

IBM DB2 Web Query 2.1 – Data Migrator Teil 5

Seit März 2015 ist das neueste Mitglied der DB2 Web Query-Familie verfügbar: **IBM DB2 Web Query for i DataMigrator ETL Extension**. In den ersten Artikeln habe ich Ihnen eine grobe Übersicht über die Möglichkeiten von DataMigrator gegeben. Dies wurde dann in Form eines einfachen Datenflusses exemplarisch gezeigt. In diesem Artikel werden weitere und komplexere Funktionalitäten von DataMigrator und insbesondere von Datenflüssen dargestellt.

Weitere Funktionen von Datenflüssen

Im letzten Artikel habe ich Ihnen die Hauptkomponenten von DataMigrator gezeigt. Dazu habe ich einen einfachen Datenfluss erzeugt, in dem die Spalten einer Tabelle eins-zu-eins an die Zieltabelle weitergereicht wurden. Im Folgenden verwende ich diesen Datenfluss als Ausgangspunkt für die Erläuterung von weiteren und komplexen Funktionalitäten. Dies erfolgt in mehreren Schritten:

- Vorbereitende Maßnahmen (wurde im letzten Artikel beschrieben)
- Einen Fluss kopieren (wurde im letzten Artikel beschrieben)
- Hinzufügen von Joins (wurde im letzten Artikel beschrieben)
- **Hinzufügen von Transformationen** (wird in diesem Artikel beschrieben)
- **Spalten und Transformationen auswählen** (wird in diesem Artikel beschrieben)
- Datenziele hinzufügen (wird im nächsten Artikel beschrieben)
- Eine Validierung hinzufügen (wird im nächsten Artikel beschrieben)
- Ein zweites Datenziel hinzufügen (wird im nächsten Artikel beschrieben)
- Fluss ausführen (wird im nächsten Artikel beschrieben)

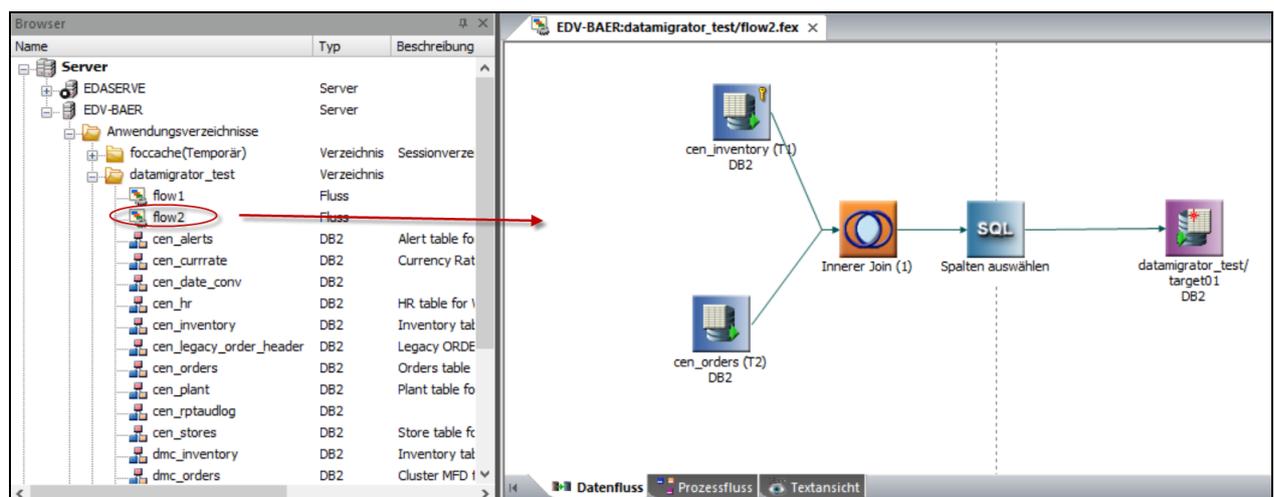


Abbildung 1: Darstellung des Flusses flow2

Wie bei flow1 besteht die Datenquelle aus der Tabelle inventory und dem Datenziel target01. Im letzten Artikel haben wir dann eine zweite Datenquelle *cen_orders* hinzugefügt. Beide Tabellen wurden durch einen Inner Join miteinander verbunden. Dieser Datenfluss wird nun weiter bearbeitet.

Hinzufügen von Transformationen

Zunächst stelle ich Ihnen noch eine interessante Methode vor, mit deren Hilfe der Inhalt der neu hinzugefügten Tabelle ORDERS geprüft werden kann. Dazu wähle ich das Kontextmenü von cen_orders und danach *Umschalten*.

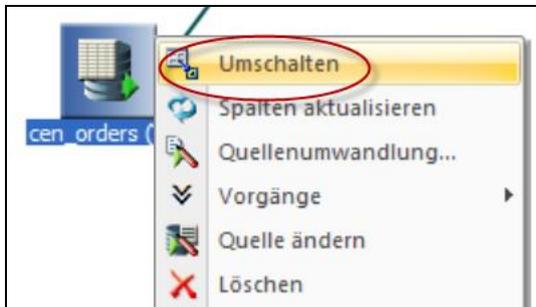


Abbildung 2: Auswahl Umschalten in Kontextmenü

Als erstes erscheint eine Liste mit den Spalten der ausgewählten Tabelle.

Name	Tabelle	Format	Beschreibung	Nullen
Order,Number	(CEN_ORDERS)	A5	Order Number	Nein
Product,Number	(CEN_ORDERS)	A4	Product Number	Nein
Order,Date	(CEN_ORDERS)	YYMD	Order Date	Nein
Store,Code	(CEN_ORDERS)	A6	Store Code	Nein
Plant,Code	(CEN_ORDERS)	A3	Plant Code	Nein
Sales Rep	(CEN_ORDERS)	A50	Sales Rep	Nein
Quantity	(CEN_ORDERS)	I11	Quantity	Ja
Revenue	(CEN_ORDERS)	P22.2	Revenue	Ja
Cost of,Goods ...	(CEN_ORDERS)	P22.2	Cost of Goods S...	Ja
Returns	(CEN_ORDERS)	I11	Returns	Ja
Warranty,Expe...	(CEN_ORDERS)	P22.2	Warranty Expe...	Ja
Shipping,Cost	(CEN_ORDERS)	P22.2	Shipping Cost	Ja
Order,Date,Year	(CEN_ORDERS)	I4	Order Date Year	Nein
Order,Date,Qu...	(CEN_ORDERS)	I1	Order Date Qua...	Nein
Order,Date,Mo...	(CEN_ORDERS)	I2	Order Date Month	Nein
Order,Date,Day	(CEN_ORDERS)	I2	Order Date Day	Nein
Order,Date,Year	(CEN_ORDERS)	YYMDy	Order Date Year	Nein
Order,Date,Y-Q	(CEN_ORDERS)	YYMDq	Order Date Y-Q	Nein
Order,Date,Y-M	(CEN_ORDERS)	YYMDm	Order Date Y-M	Nein

Abbildung 3: Auflistung der Spalten von Tabelle ORDERS

Mit der Auswahl *Beispieldaten* werden die ersten fünfzig Zeilen der ausgewählten Tabelle angezeigt. Der Standardwert von fünfzig Zeilen kann wie folgt geändert werden:

Variante-1:

Home-Tab → *Tools/Optionen* → *Optionen ausführen* → *Maximale Zeilenzahl für den Test-Report*

Variante-2:

Hauptmenü Datei (DMC) → *Optionen* → *Optionen ausführen* → *Maximale Zeilenzahl für den Test-Report*



Abbildung 4: Änderung der maximalen Zeilenanzahl für den Test-Report

Als nächstes erstellen wir eine virtuelle Spalte in der Datenquelle ORDERS. Dieser Vorgang wird auch als **Quellen-Umwandlung** bezeichnet. Diese Art von Transformation wird bei jedem gelesenen Satz ausgeführt, bevor Zeilen gefiltert oder aggregiert werden. Dies entspricht der Definition von Detail/Define-Feldern mit dem InfoAssist oder dem Synonym-Editor in DB2 Web Query (im Unterschied zu Summen/Compute-Feldern).

Die zu erstellende virtuelle Spalte soll die Differenz in Tagen zwischen den beiden Datumfeldern *tatsächliches Lieferdatum* und *gewünschtes Lieferdatum*. Dazu wähle ich aus dem Kontextmenü von *cen_orders* *Quellenumwandlung* und *Umwandlung einfügen*.

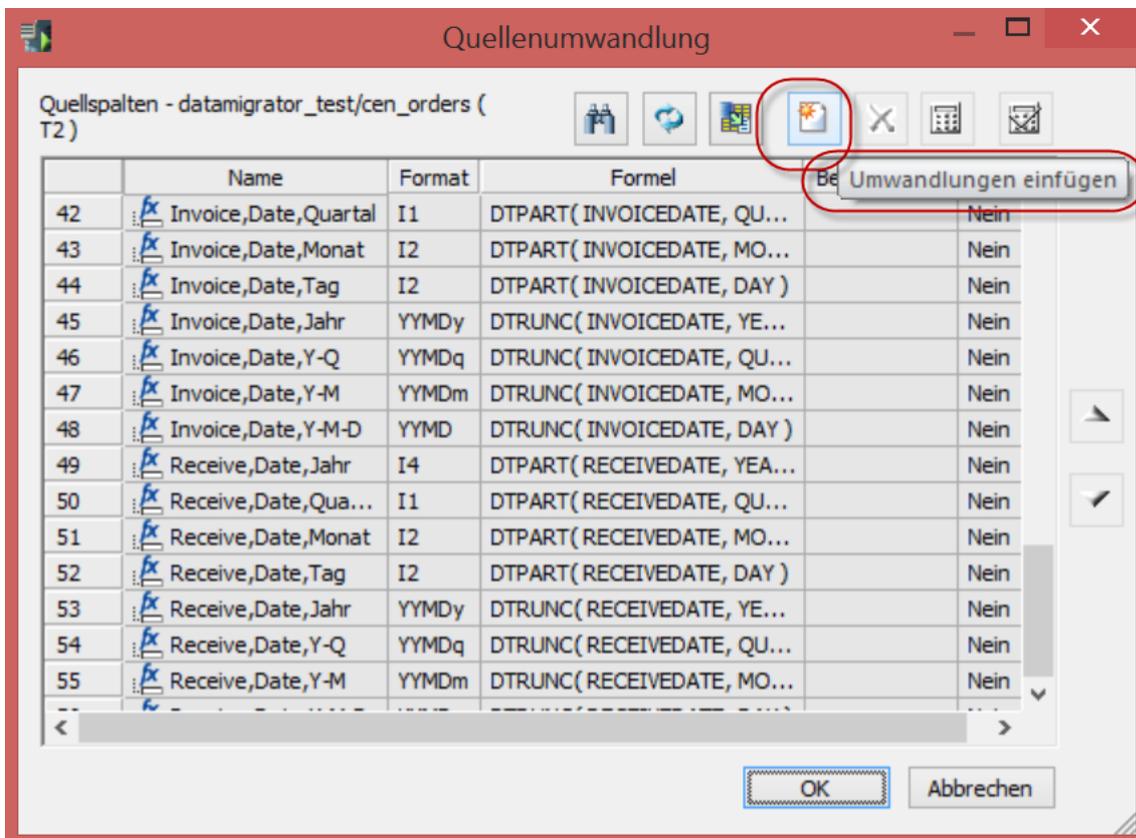


Abbildung 5: Eine Quellenumwandlung einfügen

Der Umwandlungsrechner wird geöffnet.

Für die Erstellung des Rechenfeldes sind die folgenden Schritte erforderlich (siehe entsprechende Nummerierung in der folgenden Abbildung):

1. Name für das Rechenfeld = SHIP_DIFFERENCE
2. Unter *Funktionen* → *Datum/Datum-Zeit – Vereinfacht* wähle ich die Funktion *DTDIFF*
3. end_date = SHIPDATE
4. start_date = REQUESTEDSHIPDATE
5. component = DAY

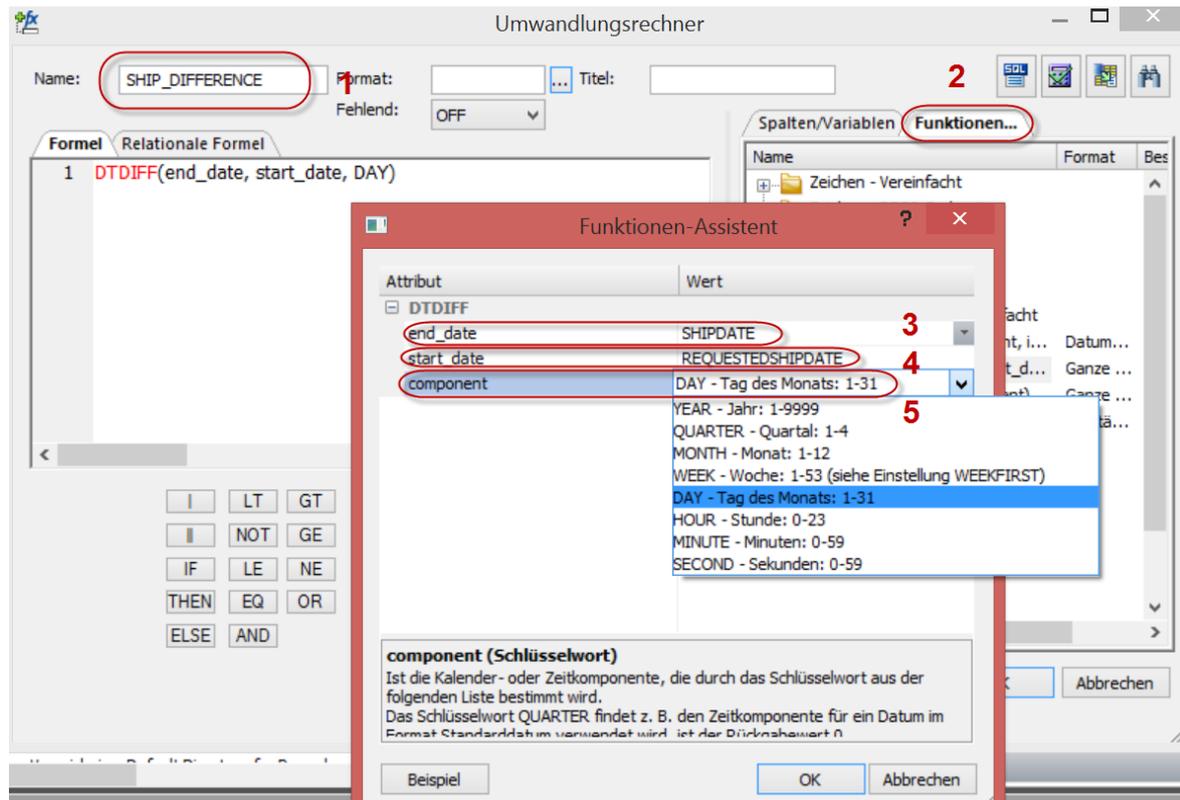


Abbildung 6: Erstellen des Rechenfeldes *SHIP_DIFFERENCE*

Nachdem ich die OK-Taste geklickt habe, erscheint das soeben erstellte Feld als neue Spalte im Fenster Quellenumwandlung.

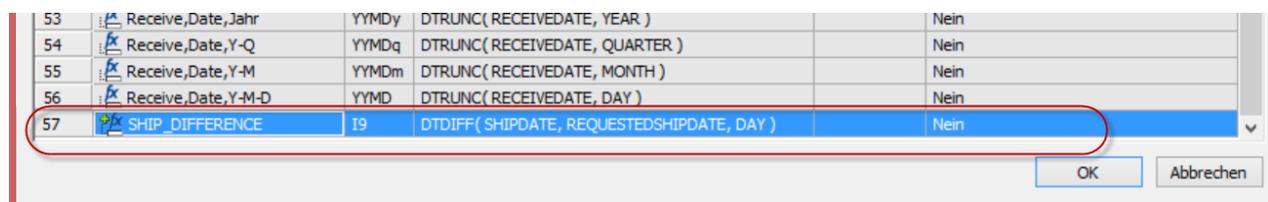


Abbildung 7: Neue Spalte *SHIP_DIFFERENCE*

Wenn wir uns nun das Ergebnis über die Anzeige als Beispieldaten anschauen, können wir feststellen, dass manche Werte negativ sind.

	Actual Ship Date	Requested Ship Date	SHIP_DIFFERENCE
1	2013/11/21	2014/01/20	-60
2	2014/01/16	2014/01/18	-2
3	2013/12/14	2013/11/27	17
4	2013/12/04	2013/11/17	17
5	2013/12/01	2013/11/26	5
6	2014/02/03	2014/01/26	8
7	2013/11/26	2013/12/04	-8
8	2013/12/06	2013/11/30	6
9	2013/12/04	2013/12/03	1
10	2013/12/04	2013/12/04	0
11	2013/12/20	2013/12/19	1

Abbildung 8: Beispielwerte für die neu erstellte Spalte

Wir sind aber nur an den absoluten Werten interessiert. Deshalb ist es noch erforderlich, unsere Formel um die Funktion ABS zu ergänzen.

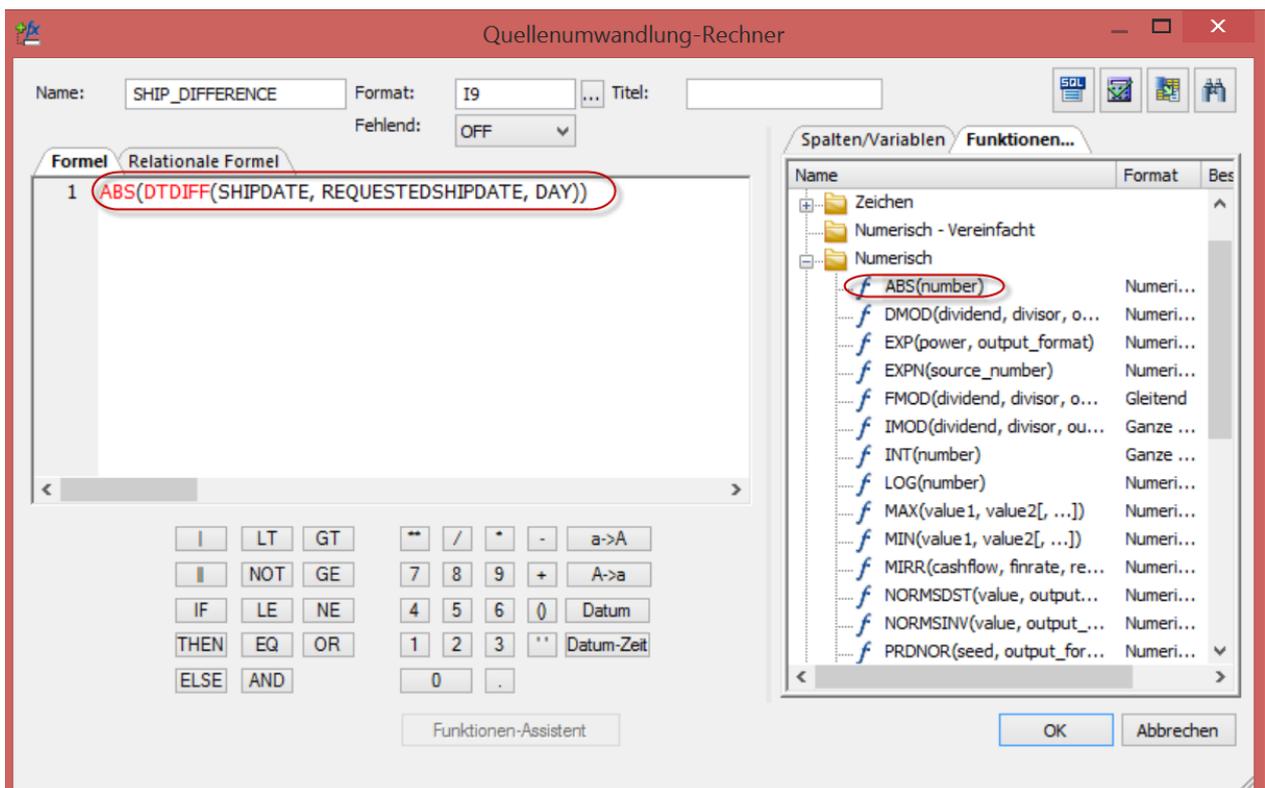


Abbildung 9: Verwendung der Funktion ABS

Mit erneutem Testen der Umwandlung erhalte ich das gewünschte Ergebnis:

The screenshot shows the IBM Data Migrator interface. On the left, a table with columns for date conversions is visible. A context menu is open over the 'SHIP_DIFFERENCE' column (row 57). The menu options are: Bearbeiten, Löschen, **Umwandlungen testen**, Fehler-Details, Umwandlungen kopieren, and Umwandlungen einfügen. A red circle highlights the 'SHIP_DIFFERENCE' column header. A red arrow points from the 'Umwandlungen testen' menu item to a separate window titled 'Testumwandlungen. Begrenzt auf 50 Zeilen.'. This window displays a table with the following data:

	Actual Ship Date	Requested Ship Date	SHIP_DIFFERENCE
1	2013/11/21	2014/01/20	60
2	2014/01/16	2014/01/18	2
3	2013/12/14	2013/11/27	17
4	2013/12/04	2013/11/17	17
5	2013/12/01	2013/11/26	5
6	2014/02/03	2014/01/26	8
7	2013/11/26	2013/12/04	8
8	2013/12/06	2013/11/30	6
9	2013/12/04	2013/12/03	1
10	2013/12/04	2013/12/04	0
11	2013/12/20	2013/12/19	1

Abbildung 10: Umwandlung mit ABS und DTDIFF testen

Spalten und Transformationen auswählen

Nun sind wir soweit, dass wir Spalten auswählen können, deren Inhalte dann in die Datenziele übertragen werden.

Schauen wir uns zunächst das Kontextmenü von *SQL-Spalten auswählen* an.

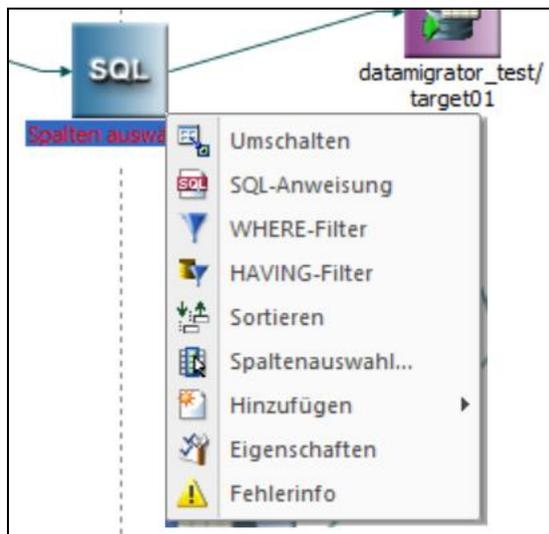


Abbildung 11: Kontextmenü von SQL

Im Einzelnen sind das:

- Umschalten
Öffnet ein Informationsfenster (wie bereits oben beschrieben)
- SQL-Anweisung
Zeigt die aktuelle SQL-Anweisung an. Im Moment werden die bereits für den Datenfluss *flow1* ausgewählten Spalten der Tabelle INVENTORY und der Join INVENTORY-ORDERS angezeigt.
- WHERE-Filter
Stellt einen Filtereditor zur Verfügung, womit Ausdrücke erstellt werden können, die die Anzahl der ausgewählten Zeilen begrenzt. Beispielsweise könnten nur Zeilen eines bestimmten Jahres erwünscht sein.
- HAVING-Filter
Stellt einen Filtereditor zur Verfügung, mit dem Ausdrücke erzeugt werden können, die die Anzahl Zeilen nach erfolgter Aggregation mit Hilfe von GROUP BY reduziert. Diese Option erscheint nur dann, wenn die Spaltenselektion bereits GROUP BY beinhaltet.
- Sortieren
Liefert einen Bildschirm mit Auswahl von Spalten für die Reihenfolge der Verarbeitung.
- Spaltenauswahl
Zeigt noch nicht ausgewählte Spalten und die für das Datenziel bereits ausgewählten Spalten an. Spalten können hinzugefügt oder entfernt werden. Weitere Funktionalitäten werden wir im nächsten Schritt kennen lernen.
- Hinzufügen
Weitere Quellen, Joins und Union-Objekt können hinzugefügt werden.
- Eigenschaften
- Fehlerinfo

Nun werden wir die erforderlichen Spalten hinzufügen und dabei weitere interessante Funktionen kennenlernen. Dazu wähle ich aus dem Kontextmenü (siehe vorherige Abbildung) *Spaltenauswahl*. Damit wird das folgende Fenster angezeigt:

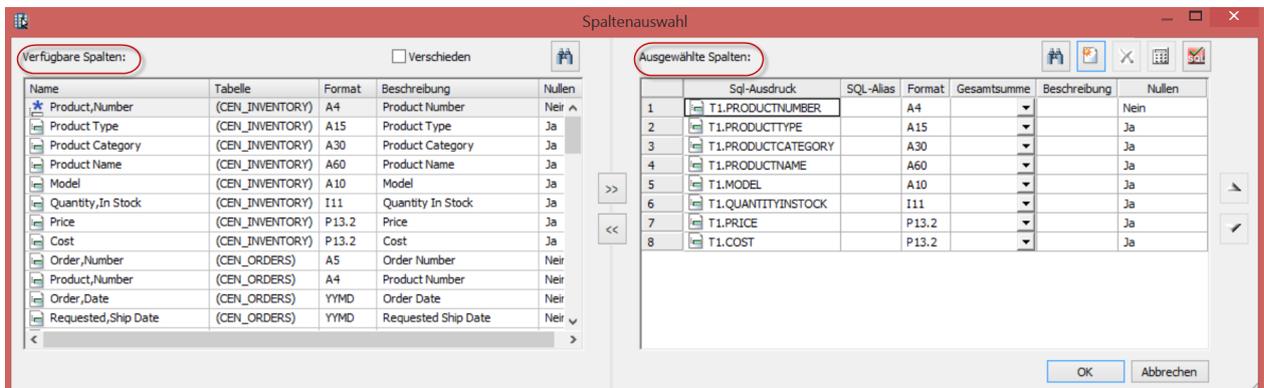


Abbildung 12: Spaltenauswahl mit verfügbaren und ausgewählten Spalten

Da wir den neuen Fluss *Flow2* durch Kopieren des zu Anfang erstellten Flusses *flow1* erstellt haben, enthält die Liste der ausgewählten Spalten alle Spalten der Tabelle INVENTORY.

Hinweis:

Die für die einzelnen Spalten angezeigten Attribute sind unter *Optionen* → *Allgemein* → *Spaltenverwaltung* hinterlegt und können jederzeit geändert werden.

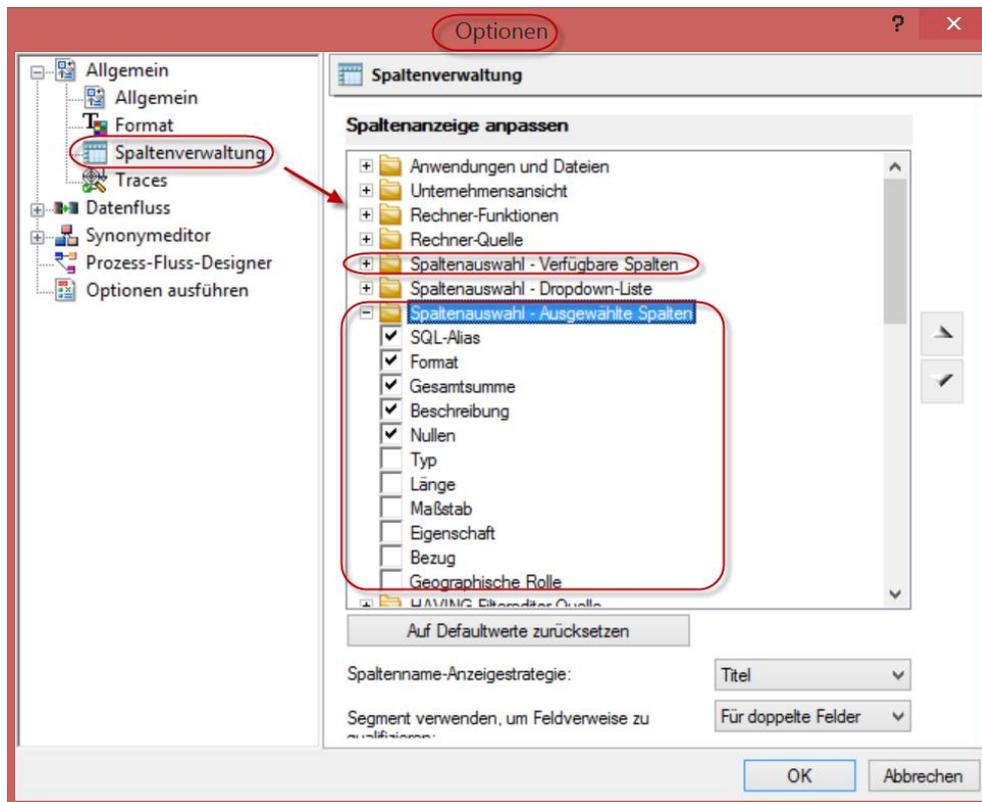


Abbildung 13: Anpassung der angezeigten Spaltenattribute

Für die Fortführung meines Beispiels verwende ich die folgenden Spalten:

- PRODUCTNAME
- PRODUCTTYPE
- PRODUCTNUMBER
- STORECODE
- SHIPDATE
- QUANTITY
- LINETOTAL
- COSTOFGOODSSOLD

Die Art und Weise, wie Spalten hinzugefügt, entfernt oder verschoben werden können, ist selbsterklärend. Das Ergebnis sollte wie in folgender Abbildung dargestellt aussehen:

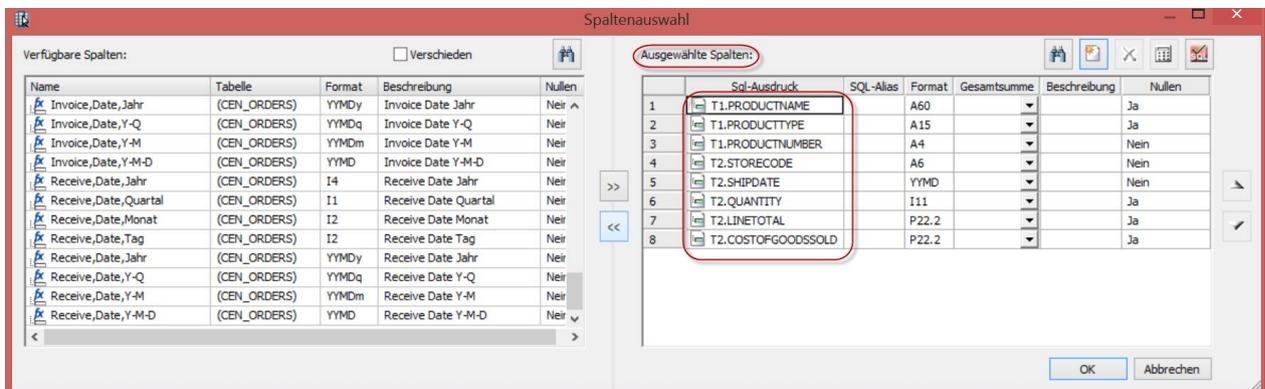


Abbildung 14: Liste der ausgewählten Spalten

Als nächstes soll der Profit als Differenz zwischen Umsatz und Kosten ermittelt werden. Dazu verwende ich den SQL-Rechner und erstelle eine neue Spalte PROFIT:

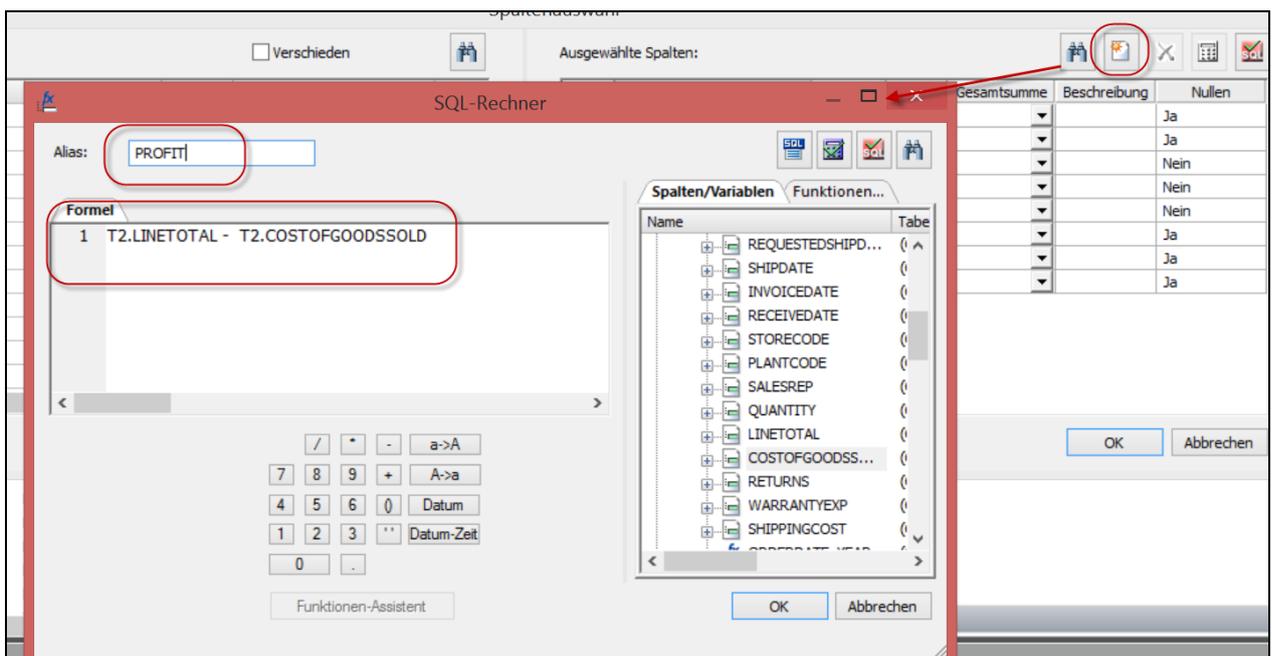


Abbildung 15: Neue Spalte PROFIT hinzufügen

Die soeben erstellte und errechnete Spalte PROFIT erscheint nun am Ende der Liste der ausgewählten Spalten.

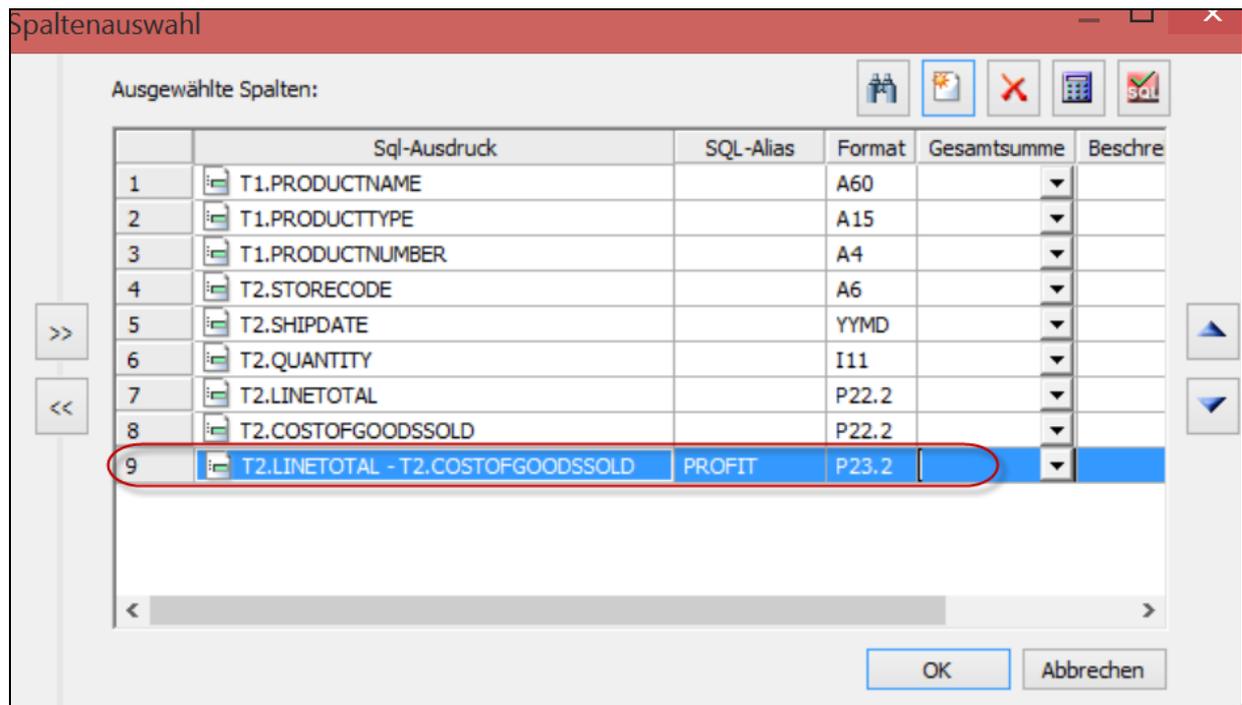


Abbildung 16: Liste der ausgewählten Spalten mit der neuen Spalte PROFIT

Nun sollen die Aufträge nach Datum zusammengefasst werden. Dazu wird nach den Dimensions-Spalten gruppiert (GROUP BY) und nach den Maßeinheiten summiert (SUM).

Für die Dimensionsspalten (=alphanumerische Spalten) wähle ich in der Spalte *Gesamtsumme* aus: *Gruppieren nach*.

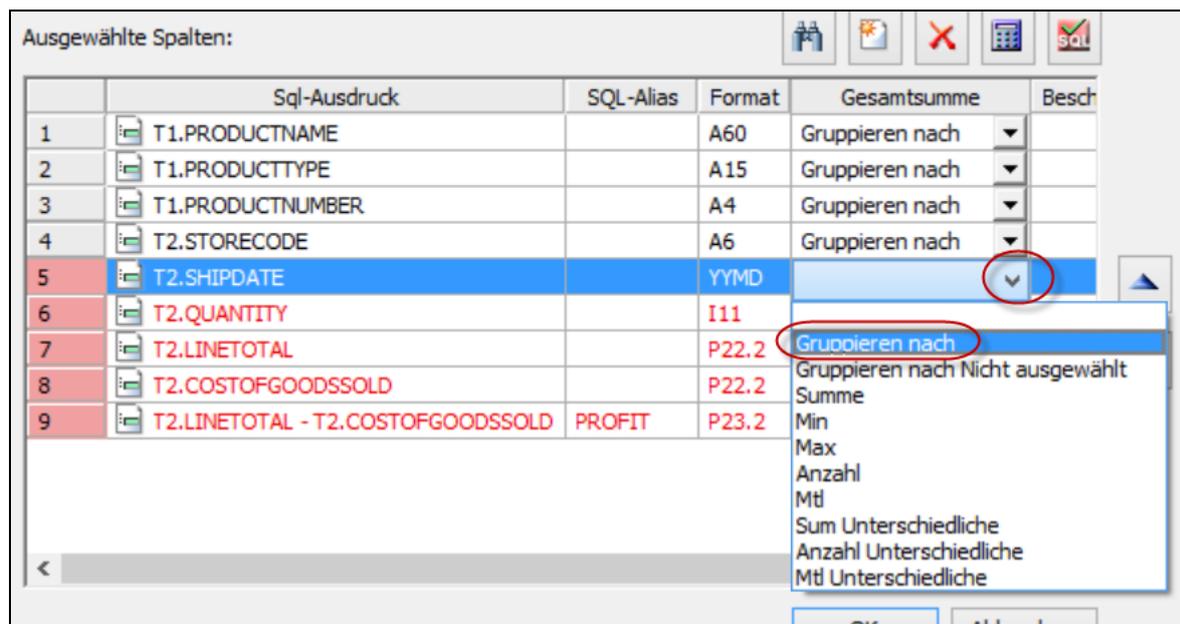


Abbildung 17: Auswahl *Gruppieren nach* für die Dimensionsspalten

Bei den numerischen Spalten (= Maßeinheiten) wähle ich auf die gleiche Weise *Summe*.

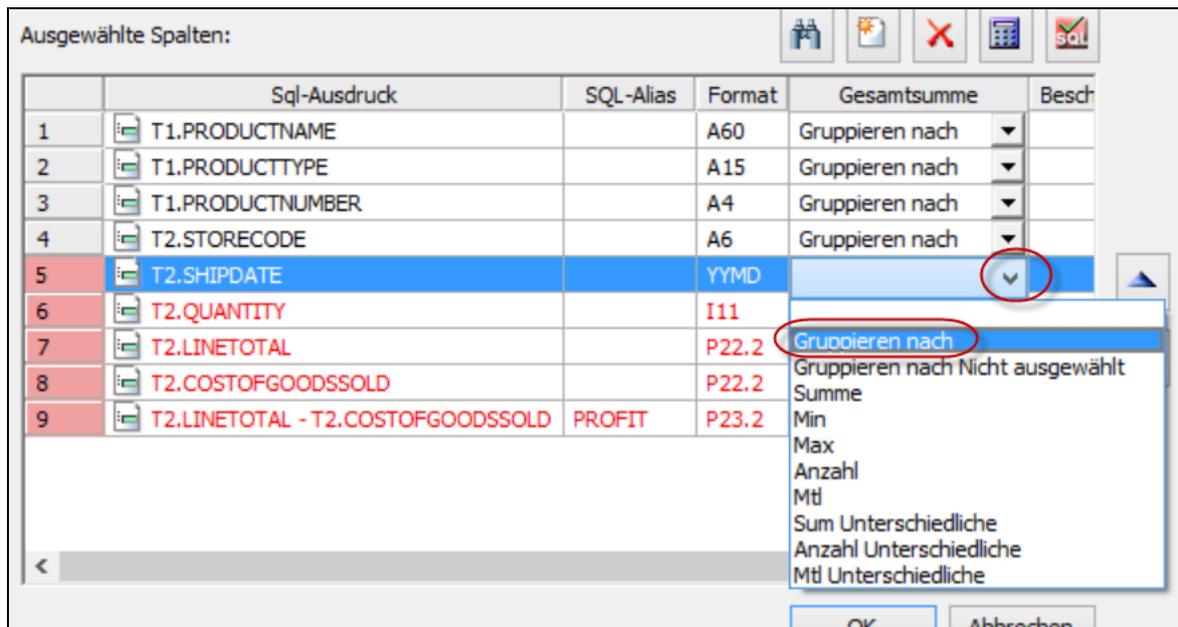


Abbildung 18: Auswahl *Summe* für die Maßeinheitsspalten

Hinweis:

Sobald für eine Spalte *Gruppieren nach* ausgewählt wurde, werden alle anderen Spalten so lange in Rot dargestellt, bis für alle anderen Spalten eine Aggregationsfunktion ausgewählt wurde.

Jetzt fehlt uns noch die zu Beginn erstellte Spalte *SHIP_DIFFERENCE*. Diese fügen wir auch noch hinzu. Hier interessiert uns nicht die Summe, sondern der jeweilige Durchschnittswert.

Hinweis:

Offensichtlich ist bei der Übersetzung der Auswahlliste ein Fehler unterlaufen. Für die Funktion *Average(Avg)* wurde der Begriff *Mtl* (??) verwendet. Ich habe das Labor in Rochester bereit darüber informiert und gebeten, *Avg* mit *Durchschnitt* zu übersetzen.

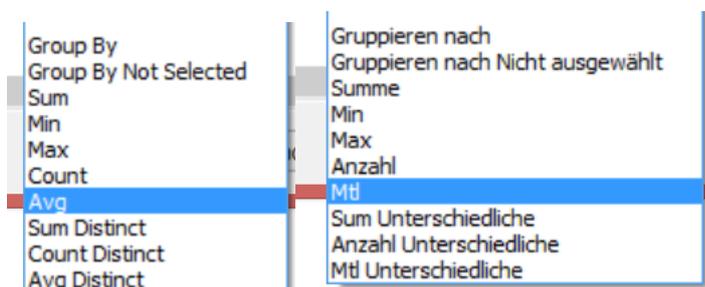


Abbildung 19: Auswahllisten für die Aggregatfunktionen in Englisch und Deutsch

Wählen Sie also bitte für die Spalte *SHIP_DIFFERENCE* die Funktion *Mtl* aus.

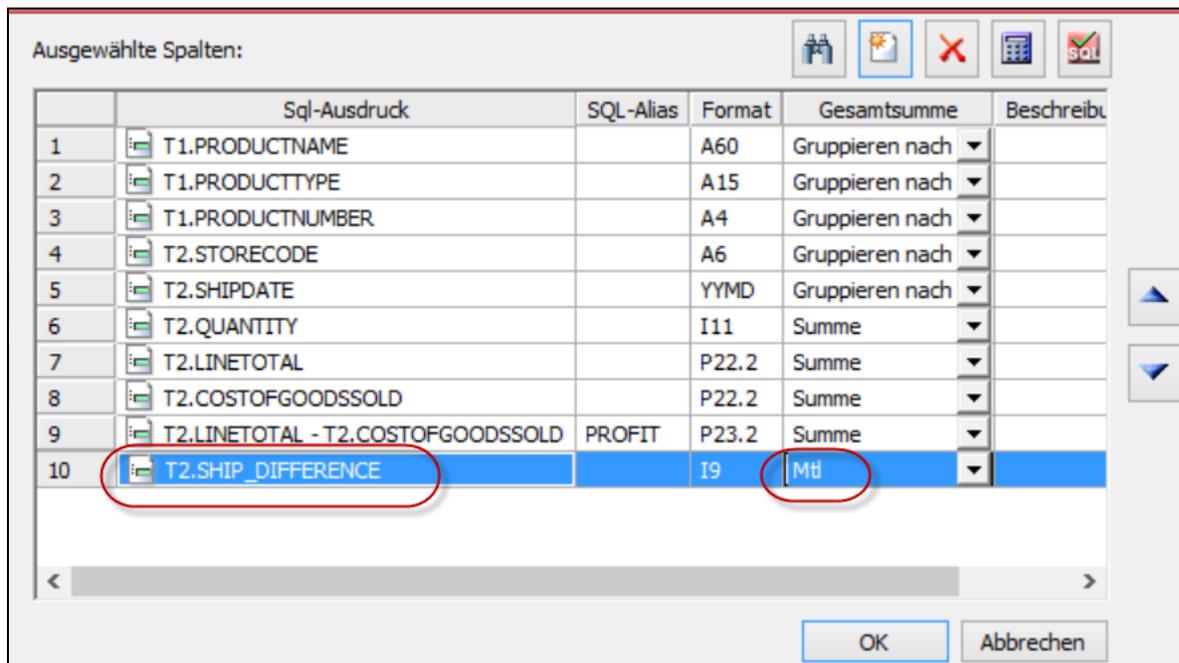


Abbildung 20: Neu ausgewählte Spalte SHIP_DIFFERENCE mit *Durchschnitt (=Mtl)*

Nun haben wir alle gewünschten Spalten der Datenquelle ausgewählt. Unsere Tabelle ORDERS enthält aber Aufträge aus mehreren Jahren. Für unsere Betrachtung sollen nur Aufträge ab dem Jahr 2013 berücksichtigt werden. Für diesen Zweck verwenden wir die Auswahl *WHERE-Filter* des Kontextmenüs von SQL.



Abbildung 21: Auswahl WHERE-Filter

Im darauf angezeigten WHERE-Filtereditor trage ich als Relationale Formel ein:

Spalte : Actual Ship Date
 Relation : >=
 Typ: Wert
 Wert: 2013-01-01'

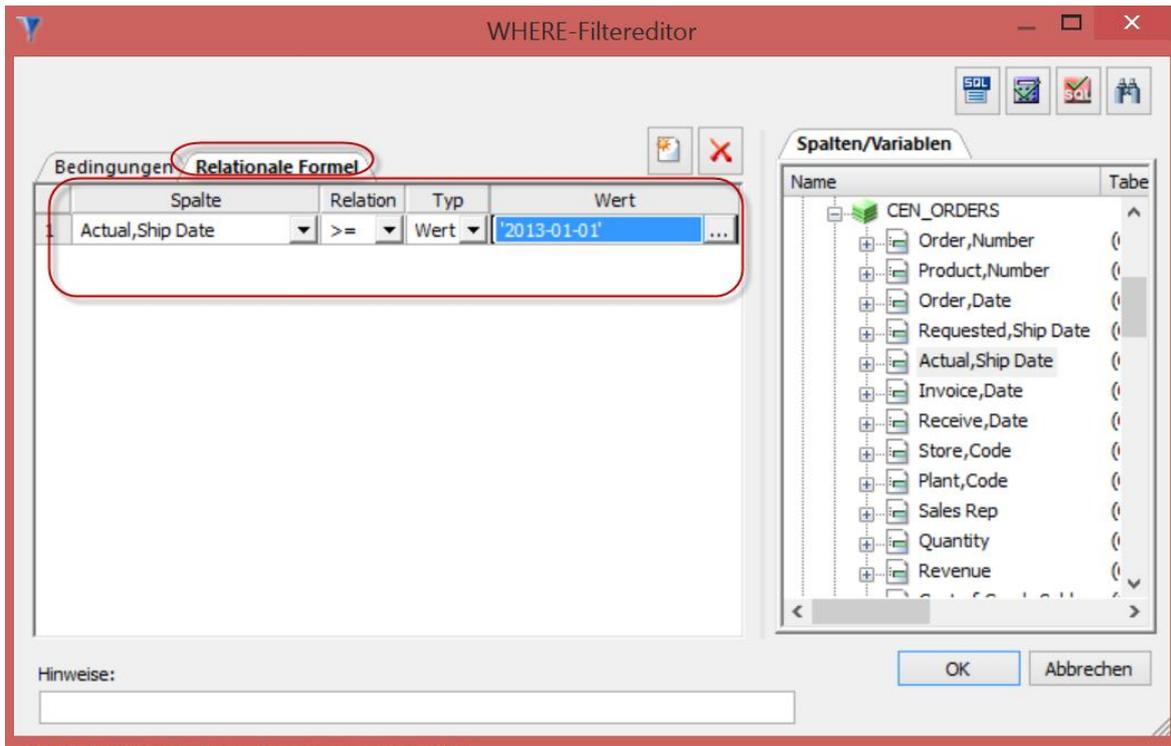


Abbildung 22: Relationale Formel im WHERE-Filtereditor

Damit ist gewährleistet, dass nur Daten aus der Tabelle ORDERS übertragen werden, deren tatsächliches Auslieferungsdatum nach dem 31.12.2012 liegt.

Hinweis:

Über den Fortsetzungs-Button können Sie sich die verfügbaren Werte der ausgewählten Spalte anzeigen lassen (siehe nachfolgende Abbildung).

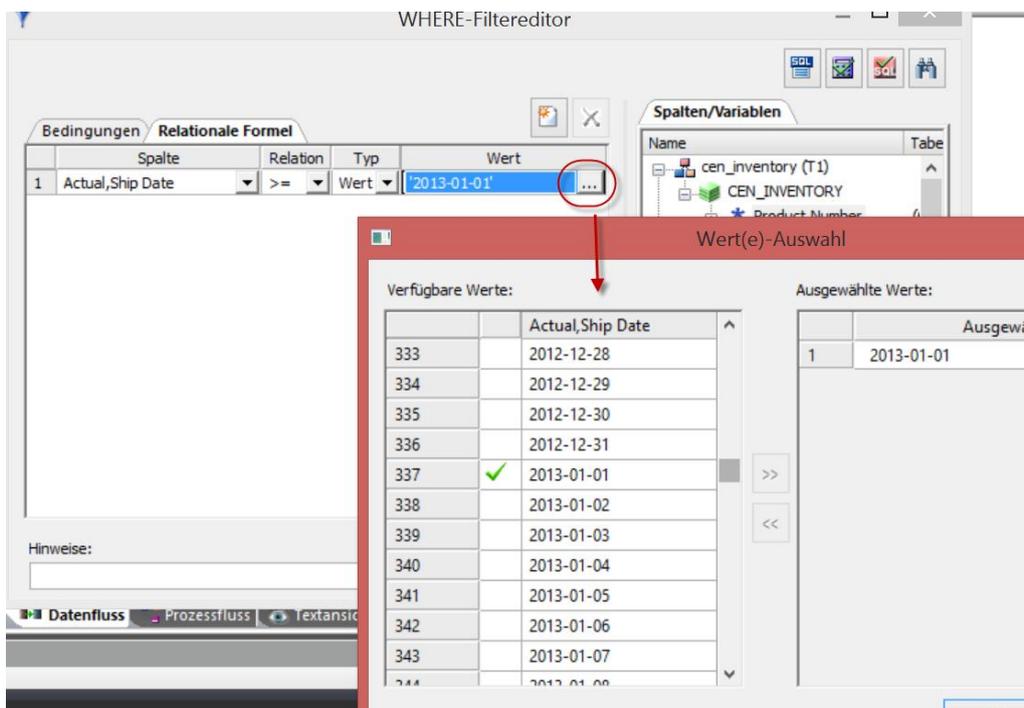


Abbildung 23: Anzeige der verfügbaren Werte der Spalte Actual Ship Date

Hinweis:

Wenn Sie den Tab von *Relationale Formel* in *Bedingungen* wechseln, erhalten Sie die folgende Darstellung:

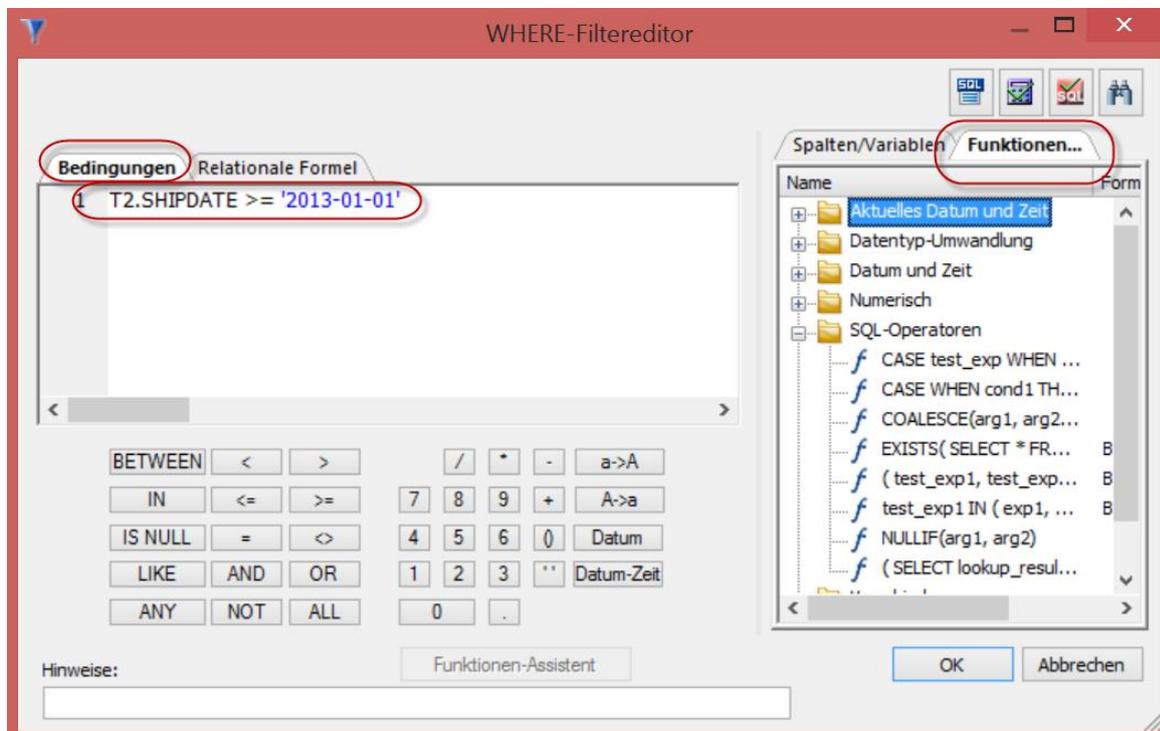


Abbildung 24: Bedingungen des WHERE-Filtereditors

Hier stehen Ihnen auch zahlreiche Funktionen über den Tab *Funktionen* zur Verfügung.

Jetzt lassen wir uns noch die inzwischen aktuell erzeugt SQL-Anweisung anzeigen. Dazu wählen wir aus dem Kontextmenü von SQL *SQL-Anweisung*.

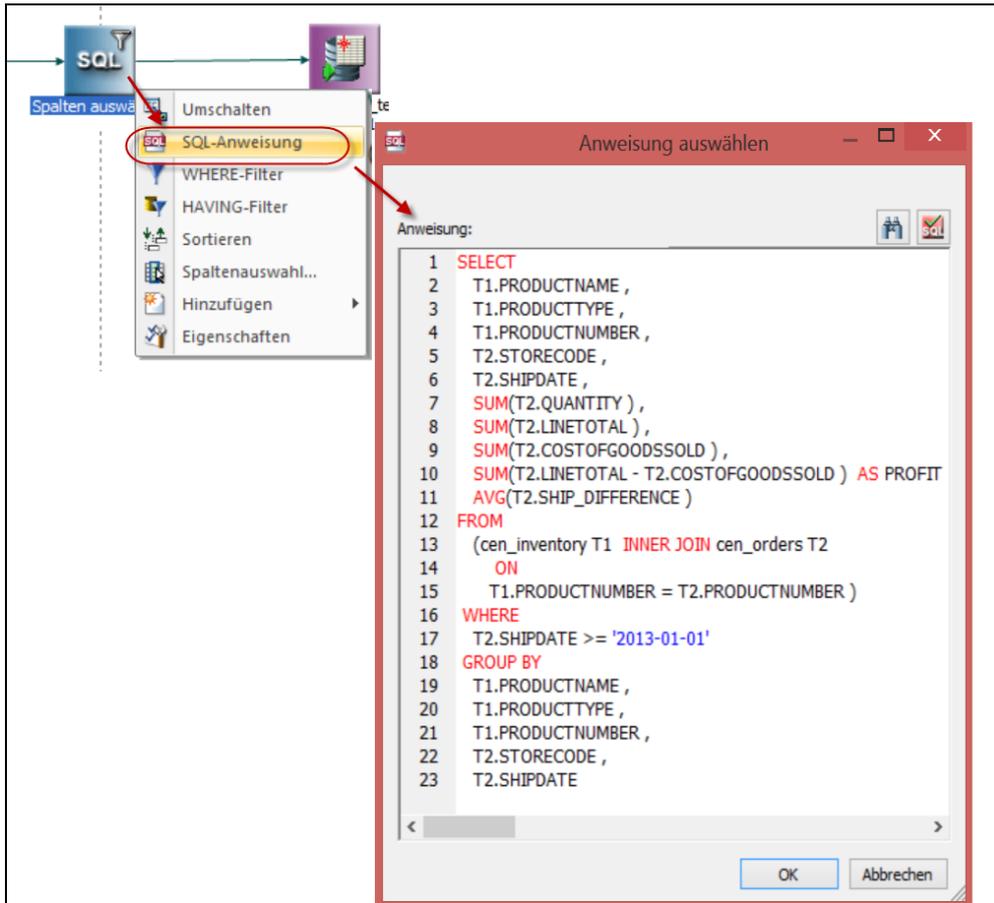


Abbildung 25: Anzeigen der generierten SQL-Anweisung

Mit der Auswahl *SQL-Anweisung überprüfen* werden Beispieldaten für die erzeugte SQL-Anweisung erstellt:

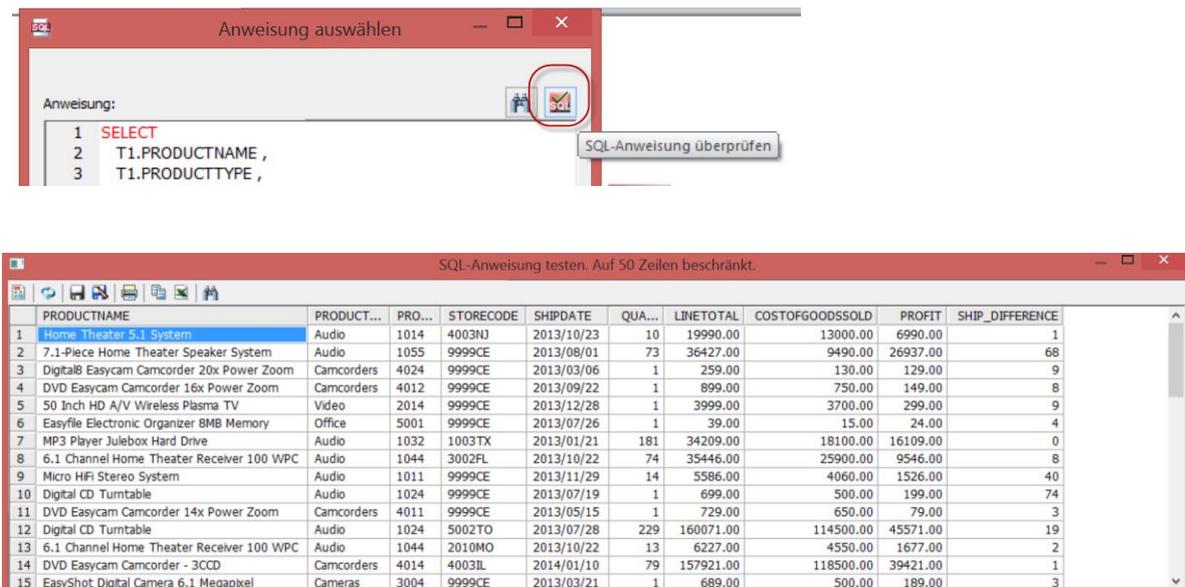


Abbildung 26: Testergebnis für die SQL-Anweisung

Vergessen Sie bitte nicht, den Stand des Flusses *flow2* zu sichern.

Vorschau:

In den nächsten Artikeln werden die weiteren Funktionalitäten vorgestellt und näher beleuchtet:

- Datenziele hinzufügen
- Eine Validierung hinzufügen
- Ein zweites Datenziel hinzufügen
- Fluss ausführen

Bis dahin wünsche ich Ihnen weiterhin viel Spaß beim Vermehren Ihrer Fertigkeiten.

Den Autor Theo Bär erreichen Sie unter EDV-Beratung Theo Bär - Ringmauerweg 1 - 69250 Schönau -
Tel. (+49) 6228 912 630 - e-Mail info@edv-baer.com