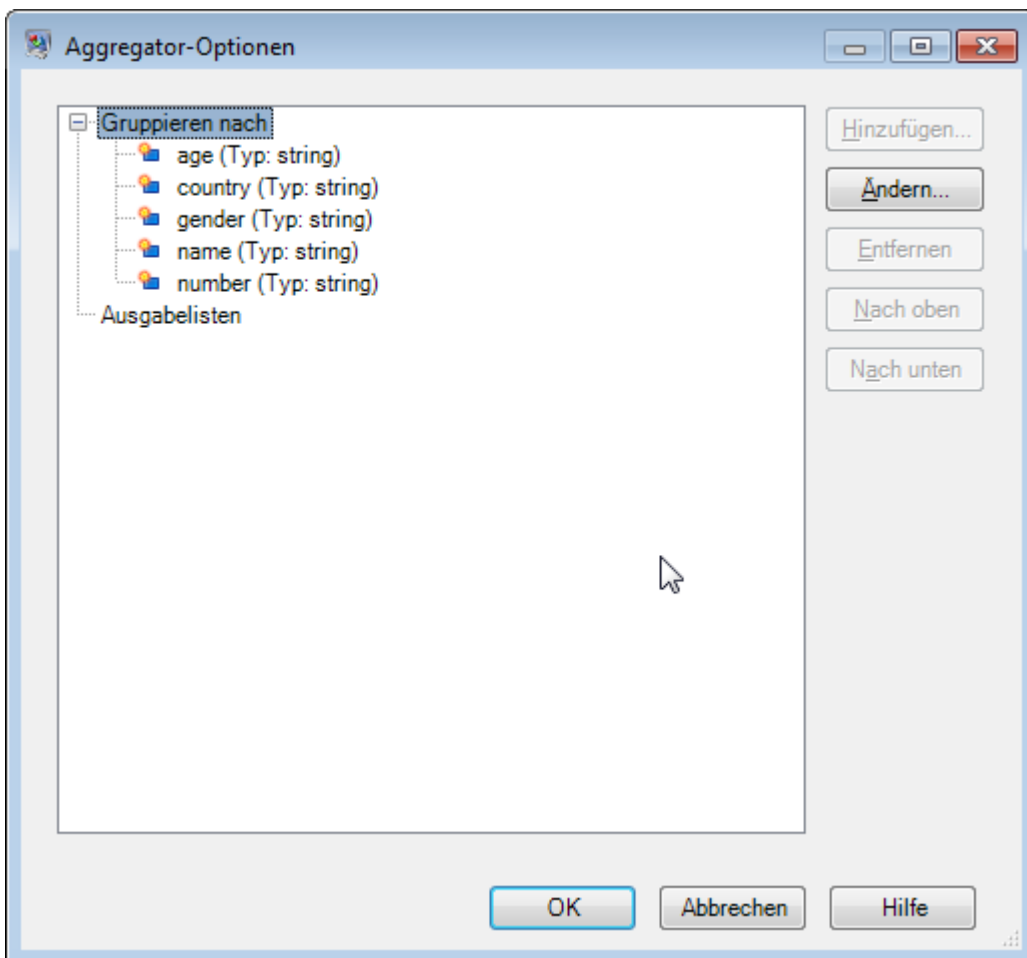
```
addressline1,age,city,country,gender,name,number,postalcode,stateprovince,type
1253 Summer St.,43,Boston,United States,M,Sam,019922,02110,MA,Savings
```

Sie könnten wie hier dargestellt die Felder mit Daten bezüglich der Adresse und Felder bezüglich des Kontos in komplexen XML-Elementen namens `<Address>` und `<Account>` gruppieren:

```
<CustomerRecord>
  <name>Sam</name>
  <age>43</age>
  <gender>M</gender>
  <country>United States</country>
  <Address>
    <addressline1>1253 Summer St.</addressline1>
    <city>Boston</city>
    <stateprovince>MA</stateprovince>
    <postalcode>02110</postalcode>
  </Address>
  <Account>
    <number>019922</number>
    <type>Savings</type>
  </Account>
</CustomerRecord>
```

1. Fügen Sie einen Aggregatorschritt an dem Punkt im Datenfluss hinzu, an dem Sie komplexe Elemente erstellen möchten.
2. Doppelklicken Sie auf den Aggregatorschritt, um die Schrittoptionen zu öffnen.
3. Wählen Sie **Gruppieren nach** aus und klicken Sie auf **Hinzufügen**.
4. Wählen Sie das Feld aus, das einen eindeutigen Wert für jeden Datensatz enthält, etwa eine Kontonummer, und klicken Sie dann auf **OK**.
5. Wählen Sie bei anderen einfachen Feldern, die Sie übergeben möchten, wieder **Gruppieren nach** aus und klicken Sie auf **Hinzufügen**. Fügen Sie dann alle einfachen Felder hinzu, die Sie einschließen möchten.

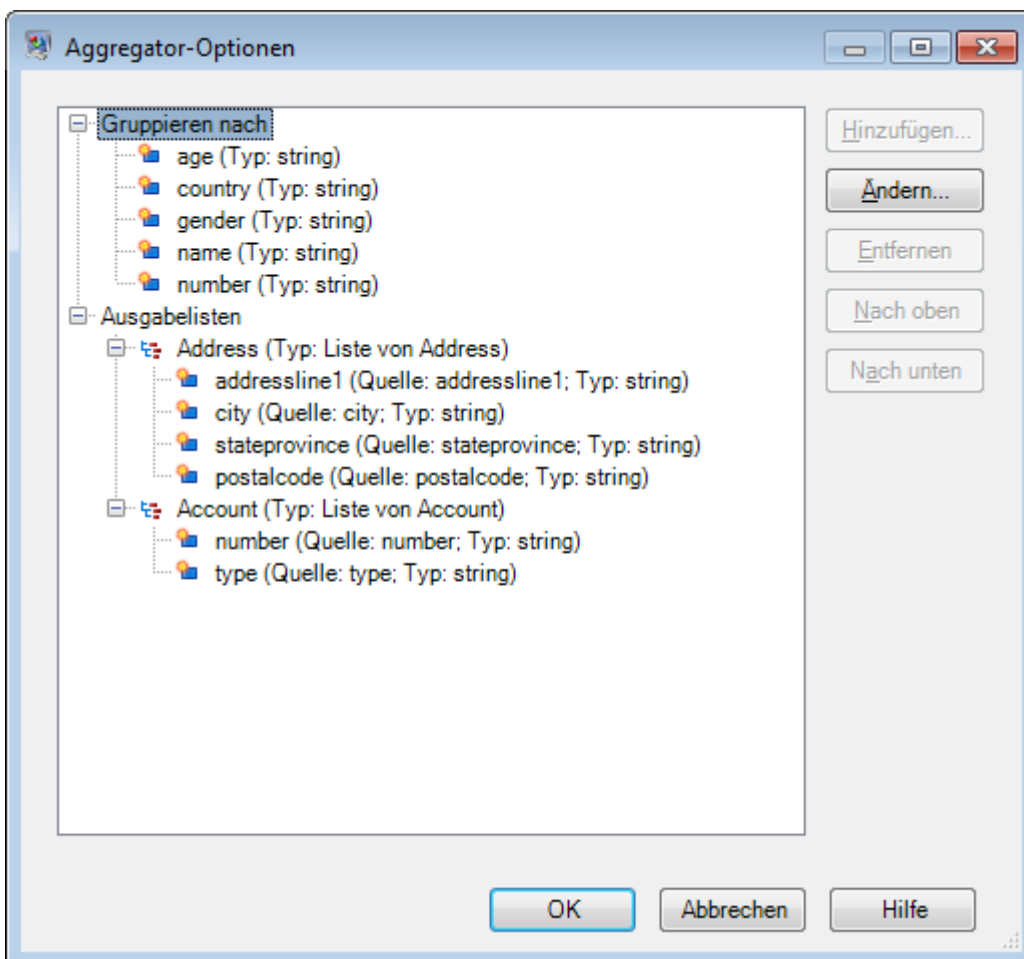
In diesem Fall beispielsweise gibt es fünf einfache Felder, die in jeden Datensatz eingeschlossen werden: Nummer, Name, Alter, Geschlecht und Land.



6. Wählen Sie **Ausgabelisten** aus und klicken Sie auf **Hinzufügen**.
7. Wählen Sie **Neuer Datentyp** aus. Dies bewirkt, dass ein neues komplexes Element definiert wird. Geben Sie eine Beschreibung für die Art von Daten ein, die dieses komplexe Element enthalten soll. Sie könnten beispielsweise „Komplex“ eingeben, da Sie ein komplexes XML-Element erstellen. Der Datentypname kann ein beliebiger Name sein.

8. Geben Sie in das Feld **Name** den Namen für das Feld ein. Dies ist zugleich der Name des XML-Elements.
9. Klicken Sie auf **OK**.
10. Wählen Sie das gerade erstellte Feld aus und klicken Sie auf **Hinzufügen**.
11. Wenn **Vorhandenes Feld** aktiviert ist, wählen Sie ein Feld, das Sie als untergeordnetes Feld zum komplexen Element hinzufügen möchten, und klicken Sie auf **OK**.
12. Wiederholen Sie die vorherigen zwei Schritte, um zusätzliche Felder zum komplexen Element hinzuzufügen.
13. Fügen Sie nach Bedarf zusätzliche komplexe Felder hinzu.

Wenn Sie fertig sind, sollten Sie über einen Aggregatorschritt verfügen, der jedes einfache und komplexe Feld auflistet, das Sie in jeden Datensatz einschließen möchten. Beispiel:



14. Klicken Sie auf **OK**.

Datums- und Zahlenmuster

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
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

Buchstabe	Beschreibung	Beispiel
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
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

Buchstabe	Beschreibung	Beispiel
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
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 dass 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: 1.234×10^3 .

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 verwendet, 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	<p>Steht für eine Stelle im Muster, bei dem Nullen zur Füllung des Musters verwendet werden. Durch Anwendung des folgenden Musters auf die Zahl 27:</p> <p>0000</p> <p>würde die Zahl wie folgt aussehen:</p> <p>0027</p>
#	<p>Steht für Stelle, für die Nullen ausgelassen werden. Durch Anwendung des folgenden Musters auf die Zahl 27:</p> <p>####</p> <p>würde die Zahl wie folgt aussehen:</p> <p>27</p>
.	<p>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.</p>
-	<p>Das Negativzeichen, das im ausgewählten Gebietsschema verwendet wird. Für die meisten Gebietsschemata ist dies das Minus (-).</p>

Symbol	Beschreibung
,	<p>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.</p> <p>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 verwendet. Alle folgenden Muster erzeugen beispielsweise das gleiche Ergebnis:</p> <pre>#, ##, ###, #### #####, ##### ##, ####, #####</pre>
E	<p>Trennt Mantisse und Exponent in wissenschaftlicher Schreibweise. Sie müssen das E in Ihrem Muster nicht in Anführungszeichen einschließen. Siehe Wissenschaftliche Schreibweise auf Seite 321.</p>
;	<p>Trennt positive und negative Untermuster. Siehe Muster für negative Zahlen auf Seite 321.</p>
%	<p>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:</p> <pre>##%</pre> <p>Folgendes Ergebnis aufweisen:</p> <pre>35%</pre>
¤	<p>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.</p>

Symbol	Beschreibung
'	<p data-bbox="535 325 1433 367">Dient der Angabe von Sonderzeichen in einem Präfix oder Suffix. Beispiel:</p> <pre data-bbox="535 378 1433 409">' '# '#'</pre> <p data-bbox="535 420 1433 451">Formatiert 123 in:</p> <pre data-bbox="535 462 1433 493">#123"</pre> <p data-bbox="535 504 1433 577">Verwenden Sie zwei aufeinander folgende Anführungsstriche, um einen einzelnen zu erzeugen:</p> <pre data-bbox="535 588 1433 619"># o' 'clock"</pre>

6 - Konfigurationen

In this section

Oracle LogMiner-Konfiguration

326

Oracle LogMiner-Konfiguration

Beim Oracle LogMiner handelt es sich um ein Back-End-Dienstprogramm, über das Spectrum diejenigen Protokolle abfragen und auf diejenigen Protokolle zugreifen kann, die für eine Oracle-Datenbank erstellt wurden.

Mithilfe dieses Dienstprogramms kann Spectrum™ die Protokolle der Oracle-Datenquelle lesen, um Änderungen an den Spalten ihrer Tabellen als Teil des Schrittes **DB Change Data Reader** nachzuverfolgen.

Nicht unterstützte Datentypen und Attribute des Tabellenspeichers

Oracle LogMiner unterstützt die folgenden Datentypen und Attribute des Tabellenspeichers nicht:

- Datentyp **BFILE**
- Einfache und geschachtelte abstrakte Datentypen (**ADT**)
- Kollektionen (geschachtelte Tabellen und VARRAYs)
- Objektverweise
- Tabellen mit aktivierter Komprimierung:
- SecureFiles

Unterstützte Datenbanken und Redo-Log-Dateiversionen

LogMiner wird auf Oracle-Datenbanken mit der Version 8.1 oder höher ausgeführt.

Mit LogMiner können Sie auch Redo-Log-Dateien aus Oracle-Datenbanken der Version 8.0 analysieren. Welche Informationen abgerufen werden, hängt jedoch von der Protokollversion statt von der Version der verwendeten Datenbank ab.

Zur optimalen Nutzung von LogMiner können die Redo-Log-Dateien für Oracle9i beispielsweise vergrößert werden, um zusätzliche Informationen zu erfassen, wenn die ergänzende Protokollierung aktiviert ist. Redo-Log-Dateien, die mit älteren Oracle-Versionen erstellt wurden, verfügen nicht über die zusätzlichen Daten und weisen deswegen gegebenenfalls Einschränkungen hinsichtlich der von LogMiner unterstützten Vorgänge und Datentypen auf.

*Einschränkungen von SQL*Loader*

Spectrum CDC kann Daten erfassen, die vom Dienstprogramm SQL*Loader in Oracle-Tabellen geladen wurden. Allerdings gelten die folgenden Einschränkungen:

1. Der Ladevorgang für die Daten muss über einen *herkömmlichen Pfad* erfolgen. Spectrum CDC kann keine Daten erfassen, die über einen *Direct-Path-Load* geladen werden, da Oracle LogMiner keine Direct-Path-Loads unterstützt.
2. Die Lademethode sollte INSERT, APPEND oder REPLACE sein.

TRUNCATE wird nicht unterstützt, da der TRUNCATE-Befehl bewirkt, dass der SQL*Loader den Befehl TRUNCATE TABLE DDL eingibt. Da das Spectrum CDC-Feature die erwähnte DDL

nicht erfasst, werden Löschvorgänge für Zeilen, die durch die Verwendung des Befehls TRUNCATE TABLE DDL entstehen, nicht erfasst.

Erforderliche Benutzerberechtigungen

In der folgenden Tabelle werden die mindestens erforderlichen Systemberechtigungen aufgeführt, über die Oracle CDC-Benutzer verfügen müssen:

Systemberechtigung	Oracle-Version
ALTER ANY TABLE	ALL
CONNECT	ALL
LOCK ANY TABLE	ALL
SELECT ANY TRANSACTION	10g oder höher

In der folgenden Tabelle werden die mindestens erforderlichen Objektberechtigungen aufgeführt, über die Oracle CDC-Benutzer verfügen müssen:

Objektname	Berechtigung
Quellentabellen	LOCK ANY TABLE OR SELECT
PUBLIC.V\$DATABASE	SELECT
PUBLIC.V\$LOGMNR_CONTENTS	SELECT
SYS.DBMS_LOGMNR	EXECUTE
SYS.DBMS_LOGMNR_D	EXECUTE

Weitere Informationen zum Oracle LogMiner finden Sie [hier](#).

7 - Optimieren der Leistung

In this section

Ermitteln einer idealen Abrufgröße

329

Ermitteln einer idealen Abrufgröße

Im Schritt **Read from DB** wird die ideale Abrufgröße berechnet, indem die Ausführungszeit zwischen einem Schritt **Read from DB** und einem Schritt **Write to Null** verwendet wird.

Stellen Sie sicher, dass Sie Ausführungszeiten des Testauftrags anhand von unterschiedlichen Abrufgrößenwerten testen und dabei Ihre eigene Anwendung mit dem Spectrum™ Data Integration-Modul verwenden.

1. Erstellen Sie einen Auftrag im Enterprise Designer.
2. Ziehen Sie einen **Read from DB**-Schritt auf die Arbeitsfläche.
3. Ziehen Sie einen **Write to Null**-Schritt auf die Arbeitsfläche.
4. Erstellen Sie einen Kanal zwischen den beiden Schritten.
5. Klicken Sie doppelt auf den Schritt **Read from DB**, um ihn so zu konfigurieren, dass er Daten aus einer Tabelle liest, die die Testdaten enthält.
 - a) Wählen Sie auf der Registerkarte **Allgemein** im Feld **Verbindung** die Datenbank aus, die die Testdaten enthält.
 - b) Klicken Sie auf **SQL erstellen...**, um die SQL-Abfrage anhand ausgewählter Schemata und Tabellen zu erstellen, aus denen die Daten gelesen werden sollen.
Stellen Sie sicher, dass die ausgewählte Tabelle über mindestens 1.000 Datensätze verfügt, um optimale Testwerte zu erzielen.
 - c) Aktivieren Sie das Kontrollkästchen **Abrufgröße** auf der Registerkarte **Laufzeit**.
 - d) Geben Sie die Anzahl der Datensätze, die in einem Durchlauf gelesen werden sollen, in das verknüpfte Feld ein.
Spectrum™ Technology Platform wurde getestet, um optimal bei einer Abrufgröße von bis zu 1.000 zu funktionieren.
 - e) Klicken Sie auf **OK**.
6. Speichern Sie den Auftrag.
7. Führen Sie den Auftrag aus.
Das Fenster **Ausführungsdetails** wird geöffnet.
8. Klicken Sie auf **Aktualisieren**.
9. Notieren Sie sich die Zeiten für die Einträge **Gestartet** und **Abgeschlossen**.
10. Wiederholen Sie die Schritte 7–9. Erhöhen Sie dabei die Abrufgröße Schritt für Schritt, um die optimale Einstellung für Ihren Server zu ermitteln.

Sie haben nun die Einstellung der Abrufgröße ermittelt, die die optimale Leistung für Ihre Umgebung bietet.

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