

## Kapitel 3

# Grundlagen für Business Intelligence

### **In diesem Kapitel:**

Mehrdimensionale Datenanalyse	64
Logische Organisation von Geschäftsdaten	65
Business Intelligence Development Studio	74
Von der Vision zum fertigen Cube	76
Zusammenfassung	90

Damit Sie effektiv Business Intelligence in Ihrem Unternehmen einsetzen können, ist eine solide Datenbasis die absolute Voraussetzung. Daher erhalten Sie in diesem Kapitel einen Überblick, wie die Daten für Business Intelligence in SQL Server 2008 R2 organisiert sind und wie diese bereitgestellt werden müssen, damit Sie überhaupt wertvolle Features wie Zeitintelligenz einsetzen können. Sie werden am Beispiel der *Adventure Works*-Beispieldatenbanken, die Sie unter der URL <http://msftdbprodsamples.codeplex.com> kostenlos zum Download erhalten, in SQL Server Analysis Services eingeführt. Daneben lernen Sie den typischen Lebenszyklus von Business Intelligence-Implementierungen kennen. Darüber hinaus erhalten Sie eine Einführung darüber, welche Werkzeuge Ihnen zur Entwicklung von Business Intelligence-Grundlagen auf der Ebene von SQL Server 2008 R2 zur Verfügung stehen.

SQL Server 2008 R2 bietet außer dem bekannten Datenbankmodul noch weitere Module, die Sie für Business Intelligence einsetzen können, bzw. die notwendig sind, damit Sie später überhaupt Business Intelligence mit SharePoint Server 2010 betreiben können. Die Module sind im Einzelnen:

- Datenbankmodul
- SQL Server Integration Services (SSIS)
- SQL Server Analysis Services (SSAS)
- SQL Server Reporting Services (SSRS)

Natürlich gibt es noch weitere Module, die SQL Server 2008 R2 ausmachen. Die genannten vier bilden allerdings die Grundlage für das Bereitstellen von Business Intelligence-Daten. SQL Server Integration Services helfen Ihnen dabei, Daten aus verschiedensten Datenquellen in ein Data Warehouse zu integrieren. SQL Server Analysis Services ist die Komponente, die dabei hilft, logische Objekte wie Measures, Dimensionen und Cubes zu erstellen. Diese können Sie später direkt in den Dashboardelementen von SharePoint Server 2010 wiederverwenden. SQL Server Reporting Services schließlich hilft Ihnen bei der detaillierten Auswertung der gesammelten Geschäftsdaten sowie bei der Darstellung und Bereitstellung für Geschäftsanwender. Sie sehen also: Von Beginn an (d.h. vom Sammeln der Daten an) unterstützt Sie SQL Server 2008 R2 bis hin zur Darstellung für den Endbenutzer. Dazu bietet Microsoft verschiedene, kostenfreie Werkzeuge an, die Sie jeweils verwenden können: Da wäre zum einen SQL Server Management Studio, mit dem Sie SQL Server Analysis Services verwalten können. Mit Business Intelligence Development Studio (BIDS) hingegen definieren Sie die genannten logischen Objekte, wie z.B. Cubes. Mit diesem Werkzeug können Sie übrigens auch die Integration von externen Datenquellen vornehmen, diese in einem Data Warehouse vereinen sowie SSRS-Berichte erstellen.

## Mehrdimensionale Datenanalyse

Im Jahr 1993 prägten B.F. Codd, E.B. Codd und C.T. Salley erstmals in einem Whitepaper den Begriff On-Line Analytical Processing<sup>1</sup> und beschäftigten sich mit der Frage, ob und wie relationale Datenbanken in Verbindung mit der Abfragesprache SQL multidimensionale Datenanalyse

---

<sup>1</sup> Codd, E.F.; Codd, S.B.; Valley, C.T.: Providing OLAP (On-Line Analytical Processing) to User Analysts: An IT Mandate.

leisten konnten. In ihrer Studie verstehen die Autoren unter mehrdimensionaler Datenanalyse das Zusammenführen heterogener Datenquellen und die nachfolgende Analyse derselben für verschiedene Anwender in einem Unternehmen. Konkret sollen damit Fragen beantwortet werden können wie beispielsweise:

*Wie hoch war die Bruttorendite im dritten Quartal 2010 in Bayern?*

Prinzipiell lassen sich mit der Abfragesprache SQL Sichten erzeugen, die eine Auswertung mehrdimensionaler Daten ermöglichen. Diese bergen aber einige Nachteile: Sobald Sie eine Dimension verändern und die Sicht aus einer anderen Perspektive betrachten möchten, müssen Sie eine neue SQL-Abfrage erstellen, was in komplexen Fällen einen großen Aufwand mit sich bringen kann. Ein weiterer Grund, wieso sich SQL für die Erstellung mehrdimensionaler Sichten schlechter eignet als OLAP, ist die Leistung. Mit SQL müssen Aggregationen bei jedem Aufruf einer Abfrage oder Sicht berechnet werden. Dies erfordert bei komplexen Abfragen nicht unerhebliche Ressourcen. Mit OLAP dagegen können Sie Aggregationen vorberechnen lassen und in dafür vorgesehenen Strukturen speichern. Damit ist eine entsprechende OLAP-Abfrage im Vergleich um Längen schneller. Im Übrigen würde eine vergleichbare SQL-Abfrage in der Regel mehr Zeilen als gewünscht liefern; dies geht zu Lasten der Leistung, aber auch zu Lasten der Übersichtlichkeit für die Anwender. Nicht zuletzt erfordern derart komplexe SQL-Abfragen einiges an technischem Fachwissen von Anwendern, das Sie für diese nicht unbedingt voraussetzen dürfen.

## Logische Organisation von Geschäftsdaten

Bevor Sie Geschäftsdaten verwenden, um diese z.B. in Analysediagrammen oder Dashboards darzustellen, sollten Sie sich mit einigen Konzepten vertraut machen. Diese Konzepte erleichtern das Verständnis für die nachfolgenden Kapitel und ermöglichen es Ihnen zudem, Business Intelligence effektiv einzusetzen. Die Modellierung von Business Intelligence-Darstellungen kann nur so gut sein wie die darunter liegenden logischen Objekte definiert sind. Wenn Sie Ihre Business Intelligence-Informationen nur aus tabellarischen Daten wie z.B. Excel-Arbeitsmappen oder SharePoint-Listen beziehen, benötigen Sie diese Informationen nicht unbedingt. Wenn Sie aber ein fortgeschrittenes Business Intelligence verwenden möchten, ist der Einsatz von mehrdimensionalen Datenquellen und SQL Server Analysis Services unumgänglich.

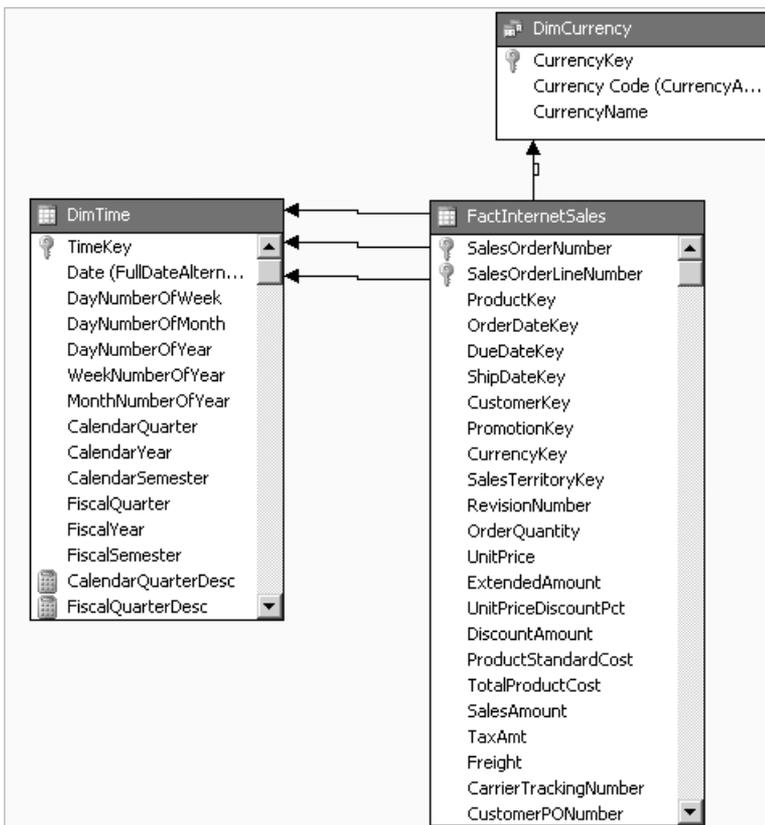
Die logischen Elemente wie Measures, Dimensionen und Cubes werden in eigenen SQL Server Analysis Services-Datenbanken gespeichert und sind diesen logisch untergeordnet. Da die Grundlage für diese Objekte immer noch eine relationale Datenbank bildet, werden diese logischen Objekte auch mithilfe einer solchen abgebildet und implementiert. Mit anderen Worten: Fakten und Dimensionen sind mithilfe von gewöhnlichen Tabellen abgebildet. Allein die Semantik der Daten entscheidet darüber, ob ein Feld einer Tabelle für die Verknüpfung mit einer Dimension oder als aggregierbare Measure verwendet werden kann.

- **Faktentabellen** In diesen Tabellen werden die tatsächlich quantifizierbaren, meist numerischen Werte gespeichert, wie z.B. die Anzahl der Verkäufe, die Lagermenge von Produkten usw. Diese Spalten dienen als Grundlage für die logischen Measures, die Sie dann für die Beantwortung von Business Intelligence-Fragen verwenden können. Die Faktentabellen ent-

halten zudem Verweise auf die Dimensionstabellen, mit denen sie verknüpft sind. Diese sind als Fremdschlüssel gespeichert. Mit Business Intelligence Development Studio erstellte Faktentabellen beginnen normalerweise mit dem Schlüsselwort *Fact*.

- **Dimensionstabellen** In diesen Tabellen werden die Informationen der Dimensionen gespeichert, mit denen eine Faktentabelle verknüpft ist. Typische Dimensionen sind beispielsweise *Zeit* und *Geografie*. Eine Faktentabelle kann mit beliebig vielen Dimensionentabellen verknüpft sein. Diese Verknüpfungen ermöglichen es Ihnen letztendlich, Ihre Fragestellungen einzuzugrenzen und z.B. die Anzahl der Verkäufe des ersten Quartals in einer bestimmten Region zu ermitteln.

Allerdings werden Sie bei der Einbindung externer Datenquellen, wie z.B. beliebiger SQL Server-Datenbanken, kaum die Zeit aufbringen können, jedes Feld zu analysieren und zu entscheiden, welches als Dimensionsverknüpfung und welches Feld fungieren kann. Basierend auf den importierten Felder und anhand der Kategorisierung der Felder werden dann Faktentabellen und Dimensionentabellen erstellt. Diese müssen Sie in der Regel nicht von Hand erstellen, sondern können sie über einen Assistenten erzeugen lassen.



**Abbildung 3.1** Sie könnten diese Faktentabelle zur Auswertung der Internetverkäufe unter Berücksichtigung der Dimensionen *Zeit* und *Währung* verwenden

Clou des Ganzen ist: Business Intelligence Development Studio kann selbstständig erkennen, welche Felder aus einer Tabelle numerisch aggregierbar sind und somit als Kandidaten für Measures in Frage kommen. Im Prinzip müssen Sie also nur festlegen, welche Tabellen Sie als Grundlage verwenden wollen, welche Felder Dimensionenverknüpfungen und welche Felder Measures darstellen. Den Rest erledigt der Assistent automatisch und Sie können sich anschließend um die Erstellung eines Cubes kümmern. Bevor Sie aber den Assistenten in Aktion sehen, erhalten Sie noch einige Grundlagen über den Aufbau der genannten logischen Objekte, beginnend mit den Measures, was zu Deutsch etwa so viel wie *Maße* bedeutet. Die in Abbildung 3.1 dargestellte Faktentabelle *FactInternetSales* enthält Fakten zu den Verkäufen des fiktiven Sportartikelherstellers *Adventure Works*. Hier finden Sie quantifizierbare Spalten wie z.B. *SalesAmount* oder *Freight*. Diese sind aggregierbar und damit Kandidaten für Measures.

## Measures und Measuregruppen

Im vorherigen Abschnitt haben Sie bereits einen Einblick darüber erhalten, dass Faktentabellen numerische Spalten enthalten, die logisch als Measures definiert werden können. Diese Daten werden Sie später, wenn Sie Business Intelligence-Objekte wie z.B. Key Performance Indicators (KPIs) oder Dashboards verwenden, aggregieren können. Für Measures können Sie also Aggregationsberechnungen durchführen, daher kommt die Voraussetzung, dass die entsprechenden Spalten in der Faktentabelle numerisch sein müssen. Da eine Faktentabelle mehrere solcher numerischer Spalten enthalten kann, wie z.B. *FactInternetSales* in Abbildung 3.1, ist es sinnvoll, diese logisch zu gruppieren. Dementsprechend werden die Measures, die auf derselben Faktentabelle basieren, logisch zu einer Measuregruppe zusammengefasst. Entsprechend strukturiert wird diese dann auch in Business Intelligence Development Studio dargestellt: Sie erhalten dort eine hierarchische Übersicht über die in einem SQL Server Analysis Services-Projekt definierten Measuregruppen und die darin enthaltenen Measures.

---

**HINWEIS**

Typische Aggregationen sind Summen, Mittelwerte, Minimal- und Maximalwerte.

---

Unter anderem befinden sich in Measuregruppen logischerweise die Gesamtheit der Measures, aber auch deren Aggregatfunktionen (z.B. ob die Summe eines bestimmten Measures gebildet werden soll oder dessen Durchschnitt). Daneben finden sich noch Informationen zu den verwendeten Dimensionen, zu denen Sie im nun folgenden Abschnitt mehr erfahren werden. Rekapitulieren wir noch einmal: Measures sind numerisch messbare Objekte wie z.B. die Anzahl der Verkäufe. Diese Werte sind normalerweise das Objekt Ihrer Business Intelligence-Analysen. Stellen Sie sich vor, Sie wollten die folgende Frage beantworten:

*Wie viele Verkäufe hatte mein Unternehmen im ersten Quartal 2011 in Norwegen zu verbuchen?*

Wenn Sie diese Frage aufschlüsseln, kommt dabei Folgendes heraus:

- Measure (Verkäufe): Wie viele Verkäufe ...
- Dimension (Zeit): ... im ersten Quartal 2011 ...
- Dimension (Geografie): ... in Norwegen ...

Werfen Sie nun also einen Blick auf die Dimensionen, um den Zusammenhang mit Measures zu verstehen und am Ende dieses konzeptuellen Abschnitts ein vollständiges Bild davon zu haben, womit Sie sich eigentlich im Rahmen mehrdimensionaler Datenanalyse beschäftigen.

## Dimensionen

Damit Sie aussagekräftige Analysen erstellen und komplexe Fragen beantworten können, benötigen Sie Dimensionen. Diese stellen bei der Analyse mehrdimensionaler Daten sozusagen Ihre Filter dar, nach denen Sie Informationen im Detail abrufen können. So ist es z.B. erforderlich, dass Sie im vorangegangenen Abschnitt eine Dimension *Zeit* definiert haben, die zudem noch das Attribut *Quartal* enthält. Ansonsten könnten Sie die gestellte Frage nach der Anzahl der Verkäufe im ersten Quartal 2011 in Norwegen nicht beantworten. Analog dazu benötigen Sie eine Dimension *Geografie*, die als Attribut das Land enthält.

Dimensionen bestehen aus sogenannten Attributen. Diese entsprechen Feldern in der zugehörigen Dimensionentabelle und stellen nichts anderes als Ihre Filter dar, nach denen Sie später Daten auswerten können. Im Beispiel von Abbildung 3.1 bedeutet dies, dass Sie z.B. das Measure *SalesAmount* (entspricht der Anzahl der Verkäufe) nach den Attributen *CalendarYear*, *CalendarSemester*, also Kalenderjahr und Semester filtern könnten. Damit könnten Sie also bereits Fragen beantworten, wie:

*Wie viele Verkäufe hatte mein Unternehmen in Summe im Kalenderjahr 2010?*

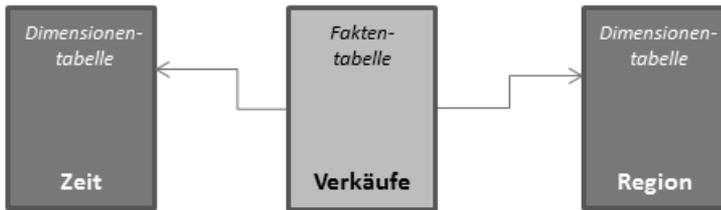
Es gilt: Je mehr Dimensionen Sie definiert haben, desto detailliertere Fragestellungen werden Sie mit Ihrem SQL Server Analysis Services-Datenbank beantworten können.

---

**HINWEIS** Hierarchien in Dimensionen ermöglichen Ihnen, detaillierte Analysen zu betreiben. Nehmen Sie dafür an, Sie filtern Ihre *Verkäufe* nach einem bestimmten *Verkäufer*. Wenn die Hierarchie vorsieht, dass *Verkäufer* ein untergeordnetes Element *Standort* hat, könnten Sie dieselbe Statistik auf einen bestimmten Verkäufer eingrenzen. In der Folge könnten Sie alle Verkäufe dieses bestimmten Verkäufers, und zwar für alle verschiedenen Standorte, sehen. Diese Aktion wird auch als Drilldown bezeichnet.

---

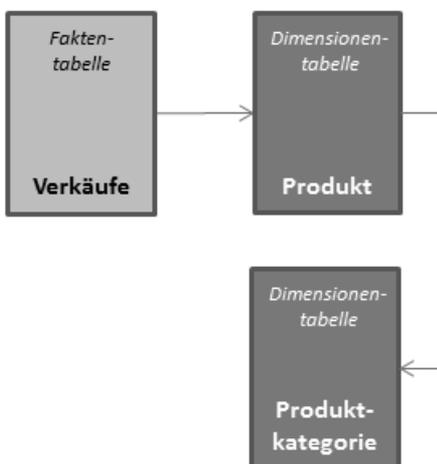
Zudem gibt es Hierarchien unter den Attributen, die entweder fest definiert oder aber benutzerdefiniert wählbar sein können. Diese Hierarchien können wiederum von Ihnen als Benutzer dafür verwendet werden, um in einem Cube zu navigieren. Für den Endverbraucher der Business Intelligence-Objekte heißt das: Er hat mehr Detailgrade und Drilldownmöglichkeiten, je mehr Hierarchien für eine Dimension definiert sind. Im Beispiel von Abbildung 3.1 ist die Faktentabelle *FactInternetSales* mit den beiden Dimensionentabellen *DimTime* und *DimCurrency* verbunden. Demzufolge ist auch die entsprechende Measuregruppe mit den beiden Dimensionen *Time* und *Currency* verbunden. In diesem Fall ist die Verbindung zu den beiden Dimensionen unabhängig, d.h., die Dimensionen weisen zueinander keinen Bezug auf.



**Abbildung 3.2** Wenn Sie Dimensionen wie hier Faktentabellen in einem Sternschema zuordnen, sollten Sie zusätzlich benutzerdefinierte Hierarchien definieren, um den Detailgrad für die Analyse durch Geschäftsanwender zu erhöhen

Je nachdem, wie Fakten- und Dimensionentabellen miteinander in Verbindung stehen, kann es sich um ein Sternschema oder ein Schneeflockenschema handeln. Ersteres zeichnet sich durch direkte Beziehungen der Faktentabellen mit den Dimensionentabellen aus. Dies hat den Vorteil, dass Sie direkt im Diagramm bereits wissen, nach welchen Kriterien Sie später den Cube durchsuchen können (im vorliegenden Fall in Abbildung 3.2 nach den Dimensionen *Zeit* und *Region*). Erweiterte Drilldownfunktionen sind nicht unmittelbar verfügbar. Sie bzw. die Business Intelligence-Entwickler können aber benutzerdefinierte Hierarchien für Attribute festlegen und so z.B. zuerst eine Einschränkung der Dimension *Zeit* auf das Jahr 2010 durchführen und dann im Detail nach den vier Quartalen 2010 den Cube durchsuchen.

Die Abbildung 3.3 zeigt die Schneeflockendarstellung des Schemas. Die Faktentabelle *Verkäufe* ist in diesem Fall nur mit der Dimensionentabelle *Produkt* verknüpft. Diese ist aber ihrerseits mit einer weiteren Dimensionentabelle, *Produktkategorie*, verknüpft. Dies hat zur Folge, dass für Sie und Ihre Anwender erweiterte Drilldowns zur Verfügung stehen. Sie können also ein Measure nicht nur auf ein Produkt, sondern auch auf dessen Produktkategorien eingrenzen.



**Abbildung 3.3** Wenn Sie Dimensionen wie hier Faktentabellen in einem Schneeflockenschema zuordnen, haben Sie damit implizit erweiterte Drilldownmöglichkeiten

Dimensionen, die im Schneeflockenschema angeordnet sind, ermöglichen Ihnen erweiterte Drilldowns und damit detailliertere Analysen.

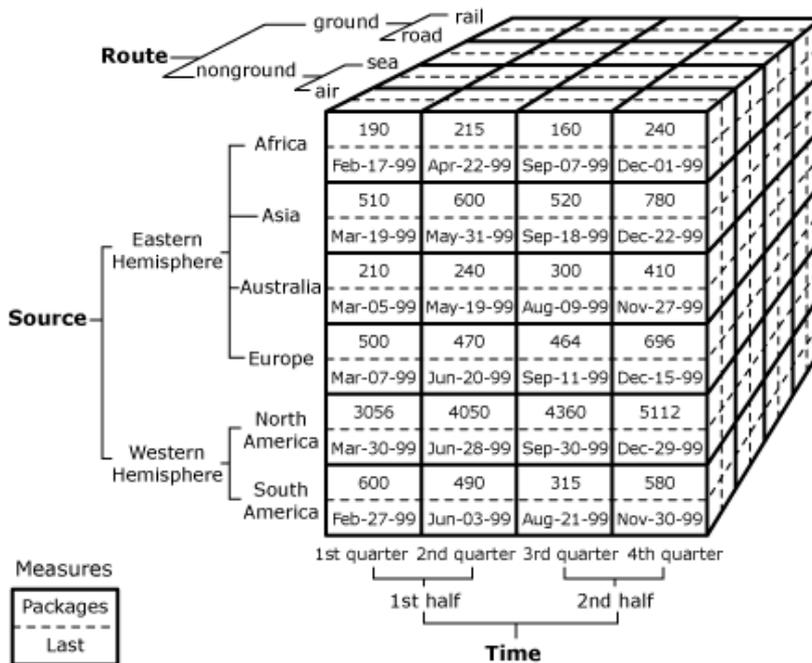
## Cubes

Eine mehrdimensionale Datenmenge, die durch Dimensionen und Measures definiert werden, nennt man *Cube*, was zu Deutsch so viel wie Würfel heißt. Cubes dienen der logischen Darstellung von Unternehmensdaten und sind eine logische Struktur in einem Data Warehouse, mit der Sie Unternehmensdaten analysieren können. Mithilfe von Cubes können Sie auf Daten einfach und flexibel zugreifen. Damit überwinden Cubes die tabellarische Darstellung relationaler Datenbanken: Während in relationalen Datenbankschemata die Zusammenhänge durch die verknüpften Tabellen praktisch nicht erfassbar sind, wird dies im Falle eines Cubes extrem erleichtert. Wie Sie in der Abbildung 3.4 sehen können, besteht ein erstellter Cube im Prinzip auch aus miteinander verknüpften Dimensionen- und Faktentabellen (die in der Abbildung einfach als Dimensionen und deren zugeordnete Werte erkennbar sind). Durch die übergeordnete und abstraktere Struktur, den Cube, werden diese leichter erfassbar und ermöglichen eine Interpretation der Daten. Der Begriff *Cube* ist übrigens nicht wörtlich zu nehmen, da in den allermeisten Fällen mehr als nur drei Dimensionen verwendet werden. Im Beispiel von Abbildung 3.4 sind es zur Veranschaulichung nur drei. Sie können aber beliebig viele Dimensionen mit Measures verknüpfen und so die Aussagekraft eines Cubes steigern.

Für umfangreiche Analysen von Geschäftsdaten stellen Cubes einige Operationen bereit, die Sie auf diesen durchführen können:

- **Slicing** Auf Seite der Endanwender versteht man diese Operation auch als *Filtern*. Stellen Sie sich vor, wie Sie bildlich eine Scheibe des Cubes abschneiden und so dessen Informationsgehalt auf eine bestimmte Dimension, z.B. die Zeit, einschränken. Im Beispiel der Beantwortung der Frage nach den Verkäufen in einem bestimmten Quartal kommt also effektiv Slicing zum Einsatz.
- **Dicing** Das so genannte Dicing (Würfeln) ist ein mehrfach ausgeführtes Slicing, das einen kleineren Cube aus dem originalen Cube heraustrennt und so eine detailliertere Sicht auf Ihre Unternehmensdaten ermöglicht.
- **Pivotierung** Mit dieser Operation wird der Würfel gedreht und ermöglicht so einen Einblick in neue Dimensionen. Sie können Ihre Unternehmensdaten hiermit aus verschiedenen Blickwinkeln betrachten oder – besser gesagt – nach anderen Dimensionen filtern.
- **Drilldown** Diese Operation ermöglicht Ihnen das Betrachten einer bestimmten Dimension im Detail. Hierfür werden die Hierarchien der Dimensionentabellen verwendet. Am konkreten Beispiel heißt dies: Wenn Sie z.B. die Abbildung 3.4 betrachten und einen Drilldown über die Dimension *Route* ausführen, könnten Sie die Measures im Detail betrachten, würden also die Anzahl der Pakete für die einzelnen Routenarten wie Luft, Wasser etc. visualisieren können. Die entgegengesetzte Operation ist ebenfalls möglich und wird Drillup genannt. Beide Methoden können Sie in SharePoint indirekt über Scorecards und Dashboards einsetzen.

Daneben gibt es noch zahlreiche andere Operationen, die auf Cubes durchgeführt werden können. Die oben genannten gehören jedoch zu den gängigsten. Cubes werden üblicherweise von Business Intelligence-Entwicklern erstellt und für Geschäftsanwender zur Verfügung gestellt. Dashboardautoren können Cubes dazu verwenden, mehrdimensionale Analysen in Form von Diagrammen, Tabellen oder anderen Visualisierungen zu ermöglichen.



**Abbildung 3.4** Verwenden Sie einen solchen Cube, um ein Measure wie die Anzahl der Pakete (*Packages*) nach den Dimensionen Route (*Route*), Herkunft (*Source*) und Zeit (*Time*) zu analysieren

Diese fertigen Dashboards und deren Elemente können dann von Geschäftsanwendern einfach verwendet werden, ohne zu wissen, wie die darunter liegenden Daten funktionieren.

Für Sie als Dashboardautor ist es allerdings unumgänglich zu wissen, wie Measures, Cubes und Dimensionen konzeptuell funktionieren. Dies ist Voraussetzung für den richtigen Einsatz von Cubes bei der Dashboarderstellung. Measures und Dimensionen müssen vor der Erstellung eines Cubes erstellt werden. Dabei helfen Ihnen Assistenten von Business Intelligence Development Studio, mit deren Hilfe Sie die geeigneten Tabellen aus externen Datenquellen importieren und darauf basierend Measures und Dimensionen erstellen. Genauso verhält es sich bei der Erstellung von Cubes: Auch hier werden Sie von einem Assistenten durch eine Reihe von Dialogfeldern geführt, mit deren Hilfe Sie die im Cube zu verwendenden Measures und Dimensionen einbinden können.

Nachdem alle Objekte wie Measures, Dimensionen und Cubes erstellt sind, lassen sich diese für eine SQL Server Analysis Services-Instanz bereitstellen. Ab diesem Zeitpunkt können Dashboardautoren diese Objekte für die Erstellung von Business Intelligence-Objekten wie Dashboards oder Scorecards verwenden.

## SQL Server Analysis Services (SSAS)

SQL Server Analysis Services (SSAS) ist der Bereich von Microsoft SQL Server 2008 R2, der sich mit multidimensionalen Daten, deren Definition und Abfragen beschäftigt und somit Online Analytical Processing (OLAP) bereitstellt. SSAS basiert seine logischen Strukturen auf Data Warehouses und ermöglicht Business Intelligence-Entwicklern die Definition von Measure, Dimensionen und Cubes. Die Cubes wiederum können bei der Erstellung von Dashboardelementen in SharePoint Server 2010 verwendet werden und liefern aufschlussreiche Daten über Zusammenhänge zwischen den Fakten und den Dimensionen. OLAP in SQL Server Analysis Services besteht aus folgenden Komponenten, die Sie für Business Intelligence einsetzen können:

- Optimierte Strukturen für schnelle Ad-hoc-Abfragen
- Berechnungsmechanismen für eine schnelle Transformation der Grunddaten
- Entdecken von Trends und Statistiken, die nicht im Data Warehouse stehen
- Bereitstellung von Cubes: Self-Service-Analyse und eigene Berichte

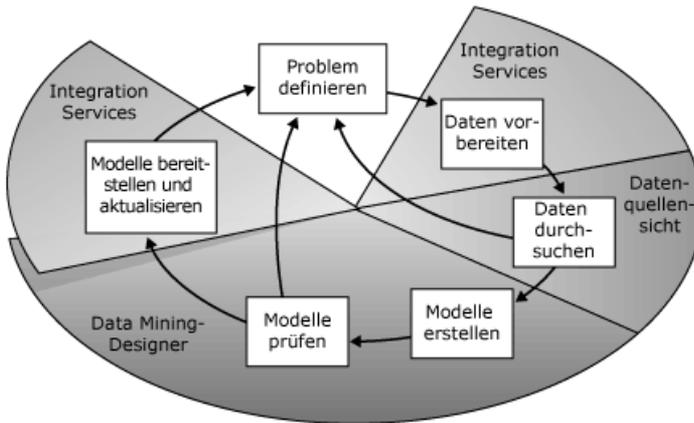
Mithilfe von Business Intelligence Development Studio (BIDS) können unter anderem folgende Objekte erstellt, verändert und verwaltet werden:

- Datenquellen
- Cubes
- Dimensionen
- Miningstrukturen
- Rollen

## Data Mining

Data Mining ist die zweite der beiden Komponenten für die Unterstützung mehrdimensionaler Daten in SQL Server 2008 R2. Mithilfe der Data Mining-Komponenten und -Werkzeuge von SQL Server 2008 können Sie Muster und Trends in Ihren Daten erkennen und so gezielt Vorhersagen erstellen, die z.B. Fragen beantworten, wie sich das Kaufverhalten der Kunden in den nächsten sechs Monaten verändern wird. Ohne den Einsatz von Data Mining wären solche Fragen nicht zu beantworten, da die Datenmengen viel zu groß sind und diese Informationen auch nicht direkt enthalten. Im Wesentlichen besteht ein Data Mining-Prozess aus den folgenden Schritten:

- Definieren der Problemstellung
- Vorbereiten der Daten
- Durchsuchen von Daten
- Erstellen von Modellen
- Durchsuchen und Validieren von Modellen
- Bereitstellen und Aktualisieren von Modellen



**Abbildung 3.5** Machen Sie sich hier mit den grundlegenden Data Mining-Prozessschritten vertraut

In Abbildung 3.5 wird der Data Mining-Prozess grafisch dargestellt und seine Werkzeuge in Zusammenhang mit den Prozessschritten gesetzt. Die Abbildung stammt aus dem Artikel »Data Mining-Konzepte (Analysis Services – Data Mining)« der Microsoft TechNet-Bibliothek. Sie können den vollständigen Artikel unter der URL <http://technet.microsoft.com/de-de/library/ms174949.aspx> einsehen. Es gibt die folgenden Data Mining-Werkzeuge, die Sie für die Verwendung von Data Mining einsetzen können:

- **Data Mining-Assistent** Dieses Werkzeug ist Bestandteil von Business Intelligence Development Studio (BIDS) und unterstützt Sie bei der Entwicklung und Erstellung von Data Mining-Implementierungen. Mit ihm definieren Sie Data Mining-Strukturen und Modelle, wählen Data Mining-Algorithmen und Datenquellen aus und definieren Falltabellen.
- **Data Mining-Designer** Dieses Werkzeug ist ebenfalls Bestandteil von Business Intelligence Development Studio. Es wird automatisch geöffnet, sobald Sie den Data Mining-Assistenten abgeschlossen haben. Mit Data Mining-Designer erhalten Sie die Möglichkeit zum Erstellen, Durchsuchen und Bereitstellen von Vorhersagen existierender Data Mining-Modelle.
- **SQL Server Management Studio** Dieses Werkzeug unterstützt Sie bei der Verwaltung einmal bereitgestellter Data Mining-Modelle. Sie können Modelle durchsuchen und Vorhersagen erstellen. Dafür gibt es eine spezielle Abfragesprache, die Data Mining Extensions (DMX), die Sie für die Erstellung komplexer Vorhersagen verwenden können.
- **SQL Server Integration Services** Auch dieses Werkzeug können Sie über Business Intelligence Development Studio verwenden. Vor allem im Bereich der Transformation des ETL-Prozesses liefert es Ihnen die Möglichkeit, Data Mining-Algorithmen einzusetzen, um zusätzliche Informationen zu gewinnen, die so explizit nicht in den Quelldatenbanken stehen.

Gängige Data Mining-Szenarien finden Sie in der folgenden Auflistung. Es handelt sich hierbei um Fragen, die vor allem für Analysten und Marketing beantwortet werden sollen und ohne Data Mining-Algorithmen nicht möglich wären. Es handelt sich im Detail um:

- **Vorhersagen (Forecasting)** Hierbei geht es um die Vorhersage von z.B. dem Kaufverhalten von Kunden unter Berücksichtigung bestimmter Dimensionen, wie z.B. Zeit und Ort

- **Direktmarketing für bestimmte Kundengruppen** Basierend auf demografischen Daten soll hier beispielsweise ermittelt werden, welche Kunden sich für Direktmarketing im Rahmen einer Kampagne besonders eignen. Auch hier gilt: Die Beantwortung derartiger Fragen wäre ohne Data Mining-Algorithmen undenkbar.
- **Market Basket Analysis** Hierbei werden Fragen beispielsweise für Empfehlungssysteme beantwortet. Basierend auf den Waren, die ein Kunde in einem Onlinegeschäft angesehen oder gekauft hat, für welche Waren wird er sich am wahrscheinlichsten als Nächstes interessieren?
- **Sequence Clustering** Hierbei werden Fragen beantwortet, wie z.B., welche Waren ein Kunde in den Einkaufswagen legt und in welcher Reihenfolge

Wie Sie bereits in der vorangegangenen Auflistung gesehen haben, beinhaltet Business Intelligence Development Studio zwei von vier Werkzeugen für die Erstellung von Data Mining-Objekten. Doch damit nicht genug: Business Intelligence Development Studio ist zusammen mit SQL Server Management Studio Ihr Schweizer Taschenmesser für Business Intelligence mit SQL Server 2008 R2. Beide Werkzeuge sind kostenlos; Business Intelligence Development Studio hat einen starken Fokus auf die Entwicklung von OLAP-Strukturen, während SQL Server Management Studio eher verwaltend benutzt werden kann.

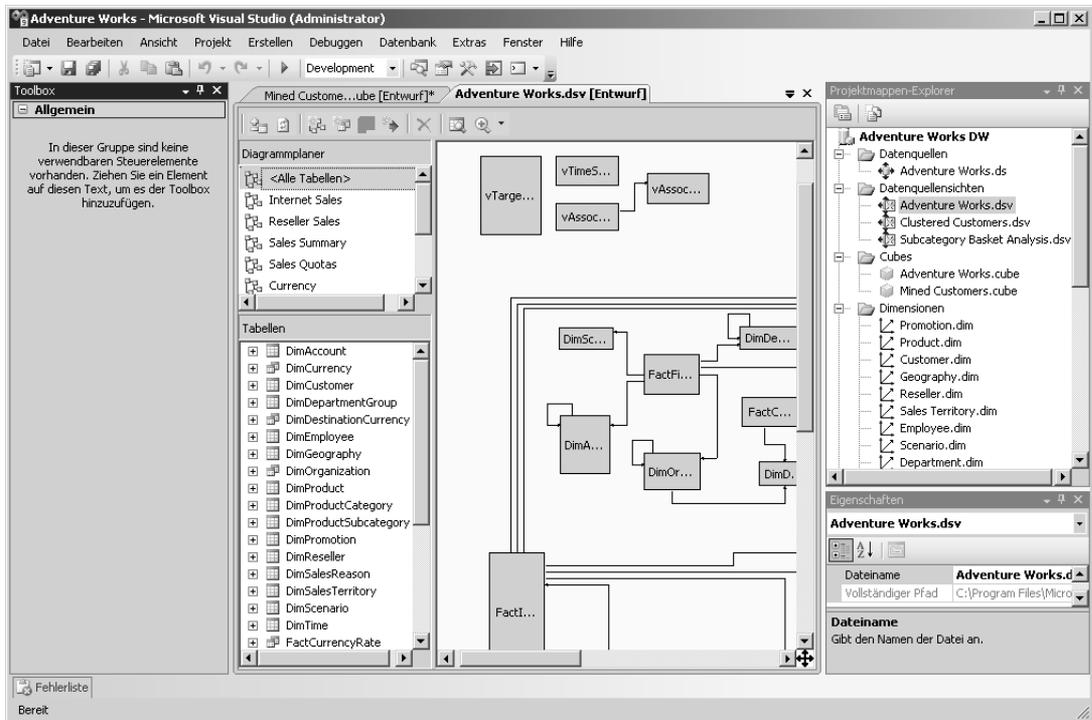
## Business Intelligence Development Studio

Business Intelligence Development Studio (BIDS) ist ein Werkzeug zur Erstellung und Bereitstellung von SQL Server Analysis Services-, SQL Server Integration Services- und SQL Server Reporting Services-Projekten. Die Entwicklungsumgebung basiert auf der Visual Studio 2008-Plattform, wodurch ihre Handhabung für Entwickler gewohnt ausfällt. Mit diesem Werkzeug erstellen Business Intelligence-Entwickler Measures, Dimensionen und Cubes, aber auch Berichte und SSIS-Projekte. Obwohl BIDS auf Visual Studio 2008 basiert, sind dessen Projektvorlagen nicht verfügbar. Stattdessen findet sich eine Reihe von Projektvorlagen, die für die Entwicklung von Business Intelligence-Objekten verwendet werden können.

BIDS besteht aus den folgenden vier Hauptfenstern:

- **Projektmappen-Explorer** In diesem Bereich sehen Sie Ihr Projekt strukturiert nach Typen. So finden Sie z.B. eine Kategorie für Dimensionen, eine Kategorie für Fakten, eine Kategorie für Datenquellen usw. Über den Projektmappen-Explorer können Sie einzelne Objekte auswählen, öffnen oder auch löschen.
- **Eigenschaftfenster** Im Eigenschaftfenster können Sie grundlegende Einstellungen für ein im Projektmappen-Explorer ausgewähltes Objekt vornehmen
- **Entwurfswindow** Im Entwurfswindow können Sie das jeweils aktuelle Objekt bearbeiten. Hier lassen sich beispielsweise die Beziehungen von Dimensionen, die Beziehungen zwischen Tabellen und Dimensionen in Cubes und Datensichten einsehen und verändern usw. Das Entwurfswindow wird je nach Objekttyp in unterschiedliche Unterfenster aufgeteilt, in denen Sie jeweils verschiedene Aktionen vornehmen können.

- **Toolbox** Aus der Toolbox können Sie Elemente in das Entwurfswindow ziehen und dort durch Loslassen der Maustaste platzieren



**Abbildung 3.6** Navigieren Sie im Projektmappen-Explorer (rechts) durch die Dimensionen und Cubes und bearbeiten Sie diese grafisch im Entwurfswindow (Mitte)

**HINWEIS** Im Artikel »Einführung in Business Intelligence Development Studio« der Microsoft TechNet-Bibliothek erhalten Sie eine ausführliche Einführung in die Bestandteile von BIDS. Sie können den Artikel unter der Adresse <http://technet.microsoft.com/de-de/library/ms173767.aspx> aufrufen.

Wenn Sie die *AdventureWorks*-Beispieldatenbanken installiert haben, erhalten Sie nach dem Durchlaufen des Assistenten im letzten Schritt eine Liste mit den bereits ausgeführten Aktionen angezeigt. Es wurden jedoch nicht alle Aktionen für Sie durch den Assistenten ausgeführt: Sie müssen in diesem Fall noch das *AdventureWorks*-Beispielprojekt laden, erstellen und für SQL Server Analysis Services bereitstellen, bevor Sie die vordefinierten Cubes verwenden können.

**TIPP** Damit Sie sich mit BIDS orientieren und die geeigneten Beispiele selbst erkunden können, empfiehlt es sich, die Beispieldatenbanken des fiktiven Unternehmens Adventure Works herunterzuladen und zu installieren, bevor Sie fortfahren. Damit gewährleisten Sie, dass Sie wichtige Business Intelligence-Strukturen wie Dimensionen und Cubes selbst erkundet und angewandt haben, bevor Sie diese zum Erstellen von Analysediagrammen und Dashboards einsetzen. Sie werden sehen, es zahlt sich aus.

Die kostenlosen *AdventureWorks*-Beispieldatenbanken finden Sie auf der Webseite <http://msftdbprodsamples.codeplex.com/>. Klicken Sie dort im rechten Bereich auf die Schaltfläche *Download* und wählen Sie die richtige Version für Ihre SQL Server-Installation aus. Nachdem der Download abgeschlossen ist, können Sie die ausführbare Datei direkt starten. Ein Assistent wird Sie durch die Installation führen. Am Ende sehen Sie noch eine Zusammenfassung der Installation zusammen mit Hinweisen, wie Sie die OLAP-Strukturen erstellen und für SSAS bereitstellen können.

---

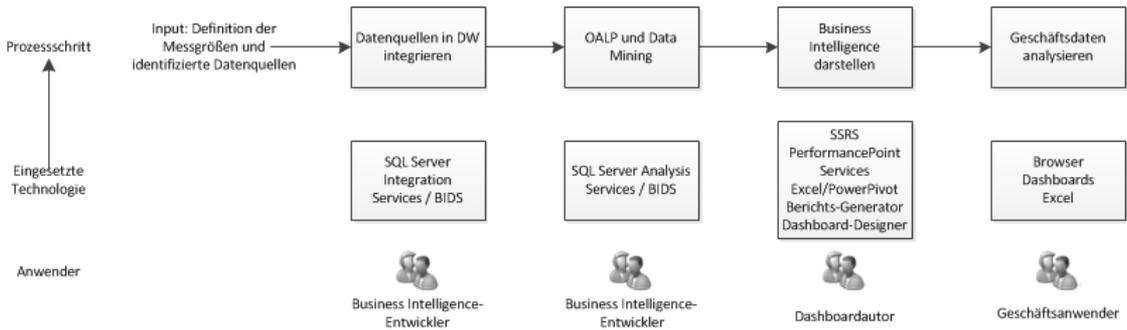
Führen Sie nach erfolgreicher Installation der *AdventureWorks*-Datenbanken folgende Schritte aus, um die SSAS-Objekte zu erstellen und für die Verwendung bereitzustellen:

1. Öffnen Sie Business Intelligence Development Studio.
2. Laden Sie die richtige *.sln*-Datei, die der Version Ihrer SQL Server-Installation entspricht.
3. Klicken Sie mit der rechten Maustaste im Projektmappen-Explorer auf den Knoten *Adventure Works DW*.
4. Wählen Sie im Kontextmenü den Eintrag *Erstellen*.
5. Klicken Sie erneut mit der rechten Maustaste auf den Knoten *Adventure Works DW*.
6. Wählen Sie im Kontextmenü den Eintrag *Bereitstellen*.

Anschließend werden die Objekte wie Fakten, Dimensionen und Cubes erstellt und bereitgestellt. Damit sind die beiden vorbereiteten Cubes für eine Verwendung z.B. in Dashboard-Designer bereit. Sie können die SQL Server Analysis Services-Instanz aber auch mit Microsoft Excel und Microsoft PowerPivot für Microsoft Excel verwenden. Auch für die folgenden Beispiele in diesem Kapitel ist es durchaus sinnvoll, wenn Sie bereits einige OLAP-Strukturen definiert haben, um diese im Detail erkunden zu können. Nicht zuletzt werden Sie am Ende dieses Buchs wieder auf die Anbindung und Analyse von SSAS-Geschäftsdaten zurückkommen. Spätestens zu diesem Zeitpunkt werden Sie die *AdventureWorks*-Beispieldatenbanken gebrauchen können.

## Von der Vision zum fertigen Cube

Von der Unternehmensvision bis hin zur Analyse von Geschäftsdaten wird ein gewisser Lebenszyklus von Business Intelligence-Implementierungen durchlaufen. Dabei ist klar ein Zusammenhang zwischen den zu erhebenden Messgrößen, der Identifizierung der dafür notwendigen bereits vorhandenen Datenquellen, der Integration dieser Datenquellen in ein Data Warehouse (das vorher erstellt werden muss), der Erstellung von OLAP-Objekten wie Measures, Dimensionen und Cubes, der Bereitstellung der aufbereiteten Geschäftsdaten und der letztendlich Analysen durch die Geschäftsanwender zu erkennen. Die Abbildung 3.7 zeigt einen Ausschnitt aus diesem Prozess. In diesem Ausschnitt wird angenommen, dass die Definition der Messgrößen und die Identifizierung der Datenquellen bereits stattgefunden hat. Diese beiden Schritte bilden in der Regel den Ausgangspunkt einer Business Intelligence-Lösung und werden sowohl von der Geschäftsführung als auch von Business Intelligence-Entwicklern aus der IT-Abteilung getragen.



**Abbildung 3.7** Sie steigen in diesem Abschnitt zum Zeitpunkt der Integration der Datenquellen in ein existierendes Data Warehouse ein

Für den Business Intelligence-Entwickler ist eine Übersicht über die zu integrierenden Datenquellen unumgänglich. Hierfür kann er bereits BIDS einsetzen und die Datenquellen direkt als solche definieren.

Es ist aber auch durchaus möglich, dass der Business Intelligence-Entwickler diese erst in einer Liste erfasst und zu einem späteren Zeitpunkt in BIDS definiert.

Die als Voraussetzung für den Einstieg in den Business Intelligence-Lebenszyklus (Abbildung 3.7) gehandelten Prozessschritte sind:

- Die Erstellung geeigneter Messgrößen
- Die Erfassung aller benötigten Datenquellen

Diese Messgrößen basieren ihrerseits auf den Unternehmenszielen (welche wiederum strategiekonform sein müssen). Würden Sie diesen Vorgang weiter nach oben treiben, würden Sie am Ende wieder zur Vision des Unternehmens gelangen. Den Einstieg in den Ausschnitt des Business Intelligence-Lebenszyklus (wie in Abbildung 3.7 dargestellt) bildet demzufolge der Schritt der Integration der Datenquellen in ein existierendes Data Warehouse. Würde zu diesem Zeitpunkt kein solches existieren, müsste der Business Intelligence-Entwickler ein neues erstellen. Der dargestellte Prozess ist stark verkürzt dargestellt und fasst einige Schritte zusammen. Daher werden die folgenden drei Schritte in den nächsten Abschnitten noch im Detail erläutert:

- Datenquellen in Data Warehouse integrieren
- Measures, Dimensionen und Cubes erstellen
- Business Intelligence darstellen

Begonnen wird dabei mit der Integration identifizierter Datenquellen in ein Data Warehouse.

## Datenquellen in Data Warehouses integrieren

Für diesen Schritt gibt es verschiedene Vorgangsweisen. Einerseits können Sie existierende Datenquellen in ein SQL Server Analysis Server-Projekt in BIDS einbinden. In der Folge müssten Sie dann die Daten importieren. Hierfür könnten Sie entweder SQL Server Integration Services verwenden (auch hierfür können Sie eigene Projekte erstellen). Da SSIS den gesamten

ETL-Prozess unterstützt, werden bei einem Import der Daten mit SSIS diese auch gleichzeitig gesäubert und für das Data Warehouse transformiert und übertragen. Sie können aber auch – falls Ihnen diese Option zu wenig Kontrolle bietet – die Daten auf eigene Faust mithilfe von T-SQL und gespeicherten Prozeduren importieren.

---

**ACHTUNG** Der Prozessschritt der Datenintegration ist üblicherweise der längste. Erwarten Sie hier Schwierigkeiten, denn heterogene Datenquellen in einem Data Warehouse unterzubringen, heißt, sich mit fehlenden Werten, unvollständigen Datensätzen und unterschiedlicher Datenqualität auseinanderzusetzen. Typischerweise ist der ETL-Prozess iterativer Natur. Mit anderen Worten: Das Säubern der Daten muss unter Umständen mehrfach wiederholt werden. Dies erfordert Zeit und finanzielle Ressourcen. Sie sollten daher im Vorfeld untersuchen, ob und welche Ressourcen Sie für den ETL-Prozess zur Verfügung haben.

---

Die Verwendung von SSIS bietet sich insofern auch an, als dass Sie damit wiederverwendbare ETL-Prozesse erstellen können, für die geringfügige Anpassungen bei der Änderung der zu importierenden und zu säubernden Datenquelle ausreichen können.

Mit SSIS können Sie Datenquellen und Sichten auf diese im Projektmappen-Explorer erstellen, genauso wie in einem SSAS-Projekt. Hierfür können Sie die identischen Assistenten zur Erstellung von Sichten und Datenquellen nutzen. Sie können im Designer einen vollständigen ETL-Prozess definieren, indem Sie Datenquellen mit Datenflusstransformationen und letztlich mit Datenzielen verbinden. Sie erstellen also ein Datenflussdiagramm und verbinden beliebige Datenquellen mit beliebigen Transformationen (z.B. Pivotierung), bevor Sie diese in die Zieldatenbank (das gewünschte Data Warehouse) laden. Diese drei Schritte entsprechen dem ETL-Prozess und sie werden auch genauso über die drei Kategorien der Toolbox widergespiegelt.

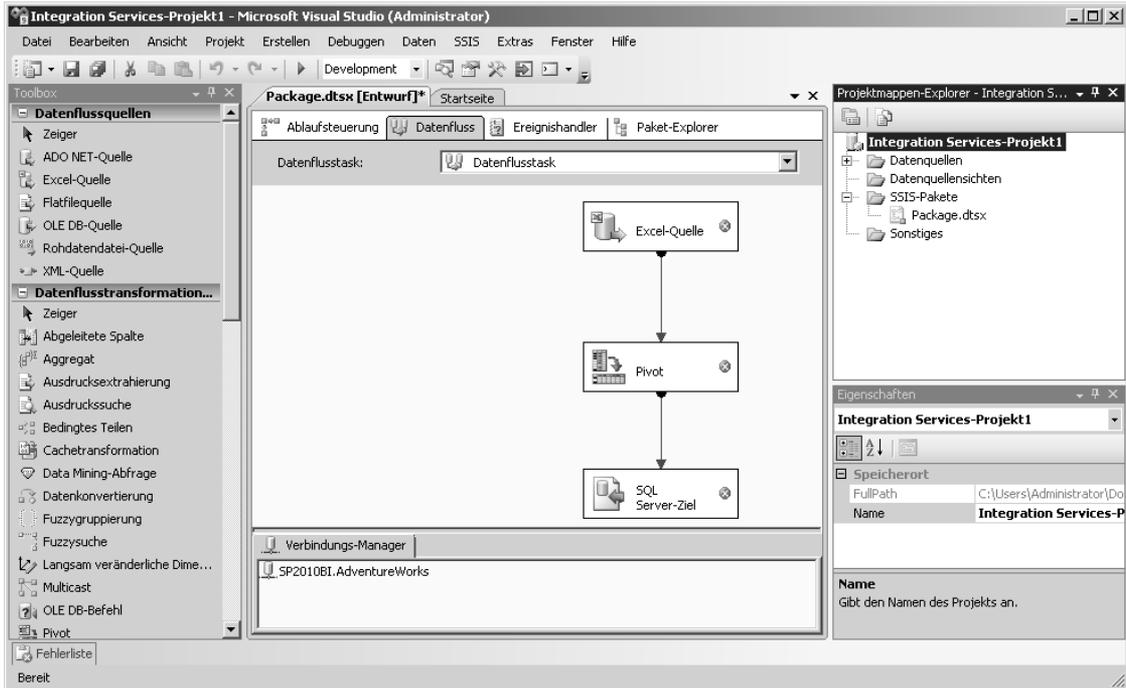
- **Datenflussquellen** Über diese Komponenten findet die Extraktion des ETL-Prozesses statt. Es sind folgenden Datenquellentypen verfügbar, aus denen Sie Daten extrahieren können:
  - ADO NET
  - Excel
  - Flatfile
  - OLE DB
  - Rohdatendatei
  - XML
- **Datenflusstransformationen** Über diese Komponenten findet die Transformation der extrahierten Daten statt. Es sind unter anderem folgende Transformationen verfügbar, die Sie auf die extrahierten Daten anwenden können:
  - Aggregat
  - Data Mining-Abfrage
  - Datenkonvertierung
  - Pivot
  - Unpivot

- **Datenflussziele** Über diese Komponenten findet das Laden der extrahierten und transformierten Daten statt:
  - ADO.NET
  - Excel
  - Flatfile
  - OLE DB
  - Rohdatendatei
  - SQL Server

Die Erstellung dieser so genannten Datenflusstasks ist sehr intuitiv, da Sie die gewünschten ETL-Komponenten einfach aus der Toolbox in das Entwurfsfenster ziehen und in der Folge bearbeiten können. Die Abbildung 3.8 zeigt einen nur teilweise konfigurierten Datenflusstask, in dem eine Datenquelle (*Excel-Quelle*) mit einer Transformation vom Typ Pivot verbunden und anschließend in eine SQL Server-Datenbank überführt wird. Dies stellt einen einfachen Fall eines ETL-Prozesses dar; in der Realität werden Sie häufig wesentlich komplexere Datenflusstasks einrichten müssen. Darüber hinaus haben Sie die Möglichkeit, mit SSIS das Verhalten im Fehlerfall zu kontrollieren. Hierfür können Sie eigene Ereignishandler einrichten: Falls ein bestimmter Fehlerfall eintritt, können Sie einen vordefinierten Arbeitsfluss ablaufen lassen. Sie sehen also: SSIS ist bereits auf Fehlerfälle eingestellt. Dies zeigt, dass das Auftreten von Fehlern im Rahmen des ETL-Prozesses der Regelfall ist. Mit Mechanismen wie dem Verwenden von Ereignishandlern können Sie das Auftreten von Fehlern aber kontrollieren und schneller darauf reagieren, als wenn Sie diese mit selbst geschriebenen T-SQL-Anweisungen und gespeicherten Prozeduren selbst behandeln müssten.

Sie können mehrere solcher Datenflusstasks innerhalb eines SSIS-Projekts definieren und damit umfangreiche ETL-Prozesse ablaufen lassen. Datenflusstasks sind zudem in SSIS-Pakete organisiert, die Sie ebenfalls über den Projektmappen-Explorer verwalten können. Am Ende der Erstellung Ihrer Datenflusstasks können Sie den gesamten ETL-Prozess über das oder die SSIS-Pakete ausführen. Klicken Sie dazu im Projektmappen-Explorer auf das gewünschte SSIS-Paket und klicken Sie im Kontextmenü auf den Eintrag *Paket ausführen*.

Nach dem erfolgreichen Durchlaufen der Extraktion, Transformation und des Ladens der Daten in das Data Warehouse können Sie sich mit der Erstellung von OLAP-Strukturen wie Measures, Dimensionen und Cubes beschäftigen. Diese basieren typischerweise auf den Daten des Data Warehouses, die Sie mithilfe von SSIS importiert haben.



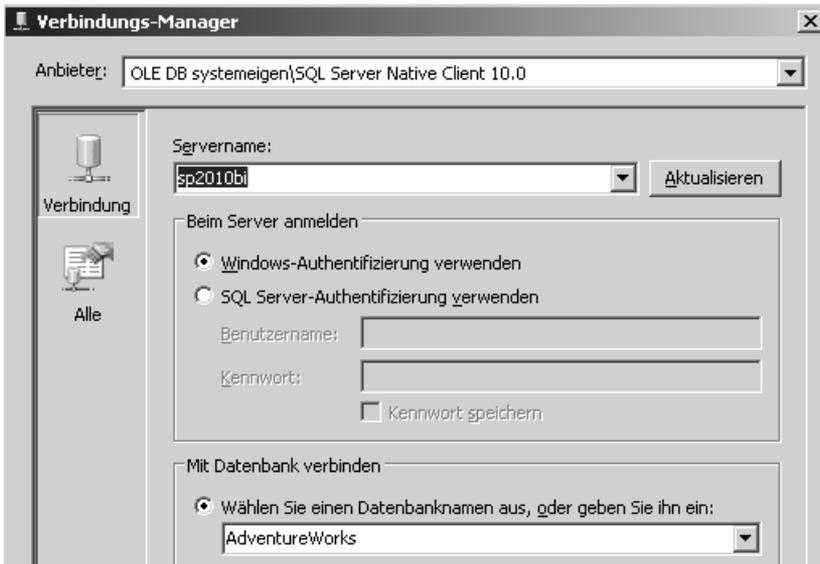
**Abbildung 3.8** Verbinden Sie im Entwurfsfenster Datenquellen mit Transformationen und schließlich mit Datenzielen

## Measures, Dimensionen und Cubes erstellen

Sie können zur Erstellung eines neuen Data Warehouse ein neues SSAS-Projekt mit BIDS starten oder aber ein bestehendes wie z.B. *Adventure Works DW* verwenden. In beiden Fällen haben Sie die Möglichkeit, neue Datenquellen, Datenquellsichten, Dimensionen und Cubes zu erstellen. Sobald Sie dieses Projekt bereitstellen, wird das Data Warehouse mit den neuen OLAP-Strukturen aktualisiert bzw. das Data Warehouse neu erstellt. Sie können Ihr Data Warehouse auch mithilfe von SSMS betrachten und im Objekt-Explorer die einzelnen Dimensionen, Cubes und deren Measuregruppen darstellen und durchsuchen. Für das Durchsuchen eines Cubes gibt es eine eigene Darstellungsweise im mittleren Bereich von SSMS, die Sie zu einem späteren Zeitpunkt in diesem Abschnitt kennenlernen werden.

Bevor Sie mit der Erstellung eines Cubes in SSIS beginnen, sollten Sie zuerst eine Datenquelle und danach eine oder mehrere Dimensionen erstellen, die Sie für die Erstellung des Cubes verwenden können. Die Datenquelle dient zum Import von Faktentabellen, die Sie später bei der Erstellung eines Cubes mit den Dimensionen zusammen verwenden können. Die Erstellung aller mehrdimensionalen Objekte beginnt im Projektmappen-Explorer, in dem Sie zuerst die Kategorie auswählen, für die Sie ein Objekt erstellen möchten.

Wenn Sie die *Adventure Works DW*-Beispieldatenbank verwenden (für die nachfolgenden Beispiele empfohlen), sehen Sie bereits eine Reihe vordefinierter Datenquellen, Datenquellsichten, Dimensionen und Cubes.



**Abbildung 3.9** Wählen Sie Datenbankserverinstanz und die Datenbank aus, die Sie analysieren möchten

Um eine neue Datenquelle zu erstellen, klicken Sie mit der rechten Maustaste auf die Kategorie *Datenquellen* im Projektmappen-Explorer. Klicken Sie im Kontextmenü auf den Eintrag *Neue Datenquelle*. Daraufhin öffnet sich der Datenquellen-Assistent, mit dem Sie Datenverbindungen herstellen können (Abbildung 3.9). Mit dem Datenquellen-Assistenten stellen Sie Verbindungen zu existierenden Datenbanken (unter Angabe des Datenbankservers und der zu verwendenden Datenbank) her und können diese später verwenden, um Faktentabellen im Data Warehouse zu erstellen.

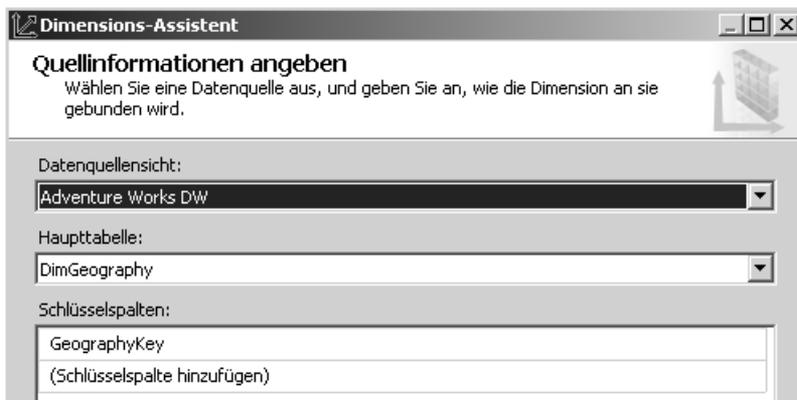
Optional können Sie für eine Datenquelle auch eine Datenquellsicht erstellen, mit der Sie Tabellen auswählen, die Sie importieren und verwenden möchten. Auf diese Weise können Sie nur eine Untermenge einer Datenquelle verwenden, wenn z.B. nicht alle Tabellen einer Datenbank für die Datenanalyse interessant sind. Vom Konzept her können Sie Datenquellsichten mit Tabellsichten in SQL vergleichen: Sie als Autor entscheiden, welche Informationen sichtbar sein sollen. Auch für Datenquellsichten gibt es einen analogen Assistenten, den Sie über die entsprechende Kategorie im Projektmappen-Explorer starten können.

Nachdem Sie eine Datenquelle und optional eine oder mehrere Datenquellsichten erstellt haben, können Sie sich dem Erstellen von Dimensionen zuwenden. Auch hier gilt: Falls Sie *Adventure Works DW* als SSAS-Beispieldatenbank verwenden, können Sie einfach die existierenden Dimensionen betrachten und diese gegebenenfalls verändern.

In der Kategorie *Dimensionen* können Sie mit der rechten Maustaste das Kontextmenü öffnen und dort den Eintrag *Neue Dimension* auswählen. Daraufhin öffnet sich der Dimensions-Assistent, der Sie in mehreren Schritten durch die Erstellung einer neuen Dimension begleitet. Beim Durchlaufen müssen Sie die folgenden Informationen der neuen Dimension festlegen:

- Typ
  - Basierend auf existierender Tabelle
  - Alternativ: Erstellen neuer Zeittabelle
- Datenquellensicht
- Zu verwendende Tabelle (Abbildung 3.10)
- Attribute (Abbildung 3.11)
- Name der neuen Dimension

Nach dem Fertigstellen des Assistenten erscheint die neue Dimension mit der Endung *.dim* in der Kategorie *Dimensionen* im Projektmappen-Explorer. Sie ist nun bereit für eine Verwendung durch den Cube-Assistenten (zugleich müssen Sie mindestens eine Dimension definiert haben, um einen Cube erstellen zu können).



**Abbildung 3.10** Wählen Sie die Datenquellensicht aus und die Tabelle, die für den Import von Dimensionsdaten dienen soll

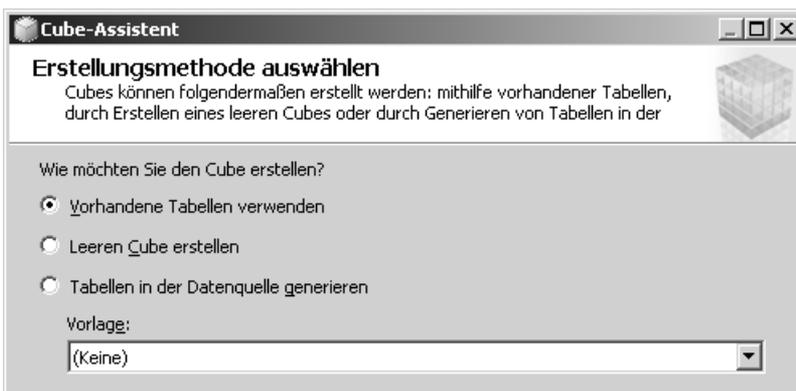
Nach dem Erstellen von neuen Dimensionen können Sie sich der Erstellung eines neuen Cubes widmen oder auch alternativ die bestehenden Cubes der *Adventure Works DW* betrachten und bei Bedarf verändern. Die Erstellung eines neuen Cubes passiert ebenfalls durch einen Assistenten und wird im Normalfall ebenfalls durch einen Business Intelligence-Entwickler vorgenommen.

Für die spätere Erstellung von Analysediagrammen und Business Intelligence-Objekten, die auf SQL Server Analysis Services basieren, werden Sie diese Objekte wiederverwenden können, wenn Sie die Assistenten für die Erstellung von Dimensionen und Cubes durchlaufen. Sie können so das erworbene konzeptuelle Wissen festigen und erhalten einen bleibenden Eindruck, wie die SSAS-Daten aufgebaut sind, und welche Wechselwirkungen sie untereinander haben.



**Abbildung 3.11** Legen Sie hier die einzelnen Attribute der Dimension fest. Sie werden zu einem späteren Zeitpunkt, im Rahmen einer Analyse, diese Attribute darstellen können.

Der finale Zweck liegt darin, dass Sie mit diesem Wissen und den Eindrücken der praktischen Beispiele die OLAP-Strukturen optimal einsetzen können, um Geschäftsanalysen in Form von Diagrammen und Dashboards vorzubereiten und für Ihre Geschäftsanwender bereitzustellen.



**Abbildung 3.12** Wählen Sie hier als Erstellungsmethode *Vorhandene Tabellen verwenden* aus

Zum Starten des Cube-Assistenten klicken Sie in gewohnter Manier im Projektmappen-Explorer von BIDS auf die Kategorie *Cubes* und wählen im Kontextmenü den Eintrag *Neuer Cube* aus. Die Abbildung 3.12 zeigt den ersten Schritt des Cube-Assistenten, mit dem Sie prinzipiell folgende Informationen festlegen:

- Typ des Cubes
  - Basierend auf existierenden Tabellen
  - Leerer Cube
  - Tabellen in der Datenquelle erstellen

- Tabellen, die als Measurgruppen verwendet werden sollen
- Spalten, die als Measures verwendet werden sollen
- Dimensionen (existierend)
- Dimensionen (neu)

Mit dem Cube-Assistenten legen Sie also die zugrunde liegende Datenquelle fest und welche Tabellen als Measuregruppen verwendet werden sollen. In den weiteren Schritten bestimmen Sie noch, welche Spalten der ausgewählten Tabellen Sie als Measures verwenden möchten. In den letzten beiden Schritten wählen Sie noch Dimensionen aus, die Sie bereits erstellt haben, und können zusätzlich auch neu zu erstellende Dimensionen festlegen.

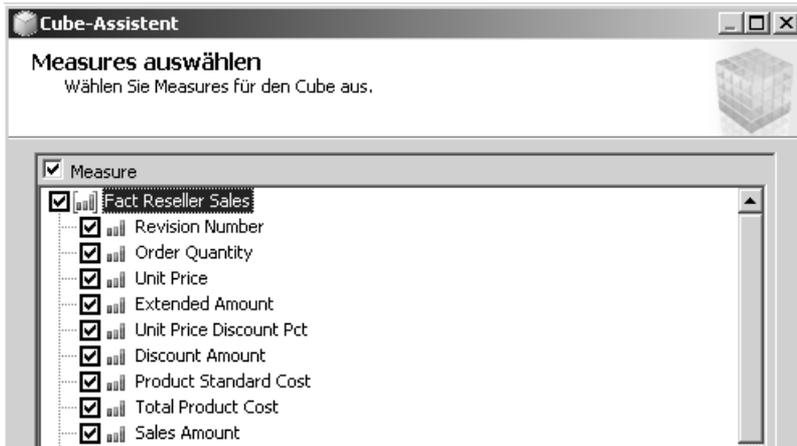
Mit dem Abschluss des Cube-Assistenten ist die Erstellung eines Cubes aber noch nicht abgeschlossen. Schließlich müssen Sie dafür noch definieren, welche Measures mit welchen Dimensionen in Verbindung stehen. Damit legen Sie direkt fest, welche Analysemöglichkeiten ein Endanwender später haben wird. Sie sehen also: Ihre Aufgabe als Dashboardautor ist stark davon abhängig, wie gut und wie detailliert die Cubes von Business Intelligence-Entwicklern zur Verfügung gestellt werden. Eine transparente Zusammenarbeit und gemeinsame Definition der zu erreichenden Analyseziele ist unabdingbar für die Erstellung der gewünschten Datenbasis und Geschäftsdatenanalysen. Nur wenn dieses gemeinsame Verständnis zwischen Business Intelligence-Entwicklern, Dashboardautoren und Geschäftsanwendern herrscht, kann die Bereitstellung und Konsumierung von Business Intelligence im Unternehmen funktionieren.

Damit nimmt auch die Kommunikation zwischen diesen Anwendergruppen eine zentrale Rolle im gesamten Lebenszyklus einer Business Intelligence-Implementierung ein. Es muss ein einheitliches Verständnis zwischen der Geschäftsführung, den Business Intelligence-Entwicklern, den Dashboardautoren und letztlich den Geschäftsanwendern herrschen. Nur so kann gewährleistet werden, dass die Informationen, die zusammengetragen, als Cubes modelliert und schließlich analysiert werden, auch die richtigen Fragen beantworten und mit der Strategie des Unternehmens einhergehen.

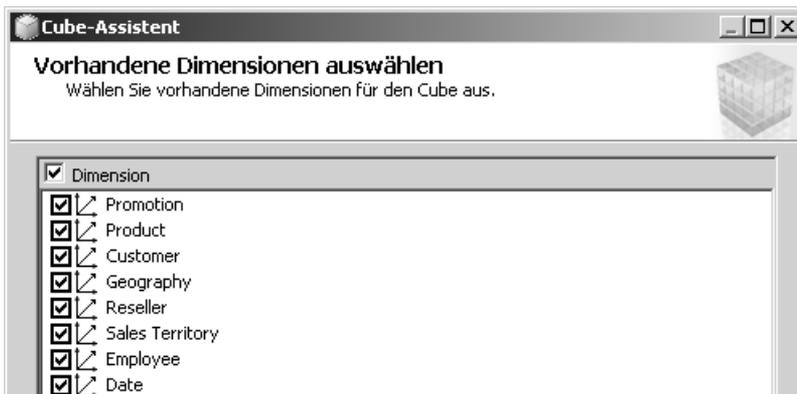


**Abbildung 3.13** Wählen Sie in diesem Dialogfeld des Cube-Assistenten die Faktentabellen aus, die als Grundlage für die Measuregruppen dienen sollen. Die Measures, die Sie in einem späteren Schritt des Cube-Assistenten auswählen können, entsprechen den Spalten dieser Tabellen.

Die Measures wählen Sie im Rahmen des Cube-Assistenten aus den verfügbaren Measuregruppen aus. Der Cube-Assistent schlägt Ihnen dabei nur diejenigen Spalten vor, die auch tatsächlich auf numerischen und damit quantifizierbaren Spalten basieren. Business Intelligence-Entwickler müssen sich also nicht darum kümmern, nicht-numerische Spalten auszublenden; BIDS übernimmt das für sie. Die existierenden Dimensionen sind so aufgelistet, wie Sie sie mit Business Intelligence Development Studio erstellt haben. Daneben können Sie noch neue Dimensionen erstellen lassen: Anhand der erstellten Tabellen stellt Business Intelligence Development Studio selbständig fest, welche nicht-numerischen Spalten der Dimensionen als Spalten fungieren können. So wird bei der Business Intelligence-Entwicklung sichergestellt, dass im Rahmen des Assistenten alle verfügbaren Dimensionen visualisiert werden und somit keine mögliche Dimension vergessen wird.



**Abbildung 3.14** Wählen Sie hier die Measures aus, die später in Berichten von Geschäftsanwendern darstellbar sein sollen



**Abbildung 3.15** Wählen Sie hier vorhandene Dimensionen aus. Im nächsten Schritt können Sie neue Dimensionen hinzufügen, die vom Assistenten vorgeschlagen werden.

Im Entwurfswindow werden Fakten- und Dimensionstabellen unterschiedlich dargestellt, damit Sie dies auf den ersten Blick als solche erkennen können und auch bei späterer Betrachtung den Überblick über die Beziehungen erhalten können. So wird die Titelleiste von Faktentabellen standardmäßig gelb und die der Dimensionentabelle blau hinterlegt. Zudem können Sie bei korrekter Benennung auch anhand des Namens der Tabellen ableiten, um welchen Typ es sich handelt. Faktentabellen sollten mit dem Präfix *Fact* beginnen, Dimensionentabellen dagegen mit dem Präfix *Dim*.

Die eigentlichen Beziehungen zwischen den beiden Tabellentypen werden mit Pfeilen dargestellt. Sie können eine solche Beziehung einfach herstellen, indem Sie ein Feld aus einer Tabelle per Drag & Drop auf das zugehörige Feld einer anderen Tabelle ziehen.

Um ganz sicher zu gehen, sollten Sie nach Erstellung einer solchen Beziehung per Doppelklick auf den Verbindungspfeil in der Datenquellsicht klicken. Es öffnet sich das Dialogfeld *Bezie-*

*hung bearbeiten.* Hier können Sie festlegen, welche beiden Tabellen für die Beziehung herangezogen werden und über welche Felder diese zusammenhängen.

**ACHTUNG** In der Cube-Entwurfsansicht sehen Sie die verschiedenen verknüpften Fakten- und Dimensionentabellen zwar zusammen mit deren Beziehungen, können diese aber dort nicht verändern.

Die Änderungen in der Entwurfsansicht der Datenquellensicht wirken sich unmittelbar auf die Entwurfsansicht des Cubes aus: Fügen Sie beim Entwurf der Datenquellensicht eine Beziehung hinzu, ist diese auch in der Ansicht des Cubes sofort sichtbar (wenngleich Sie die Beziehung dort auch nicht bearbeiten können, sondern nur beim Entwurf der Datenquellensicht).

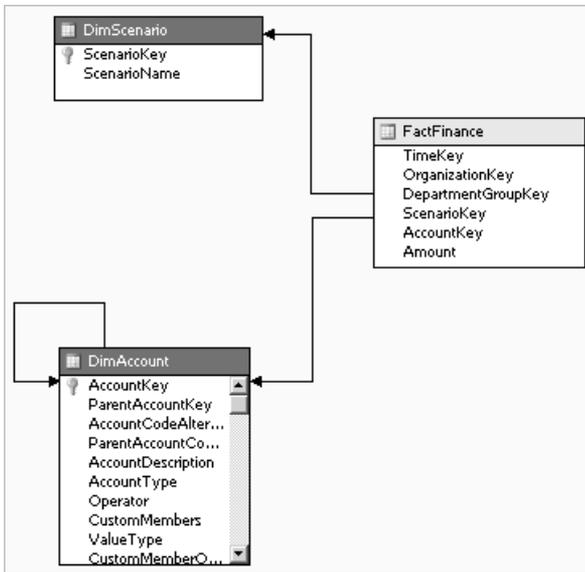
Was hingegen sehr wohl in der Entwurfsansicht eines Cubes möglich ist, ist die Erstellung neuer Measures aufgrund vorhandener Spalten. Das Entwurfswindow des Cubes ist zudem aufgeteilt in drei Bereiche:

- **Measures** Hier können Sie neue Measures definieren und existierende auswählen, bearbeiten und in das Diagramm des Cubes einfügen
- **Dimensionen** Hier können Sie vorhandene Dimensionen sehen und weitere zum Diagramm in der Entwurfsansicht hinzufügen, um diese darzustellen
- **Datenquellenansicht** Hier sehen Sie die verwendeten Measures und Dimensionen bzw. deren Tabellenansicht und die Beziehungen zwischen den Tabellen

Daneben gibt es noch unzählige Möglichkeiten, die Analysefähigkeiten von Cubes zu erweitern und bereits vordefinierte Business Intelligence-Objekte zu erstellen:

- **Key Performance Indicators** Kennzahlen mit Zielwerten, die grafisch einen schnellen Überblick über die Leistung einer bestimmten Kennzahl bereitstellen. KPIs werden in Business Intelligence Development Studio mithilfe der multidimensionalen Abfragesprache Multidimensional Expressions (MDX) erstellt.
- **Aktionen** Hierbei handelt es sich um in SQL Server Analysis Services gespeicherte Aktionen in Form von MDX-Befehlen. Diese können in Clientanwendungen integriert und von den Endanwendern verwendet werden.
- **Aggregationen** Im Vorfeld berechnete und zusammengefasste numerische Measures. Durch die Definition von Aggregationen bestimmter Measures unter Einschränkung auf bestimmte Dimensionen verbessern Sie die Abfrageleistung für diese berechneten Werte enorm.
- **Perspektiven** Perspektiven ermöglichen die anfängliche Einschränkung eines komplexen Cubes auf einige bestimmte Dimensionen und Measures. Auf diese Weise wird den Anwendern die Möglichkeit geboten, nicht den ganzen Cube zu durchsuchen, sondern lediglich bestimmte Bereiche, die für sie von Interesse sind – ohne sich dabei in der Komplexität des gesamten Cubes zu verlieren.
- **Übersetzungen** Sie können für von Ihnen gewählte Bezeichnungen und Beschriftungen eines Cubes Übersetzungen erstellen und diese so für mehrere Sprachen zugänglich machen und entsprechend in Berichten und Dashboards darstellen. Dies ist insbesondere für Unter-

nehmen interessant, die Berichte für Geschäftsanwender aus mehreren Ländern veröffentlichen wollen, also multinationale Konzerne.



**Abbildung 3.16** Verändern Sie die Beziehungen von Fakten- und Dimensionentabellen nicht in der hier dargestellten Cubeansicht, sondern in der Entwurfsansicht der Datenquellensicht

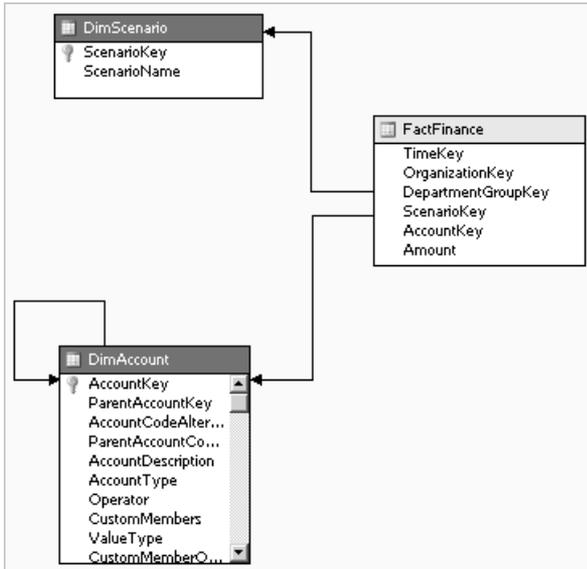
#### TIPP

Wenn Sie sich näher mit den Fähigkeiten von Business Intelligence Development Studio und SQL Server Analysis Services auseinandersetzen möchten, finden Sie unter <http://msdn.microsoft.com/de-de/library/ms170208.aspx> das SQL Server Analysis Services-Lernprogramm. Hier erfahren Sie Schritt für Schritt, wie Sie mit Measures, Dimensionen, Cubes, Perspektiven, Übersetzungen usw. arbeiten.

Der Business Intelligence-Lebenszyklus endet noch nicht mit der Definition der OLAP-Strukturen. Damit stellen Sie lediglich die Datenbasis für aussagekräftige Geschäftsdatenanalysen bereit. Im nächsten Schritt geht es nun darum, die gewonnenen Metadaten auszuwerten und für Geschäftsanwender darzustellen. Diese Auswertung der Daten kann auf vielfältige Art und Weise erfolgen, unter anderem mit:

- Microsoft Excel
- Dashboards
- Diagrammen
- SSRS-Berichten
- Exportierte Daten in Form von *.pdf*- oder *.xls*-Dateien usw.

Sie können aber auch SQL Server Management Studio dazu verwenden, um Cubes schnell zu durchsuchen und nach bestimmten Informationen zu suchen. Öffnen Sie hierfür SSMS und verbinden Sie sich mit der *Adventure Works DW*-Beispieldatenbank.



**Abbildung 3.17** Durchsuchen Sie Cubes mit SQL Server Management Studio und exportieren Sie die Resultate in Standardformate

Wählen Sie für den Verbindungstyp im Dialogfeld *Verbindung herstellen* über das entsprechende Dropdownlistenfeld den Verbindungstyp *Analysis Services* aus.

Im Objekt-Explorer finden Sie nun eine Liste der Measures, Dimensionen und Cubes. Die Cubes können Sie durchsuchen. Sie sollten einen Cube aus dem Objekt-Explorer auswählen und mit der rechten Maustaste darauf klicken. Anschließend wählen Sie im Kontextmenü den Eintrag *Durchsuchen* aus. Daraufhin öffnet sich ein Fenster entsprechend der Abbildung 3.17, in dem Sie verschiedene Measures und Dimensionen zur Darstellung eines Ad-hoc-Berichts verwenden können. Sie können diese auf die folgenden Elemente des Ad-hoc-Berichts ziehen:

- Filter
- Spalten
- Zeilen
- Numerische Werte

Die Informationen werden sofort auf tabellarische Art und Weise dargestellt und bei Änderungen aktualisiert. Diese Berichte sind sicherlich nicht dafür gedacht, ausführliche Geschäftsdatenanalysen durchzuführen. Das schnelle Durchsuchen und Filtern nach bestimmten Dimensionen und Measures ist aber eine gute Möglichkeit, mit wenig Zeitaufwand die Inhalte eines Cubes anzuzeigen und eine schnelle Auswertung von Geschäftsdaten vorzunehmen. Obwohl die Ansicht der Cubedaten über den Menübefehl *Durchsuchen* nicht die Weitergabe oder gar Veröffentlichung dieser Ansichten ermöglicht, können Sie die Resultate trotzdem in verschiedene Dateiformate exportieren und anschließend per E-Mail oder Dateifreigabe verteilen.

Hier gilt wie für alle Business Intelligence-Objekte: Ob Sie KPIs erstellen, Berichte oder Dashboards, wie im vorangegangenen Kapitel beschrieben, können Sie verschiedene Werkzeuge zur Erstellung der gleichen Business Intelligence-Objekte verwenden. Diese Werkzeuge können Sie für Geschäftsanwender transparent darstellen, d.h., diese merken nicht notwendigerweise,

womit Sie Ihre Berichte und KPIs erstellt haben. Je nachdem, wie komplex Ihre Anforderungen sind, kann es aber unter Umständen von Vorteil sein, ein Erstellungswerkzeug einem anderen vorzuziehen. Dies wird besonders am Beispiel von KPIs deutlich.

	Business Intelligence Development Studio	Dashboard-Designer
Kennzahlauswahl aus Datenquelle	Ja	Ja
Kennzahl und Zielwertdefinition	MDX	Grafisch
Vergleichsdefinition	MDX	Grafisch
Auswahl Statusindikatoren	Wenige	Viele
Verwendung in PerformancePoint-Dashboards	Ja	Ja

**Tabelle 3.1** Wählen Sie Ihr Werkzeug für die Erstellung von KPIs anhand des Komplexitätsgrads aus

Sie sehen also in Tabelle 3.1, dass die Erstellung von KPIs je nach gewähltem Werkzeug unterschiedlich komplex ausfallen können. Dementsprechend ist die Definition eines KPI per MDX um einiges schwieriger als die grafische Zuweisung von Vergleichen und Indikatoren. Andererseits aber ist die Abfragesprache MDX hinsichtlich komplexer Ausdrücke wesentlich flexibler. Es kann also sein, dass Sie für die Erstellung eines KPIs den Dashboard-Designer einsetzen können, für die Erstellung eines anderen KPIs hingegen Business Intelligence Development Studio einsetzen *müssen*, da Sie sonst einen komplexen Vergleich nicht ziehen könnten.

Dieses Beispiel spiegelt perfekt die Wahlmöglichkeiten für die Erstellung von Business Intelligence-Objekten für die Geschäftsdatenanalyse wieder: Sie müssen stets anhand Ihrer Anforderungen entscheiden, welches Werkzeug Sie für welches Objekt nutzen. Die geschickte Wahl des richtigen Werkzeugs kann Ihnen bei der Erstellung und Bereitstellung viel Zeit ersparen. Und gerade diese Zeitersparnis steht bekanntlich im Sinne der zeitgerechten Informationsverteilung besonders im Vordergrund, damit Geschäftsentscheider schnell auf Änderungen im Unternehmen reagieren können.

Im weiteren Verlauf dieses Buchs werden wir uns mit den Erstellungswerkzeugen für die Auswertung von Geschäftsdaten beschäftigen und Ihnen einen Einblick geben, wie Sie diese am effizientesten einsetzen können.

## Zusammenfassung

Im vorangegangenen Kapitel haben Sie einen Einblick darüber erhalten, welche Möglichkeiten für die Erstellung von Business Intelligence-Objekten in der Microsoft Business Intelligence-Plattform zur Verfügung stehen und wann Sie idealerweise welches Werkzeug einsetzen sollten. In diesem Kapitel haben Sie nun einen konzeptuellen Überblick darüber erhalten, wie mehrdimensionale Daten in SQL Server Analysis aufgebaut sind, und wie Business Intelligence-Entwickler grundsätzlich vorgehen, um diese im Rahmen des Business Intelligence-Lebenszyklus zu

erstellen. Die verschiedenen behandelten Themen finden Sie hier noch einmal in der Schnellübersicht:

- Überblick über die logische Organisation von Geschäftsdaten
- Konzepte von Measures, Dimensionen und Cubes
- Aufbau von SQL Server Analysis Services (SSAS)
- Business Intelligence-Lebenszyklus
- Definition von Measures, Dimensionen und Fakten

Nachdem Sie nun in den ersten drei Kapiteln einen Einblick in die grundlegenden Möglichkeiten der Microsoft Business Intelligence-Plattform, in den Aufbau mehrdimensionaler Daten und in die Auswahl der geeigneten Business Intelligence-Werkzeuge erhalten haben, ist es nun an der Zeit, dass Sie selbst die Business Intelligence-Erstellungswerkzeuge praktisch einsetzen. In den folgenden Kapiteln erfahren Sie, wie Sie Erstellungswerkzeuge wie Microsoft Excel 2010 und Excel Services einrichten und einsetzen, um Business Intelligence für Geschäftsanwender greifbar zu machen.