

IBM DB2 Web Query 2.1 – Data Migrator Teil 3

Seit März 2015 ist das neueste Mitglied der DB2 Web Query-Familie verfügbar: *IBM DB2 Web Query for i DataMigrator ETL Extension*. Es bietet Möglichkeiten zum Extrahieren, Transformieren und Laden von Daten. Damit können auch Prozesse wie das Konsolidieren, Bereinigen, Übertragen und Umsetzen von Daten für Reporting-Aufgaben zeitgesteuert definiert und angestoßen werden.

Eine grundlegende Funktionalität von DataMigrator for i ist die Erzeugung von Datenflüssen. In diesem Artikel zeige ich Ihnen die einzelnen Schritte, die zum Erstellen eines Datenflusses erforderlich sind

Im Folgenden erläutere ich Ihnen die einzelnen Schritte, die zum Erstellen und Testen eines Datenflusses erforderlich sind:

- DataMigrator-Server erstellen
- Synonyme definieren
- **Fluss erstellen**
- **Datenziel hinzufügen**
- **Fluss ausführen**
- **Fluss sichern**

Die beiden ersten Schritte habe ich Ihnen im letzten Artikel beschrieben. Hier folgen nun die noch ausstehenden Schritte.

Fluss erstellen

Mit einem Rechtsklick auf den bereits erstellten Ordner *datamigrator_test* (=Kontextmenü) und der Auswahl *Neu* → *Fluss* kann ein neuer Fluss erstellt werden.

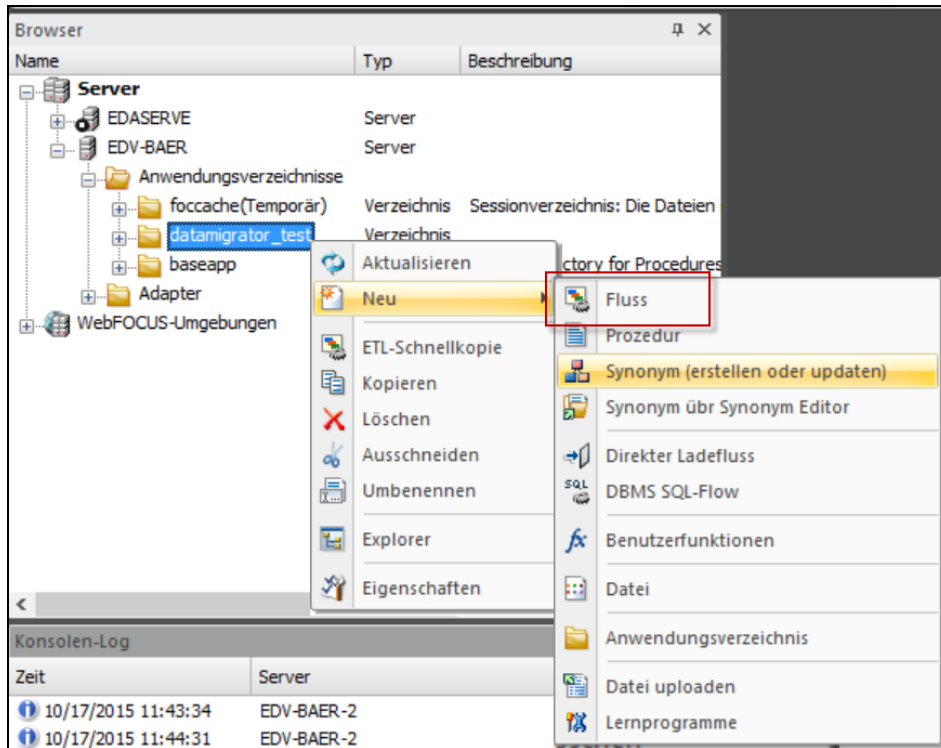


Abbildung 1: Auswahl zur Erstellung eines neuen Flusses

Falls das Kontextmenü die Auswahl *Fluss* nicht enthält, kann das darauf hinweisen, dass während der Installation etwas nicht korrekt abgelaufen ist. Entweder ist die Option 8 des Produktes nicht installiert, wurde das aktuelle Hotfix nicht aktiviert, wurde Web Query nicht erneut gestartet oder der in DMC eingeloggte Benutzer hat keine Berechtigung für Developer Workbench.

Nach Überprüfung und Klärung der durchzuführenden Tätigkeiten sollte im DMC-Arbeitsbereich der Daten-Fluss in einem Daten-Fluss-Tab mit dem SQL-Icon auf der Palette angezeigt werden

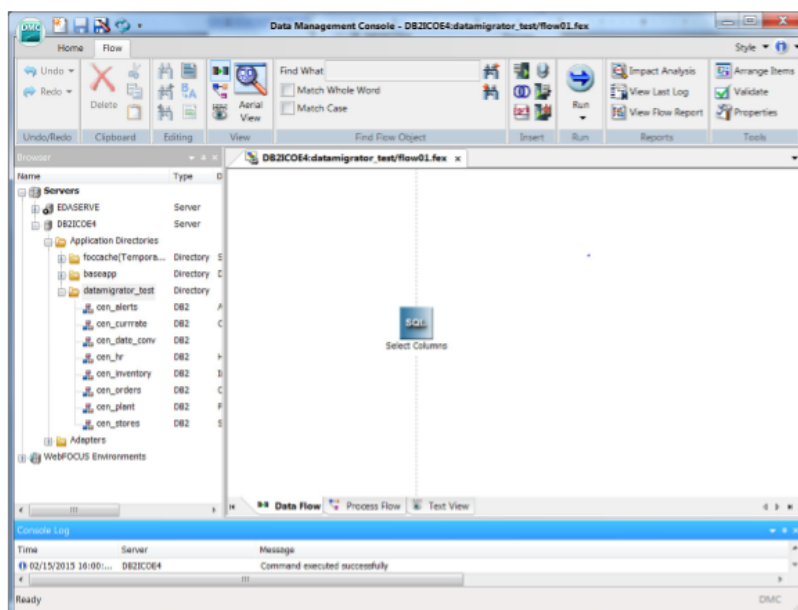


Abbildung 2: Neue Datenfluss-Palette

An dieser Stelle werden die Datenquellen als Navigationsbaum auf der linken Seite und der Datenfluss auf der rechten Seite angezeigt.

Hinweis:

Falls die Datenquellen nicht angezeigt werden, dann müssen Sie die Navigationsordner durch einen Links-Klick auf das Plus-Zeichen (+) erweitern.

Das SQL-Icon und die durch das Icon geführte vertikale gestrichelte Linie stellen den Hauptbereich des Datenflusses dar. Objekte, die sich links von der Linie befinden, sind Datenquellen. Objekte auf der rechten Seite sind Datenziele.

Für die Erstellung eines Datenflusses ziehe ich aus dem Navigationsbereich eine Datenquelle in den linken Bereich der Datenfluß-Palette. In meinem Beispiel verwende ich die Datenquelle *cen_inventory*.

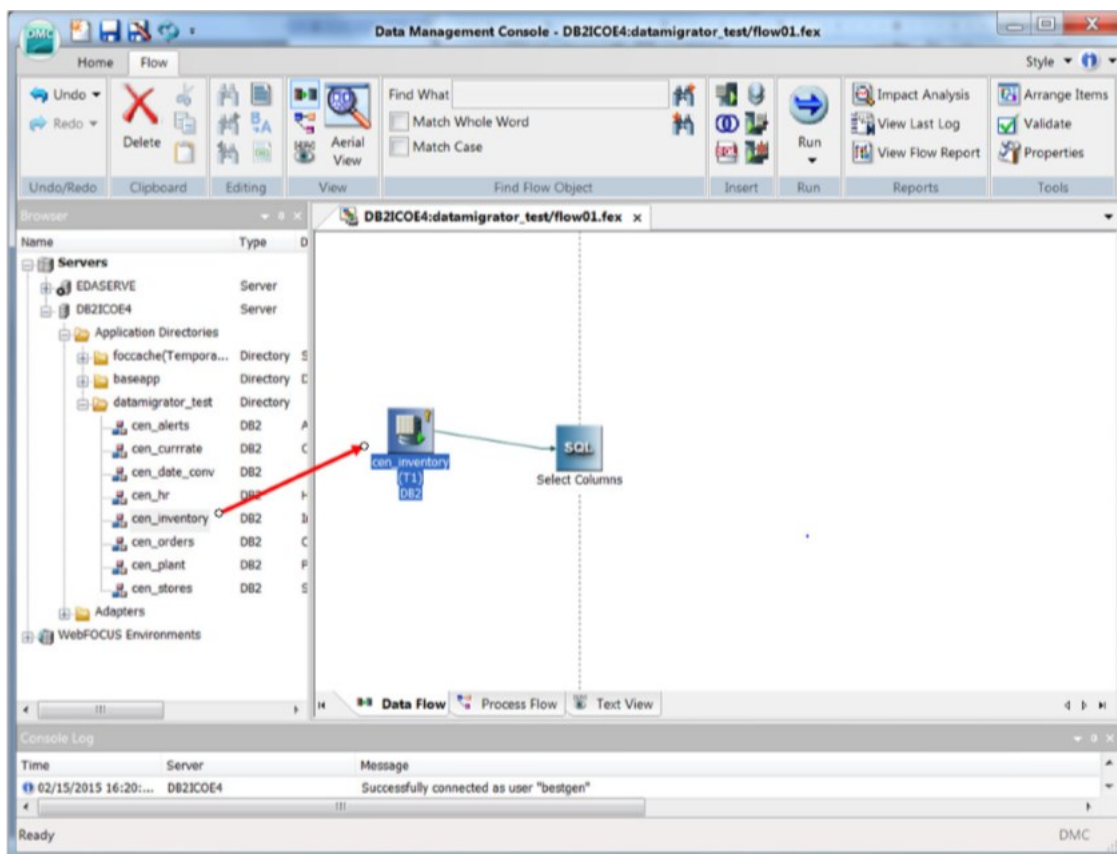


Abbildung 3: Auswahl der Datenquelle cen_inventory

Sobald die Datenquelle (oder das Synonym) auf der Palette abgelegt wird, verbindet es die DMC mit dem SQL-Icon in Form eines Pfeiles. Damit wird angezeigt, dass der Datenfluss von der Datenquelle in die SQL-Operation erfolgt.

Hinweis:

Falls Sie das Synonym versehentlich auf die rechte Seite des SQL-Icons ziehen, wird dieses als Datenziel interpretiert. Dies wird wiederum in Form eines Pfeiles dargestellt. Das würde bedeuten, dass die Daten des Datenflusses in das angegebene Ziel übertragen werden. Wie Sie sehen, ist es von großer Bedeutung, auf welche Seite die Datenquelle platziert wird.

Alternativ zu der oben beschriebenen Art und Weise per Drag-and Drop können Sie auch mit einem Rechtsklick in die linke Seite der Palette oder auf das SQL-Icon das Kontextmenü einblenden und daraus auswählen: *Hinzufügen* → *Quelle* und danach das Synonym für die Datenquelle auswählen.

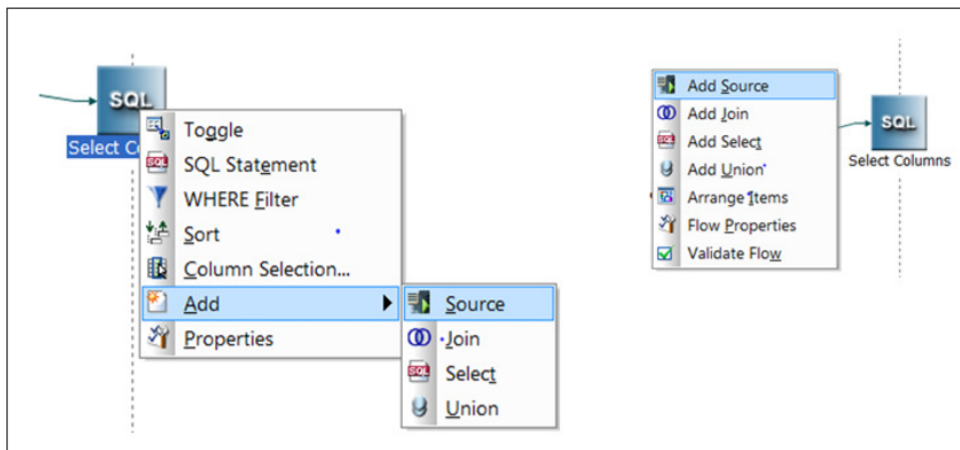


Abbildung 4: Alternative Methoden für die Auswahl einer Datenquelle

An jeder Stelle der Entwicklung eines Datenflusses können Sie die Beispieldaten einblenden um zu prüfen, ob dies die korrekte Auswahl wie von Ihnen erwartet ist. Mit einem Rechtsklick auf die Datenquelle und der Auswahl *Operationen* → *Beispieldaten* wird ein neuer Tab mit den Beispieldaten der ausgewählten Datenquelle erzeugt. Zum Schließen des Tabs klicken Sie auf das X, wie in Abbildung 6 gezeigt.

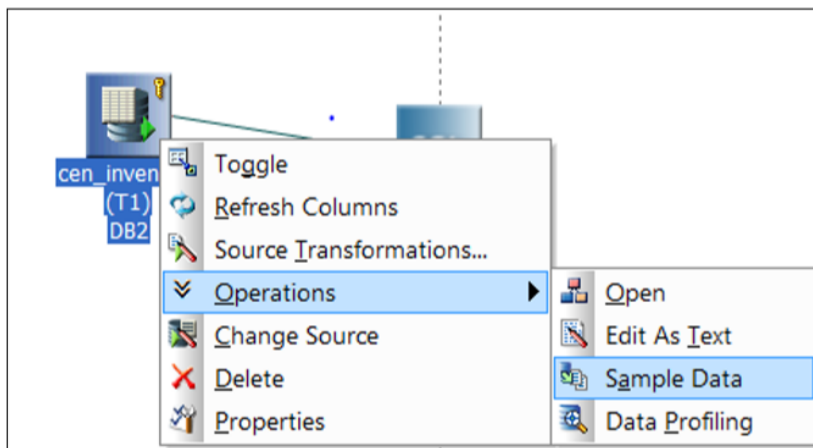


Abbildung 5: Beispieldaten für eine Datenquelle

Product Number	Product Type	Product Category	Product Name	Model	Quantity In Stock	Price	Cost
1	1001	Audio	Power Amplifier	PA-100	1068	249.00	180.00
2	1002	Audio	PA4000 Stereo & Surround Power Amplifier	PA-2000L	1527	299.00	220.00
3	1003	Audio	Modular Components Series Preamp 5.1	PA-MCS1	989	399.00	330.00
4	1004	Audio	PreAmp/Tuner Two	PT-1500	1758	499.00	250.00
5	1005	Audio	AM / FM Stereo Tuner	TU-20	1972	199.00	100.00
6	1011	Audio	Micro HiFi Stereo System	MS-H100	990	399.00	290.00
7	1012	Audio	Micro 5.1 System	MS-H200	707	499.00	400.00
8	1013	Audio	Home Theater Surround System	HT-1000S	573	999.00	750.00
9	1014	Audio	Home Theater 5.1 System	HT-2000S	961	1999.00	1300.00
10	1015	Audio	Home Theater 7.1 THX System	HT-3000S	400	2999.00	2000.00
11	1021	Audio	CD Changer / CD Player	CD-100CP	4000	199.00	120.00
12	1022	Audio	CD Recorder with 50GB Hard Disc Drive	CDN-200	3444	799.00	600.00
13	1023	Audio	400 Disc Super Audio CD Changer	CD-400C	7499	899.00	700.00
14	1024	Audio	Digital CD Turntable	CD-500DT	1020	699.00	500.00
15	1025	Audio	Multichannel Super Audio CD Player	CD-505A	1990	1999.00	1300.00
16	1031	Audio	MP3 Player	MP-20	7050	129.00	60.00
17	1032	Audio	MP3 Player Jukebox Hard Drive	MP-20H	9480	189.00	100.00
18	1033	Audio	MP3 Digital Audio Computer	MP-10C	3068	279.00	150.00
19	1034	Audio	MP3 Digital Audio Computer 4GB	MP-20G	3527	329.00	250.00

Abbildung 6: Anzeige der Beispieldaten

Nun vervollständigen wir unseren einfachen Datenfluss, indem wir ein Datenziel hinzufügen

Datenziel hinzufügen

Zunächst erstelle ich eine neue Bibliothek bzw. ein neues Schema. Diese(s) soll die neu zu erstellende Datentabelle enthalten. Dazu verwende ich den System i Navigator oder STRSQL. Der Name des Schemas soll sein: QWQDMTEST.

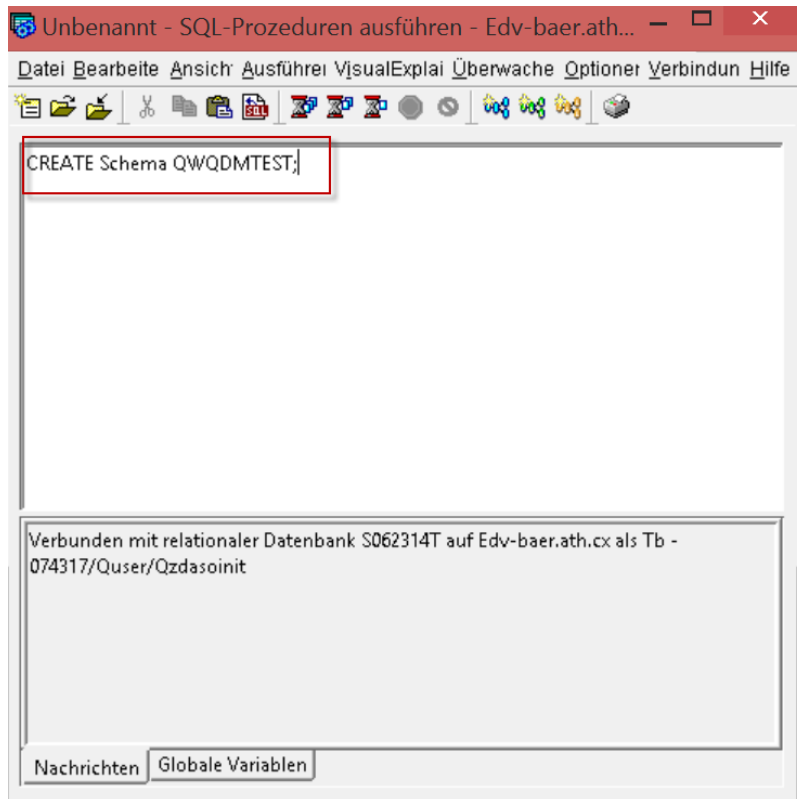


Abbildung 7: Das Schema QWQDMTEST mit dem System i Navigator erstellen

Jetzt kehren wir wieder zurück zur DMC.

Hinweis:

Mit DataMigrator ist es möglich, eine neue Datenziel-Tabelle zu erstellen oder Sätze in eine bestehende Tabelle einzufügen. Abhängig von der beabsichtigten Verwendungsart kann es sich dabei um ein einfaches Kopieren einer Datei oder um eine sehr komplexe Abfolge von Operationen handeln, die als Ergebnis eine hoch technisierte Datenbank als Data Warehouse zur Folge haben.

Das Hinzufügen eines Datenziels kann über einen Rechtsklick in den rechten Bereich der Palette erfolgen. Aus dem dann folgenden Kontextmenü wählen Sie *Ziel hinzufügen* → *Neu*. Alternativ können Sie auch auf das Icon *Neues Ziel* im Menüband klicken.

Sobald das Datenziel im Arbeitsbereich platziert ist, wählen Sie aus dem zugehörigen Kontextmenü *Eigenschaften* aus.

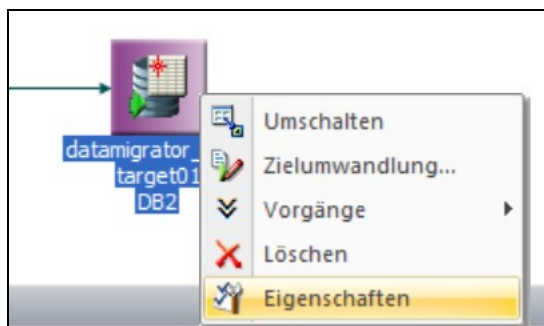


Abbildung 8: Eigenschaften des Datenziels auswählen

Das Eigenschaftsfenster bietet die Möglichkeit, das Erscheinungsbild der dem Datenziel zugrunde liegenden Tabelle zu kontrollieren und festzulegen, wie diese gepflegt wird. Ich möchte mich hier auf drei wesentliche Elemente beschränken:

1. Name des Synonyms
Standardmäßig wird während der Ausführung des Datenflusses ein Synonym mit dem Namen *targetxx* erstellt (wobei xx eine fortlaufende Nummer ist). Dieses Synonym wird im selben Ordner abgelegt, in dem der Datenfluß erstellt wurde. Eine Änderung kann über die Auswahl der Auslassungspunkte rechts neben dem Text erfolgen. In meinem Beispiel belasse ich den Standard.
2. Tabelle
Gibt den Namen der zugrunde liegenden Tabelle und die Lokation an. Standardmäßig wird hierfür derselbe Name wie für das Synonym vorgeschlagen, in unserem Fall *targetxx*. Desweiteren wird keine Bibliothek angegeben. Das würde bedeuten, dass die Tabelle in der ersten beschreibbaren Bibliothek aus dem Benutzerteil der Bibliotheksliste erstellt würde. um das zu vermeiden, verwenden wir die Bibliothek/Datei-Notation der IBM i. In unserem Beispiel ändere ich den Namen in *qwqdmtest/dmtarget01*.
3. Ladetyp
Hiermit wird festgelegt, wie die Sätze aus der Datenquelle gelesen und in die Zieltabelle eingefügt werden sollen. Standard ist Insert/Update. In unserem Beispiel wähle ich *Datensätze vom Speicher einfügen*. Dadurch werden mehrere Sätze zusammengefasst. Damit wird auch eine bessere Performance erzielt.

Attribut	Wert
Allgemeines	
Notizen	
Zieloptionen	
Typ	Neu
Adapter	DB2
Verbindung	*LOCAL
Synonym	datamigrator_test/target01
Tabelle	qwqdmtest/dmtarget01
Schlüssel	
Ziel-Ladeoptionen	
*Ladetyp	Datensätze vom Speicher einfügen
*Commit für jede Zeile	500000
*Blockgröße	1000

Abbildung 9: Angaben für die Eigenschaften des Datenziels qwqdmtest/dmtarget01

Fluss ausführen

Nun können wir den soeben erstellten einfachen Fluß ausführen. Dazu klicken Sie auf die Auswahl *Run* im Menüband.

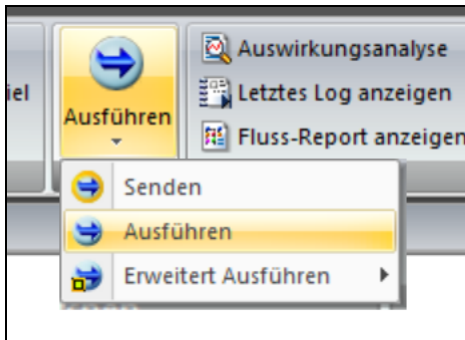


Abbildung 10: Mögliche Optionen für das Ausführen eines Flusses

Genau wie ich sollten Sie auch eine Fehlermeldung erhalten, welche besagt, dass der Fluss nicht vollständig ist, weil keine Spalten ausgewählt wurden:

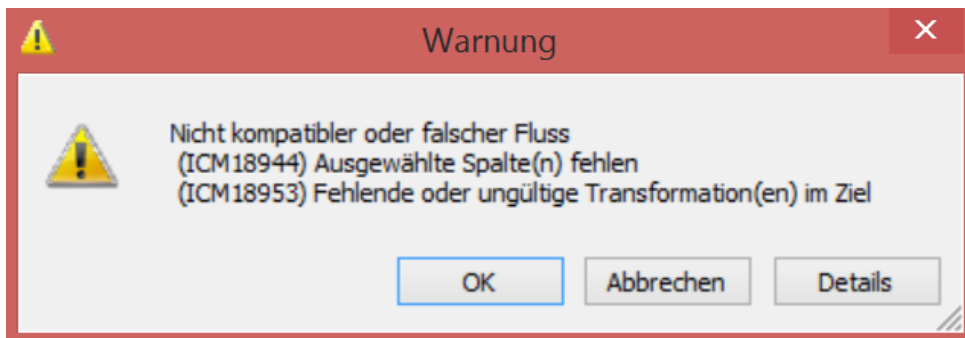


Abbildung 11: Fehlermeldung wegen fehlender Spalten

Diese Warnung ist auch ein Hinweis dafür, dass der DataMigrator nicht einfach ein Tool zum Kopieren von Dateien ist. Weil es ein ETL-Produkt ist und nicht ein einfaches Kopier-Tool erwartet DataMigrator, dass bestimmte Umformungen und Selektionen Bestandteil des Datenflusses sind. Dafür haben wir bisher eine Quelle und ein Ziel ausgewählt, aber noch keine Spalten oder Daten-Transformationen. DataMigrator kann zwar auch für einfaches Kopieren von Dateien verwendet werden, aber es kann weitaus mehr als dies.

Das SQL-Icon ist ein Kennzeichen für die Überleitung von der Quelle zum Ziel. Hier befindet sich der Hauptteil, in dem Zuordnung und Umsetzung definiert werden. Dies werden wir nun anwenden.

Hierzu gehen wir zurück zu unserer Flußdefinition und fügen eine Liste von Spalten hinzu, die von der Quelle in das Ziel übertragen werden sollen.

Dazu aktivieren wir das Kontextmenü des SQL-Icons und wählen *Spaltenauswahl*.

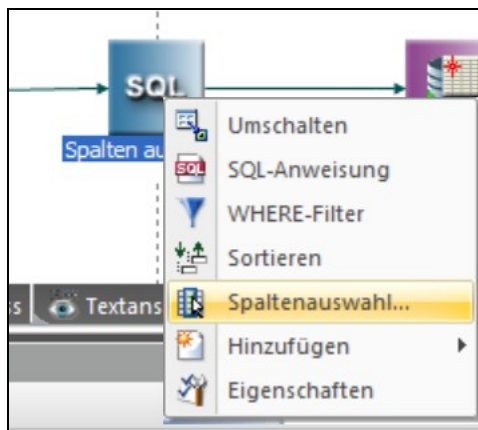


Abbildung 12: Spaltenauswahl treffen

Das Fenster für die Spaltenauswahl erscheint. Auf der linken Seite befinden sich die potenziellen Spalten, die aus der Datenquelle verfügbar sind. Ich markiere alle Spalten und wähle *Spalten nach rechts verschieben*:

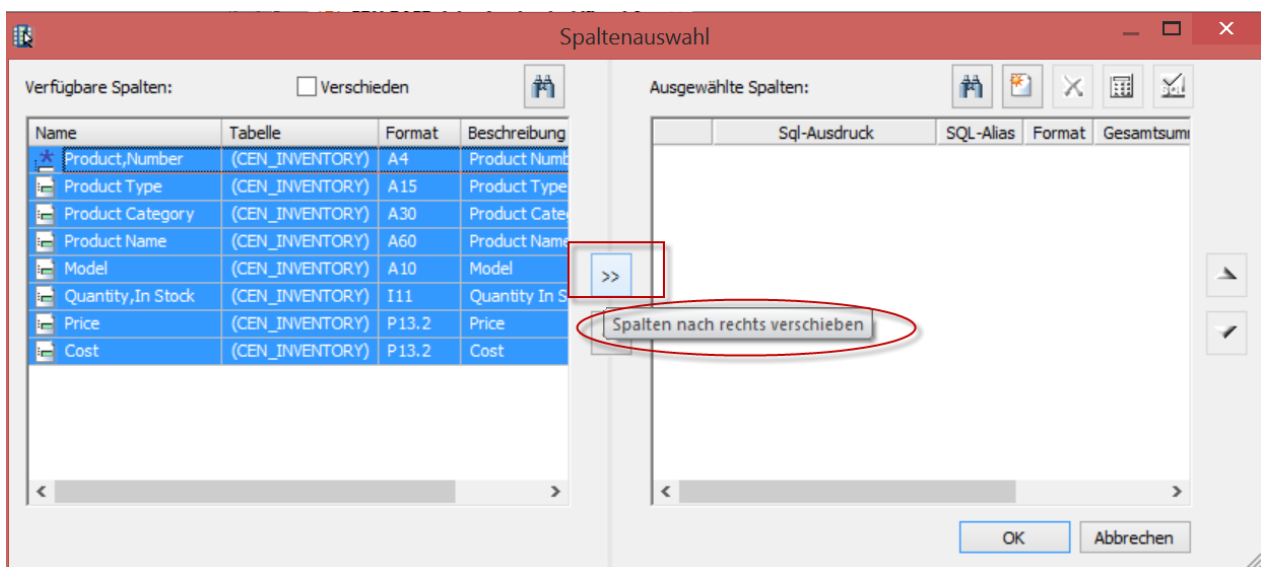


Abbildung 13: Auswahl der zu übertragenden Spalten

Natürlich ist es auch möglich, einzelne Spalten auszuwählen.

Bevor wir den Datenfluss-Prozeß weiter bearbeiten, möchte ich Sie noch auf einige interessante Funktionalitäten des Spaltenauswahl-Fenster aufmerksam machen (siehe Abbildung 14):

- 1) Mit der Auswahl *Verschieden* können bei Ausführung des Datenflusses doppelte Spalten unterdrückt werden (Distinct)
- 2) Mit der Auswahl *Spalten einfügen* können Rechenfelder als neue Spalte erstellt werden.
- 3) Die Auswahl *SQL* überprüft die Spaltenauswahl auf Korrektheit und zeigt die ersten 50 Zeilen gemäß der Auswahl an.
- 4) Die Reihenfolge der ausgewählten Spalten kann mit den Pfeilen nach oben/unten geändert werden.

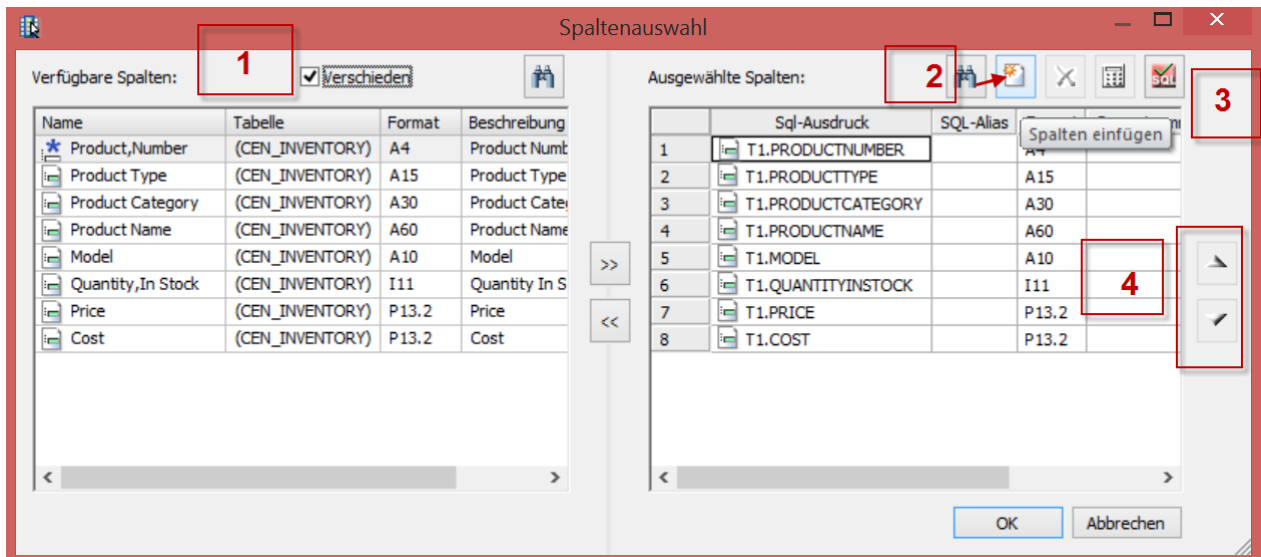


Abbildung 14: Weitere Funktionalitäten des Fensters Spaltenauswahl

Weitere komplexe Operationen wie das Ermitteln von Ausdrücken und Aggregationen (Group By) können auch durchgeführt werden. Diese werde ich Ihnen im nächsten Artikel darstellen.

Nachdem alle Spalten selektiert wurden, wird die Spaltenauswahl mit OK bestätigt.

Wenn wir jetzt erneut *Ausführen* wählen, sollte der Datenfluss erfolgreich ablaufen. Sie können das verifizieren, indem Sie den Konsolen-Log im unteren Bereich der DMC prüfen:

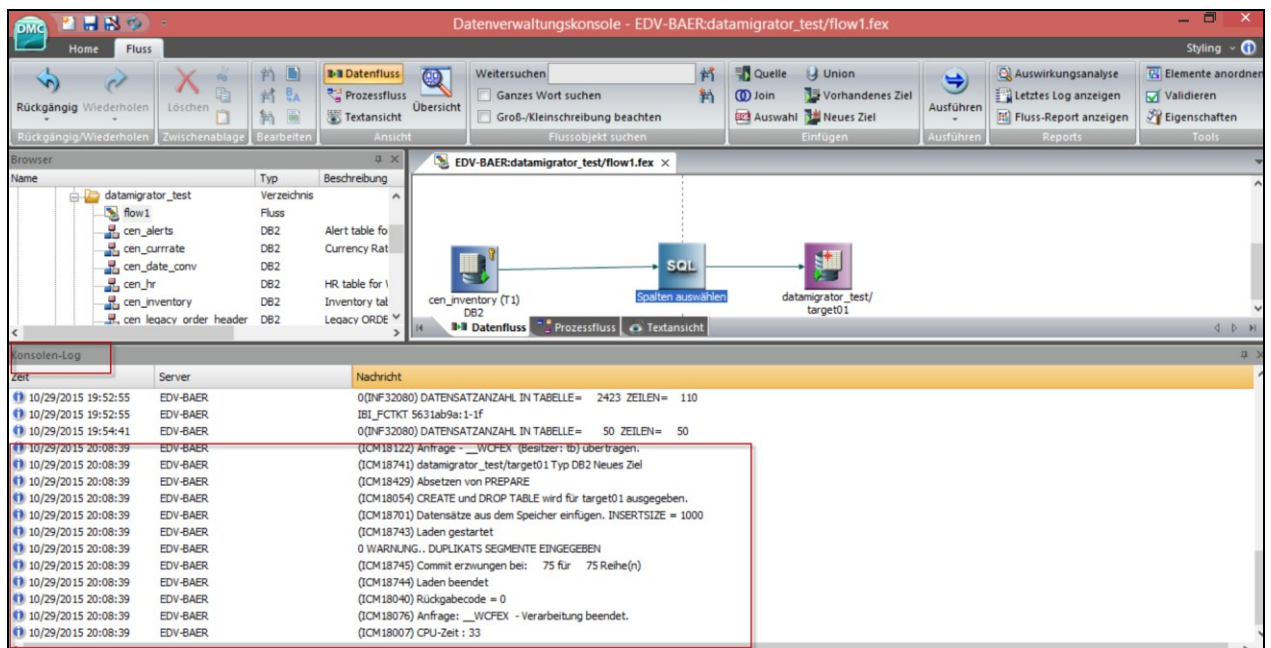


Abbildung 15: Konsolen-Log mit erfolgreicher Ausführung des Datenflusses

Nun können wir noch prüfen, ob die Zieltabelle in der angegebenen Bibliothek tatsächlich erstellt und korrekt gefüllt wurde. Dies kann auf drei verschiedene Arten durchgeführt werden:

1. Prüfen Sie in der Bibliothek QWQDMTEST, ob die Tabelle DMTARGET01 erstellt und gefüllt wurde.
2. Verwenden Sie eine SQL-Anweisung außerhalb von DMC:
SELECT * FROM QWQDMTEST.DMTARGET01
3. Verwenden Sie die Option *Vorgänge* → *Beispieldaten* des Kontextmenüs des Datenziel-Icons

The screenshot shows the IBM Data Migration Center (DMC) interface. At the top, a diagram shows an 'SQL' object connected to a 'datamigrator_target01_DB2' object. A context menu is open over the 'datamigrator_target01_DB2' object, with the 'Beispieldaten' option selected. Below the menu, a table displays sample data for audio products. The table has columns: Product Number, Product Name, Product Category, Product Name, Model, Quantity, Price, and Cost. The data is limited to 50 rows.

	Product Number	Product...	Product Category	Product Name	Model	Quantity	Price	Cost
1	1001	Audio	Amplifiers/PreAmps/Tun...	Power Amplifier	PA-100	1068	249.00	180.00
2	1002	Audio	Amplifiers/PreAmps/Tun...	PA4000 Stereo & Surround Po...	PA-200XL	1527	299.00	220.00
3	1003	Audio	Amplifiers/PreAmps/Tun...	Modular Components Series Pre...	PA-MC51	989	399.00	330.00
4	1004	Audio	Amplifiers/PreAmps/Tun...	PreAmp/Tuner Two	PT-1500	1758	499.00	250.00
5	1005	Audio	Amplifiers/PreAmps/Tun...	AM / FM Stereo Tuner	TU-20	1972	199.00	100.00
6	1011	Audio	Audio Systems	Micro HiFi Stereo System	MS-H100	990	399.00	290.00
7	1012	Audio	Audio Systems	Micro 5.1 System	MS-H200	707	499.00	400.00
8	1012	Audio	Audio Systems	Home Theater Surround System	HT-1000S	572	999.00	750.00

Abbildung 18: Prüfen des Datenflusses anhand der Beispieldaten

Fluss sichern

Zu guter Letzt sollte der zuvor erstellte Fluss auch gesichert werden. Ich nenne ihn *flow1* und sichere ihn im Ordner *datamigrator_test*:

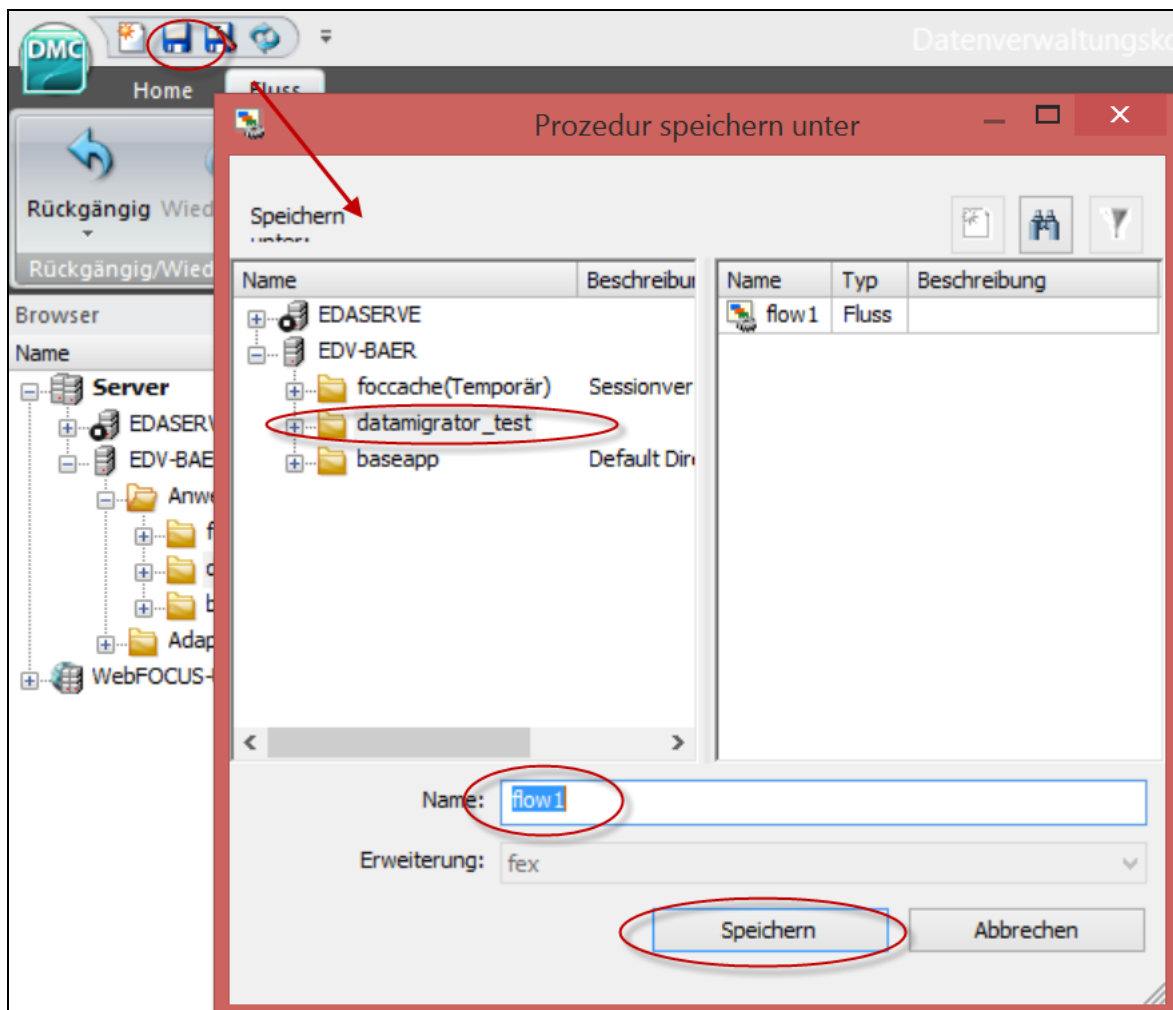


Abbildung 19: Sichern des Flusse unter datamigrator_test/flow1

Vorschau:

Im nächsten Artikel werde ich Ihnen weitere Funktionen des DataMigrator for i von DB2 Web Query vorstellen.

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