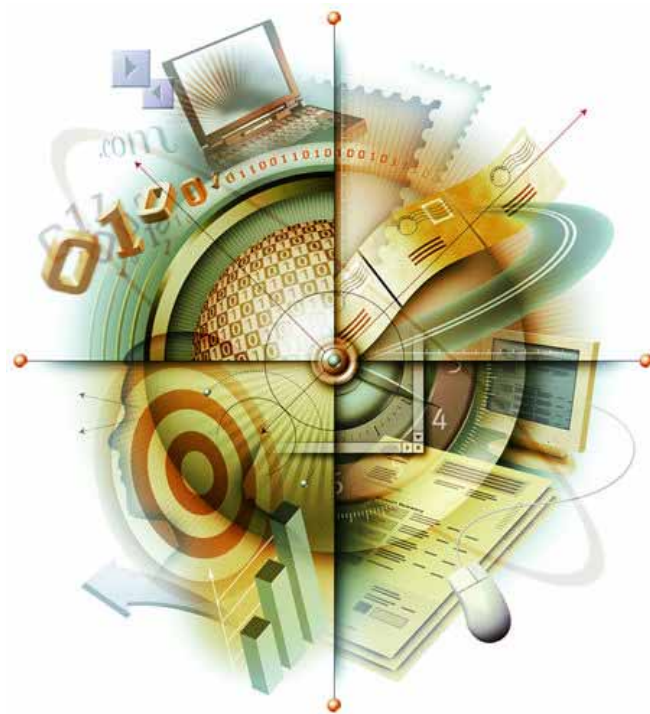


# Einsatz und Anwendung von Subplans

März 2005



Einführung .....	3
Übersichtlichkeit .....	3
Iterative Subplans (Schleifen) .....	5
Batchverarbeitung .....	10
Performanz und Skalierbarkeit.....	13
Zusammenfassung .....	16
Abbildungen .....	17

Dieses Dokument dient dazu die Möglichkeiten, die sich durch die Einführung von Subplans in Sagent ergeben, zu beschreiben.

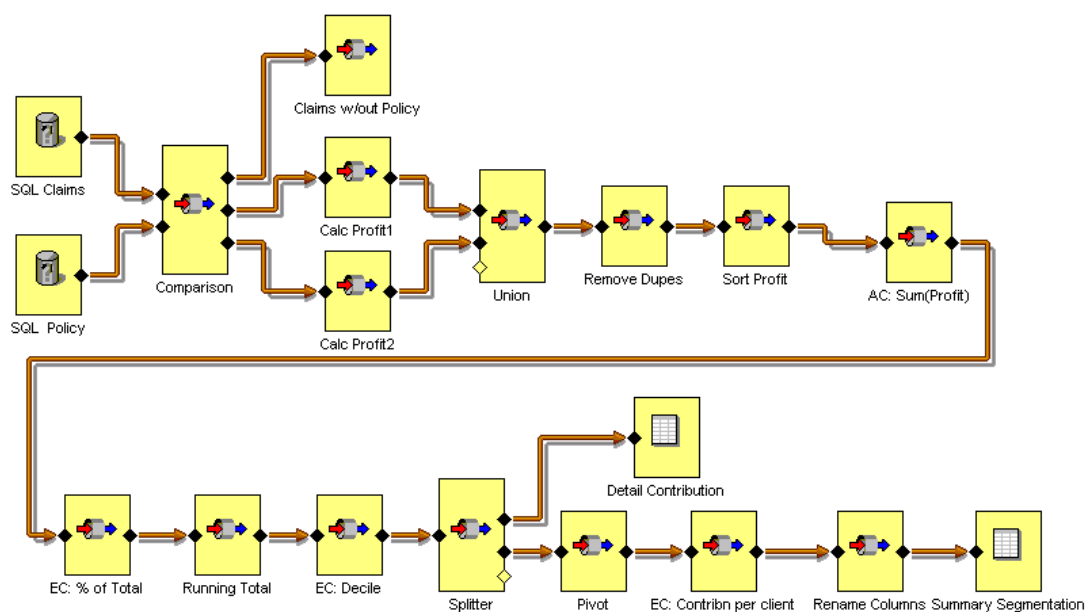
Die Ausführungen erstrecken sich, von der einfachen Art komplexe Pläne zu strukturieren und damit ihre Lesbarkeit zu erhöhen, bis hin zur prozeduralen Steuerung und der dynamischen Planausführung.

## Einführung

Subplans wurden als neue Komponente in der Sagent Version 5 eingeführt. Neben der grundsätzlichen Funktionalität, Pläne besser zu strukturieren, eröffnen sie noch weit größere Möglichkeiten, die hier beispielhaft aufgezeigt werden sollen.

## Übersichtlichkeit

Ein Plan, der auf gewohnte Art und Weise erstellt wurde, kann auf den ersten Blick etwas verwirrend wirken:



**Abbildung 1:** Kundensegmentierungsplan ohne Subplans

Durch Markieren mehrerer Transforms und der Auswahl des Punktes „Move into new Subplan“ wird dieser Plan übersichtlicher gestaltet. So werden durch den Einsatz von Subplans einige Detailschritte in Untergruppen eingeteilt und zu logischen Blöcken zusammengefasst.

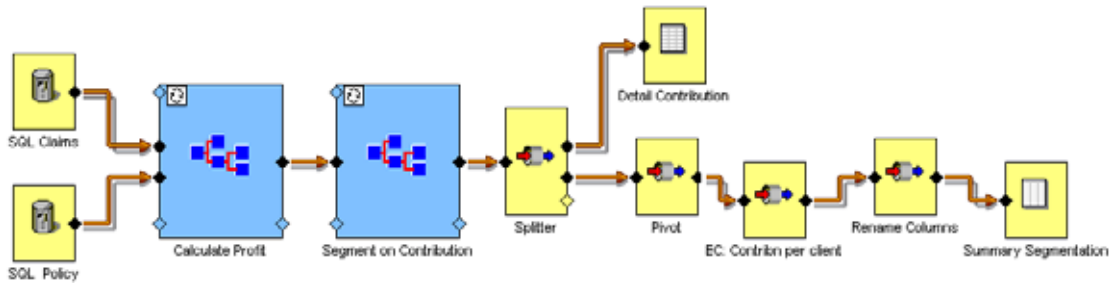


Abbildung 2: Kundensegmentierungsplan mit Subplans

Das macht den Plan einerseits besser lesbar, andererseits ermöglicht es das gleichzeitige Arbeiten an verschiedenen Teilen des Plans – wobei die Schnittstellen bzw. die Übergabepunkte genau definiert sind.

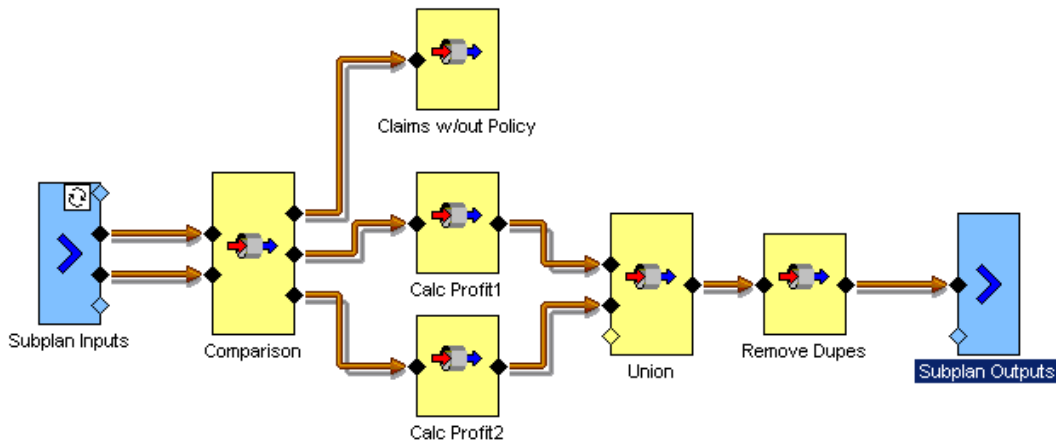


Abbildung 3: Subplan zur Gewinnerrechnung

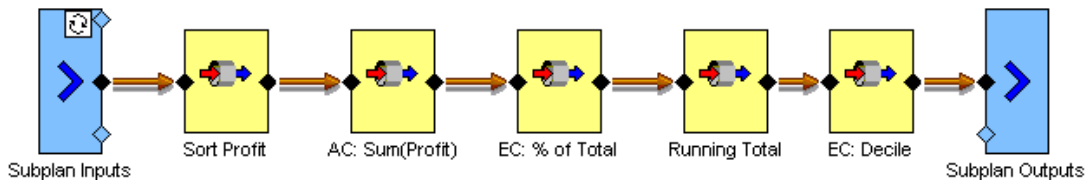


Abbildung 4: Subplan zur Segmentierung

## Iterative Subplans (Schleifen)

Um Pläne dynamischer zu gestalten, greift man auf das Konzept der Iteration zurück. Hat man beispielsweise eine Liste gleich formatierter Flat Files, die eingelesen werden müssen, so kann man das mit Hilfe von Subplans einfach bewerkstelligen.

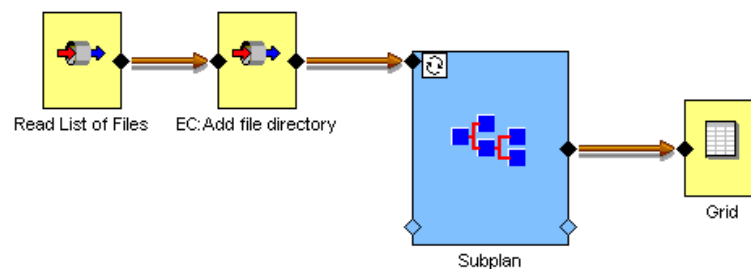


Abbildung 5: Iterativer Plan

Die Logik hinter diesem Plan ist, dass der Transform „Read List of Files“ eine Textdatei mit den Namen der auszulesenden Files verarbeitet. Diese Liste kann zum Beispiel mit dem „DIR Befehl“ unter Windows oder mit dem „ls Befehl“ unter Unix erstellt worden sein.

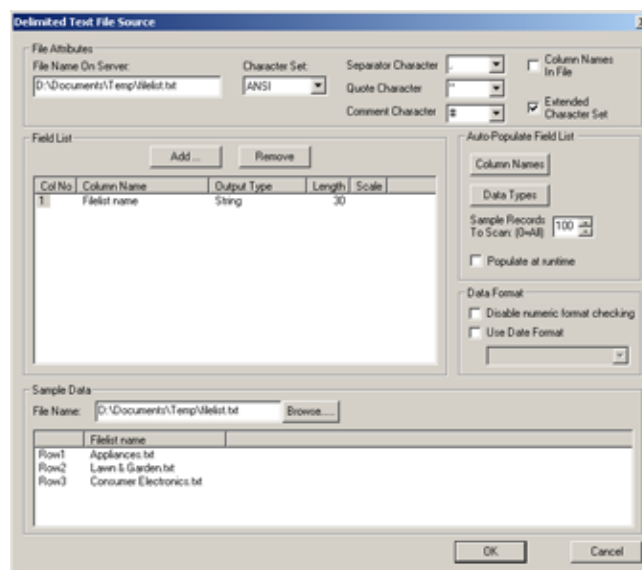


Abbildung 6: Datei mit Dateinamen

Nachdem diese Liste eingelesen wurde, wird mittels eines Expression Calculators noch der Pfadname hinzugefügt.

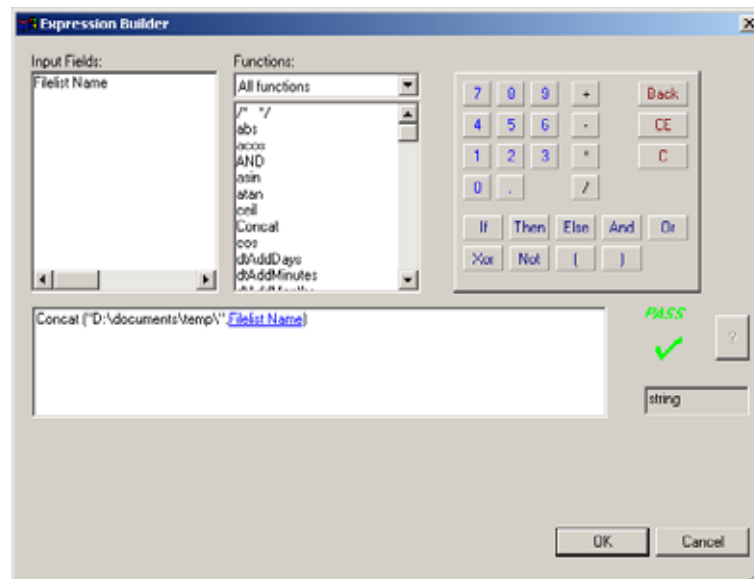


Abbildung 7: Zusammensetzen des Pfades mit dem Expression Calculator

Innerhalb des Subplans befindet sich dann eine File Quelle, welche die eigentlichen Daten liest.

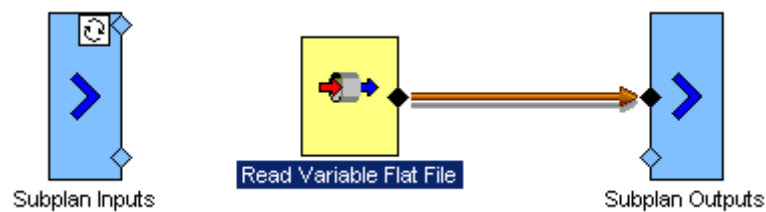


Abbildung 8: Inhalt des Subplans

Wenn man nun im Kontextmenü die Settings des Transforms aufruft, sieht man, dass der File Name (das Server File) zu den offen gelegten Eigenschaften des Transforms gehört.

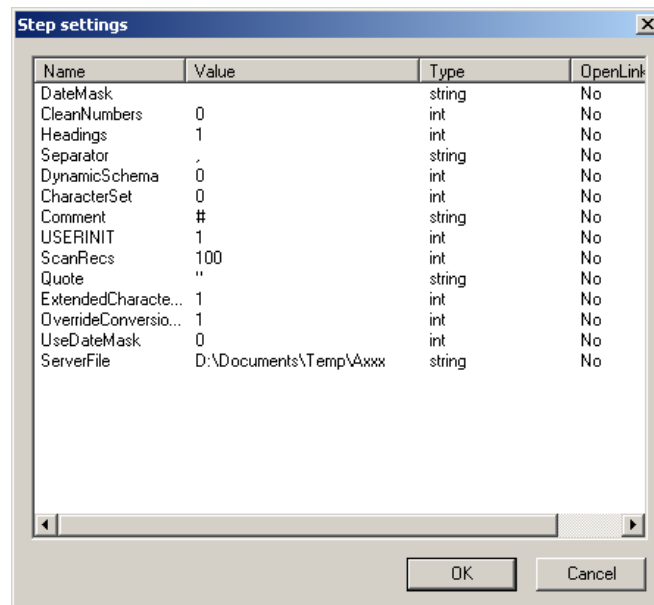


Abbildung 9: Flat File Transform Eigenschaften

Diese und jede andere offen gelegte Eigenschaft kann nun durch die Iterationseinstellung des Subplans verändert werden.

Durch die Auswahl des Punktes Iteration im Kontextmenü (rechte Maustaste) des Subplans kommt man in die Maske zur Festlegung der Iterationseigenschaften.

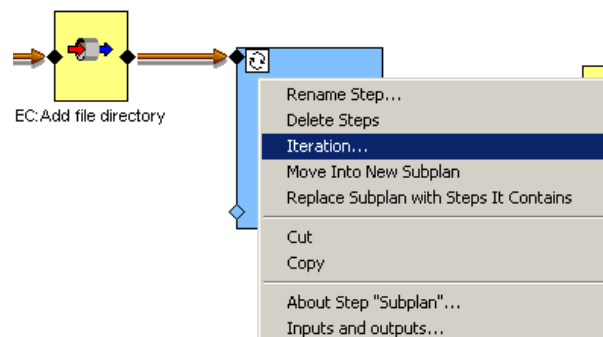


Abbildung 10: Iterationseigenschaften des Subplans wählen

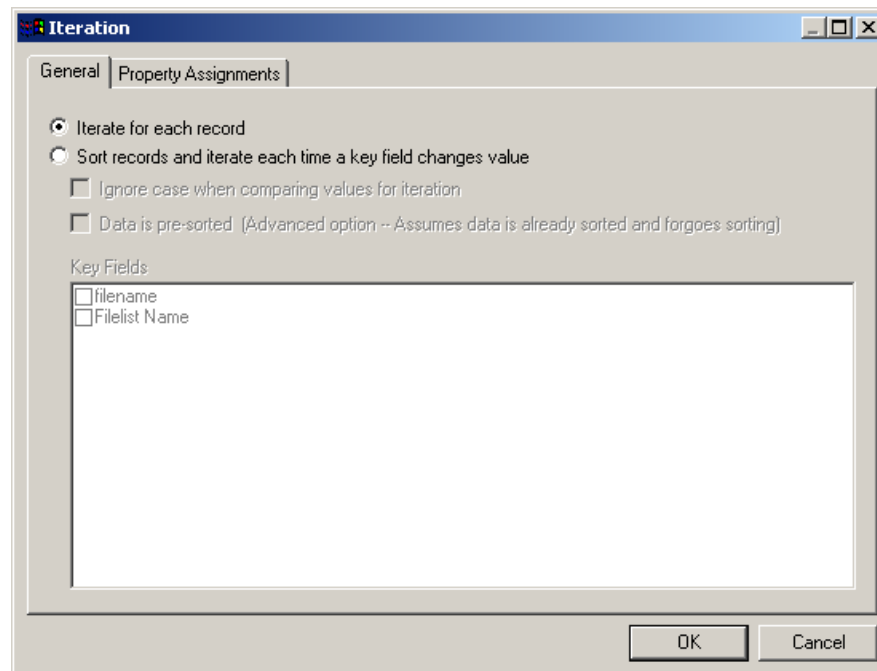


Abbildung 11: Eigenschaften der Iteration

Wenn nun das Property Assignment Tab gewählt wird, hat man die Möglichkeit, den Datenfeldern, die mit dem Iterationseingang des Subplans verbunden sind, Eigenschaften von Transforms innerhalb des Subplans zuzuordnen. Dies bedeutet, dass für jedes Durchlaufen des Subplans den Transforms ein neuer Wert zugeordnet wird. So werden dem Subplan dynamisch Daten übergeben beziehungsweise es wird dynamisch Einfluss auf die Logik des Planes genommen.

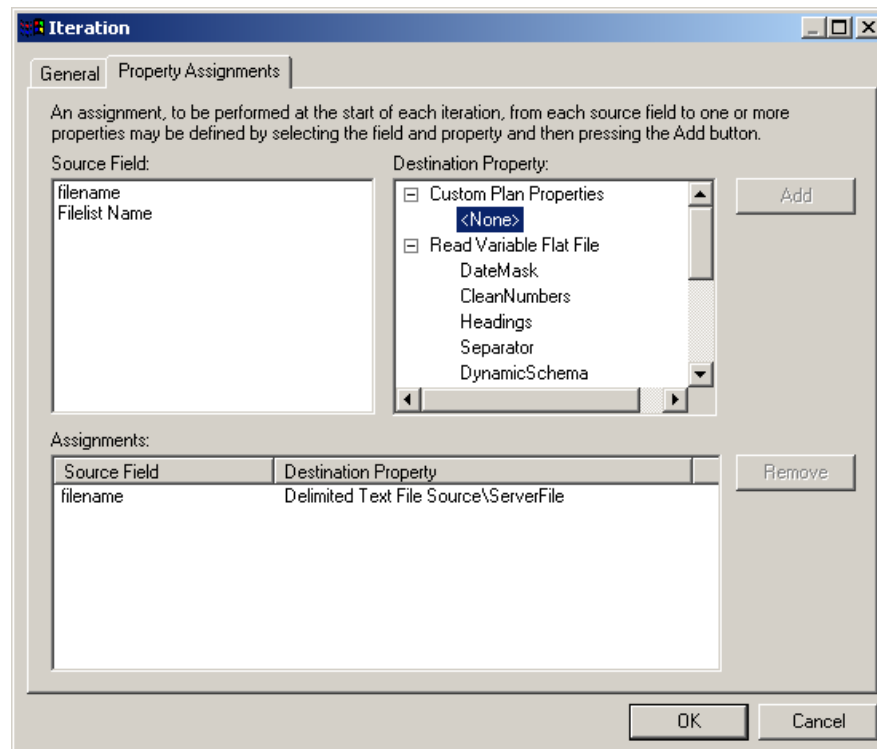


Abbildung 12: Zuordnung der Iterationseigenschaften

Während der Laufzeit liest der Data Flow Service nun die Dateinamen aus und übergibt sie nach einander dem Subplan, der die Dateinamen dynamisch dem in ihm enthaltenen Transform übergibt. Alle Daten werden so in einer Schleife eingelesen und nacheinander dem Ausgang des Subplans übergeben. In diesem Fall ist der Ausgang ein „Grid“, so dass die Daten nach Quelle geordnet, angezeigt werden. Das Ergebnis wird schließlich im „Grid“ für alle Eingangs-Dateien angezeigt.

## Batchverarbeitung

Eine andere Möglichkeit den iterativen Eingang von Subplans zu nutzen, ist die Batchverarbeitung.

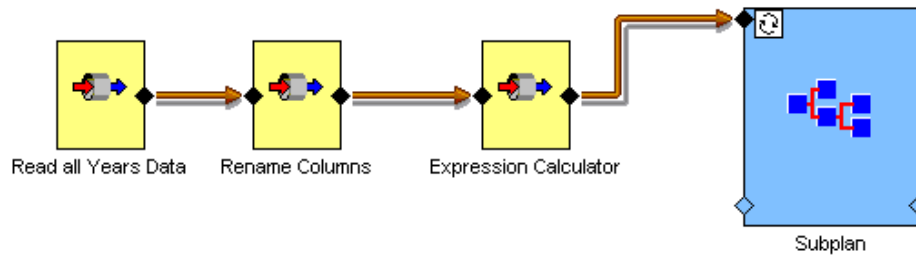


Abbildung 13: Beispiel Batchverarbeitung

In diesem Fall kommen die Daten vorsortiert z.B. aus einer Datenbankabfrage und sollen in verschiedenen Dateien gemäß dem Jahr in dem sie angefallen sind abgelegt werden.

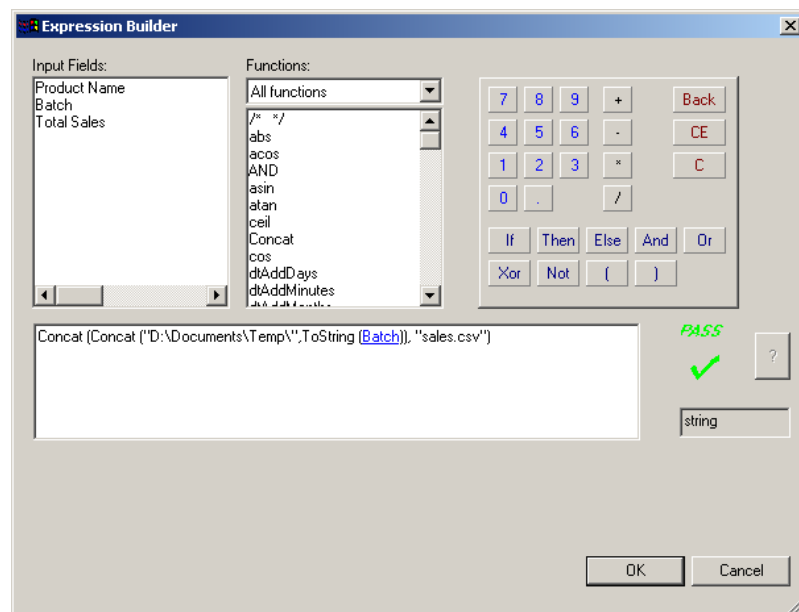


Abbildung 14: Erstellen des Dateinamen

Im Expression Calculator wird der Name des Files in dem die Daten abgelegt werden sollen, zusammengesetzt.

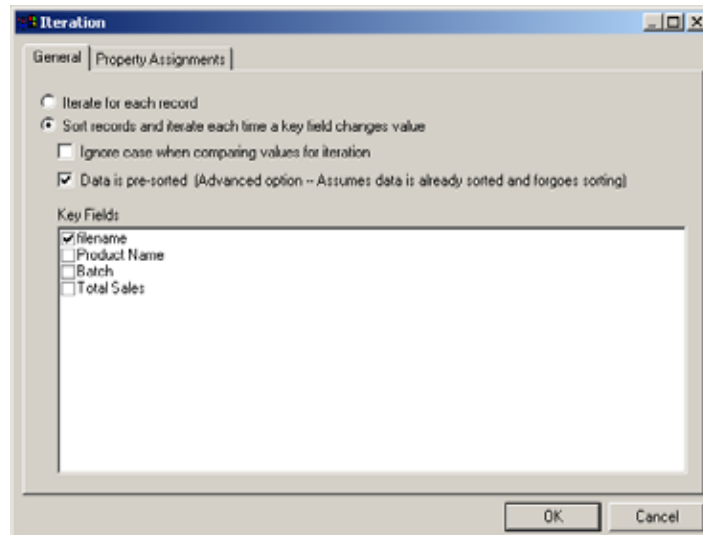


Abbildung 15: Subplan Tab General Properties

Dieser zusammengesetzte Dateiname wird nun dem Subplan übergeben. Im General Tab wird die Iterationsstufe eingestellt. Hier wird der Subplan nicht für jeden Datensatz aufgerufen, sondern nur wenn sich eine Änderung im Dateinamen (in unserem Fall das Jahr) ergibt. Außerdem wird hier eingestellt, ob die Daten vorsortiert geliefert werden oder der Subplan selbst sortiert.

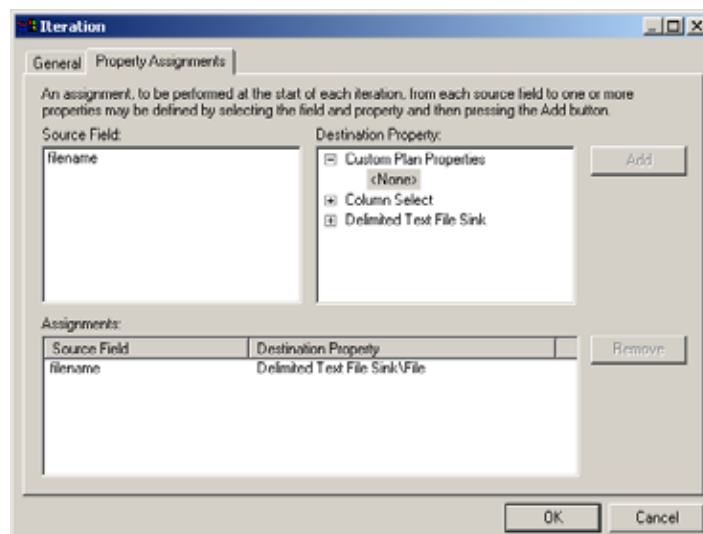


Abbildung 16: Subplan Eigenschaften Zuordnung

Im Property Assignments Tab wird nun der Dateiname (filename) der Eigenschaft ‚File‘ des Delimited Text File Sink zugeordnet.

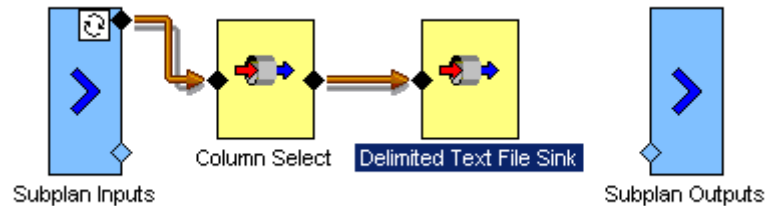


Abbildung 17: Fluss im Subplan

Im Subplan wird schließlich über einen Column Select der Dateiname aus dem Datenfluss herausgefiltert und die Daten in den jeweiligen Dateien abgelegt.

Product Name	Batch	Total Sales
1 Audio Experience	1996	417349.72
2 Carousel	1996	529966.41
3 Carousel Plus	1996	421900.53
4 Clear Vision	1996	723714.56
5 ColdStar	1996	3166136.17
6 ColdStar II	1996	3216761.26
7 Crystal Vision	1996	1377907.66
8 Deluxe Audio Experience	1996	323033.94
9 Deluxe Film Studio	1996	1112584.11
10 Deluxe Laundry Master	1996	1442626.53
11 Deluxe Redwood	1996	506512.48
12 Deluxe Satellite Dish	1996	310159.08
13 Diamond Vision	1996	826670.98
14 Family BBQ	1996	564334.72
15 Family Seating	1996	714700.00
16 Family Theatre	1996	836051.50
17 Fancy Oak	1996	1149664.29
18 Film Studio	1996	979099.48
19 Flame	1996	501529.77
20 Flame II	1996	627500.70
21 Gourmet	1996	1540988.73
22 Gourmet 2000	1996	1570919.36
23 Intimate Seating	1996	686264.61
24 Komplexe	1996	736093.15
25 Laundry Master	1996	1160203.52
26 Laundry Master 2000	1996	1237993.28
27 Mini BBQ	1996	1000700.33

Abbildung 18: Ergebnis der Batchverarbeitung

Als Ergebnis erhält man mehrere nach Jahren sortierte Dateien, die von *einem* Plan mit *einem* Planaufruf erstellt wurden. Ein weiteres Beispiel ist die Erzeugung von Dateien für verschiedene Abteilungen oder Produktgruppen.

## Performanz und Skalierbarkeit

Um zu verstehen, wie man mit Subplans die Performanz und Skalierbarkeit eines Plans verbessert, muss man zuerst verstehen in welchen Teilen des Planes es zu Engpässen kommen kann.

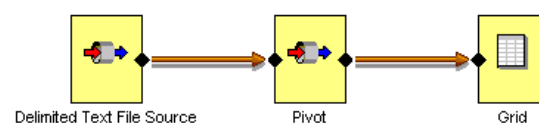
Es gibt, grob gesehen, zwei Typen von Transforms: Blocking und Non-Blocking Transforms. Non-Blocking Transforms arbeiten einfach, indem sie einen Datensatz manipulieren, weitergeben und danach den nächsten Datensatz bearbeiten. Blocking Transforms hingegen müssen erst alle Daten sammeln bevor sie einen Datensatz bearbeiten und weitergeben können. Deshalb konzentriert man sich im Tuning-Prozess oft auf diese Blocking Transforms:

- Analytical Calculator
- Comparison
- Disk Sort
- Forecaster
- Join
- Memory Sort
- Percent of Total
- Pivot
- Rank
- Remove Duplicate Rows

Alle diese Transforms müssen Daten sortieren, bevor sie mit der Manipulation anfangen können.

Die Sortierung der Daten erfordert aber desto mehr Zeit, je größer die Datenmenge ist und je mehr Felder sortiert werden müssen.

So zeigt ein einfacher Plan mit einem Pivot unterschiedliche Laufzeiten wenn die Daten vorsortiert bzw. nicht vorsortiert sind.



**Abbildung 19:** Beispiel Performanz

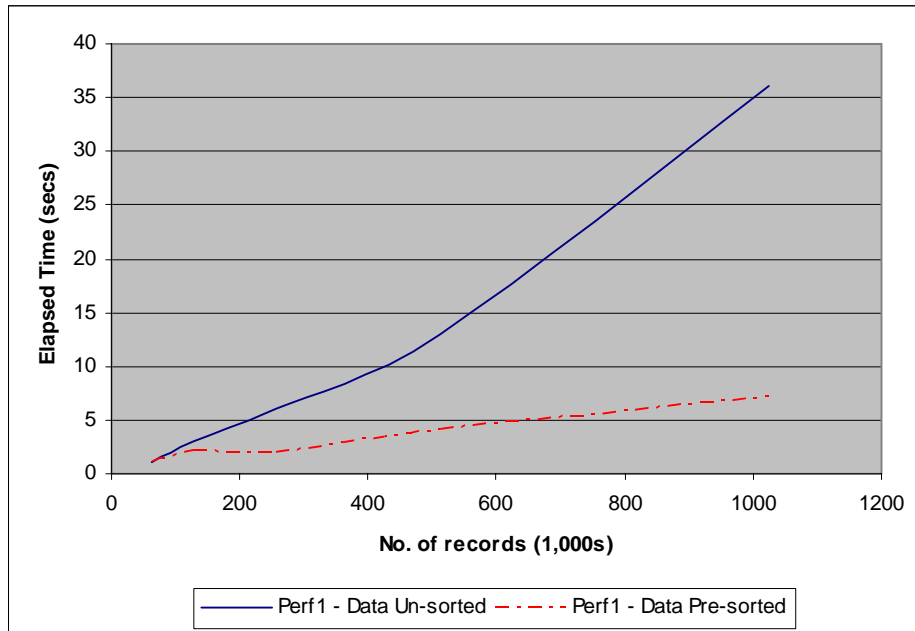


Abbildung 20: Laufzeitdiagramm in Abhängigkeit der Datenmenge

Abhilfe wird geschaffen, wenn der in Frage kommende Transform, in einen Subplan integriert und in mehreren kleinen Schritten ausgeführt wird.

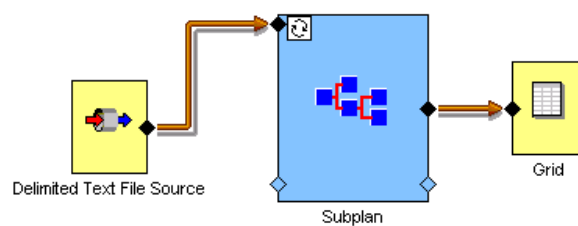


Abbildung 21: Performanzsteigerung mit Subplan

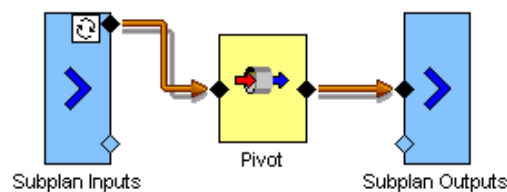


Abbildung 22: Inhalt des Subplans

Der Subplan ermöglicht es hier eine große Datenmenge in kleinere Abschnitte zu unterteilen und diese dem kritischen Transform zu übergeben.

Dieses Vorgehen ermöglicht eine Performanz-Steigerung für den Fall, dass es nicht möglich ist, die Daten in der Art zu sortieren, die der Transform benötigt.

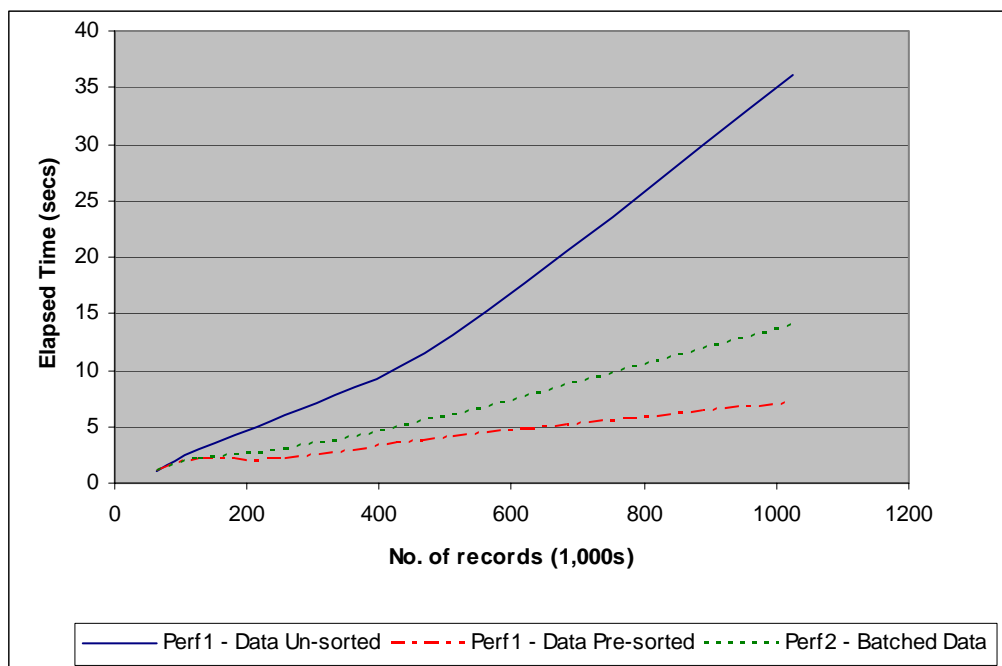


Abbildung 23: Laufzeitdiagramm in Abhängigkeit der Datenmenge

## Zusammenfassung

Durch die Einführung von Subplans in die Sagent Solution sind erhebliche Vorteile zu verzeichnen:

- Pläne werden übersichtlicher gestaltet und ihre Lesbarkeit wird erhöht.
- Manipulationen, die sich in immer wiederkehrenden Prozessen mit sich unterscheidenden Quellen beschäftigen, werden durch die Iteration in einem einzelnen Plan abgehandelt, anstatt mehrere Pläne zu schreiben oder einen Plan durch verschiedene Aufrufe immer wieder zu nutzen.
- Pläne, die Daten auf mehrere Senken verteilen, werden durch Batchverarbeitung konsolidiert, was die Fehleranfälligkeit verringert.
- Subplans ermöglichen ein elegantes Performanz-Tuning.
- Durch die dynamische Belegung der Properties kommt es zu einer Vorverarbeitung der Daten bevor der eigentliche Plan als Subplan startet.

Unsere Kunden entdecken hier immer weitere und größere Einsatzmöglichkeiten – Welche sehen Sie?

## Abbildungen

Abbildung 1: Kundensegmentierungsplan ohne Subplans .....	3
Abbildung 2: Kundensegmentierungsplan mit Subplans .....	4
Abbildung 3: Subplan zur Gewinnerrechnung .....	4
Abbildung 4: Subplan zur Segmentierung .....	4
Abbildung 5: Iterativer Plan .....	5
Abbildung 6: Datei mit Dateinamen .....	5
Abbildung 7: Zusammensetzen des Pfades mit dem Expression Calculator .....	6
Abbildung 8: Inhalt des Subplans .....	6
Abbildung 9: Flat File Transform Eigenschaften .....	7
Abbildung 10: Iterationseigenschaften des Subplans wählen .....	7
Abbildung 11: Eigenschaften der Iteration .....	8
Abbildung 12: Zuordnung der Iterationseigenschaften .....	9
Abbildung 13: Beispiel Batchverarbeitung .....	10
Abbildung 14: Erstellen des Dateinamen .....	10
Abbildung 15: Subplan Tab General Properties .....	11
Abbildung 16: Subplan Eigenschaften Zuordnung .....	11
Abbildung 17: Fluss im Subplan .....	12
Abbildung 18: Ergebnis der Batchverarbeitung .....	12
Abbildung 19: Beispiel Performanz .....	13
Abbildung 20: Laufzeitdiagramm in Abhängigkeit der Datenmenge .....	14
Abbildung 21: Performanzsteigerung mit Subplan .....	14
Abbildung 22: Inhalt des Subplans .....	14
Abbildung 23: Laufzeitdiagramm in Abhängigkeit der Datenmenge .....	15