

MICHAEL H.

# SAP Business I Warehouses

Mehrdimensionale

# **SAP Business Information Warehouse**

Mehrdimensionale Datenmodellierung

Michael Hahne

# **SAP Business Information Warehouse**

Mehrdimensionale Datenmodellierung

Mit 120 Abbildungen und 19 Tabellen



Springer

Dr. Michael Hahne  
cundus AG  
Freiherr-vom-Stein-Straße 13a  
55559 Bretzenheim  
michael.hahne@cundus.de

Bibliografische Information Der Deutschen Bibliothek  
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;  
detaillierte bibliografische Daten sind im Internet über <<http://dnb.ddb.de>> abrufbar.

ISBN 3-540-22015-1 Springer Berlin Heidelberg New York

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland vom 9. September 1965 in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtsgesetzes.

Springer ist ein Unternehmen von Springer Science+Business Media  
[springer.de](http://springer.de)

© Springer-Verlag Berlin Heidelberg 2005  
Printed in Germany

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Umschlaggestaltung: Erich Kirchner, Heidelberg

SPIN 11008743

42/3130-5 4 3 2 1 0 – Gedruckt auf säurefreiem Papier

# Vorwort

Im Kern von Business Intelligence und analyseorientierten Informationssystemen stehen oftmals Konzepte des Data Warehousing, OLAP und Data Mining. Die damit verbundene konsequente Nutzung des Informationspotenzials einer Unternehmung eröffnet für strategische und dispositive Prozesse u. a. in den Bereichen Controlling, Vertrieb und Marketing völlig neue Perspektiven. Beispiele hierfür sind Anwendungen des Risikomanagements, der Prozessoptimierung, der Konsolidierung von Finanzdaten und des Customer Relationship Management.

Das SAP Business Information Warehouse (BW) als Data Warehouse-Lösung ermöglicht eine zeitnahe Versorgung betrieblicher Entscheidungsträger mit relevanten Informationen zu Analysezwecken. Basis hierfür sind die mehrdimensionalen Info-Cubes, deren Modellierung maßgeblich die Leistungsfähigkeit und erfolgreiche Nutzung eines BW-Systems beeinflusst.

Es gibt schon eine Reihe von zumeist englischen Büchern zum SAP Business Information Warehouse sowie generell zum Thema Datenmodellierung. Dieses Buch setzt folgende besondere Akzente:

- Es wird ein durchgängiges Beispiel zur Illustration verwendet, das einerseits komplex genug ist, um alle Facetten der Modellierung beleuchten zu können, andererseits aber aufgrund seiner Einfachheit sehr einprägsam ist.
- Der Modellierungsprozess wird über die semantische, logische und physische Ebene hinweg ganzheitlich betrachtet.
- Die ausführlichen Gestaltungsempfehlungen bieten reichhaltiges Potenzial zur Optimierung von bestehenden mehrdimensionalen Modellen und dem zielgerichteten Aufbau neuer Datenmodelle des Business Information Warehouse.
- Die Anforderungen der potenziellen Benutzer eines BW-Systems bilden einen zentralen Ausgangspunkt der Betrachtung, insbesondere erfolgt eine detaillierte Diskussion der Aspekte der Zeitabhängigkeit in der Modellierung und der möglichen temporalen Berichtsszenarien.

- Die meisten Kapitel schließen mit einer Zusammenfassung der dargestellten Konzepte und Prinzipien sowie einem Verweis auf interessante weiterführende Literatur.
- Die Darstellung wird bewusst nicht mit Screenshots aus dem System aufgebläht, sondern konzentriert sich auf die Vermittlung von Konzepten, Prinzipien und Zusammenhängen. Das Handling innerhalb des Business Information Warehouse u. a. in der Administrator Workbench ist Bestandteil einer Einweisung in die Systembedienung und darüber hinaus stellt die umfangreiche Dokumentation des Systems hierzu eine zentrale Informationsquelle bereit.

Dieses Buch verfolgt das Ziel, den gesamten Prozess der mehrdimensionalen Modellierung von Informationssystemen auf Basis des Business Information Warehouse darzustellen, adäquate Methoden zu präsentieren und Gestaltungsempfehlungen zu geben.

Das Buch wendet sich an SAP-Berater und IT-Mitarbeiter, die mit der Implementierung eines BW-Systems betraut sind. Es ist als praktischer Leitfaden konzipiert und bietet neben dem Einstieg in das Thema auch ein fundiertes Nachschlagewerk für die tägliche Arbeit.

Zahlreiche Data Warehouse-Projekte und Diskussionen mit Kolleginnen und Kollegen eröffneten mir neue Sichtweisen und neue Erkenntnisse, die in die Erstellung dieses Buches mit einfließen. Meinen Kollegen Ingo Claas und Frank Hein danke ich für die konstruktiven Verbesserungsvorschläge und die Durchsicht des Manuskriptes.

Mein besonderer Dank gilt meiner Frau Sabine und meinen Töchtern Julia und Leonie, welche diese Arbeit in vielfältiger Weise begleitet haben.

Michael Hahne

# Inhaltsverzeichnis

<b>Vorwort.....</b>	<b>V</b>
<b>1 Einleitung.....</b>	<b>1</b>
<b>2 Analyseorientierte Informationssysteme .....</b>	<b>5</b>
2.1 Management Support Systeme .....	5
2.2 Data Warehouse.....	7
2.3 OLAP und mehrdimensionale Datenbanken .....	10
2.4 Weiterführende Literatur .....	12
2.5 Zusammenfassung .....	13
<b>3 Datenmodellierung und Mehrdimensionalität .....</b>	<b>15</b>
3.1 Datenmodelle und Datenmodellierung .....	15
3.2 Mehrdimensionale Datenstrukturen.....	18
3.2.1 Grundbestandteile mehrdimensionaler Datenstrukturen.....	19
3.2.2 Dimensionshierarchien .....	23
3.2.3 Hierarchie-Attribute.....	28
3.3 Kennzahlen und deren Berechnung .....	30
3.3.1 Kennzahlen und Kennzahlensysteme .....	31
3.3.2 Kennzahlen im mehrdimensionalen Modell .....	32
3.3.3 Additivitätseigenschaften .....	34
3.4 Temporale Aspekte .....	34
3.5 Weiterführende Literatur .....	38
3.6 Zusammenfassung .....	39
<b>4 Architektur des SAP Business Information Warehouse .....</b>	<b>41</b>
4.1 Komponenten der Datenspeicherung im BW .....	43
4.1.1 Info-Objekte.....	43
4.1.2 Persistant Staging Area.....	44
4.1.3 Extraktion und Transformation.....	45
4.1.4 Operational Data Store .....	47
4.1.5 Info-Cubes .....	48
4.1.6 Varianten der Datenbereitstellung .....	49

4.2 Auswertung und Analyse auf Basis vom BW.....	52
4.2.1 Queries als mehrdimensionale Schemata .....	53
4.2.2 Business Explorer Analyzer .....	57
4.2.3 BEx Web Applications .....	58
4.3 Weiterführende Literatur .....	62
4.4 Zusammenfassung .....	62
<b>5 Anwendungsbeispiel .....</b>	<b>65</b>
5.1 Ausgangssituation.....	65
5.2 Projektdefinition und Projektziel .....	66
5.3 Fachkonzept der Fallstudie .....	67
5.3.1 Beschreibung der Geschäftsprozesse.....	68
5.3.2 Datenquellenbeschreibung.....	69
5.3.3 Skizzierung des Informationsbedarfs.....	72
<b>6 Semantische mehrdimensionale Modellierung.....</b>	<b>75</b>
6.1 Methoden mehrdimensionaler semantischer Datenmodellierung...75	
6.1.1 Entity Relationship-Modelle.....	75
6.1.2 Mehrdimensionales ER-Modell (ME/R) .....	77
6.1.3 Dimensional Fact Modeling .....	79
6.1.4 Kennzahlendatenmodell .....	81
6.1.5 Multidimensional Data Model ( $\mathcal{MD}$ ) .....	82
6.2 Semantische Modellierung mit ADAPT.....	83
6.2.1 Dimensionsmodellierung.....	84
6.2.2 Modellierung von Cubes.....	89
6.3 Erweiterung von ADAPT zu T-ADAPT.....	90
6.4 Semantisches Modell des Beispiels .....	94
6.5 Weiterführende Literatur .....	101
6.6 Zusammenfassung .....	102
<b>7 Allgemeines Star Schema .....</b>	<b>105</b>
7.1 Grundform des Star Schemas .....	105
7.1.1 Modellierung von Dimensionshierarchien.....	106
7.1.2 Attribute in Dimensionen .....	111
7.1.3 Normalisierung von Dimensionen .....	112
7.1.4 Abbildung von Kennzahlen und Kennzahlensystemen .....	114
7.2 Aggregate.....	115
7.2.1 Dynamische Berechnung.....	116
7.2.2 Vorberechnete Aggregate in der Faktentabelle .....	116
7.2.3 Fact Constellation Schema .....	118
7.2.4 Snow Flake Schema.....	119



---

7.3 Weitere Modellierungsvarianten.....	121
7.3.1 Galaxien.....	121
7.3.2 Faktenlose Faktentabellen .....	122
7.3.3 Minidimensionen .....	123
7.3.4 Hierarchien über rekursive Beziehungen.....	125
7.3.5 Temporale Aspekte im Star Schema.....	127
7.4 Weiterführende Literatur .....	128
7.5 Zusammenfassung .....	129
<b>8 Logisches Datenmodell des BW .....</b>	<b>131</b>
8.1 Erweitertes Star Schema der SAP.....	131
8.2 Modellierungsvarianten hierarchischer Dimensionsstrukturen ....	136
8.2.1 Hierarchische Beziehungen zwischen Merkmalen .....	137
8.2.2 Navigationsattribute als Basis für hierarchische Strukturen..	138
8.2.3 Externe Hierarchien in den Stammdaten .....	139
8.3 Abbildung zeitabhängiger Berichtsszenarien .....	142
8.3.1 Aktuelle Struktur .....	142
8.3.2 Historische Struktur .....	145
8.3.3 Transaktionsorientierte Sichtweise.....	148
8.3.4 Vergleichbare Resultate.....	150
8.4 Modellierung von Kennzahlen.....	152
8.5 Grafische Repräsentation logischer BW-Modelle .....	153
8.5.1 Darstellung von Dimensionen .....	154
8.5.2 Abbildung von hierarchischen Strukturen .....	157
8.5.3 Modellierung von Info-Cubes.....	158
8.6 Ableitung eines BW-Modells für das Beispiel .....	160
8.6.1 Modell für den Controllingbereich .....	160
8.6.2 Aggregierte Marketingsicht.....	163
8.6.3 Modell für das Partner-Marketing .....	166
8.6.4 Modell des kundenbezogenen Marketings .....	169
8.7 Weiterführende Literatur .....	171
8.8 Zusammenfassung .....	172
<b>9 Aspekte des physischen Modells.....</b>	<b>175</b>
9.1 Multi-Provider .....	175
9.2 Komprimierung von Faktentabellen .....	178
9.3 Partitionierung von Faktentabellen .....	180
9.4 Aggregate.....	183
9.5 Weiterführende Literatur .....	185
9.6 Zusammenfassung .....	186

<b>10 Gestaltungsempfehlungen .....</b>	<b>189</b>
10.1 Modellierung von Dimensionsstrukturen .....	190
10.1.1 Aspekte der semantischen Modellierung .....	190
10.1.2 Empfehlungen auf logischer Modellebene .....	191
10.1.3 Aspekte des physischen Modells .....	194
10.2 Abbildung von Kennzahlen .....	196
10.3 Gestaltung von Info-Cubes .....	197
10.3.1 Grundprinzip der Cube-Modellierung .....	197
10.3.2 Modellierung auf der Fachanwender orientierten Ebene.....	199
10.3.3 Cube-Modellierung auf der physischen Ebene .....	202
 <b>Abkürzungsverzeichnis .....</b>	 <b>203</b>
 <b>Abbildungsverzeichnis.....</b>	 <b>205</b>
 <b>Tabellenverzeichnis.....</b>	 <b>209</b>
 <b>Literaturverzeichnis .....</b>	 <b>211</b>
 <b>Stichwortverzeichnis.....</b>	 <b>215</b>

# 1 Einleitung

Analyseorientierte Informationssysteme zielen auf die Unterstützung der dispositiven und strategischen Prozesse in Unternehmen ab und haben eine zeitnahe Versorgung betrieblicher Entscheidungsträger mit relevanten Informationen zu Analyse Zwecken im Fokus. Das *Business Information Warehouse (BW)*<sup>1</sup> als zentrale Data Warehouse-Lösung adressiert genau diesen Bereich und ist die Grundlage für vielfältige analytische Anwendungen.

Mit dem Business Information Warehouse ist die SAP AG in den Markt für Business Intelligence eingestiegen. Besonderes Leistungsmerkmal des BW ist der *Business Content* als eine Sammlung vorgefertigter betriebswirtschaftlicher Lösungen, die im BW neben den umfassenden Funktionalitäten zum Aufbau und Betrieb eines Data Warehouse zur Verfügung gestellt werden. Besonderes Augenmerk haben dabei Systemumgebungen, in denen SAP R/3 als vorgelagertes Quellsystem vorhanden ist.

Unter dem Begriff *mySAP Business Intelligence* sind Anwendungskomponenten und Dienstleistungen für die Entscheidungsunterstützung in Unternehmen zusammengefasst, deren Hauptkomponente das Business Information Warehouse als Basistechnologie zur Datenhaltung ist. Weitere Bestandteile sind u. a. der Business Explorer zur Datenauswertung und das Enterprise Portal als zentraler Zugangspunkt mit Single-Sign-On-Funktionalität zur Verteilung von Berichten und Informationen. Dies ist mittlerweile integraler Bestandteil von SAP NetWeaver. Das Strategic Enterprise Management, neuerdings Bestandteil der Lösung mySAP ERP (vormals mySAP Financials), beinhaltet Funktionen für Unternehmensplanung, Performance Management und Konsolidierung. Die Datenhaltungskomponente des SEM ist ebenfalls das BW.<sup>2</sup>

---

<sup>1</sup> SAP und Business Information Warehouse sind ebenso wie SAP R/3, NetWeaver und SEM (eingetragene) Warenzeichen der SAP Aktiengesellschaft Systeme, Anwendungen, Produkte in der Datenverarbeitung, Neurottstraße 16, D-69190 Walldorf. Die SAP AG ist nicht Herausgeber des vorliegenden Buches und auch nicht presserechtlich dafür verantwortlich.

<sup>2</sup> Die Gruppierung und Begriffsbildung unterliegt häufiger Änderungen von Seiten der SAP AG.

Für viele Lösungen aus dem Hause SAP ist das Business Information Warehouse die wesentliche technologische Basiskomponente zur Datenspeicherung, auf deren Grundlage unterschiedlichste analytische Anwendungen wie beispielsweise die Bewertung und Optimierung von Logistikketten oder Kundenbeziehungen aufbauen. Einen signifikanten Einfluss auf die Leistungsfähigkeit und Akzeptanz solcher Systeme hat die Modellierung, deren Möglichkeiten durch die Grenzen des zugrunde liegenden Modells limitiert sind.

Als zentrales Charakteristikum gewährleisten multidimensionale Sichtweisen auf unternehmensinterne und -externe Datenbestände brauchbare Näherungen an das mentale Unternehmensbild des Managers, denn in der Tat untersuchen Manager betriebswirtschaftlich relevante Sachverhalte unter Berücksichtigung zahlreicher Einflussfaktoren. So wird beispielsweise nach dem Absatz einer Produktgruppe in bestimmten Verkaufsregionen über einen definierten Zeitraum hinweg gefragt. Die Anordnung betriebswirtschaftlicher Variablen bzw. Kennzahlen, wie z. B. Umsatz oder Kostengrößen, erfolgt entlang unterschiedlicher Dimensionen, beispielsweise Kunden, Artikel und Regionen, und diese Strukturierung gilt als geeignete entscheidungsorientierte Sichtweise auf betriebswirtschaftliche Tatbestände. Bildlich gesprochen werden die quantitativen Kenngrößen in mehrdimensionalen Würfeln gespeichert, deren Kanten durch die einzelnen Dimensionen definiert und beschriftet sind.

Online Analytical Processing (OLAP) als Grundprinzip für den Aufbau von Systemen zur Unterstützung von Fach- und Führungskräften basiert im Kern auf einer mehrdimensionalen konzeptionellen Sicht auf die Daten mit Möglichkeiten der Navigation in den Würfeln mit beliebigen Projektionen und auf verschiedenen Verdichtungsstufen. Die Analysemöglichkeiten werden stark durch die Möglichkeiten zur Abbildung von Strukturen für Konsolidierungspfade beeinflusst. Der Modellierung dieser hierarchischen Dimensionsstrukturen kommt im Kontext analyseorientierter Informationssysteme eine besondere Bedeutung zu. Relevant sind ebenfalls der Umgang mit Aspekten der Zeitabhängigkeit und die Behandlung von strukturellen Änderungen in Dimensionshierarchien.

Eine geläufige Strukturierung des Modellierungsvorganges unterscheidet die Ebenen der semantischen, logischen und physischen Datenmodellierung. Ein Datenmodell soll die Bedeutung und Repräsentation von Daten beschreiben. Eine werkzeuggestützte Modellierung basiert auf der Möglichkeit, ausgehend von der semantischen Modellebene und losgelöst von möglichen Zielplattformen die fachkonzeptuelle Ebene abzubilden. In den weiteren Schritten wird dieses Modell dann auf die Ebenen des logischen und physischen Modells herunter gebrochen. Letztendlich ist es das Ziel, in festgelegten Datenbanksystemen Strukturen zu generieren.

Zur Modellierung von Informationssystemen auf Basis des SAP Business Information Warehouse (BW) stehen auf semantischer Ebene allgemein anerkannte Methoden zur Modellierung und grafischen Repräsentation zur Verfügung. Für die Abbildung des logischen Datenmodells des BW sind aber keine grafischen Darstellungsmethoden verfügbar. Oftmals erfolgt eine Modelldarstellung in tabellarischer Form in einer Tabellenkalkulation.

Das vorliegende Buch verfolgt das Ziel, den gesamten Prozess der mehrdimensionalen Modellierung von Informationssystemen auf Basis des Business Information Warehouse darzustellen, adäquate Methoden zu präsentieren und Gestaltungsempfehlungen zu geben. In Kapitel 2 erfolgt zunächst die Einordnung der Konzepte des Data Warehouse und OLAP. Eine begriffliche Aufarbeitung mehrdimensionaler Datenstrukturen ist Gegenstand des Kapitels 3. Eine einführende Darstellung der Architektur des SAP Business Information Warehouse steht in Kapitel 4 im Vordergrund.

Aufbauend auf dem in Kapitel 5 entwickelten Anwendungsbeispiel erfolgt die Diskussion der Modellierung auf den unterschiedlichen Ebenen, indem das Szenario immer wieder zur Verdeutlichung der Zusammenhänge und zur Illustration herangezogen wird. Nach der Darstellung der semantischen Modellierungsebene in Kapitel 6 ist zunächst ein Exkurs zur Darstellung der Bestandteile und Varianten des Star Schemas Betrachtungsgegenstand in Kapitel 7. Daran schließt sich die Darstellung des logischen BW-Modells mit dem erweiterten Star Schema in Kapitel 8 an. Der Ausblick auf die physische Ebene in Kapitel 9 gestattet die Diskussion der für Abfrageverhalten und Benutzerakzeptanz wichtigen Aspekte. Die Gestaltungsempfehlungen im abschließenden Kapitel 10 fassen alle Ebenen der Modellierung zusammen und bieten dem mit der Modellierung befassten Entwickler eine Hilfe bei der Entscheidung, wie ein BW-Modell im konkreten Einzelfall aufzubauen ist.

## 2 Analyseorientierte Informationssysteme

Wesentliche Aspekte dieses Buches widmen sich der mehrdimensionalen Datenmodellierung. Diese bildet die theoretische Grundlage für Datenbanksysteme, die ihren Schwerpunkt in der Unterstützung analyseorientierter Informationssysteme haben.

Der Fokus analyseorientierter Informationssysteme liegt in der zeitnahen Versorgung betrieblicher Entscheidungsträger mit relevanten Informationen zu Analysezielen. Diese Systeme zielen somit auf die Unterstützung der dispositiven und strategischen Prozesse in einem Unternehmen ab und bilden damit ein logisches Pendant zu den operativen Systemen, die zumeist in Form einer integrierten betriebswirtschaftlichen Standardsoftware wie z. B. SAP R/3 eingesetzt werden.

Im Zusammenhang der systemseitigen Unterstützung analytischer Prozesse wird auch oft der Begriff des Management Support Systems genannt und seit einigen Jahren werden zunehmend Konzepte des Data Warehouse und OLAP diskutiert, die im Folgenden unter dem Begriff des analyseorientierten Informationssystems zusammengefasst werden sollen und deren kurze begriffliche Präzisierung in diesem Kapitel erfolgt.

### 2.1 Management Support Systeme

Unter dem Begriff Management Support System (MSS) sollen alle Varianten der elektronischen Unterstützung betrieblicher Entscheidungsträger bei der Abwicklung ihrer Aufgaben zusammengefasst werden. Insbesondere gehören hierzu die folgenden Systeme zur informationstechnischen Unterstützung von Fach- und Führungskräften:

- Managementinformationssysteme (*management information system*, MIS)
- Entscheidungsunterstützungssysteme (*decision support system* (DSS), EUS)
- Führungsinformationssysteme (*executive information system* (EIS), FIS)

Der Ansatz des *Managementinformationssystems* geht auf die ersten integrierten Systeme mit dem Ziel der Informationsversorgung des Managements in den sechziger Jahren zurück. Charakteristisch für diese Systeme war bereits ein periodisch aktualisiertes Zahlenmaterial, auf welchem ein Berichtswesen für das Management aufsetzte. Aufgrund der Restriktionen der damaligen technischen Möglichkeiten waren diese Systeme jedoch nur bedingt praxistauglich. Die Erstellung des Datenmaterials und der darauf aufbauenden Berichte war sehr zeitaufwendig und starr und lief im damals üblichen Batchverfahren in den Rechenzentren. Darüber hinaus führte das umfangreiche ungefilterte Zahlenmaterial zu einem Überangebot an Informationen im Management.

Die fehlende Interaktivität dieser Systeme wurde mit den in den siebziger Jahren entstehenden *Entscheidungsunterstützungssystemen* weitgehend aufgehoben. Diese verfolgten den Ansatz, nicht nur ein starres Berichtswesen zur Verfügung zu stellen, sondern den Manager in seiner Aufgabenstellung durch Modelle und Methoden zu unterstützen. Als Kernziel wurde die Verbesserung der Entscheidungsqualität postuliert.

Insbesondere die aufkommenden Tabellenkalkulationsprogramme ermöglichten den Aufbau von DSS-Systemen, die damit gerade in den Fachabteilungen und Stabsstellen zu einem breiten Einsatzfeld führten. In dieser Konstruktion liegt auch der wesentliche Kritikpunkt an den frühen Entscheidungsunterstützungssystemen begründet, da die Ausrichtung auf den einzelnen Benutzer zu schlecht integrierbaren dezentralen Lösungen weitgehend losgelöst von anderen Systemen führte.

Entscheidungsunterstützungssysteme leisten auch heute noch gute Dienste bei begrenzten Problemfeldern. Die Erkennung besonderer Situationen und Warnsignale ist eine Schwäche dieser Systeme, die mit den Führungsinformationssystemen abgebaut wurde.

Durch zunehmend bessere Möglichkeiten der Entwicklung graphischer Benutzungsoberflächen entstand Mitte der achtziger Jahre die Systemkategorie der *Führungsinformationssysteme*, die mit ihren innovativen Möglichkeiten der Präsentation und der hohen Ergonomie und intuitiven Benutzbarkeit insbesondere das obere Management adressieren. Neue Techniken zur Auswertung von Datenbeständen wie das Navigieren von aggregierten zu detaillierten Daten (drill down) sowie Funktionen des Generierens von Warnmeldungen (exception reporting) zur frühzeitigen Erkennung und Kennzeichnung von auffälligen Abweichungen ermöglichen die Analyse von Problemen und Hervorhebung wesentlicher Informationen.

Gleichwohl diese Systeme aufgrund ihrer proprietären und starren Struktur als zu inflexibel zur Anpassung an die zunehmend schneller wechselnden betrieblichen Anforderungen von Fach- und Führungskräften

galten, ergaben sich hierdurch wesentliche Impulse auf heutige Konzepte und Systeme.

Zusammenfassend werden die Varianten von Management Support Systemen in der folgenden Abbildung 1 typisiert. Neben den beschriebenen Systemen MIS, DSS und EIS gehören auch Werkzeuge des Personal Information Management (PIM) wie z. B. elektronische Kalender zu der Kategorie der Management Support Systeme.

Management Support Systeme (MSS)		
Basissysteme	Executive Information Systeme (EIS)	Decision Support Systems (DSS)
<ul style="list-style-type: none"> <li>• Textverarbeitung</li> <li>• Tabellenkalkulation</li> <li>• Grafikverarbeitung</li> <li>• Terminplanung</li> </ul>	Communication Support <ul style="list-style-type: none"> <li>• E-Mail</li> </ul> Data Support <ul style="list-style-type: none"> <li>• Standard Reporting (MIS)</li> <li>• Ad-hoc-Reporting</li> </ul>	<ul style="list-style-type: none"> <li>• Simulation</li> <li>• Prognose</li> <li>• Optimierung</li> </ul>

**Abb. 1.** Klassifizierung von Management Support Systemen

## 2.2 Data Warehouse

Allen analyseorientierten Informationssystemen gemeinsam ist eine geeignete zugrunde liegende Datenbasis. Diese bildet damit eine wesentliche Komponente auf deren Grundlage die verschiedenen Auswertungssysteme aufsetzen. Dem Aufbau dieser zentralen Datenbasis widmet sich die Diskussion seit einigen Jahren unter dem Stichwort Data Warehouse. Hierunter soll im Folgenden ein unternehmensweites Konzept verstanden werden dessen Ziel es ist, eine logisch zentrale, einheitliche und konsistente Datenbasis für die vielfältigen Anwendungen zur Unterstützung der analytischen Aufgaben von Führungskräften aufzubauen, die losgelöst von den operativen Datenbanken betrieben wird.

Der Begriff *Data Warehouse* geht auf Inmon zurück. Inmon beschreibt ihn mit der Aufgabe, Daten zur Unterstützung von Managemententscheidungen bereitzustellen, die die folgenden vier wesentlichen Eigenschaften aufweisen:

- Themenorientierung



- Vereinheitlichung
- Zeitorientierung
- Beständigkeit

Die in einem Data Warehouse abzulegenden Daten orientieren sich an dem Informationsbedarf von Entscheidungsträgern und beziehen sich demnach auf Sachverhalte, die das Handeln und den Erfolg eines Unternehmens bestimmen. Die Daten fokussieren sich daher auf die Kernbereiche der Organisation. Diese datenorientierte Vorgehensweise unterscheidet sich deutlich von den prozessorientierten Konzepten der operativen Anwendungen.

Eine wesentliche Eigenschaft eines Data Warehouse ist ein konsistenter Datenbestand, der durch eine Vereinheitlichung der Daten vor der Übernahme entsteht. Diese bezieht sich sowohl auf die Struktur wie auch auf die Formate, häufig müssen die verwendeten Begriffe, Kodierungen und Maßeinheiten zusammengeführt werden.

Für die Managementunterstützung werden Daten benötigt, welche die Entwicklung des Unternehmens über einen bestimmten Zeitraum repräsentieren und zur Erkennung und Untersuchung von Trends herangezogen werden. Dazu wird der Data Warehouse Datenbestand periodisch aktualisiert und der Zeitpunkt der letzten Aktualisierung definiert damit einen Schnappschuss des Unternehmensgeschehens, der je nach Ladezyklus Tage, Wochen oder Monate zurückliegen kann.

Das vierte wesentliche Charakteristikum bezieht sich auf die Beständigkeit der Daten in einem Data Warehouse. Da diese in der Regel nur einmal geladen und danach nicht mehr geändert werden, erfolgt ein Datenzugriff im Allgemeinen nur lesend. Einmal erstellte Berichte auf Basis dieses Datenbestandes sind daher reproduzierbar, da auch in späteren Perioden die Datenbasis die gleiche ist. Diese Eigenschaft wird mit dem Begriff der Nicht-Volatilität umschrieben.<sup>3</sup> Die Beständigkeit bezieht sich aber auch auf ein verlässliches annähernd gleich bleibendes Antwortzeitverhalten.

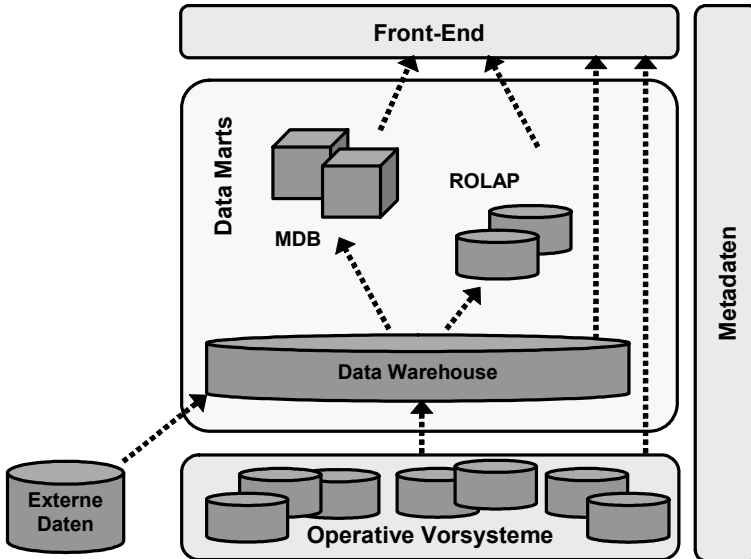
Die Einordnung eines Data Warehouse in die IT-Struktur eines Unternehmens ergibt sich aus der in Abbildung 2 dargestellten Referenzarchitektur. Ausgangsbasis dieser Architektur sind die operativen Vorsysteme, aus denen periodisch Datenextrakte generiert werden. Im Rahmen des ETL-Prozesses (*extract transform load*, ETL) erfolgt die Bereinigung und Transformation der Daten aus den verschiedenen Vorsystemen sowie externen Datenquellen zu einem konsistenten einheitlichen Datenbestand und der Transport in das Data Warehouse. Hierbei sind die beiden Phasen des

---

<sup>3</sup> Inmon beschreibt diese vierte Eigenschaft mit dem Begriff *non-volatile*, der sich auf die Änderungshäufigkeit bezieht (Inmon 1996, S. 35ff.)

erstmaligen Befüllens sowie der regelmäßigen periodischen Aktualisierungen zu unterscheiden.

Dieser ETL-Komponente kommt beim Aufbau eines Data Warehouse eine zentrale Bedeutung zu, denn ein hoher Anteil des Aufwandes beim Aufbau eines Data Warehouse resultiert aus der Implementierung von Zugriffsstrategien auf die operativen Datenhaltungseinrichtungen.



**Abb. 2.** Data Warehouse Referenzarchitektur

Aus diesem Datenbestand können des Weiteren kleinere funktions- oder bereichsbezogene Teilsichten in sog. *Data Marts* extrahiert werden. Diese müssen wiederum periodisch aus dem Data Warehouse Datenbestand aktualisiert werden. Für diese Teil-Datenbestände kommen im Allgemeinen so genannte *OLAP-Datenbanken* zum Einsatz, deren Diskussion Gegenstand des nächsten Abschnittes ist.

Die Auswertung über die Front-End-Applikationen kann sowohl direkt auf dem zentralen Data Warehouse erfolgen als auch auf den einzelnen Data Marts aufsetzen. In Data Warehouse-Konzepten können auch Applikationen wie beispielsweise Management Support Systeme auf diesen Datenbeständen basieren, d. h. die Datenbasis für diese Systeme kann auch in einem Data Warehouse liegen. Hier verbinden sich also bekannte Konzepte des Management Supports mit dem neuen Konzept des Data Warehouse zu einer neuen Systemkategorie. Eine weitere wesentliche Erweiterung ergibt sich aus dem Ansatz des On-Line Analytical Processing (OLAP), der im folgenden Abschnitt dargestellt wird.

## 2.3 OLAP und mehrdimensionale Datenbanken

Der Begriff OLAP beschreibt ein Leitbild für eine Endanwender-orientierte Analysetechnik und wird häufig konträr zum sog. On-Line Transaction Processing (OLTP) gesehen. Online Analytical Processing (OLAP) ist ein mittlerweile anerkannter Bestandteil für eine angemessene DV-Unterstützung betrieblicher Fach- und Führungskräfte und bietet einen Endanwender-orientierten Gestaltungsrahmen für den Aufbau von Systemen zur Unterstützung dispositiver bzw. analytischer Aufgaben.

Als zentrales Charakteristikum gewährleisten multidimensionale Sichtweisen auf unternehmensinterne und -externe Datenbestände brauchbare Näherungen an das mentale Unternehmensbild des Managers. Betriebswirtschaftliche Variablen bzw. Kennzahlen (wie z. B. Umsatz oder Kostengrößen) werden entlang unterschiedlicher Dimensionen (wie z. B. Kunden, Artikel, Regionen) angeordnet, und diese Strukturierung gilt als geeignete entscheidungsorientierte Sichtweise auf betriebswirtschaftliches Zahlenmaterial. Bildlich gesprochen werden die quantitativen Kenngrößen in mehrdimensionalen Würfeln gespeichert, deren Kanten durch die einzelnen Dimensionen definiert und beschriftet sind.

OLAP soll es Benutzern ermöglichen, flexible komplexe betriebswirtschaftliche Analysen wie auch Ad-hoc-Auswertungen mit geringem Aufwand eigenständig durchführen zu können. Um dieses Ziel zu erreichen, wurden von Codd, Codd und Sally 12 Regeln als Anforderung an OLAP-Lösungen definiert:<sup>4</sup>

1. Die *mehrdimensionale konzeptionelle Sicht* auf die Daten wird als elementarstes Wesensmerkmal für OLAP postuliert. Diese Darstellungsform ermöglicht eine Navigation in den Datenwürfeln mit beliebigen Projektionen und Verdichtungs- und Detail-Darstellungen.
2. *Transparenz* beschreibt die nahtlose Integration in Benutzerumgebungen.
3. Eine offene Architektur gewährleistet *Zugriffsmöglichkeiten* auf heterogene Datenbasen, eingebunden in eine logische Gesamtsicht.
4. Ein *gleichbleibendes Antwortzeitverhalten*, selbst bei vielen Dimensionen und sehr großen Datenvolumina, ist ein wesentlicher Aspekt.
5. Postuliert wird auf Basis einer *Client-Server-Architektur* die Möglichkeit verteilter Datenhaltung sowie der verteilten Programmausführung.

---

<sup>4</sup> Die zwölf Regeln wurden von Codd, Codd und Sally 1993 postuliert (Codd et al. 1993).

6. Aufgrund der *generischen Dimensionalität* stimmen alle Dimensionen in ihren Verwendungsmöglichkeiten überein.
7. Betriebswirtschaftliche mehrdimensionale Modelle sind oft sehr gering besetzt. Das *dynamische Handling* „*dünnbesetzter Würfel*“ ist elementar für eine optimale physikalische Datenspeicherung.
8. Unter *Mehrbenutzerfähigkeit* in OLAP-Systemen wird der gleichzeitige Zugriff verschiedener Benutzer auf die Analysedatenbestände, verbunden mit einem Sicherheits- und Berechtigungskonzept, verstanden.
9. Der Kennzahlenberechnung und Konsolidierung dienen *unbeschränkte dimensionsübergreifende Operationen* innerhalb einer vollständigen integrierten Datenmanipulationssprache.
10. Eine ergonomische Benutzerführung soll *intuitive Datenmanipulation* und Navigation im Datenraum ermöglichen.
11. Auf Basis des mehrdimensionalen Modells soll ein leichtes und *flexibles Berichtswesen* generiert werden können.
12. Die Forderung nach einer *unbegrenzten Anzahl an Dimensionen und Aggregationsebenen* ist in der Praxis schwer realisierbar.

Dieses Regelwerk ist nicht unumstritten, und von der Gartner Group sind 9 weitere Regeln in die Diskussion gebracht worden. Auch dieses erweiterte Regelwerk wurde teilweise um spezielle Regeln erweitert. Eine etwas pragmatischere und technologiefreie Variante zur Definition der konstituierenden Charakteristika von OLAP stammt von Pendse und Creeth, die ihren Ansatz mit FASMI benennen:

1. *Fast*: Ganz konkret wird für das Antwortzeitverhalten ein Grenzwert von zwei Sekunden für Standardabfragen und zwanzig Sekunden für komplexe Analysen festgelegt.
2. *Analysis*: Benutzern muss es ohne detaillierte Programmierkenntnis möglich sein, analytische Berechnungen und Strukturuntersuchungen auf Basis definierter Verfahren und Techniken ad-hoc zu formulieren.
3. *Shared*: Für den Mehrbenutzerbetrieb werden Berechtigungsmöglichkeiten bis auf Datenelementebene sowie Sperrmechanismen bei konkurrierenden Schreibzugriffen gefordert.
4. *Multidimensional*: Die mehrdimensionale Sichtweise ist ein elementares Wesensmerkmal analytischer Systeme.
5. *Information*: Für OLAP-Systeme ist die verwaltbare Informationsmenge bei stabilem Antwortzeitverhalten ein kritischer Bewertungsfaktor.

Verschiedene Ansätze zur Definition dessen, was OLAP ausmacht, resultieren in der Anforderung nach Vereinheitlichung und dem Setzen von

Standards. Dieses hat sich der OLAP-Council zum Ziel gesetzt.<sup>5</sup> Diese Diskussion ist losgelöst von Implementierungsaspekten, für die es jedoch gleichermaßen verschiedenste Architekturansätze gibt.

## 2.4 Weiterführende Literatur

Zum Begriff des analyseorientierten Informationssystems und deren Management-Bezug siehe (Gabriel et al. 2000), zur begrifflichen Einordnung von Data Warehouse, OLAP und Management Support System siehe (Chamoni u. Gluchowski 1999a) und sehr umfassend dargestellt in (Gluchowski et al. 1997). Eine kritische Beurteilung von Management Support Systemen findet sich in (Vetschera 1995).

Eine detaillierte Darstellung des Data Warehouse Konzeptes findet sich in (Mucksch u. Behme 2000) sowie in (Günzel u. Bauer 2001). Die Klassiker zum Thema Data Warehouse sind (Inmon 1996) und (Kimball 1996).

Zu den Aspekten des On-Line Analytical Processing (OLAP) siehe (Chamoni u. Gluchowski 1999b), (Codd et al. 1993) sowie (Pendse u. Creeth 1995).

Chamoni P, Gluchowski P (1999a) Analytische Informationssysteme – Einordnung und Überblick. In: Chamoni P, Gluchowski P (Hrsg) Analytische Informationssysteme: Data Warehouse, On-Line Analytical Processing, Data Mining, 2. Aufl. Springer, Berlin, S 3–25

Chamoni P, Gluchowski P (1999b) Entwicklungslinien und Architekturkonzepte des OLAP. In: Chamoni P, Gluchowski P (Hrsg) Analytische Informationssysteme: Data Warehouse, On-Line Analytical Processing, Data Mining, 2. Aufl. Springer, Berlin, S 261–280

Codd EF, Codd SB, Sally CT (1993) Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT-Mandate. White Paper

Gabriel R, Chamoni P, Gluchowski P (2000) Data Warehouse und OLAP – Analyseorientierte Informationssysteme für das Management. Zeitschrift für betriebswirtschaftliche Forschung 52:74–93

Gluchowski P, Gabriel R, Chamoni P (1997) Management-Support-Systeme: computergestützte Informationssysteme für Führungskräfte und Entscheidungsträger. Springer, Berlin

Günzel H, Bauer A (2001) Data-Warehouse-Systeme – Architektur, Entwicklung, Anwendung. dpunkt, Heidelberg

---

<sup>5</sup> Der OLAP-Council wurde 1995 als Informationsforum und Interessenvertretung für OLAP-Anwender gegründet. Eine Zusammenstellung der Definitionen gängiger verwendeter OLAP-Begriffe findet sich auf der Homepage des OLAP-Councils.

- Inmon WH (1996) *Building the Data Warehouse*, 2nd edn. John Wiley & Sons, New York
- Kimball R (1996) *The Data Warehouse Toolkit*. John Wiley & Sons, New York
- Mucksch H, Behme W (2000) *Das Data Warehouse-Konzept – Architektur-Datenmodelle-Anwendungen*, 4. Aufl. Gabler, Wiesbaden
- Pendse N, Creeth R (1995) *The OLAP-Report; Succeeding with On-Line Analytical Processing*. Business Intelligence 1
- Vetschera W (1995) *Informationssysteme der Unternehmensführung*. Springer, Berlin

## 2.5 Zusammenfassung

Unter dem Sammelbegriff *Business Intelligence* werden zunehmend Konzepte des Data Warehouse, OLAP und Data Mining diskutiert. Durch die zunehmend strategische Ausrichtung der Informationsverarbeitung erhalten diese Konzepte einen neuen Stellenwert in der Praxis. Unter dem Begriff *Analyseorientiertes Informationssystem* werden Systemlösungen im Bereich Business Intelligence mit der Ausrichtung an der Analyseanforderung zusammengefasst.

Analyseorientierte Informationssysteme, deren Fokus in der zeitnahen Versorgung betrieblicher Entscheidungsträger mit relevanten Informationen zu Analysezwecken liegt, zielen auf die Unterstützung der dispositiven und strategischen Prozesse in Unternehmen ab.

Der Begriff *Data Warehouse* beschreibt ein unternehmensweites Konzept, dessen Ziel die Bereitstellung einer einheitlichen konsistenten Datenbasis für die vielfältigen Anwendungen zur Unterstützung der analytischen Aufgaben von Führungskräften ist. Diese ist losgelöst von operativen Datenbanken zu betreiben.

*Online Analytical Processing (OLAP)* als Grundprinzip für den Aufbau von Systemen zur Unterstützung von Fach- und Führungskräften in ihren analytisch geprägten Aufgaben basiert im Kern auf einer mehrdimensionalen konzeptionellen Sicht auf die Daten mit Möglichkeiten der Navigation in den Würfeln mit beliebigen Projektionen und auf verschiedenen Verdichtungsstufen.

Unter dem Oberbegriff *Management Support Systeme* werden unterschiedlichste Softwaresysteme für Führungskräfte und Entscheidungsträger subsumiert, die mit einer bunten Begriffsvielfalt von MIS, DSS und EIS in der Vergangenheit aufwarteten und deren Aufgabe in der informationstechnischen Unterstützung betrieblicher Entscheidungsträger bei der Abwicklung ihrer Aufgaben liegt.

## 3 Datenmodellierung und Mehrdimensionalität

Für Systemlösungen auf Basis des Data Warehouse Konzeptes und für OLAP Anwendungen haben zwei wesentliche Begriffe zentralen Charakter: Datenmodellierung und Mehrdimensionalität. Da die Modellierung von Systemen zur Unterstützung von Fach- und Führungskräften in ihren analytischen Fragestellungen einen entscheidenden Einfluss auf deren Akzeptanz und erfolgreiche Nutzung hat, steht dieser Aspekt in diesem Buch im Vordergrund. Zur begrifflichen Klärung geht Abschnitt 3.1 daher kurz auf die Begriffe des Datenmodells und der Modellierung ein. Die Mehrdimensionalität ist das zugrunde liegende Paradigma dieser analyseorientierten Systemlösungen und die Vielfalt an Ausgestaltungsmöglichkeiten mehrdimensionaler Strukturen wird in Abschnitt 3.2 aufgearbeitet. Abschnitt 3.3 befasst sich mit Kennzahlen und Kennzahlensystemen. Auf die Aspekte der Zeitabhängigkeit und der Veränderungen in Konsolidierungshierarchien geht Abschnitt 3.4 ein.

### 3.1 Datenmodelle und Datenmodellierung

Die Begriffe *Datenmodell* und *Datenbankdesign* werden in der Literatur mit unterschiedlicher Bedeutung eingesetzt. Um Missverständnissen bezogen auf den Datenmodell-Begriff vorzubeugen, muss die verwendete Definition klar dargestellt werden. Ganz allgemein ist ein Modell ein *Objekt*, das von einem *Subjekt* auf der Grundlage einer Struktur-, Funktions- oder Verhaltensanalogie zu einem *Original* eingesetzt und genutzt wird, um Aufgaben zu lösen, deren Durchführung unmittelbar am Original selbst nicht möglich bzw. zu aufwendig ist.

In der metamathematischen Modelltheorie, einem Teilbereich der mathematischen Logik, wird unter einem Modell eine Interpretation eines Axiomensystems verstanden, wobei alle Axiome dieses Systems wahre Aussagen sind. Auf diesem Modellbegriff basieren grundlegende Verfahren zur Beurteilung von Vollständigkeit, Widerspruchsfreiheit und Definierbarkeit. Diese Interpretation des Modellbegriffs wird in der Datenbanktheorie benutzt.

Ein Datenmodell kann als formaler Rahmen zur Beschreibung von Datenstrukturen und Operationen auf Daten bezeichnet werden. Dies korrespondiert mit der Definition, wonach zur Darstellung von Datenmodellen die Unterscheidung in Strukturteil und Operationenteil vorgenommen wird (statischer und dynamischer Aspekt).

Datenmodelle dienen der Beschreibung aller in einer Datenbank enthaltenen Daten und im Allgemeinen wird angenommen, dass in einem Datenmodell Objekte, deren Eigenschaften (Attribute) sowie Beziehungen zwischen Objekten modelliert werden. Dies ist auch Basis des weitreichenden Datenmodellbegriffs nach Brodie, welcher diesem Buch zugrunde liegt:

Ein Datenmodell ist eine Menge mathematisch wohldefinierter Konzepte, die alle statischen und dynamischen Eigenschaften der Anwendungswelt erfassen soll, und zwar

- statische Eigenschaften wie Objekte, Eigenschaften von Objekten und Beziehungen zwischen Objekten,
- dynamische Eigenschaften wie Operationen auf Objekten, Eigenschaften dieser Operationen und Beziehungen zwischen Operationen,
- Integritätsbedingungen über Objekten (statische Integritätsbedingungen) sowie über Operationen (dynamische Integritätsbedingungen).

Datenmodelle sollen mithin die Bedeutung und Repräsentation von Daten beschreiben, d. h. ein Datenmodell geht aus Abstraktion eines zu modellierenden Realitätsausschnittes hervor. Datenmodelle können daher nach ihrer Nähe zur Realwelt klassifiziert werden. Diese weit verbreitete Strukturierung des Modellierungsvorganges ist in Abbildung 3 dargestellt. Danach werden die Ebenen der semantischen, logischen und physischen Datenmodellierung unterschieden. Der Realwelt am nächsten ist dabei die semantische Ebene. Die Aufgabe des semantischen Datenmodells ist es, eine Brücke zwischen der Realwelt einerseits und dem logischen Datenmodell andererseits zu schlagen. Dieses ist noch losgelöst von dem einzusetzenden Datenbanksystem und soll den zu betrachtenden Realitätsausschnitt abstrahierend in einem Modell abbilden. Die Wahl eines geeigneten semantischen Datenmodells hängt somit von dem betrachteten Realitätsausschnitt ab. Zur semantischen Modellierung für herkömmliche OLTP-Anwendungen und zum Einsatz von relationalen Datenbanksystemen hat sich in der Praxis das Entity Relationship-Modell (ERM) in weiten Bereichen bewährt.

In der zweiten Ebene, der logischen Modellierung, sind die Modelle ebenfalls noch unabhängig von der physischen Repräsentation, richten sich jedoch an der für die Speicherung einzusetzenden Datenbanktechnologie



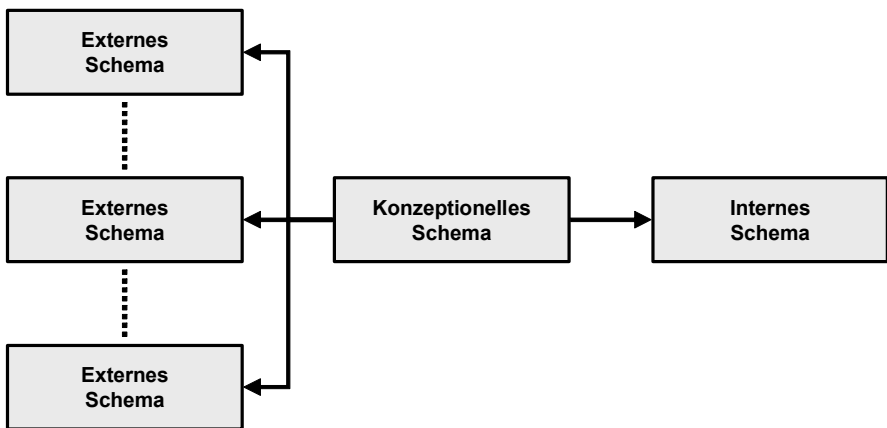
aus. Ein solches Modell heißt auch Datenbankschema. Bei relationalen Datenbanksystemen wird vom Relationenmodell gesprochen. Aufgrund des einfachen Übergangs von einem ERM zu einem Relationenmodell kommen diese Modelle oft gemeinsam zum Einsatz. Die Aspekte der physischen Speicherung und der Speicheroptimierung sind Bestandteil des physischen Datenmodells, welches die dritte Ebene bildet.



**Abb. 3.** Ebenen der Modellierung

In der Sprechweise der ANSI-3-Ebenen-Architektur entsprechen die semantische und logische Ebene dem konzeptionellen Schema und die physische Ebene korrespondiert mit dem internen Schema. Die externen Schemata in der ANSI-Architektur werden oft mit den Methoden und Darstellungsformen der semantischen und logischen Ebene modelliert.

Nach diesem Konzept erfolgt die Beschreibung der Daten einer Datenbank auf drei verschiedenen Ebenen, die dadurch gekennzeichnet sind, dass sie auf einer anderen Sichtweise, welche durch Schemata formalisiert werden, basieren. Auf der konzeptionellen Ebene steht die logische Gesamtstruktur der Daten, ihrer Eigenschaften und ihrer Beziehungen untereinander im Vordergrund. Ein externes Schema beschreibt eine benutzer-spezifische Sicht (engl. *view*), d. h. einen Ausschnitt des konzeptionellen Schemas. Das interne Schema behandelt das physische Design der Datenbank. Die Separation in diese drei Schemata ist in Abbildung 4 dargestellt.



**Abb. 4.** ANSI 3-Schichten-Konzept

Die beiden Hauptvorteile der Differenzierung in externe, konzeptionelle und interne Schicht sind:

- **Physische Datenunabhängigkeit:**

Die Trennung von konzeptioneller und physischer Ebene erlaubt die logische Beschreibung der Datenbank, ohne die genaue physische Speicherung der Daten betrachten zu müssen.

- **Logische Datenunabhängigkeit:**

Der Vorteil der Trennung von externer und konzeptioneller Ebene liegt darin, dass die konzeptionelle Sicht, d. h. das konzeptionelle Datenmodell, geändert werden kann, ohne Änderungen an den Anwendungsprogrammen vornehmen zu müssen.

Basierend auf der Strukturierung nach der 3-Ebenen-Architektur können die folgenden Phasen des Datenbankdesigns identifiziert werden:

- Analyse des Informationsbedarfs
- Entwicklung des semantischen Datenmodells
- Entwicklung des logischen Datenmodells
- Entwicklung des internen Schemas

Bei der Analyse des Informationsbedarfs wird festgelegt, welche Informationen für das geplante Datenbanksystem von Bedeutung sind. Diese Ergebnisse werden im konzeptionellen Schema durch ein semantisches Datenmodell formalisiert. In der logischen Designphase wird das konzeptionelle Schema verfeinert. In Abhängigkeit von dem zu verwendenden Datenbanksystem wird das konzeptionelle Modell in das dem Datenbanksystem zugrunde liegenden Datenbankmodell übertragen. Dieses logische Modell stellt das Bindeglied zwischen konzeptionellem und physischem Design dar.

### 3.2 Mehrdimensionale Datenstrukturen

Der Begriff *Datenstruktur* wird oft gleichbedeutend mit dem Begriff *Datentyp* verwendet. Im engeren Sinne versteht man unter Datenstrukturen den Aufbau von allgemeinen Wertebereichen aus elementaren mit Hilfe von Konstruktoren, also auf Basis von ganzen Zahlen, reellen Zahlen, Wahrheitswerten, der Menge aller Zeichen, selbst definierten endlichen Mengen und Unterbereichen der aufgeführten Wertebereiche. Konstruktor dienen dazu, die Struktur der Wertebereiche und die Zusammensetzung bzw. Zerlegung ihrer Elemente festzulegen. Üblich sind die fünf

Konstruktoren *Aggregation*, *Generalisation*, *Rekursion*, *Potenzmengenbildung* und die *Bildung von Funktionsräumen*.

Die Aggregation bildet das kartesische Produkt von Wertebereichen, mit deren Hilfe Tupel gebildet werden können. Die Generalisation bewirkt die disjunkte Vereinigung von Wertebereichen. Die Rekursion bewirkt den Übergang zu abzählbar unendlichen Wertebereichen. Die Potenzmengenbildung beschreibt die Bildung der Menge aller Teilmengen.<sup>6</sup> In Datenstrukturen vom Typ eines Funktionsraumes können die Variablen Funktionen als Werte annehmen, was jedoch aufgrund von Implementierungsproblemen eine eher untergeordnete Rolle spielt.

Um auf die einzelnen Datenelemente, aus denen eine Datenstruktur gebildet wurde, zugreifen zu können, ist jedem Konstruktor ein umgekehrt wirkender Selektor zugeordnet. Auf Datenstrukturen, die durch mehrfache Anwendung von Konstruktoren gebildet wurden, werden die zugehörigen Selektoren nacheinander in umgekehrter Reihenfolge angewendet.

Unter einem Datentyp versteht man die Zusammenfassung von Wertebereichen und Operationen zu einer Einheit. Liegt dabei der Schwerpunkt auf den Eigenschaften, die die Operationen und Wertebereiche besitzen, dann wird von abstrakten Datentypen gesprochen.

### 3.2.1 Grundbestandteile mehrdimensionaler Datenstrukturen

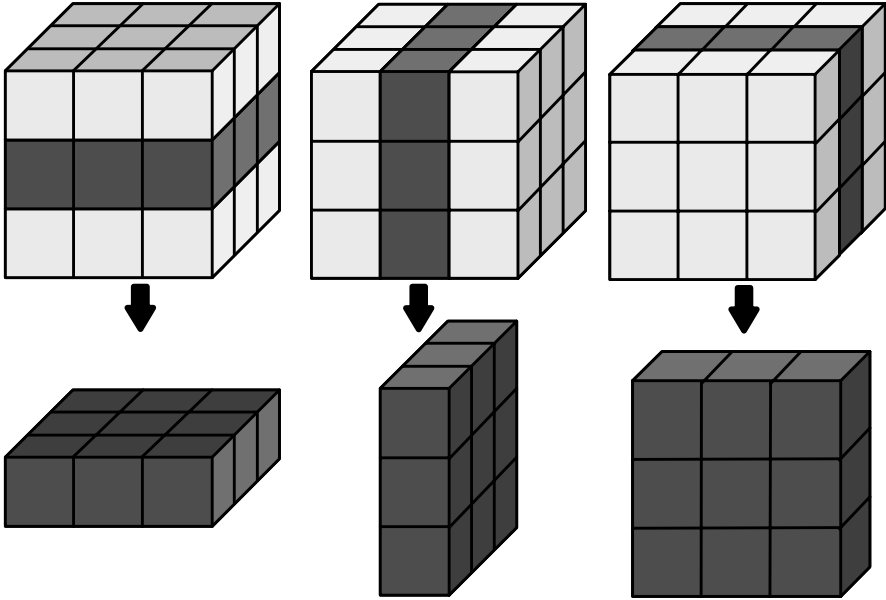
In diesem Buch liegt der Fokus auf einer speziellen Art von Datenstrukturen, den so genannten mehrdimensionalen Datenstrukturen. In diesem Ansatz werden Datenwerte in einem mehrdimensionalen Datenwürfel aufgespannt betrachtet. Die Würfelzellen beinhalten die Werte, die durch die Würfelkanten strukturiert werden. Die Würfelkanten bilden im mehrdimensionalen Sprachgebrauch die Dimensionen. Dieser Ansatz kann insofern als strukturgetrieben betrachtet werden, da durch die Struktur resultierend aus den Dimensionen, die den Würfel aufspannen, der Datenraum festgelegt ist. Zu jeder Kombination von jeweils einem Element aus jeder Dimension gibt es im Datenwürfel genau eine Zelle, deren Inhalt auch leer sein kann.

Auf diesen Datenwürfeln sind verschiedene Operationen definierbar. In der Literatur erfolgt die Charakterisierung mehrdimensionaler Datenstrukturen über die elementaren Operationen des *slice* und *dice*. Die Operation

---

<sup>6</sup> Die Aggregation könnte in Pascal beispielsweise eine Array- oder Record-Definition sein, die Generalisation kann zum Beispiel ein varianter Record in Pascal sein. Zeigerstrukturen in Pascal bilden z. B. rekursive Datenstrukturen und Potenzmengen werden durch die *set of*-Klausel gebildet.

*slice* bezeichnet eine Reduktion des Datenwürfels auf eine Dimension weniger durch die Selektion eines festen Elementes in einer Dimension. Die Begriffsbildung ist anhand der bildlichen Vorstellung des „Herausschneidens“ einer Scheibe aus einem gegebenen Datenwürfel verständlich. Abbildung 5 verdeutlicht diese Operation.

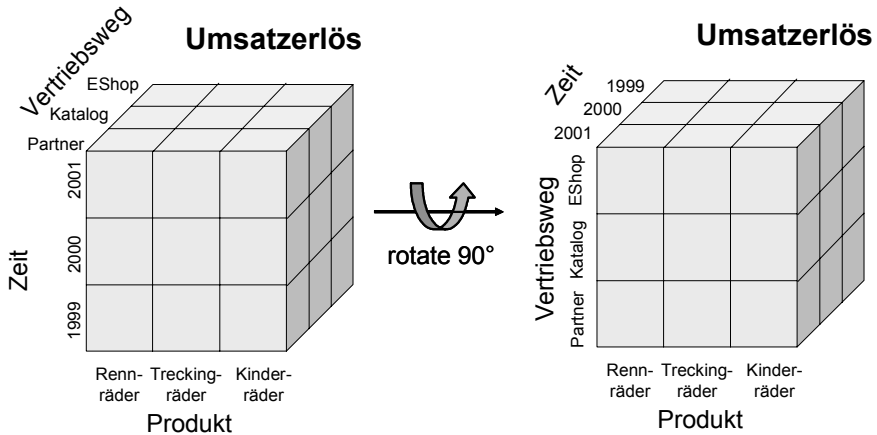


**Abb. 5.** Slicing: Herausschneiden von Ebenen aus Würfeln

Der Begriff *dice* aus dem Englischen kommend bezeichnet eigentlich das Würfeln an sich und ist als ein Drehen des Würfels interpretierbar. In der Terminologie mehrdimensionaler Operationen bezeichnet aber *rotation* diese Funktion. Die Selektion einer Teilmenge in einer Dimension, also der Übergang zu einem Teilwürfel, der dadurch entsteht, dass in einer Dimension nur eine Teilmenge der vorhandenen Elemente gewählt werden, wird als *ranging* oder auch *dice* bezeichnet.

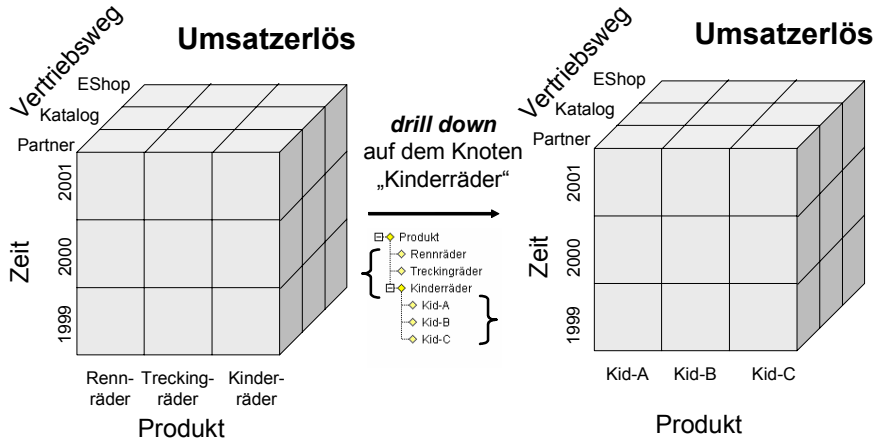
Das Drehen eines Datenwürfels leitet sich anschaulich von der Vorstellung einer möglichen Visualisierung eines mehrdimensionalen Datenraumes ab. Ein zweidimensionaler Würfel kann als Tabelle betrachtet werden. Während ein dreidimensionales Objekt durch die menschliche Vorstellungskraft noch abgebildet werden kann, braucht man bei höherdimensionalen Gebilden eine geeignete Repräsentationsform, die durch eine Reduktion auf ein zweidimensionales Teilgebilde gegeben ist. Die Rotation bezeichnet das Vertauschen von Dimensionen des Datenwürfels. Im zweidimensionalen Fall bedeutet dies das Vertauschen von Zeilen und Spalten,

was oft durch den Begriff *Transponieren* bezeichnet wird. Der dreidimensionale Fall beschreibt das Drehen eines Würfels, was in Abbildung 6 visualisiert ist.



**Abb. 6.** Drehen von Würfeln

In dieser Betrachtungsweise unterscheiden sich die Elemente innerhalb einer Dimension nur durch ihren unterschiedlichen Namen, auf ihnen ist aber keine Ordnung festgelegt. Beziehungen dieser Elemente untereinander werden aber bei mehrdimensionalen Datenstrukturen ebenfalls berücksichtigt. Diese werden durch Hierarchien auf den Dimensionen zum Ausdruck gebracht. Ganz allgemein werden diese durch hierarchische Strukturen in Form von Vorgänger- und Nachfolger-Relationen abgebildet. Entlang dieser Hierarchien werden dann Operationen des *drill-down* und des *roll-up* festgelegt, die den Übergang zu den Nachfolger-Elementen bzw. zu den Vorgänger-Elementen definieren. In diesem Sinne erfolgt auch hierdurch eine Selektion von Elementen in einer Dimension, die aber nicht durch explizite einzelne Auflistung festgelegt ist, sondern sich anhand der hierarchischen Beziehungen der Elemente untereinander ergibt. Dies wird an dem folgenden Beispiel in Abbildung 7 verdeutlicht.



**Abb. 7.** Navigieren in Dimensionsstrukturen

Erkennbar ist im Beispiel anhand der Kantenbeschriftungen in der Dimension Produkt, dass durch die skizzierte Operation ein drill-down von einer Produktgruppe zu den Produkten erfolgte.

Für die bisher dargestellten mehrdimensionalen Datenstrukturen gibt es unterschiedliche Anwendungsbereiche. In diesem Buch spielt der Bezug zur Wirtschaftsinformatik eine wesentliche Rolle. Als Anwendungsbereich werden im Folgenden Systeme zur Unterstützung von Managementanforderungen betrachtet. Fragestellungen aus dem Management basieren oftmals auf einer gewissen natürlichen Mehrdimensionalität. So werden beispielsweise Fragen nach der Entwicklung des Deckungsbeitrages bezogen auf weitere Merkmale wie etwa die Zeit, den Kunden, das Produkt gestellt.

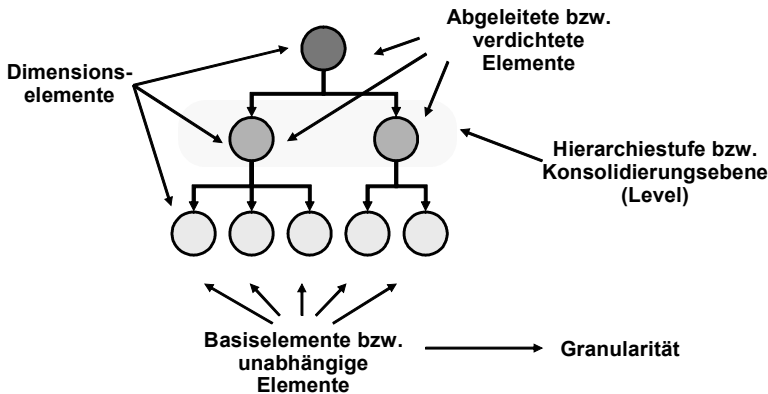
Auch für die dargestellten mehrdimensionalen Datenstrukturen ist der Prozess der Modellierung nach den drei Ebenen der semantischen, logischen und physischen Datenmodellierung, wie in Abschnitt 3.1 dargestellt, differenzierbar. Auf semantischer Ebene gibt es verschiedene Modellierungstechniken, die in der Literatur in die Diskussion gebracht wurden. Eine ganze Gruppe von Methoden fußt auf dem Paradigma der ER-Modellierung, welches in verschiedenen Facetten erweitert wurde, um die speziellen Anforderungen mehrdimensionaler Datenmodellierung zu berücksichtigen. Des Weiteren gibt es diverse Modellierungstechniken, die mit der Ausrichtung auf die semantische Modellierung mehrdimensionaler Datenstrukturen in die Diskussion gebracht wurden.

Dem Bereich der Modellierung auf physischer Ebene wird in der wissenschaftlichen Diskussion ebenfalls ein breiter Raum eingeräumt, deren Schwerpunkte auf Aspekten des Umgangs mit dünnbesetzten Datenwür-

fein einerseits und einer dauerhaft performanten Abfrage von Data Warehouse Datenbanken andererseits liegen.

### 3.2.2 Dimensionshierarchien

Die Dimensionselemente entlang einer Kante des Würfels lassen sich in einzelne Gruppen zusammenfassen. In einer Produktdimension ist beispielsweise eine Gruppierung zu Basisprodukten und Produktgruppen denkbar. Solch eine Gruppierung basiert auf einer 1:n-Beziehung zwischen Produktgruppen und Basiselementen. Mehrere derartige Beziehungen bilden die Konsolidierungs- oder Navigationspfade in der Dimension und formen die Dimensionsstruktur. Die Basiselemente auf der untersten Ebene der Struktur definieren die Granularität des Würfels und damit die detaillierteste zur Verfügung stehende Informationsebene. Die anderen Dimensionselemente werden als abgeleitete oder verdichtete Elemente bezeichnet. Zusammenfassend sind die verschiedenen Bestandteile, die eine Dimensionsstruktur formen, in Abbildung 8 veranschaulicht.



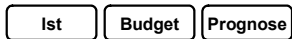
**Abb. 8.** Strukturbestandteile in Dimensionen

Die Struktur der Elemente einer Dimension untereinander wird als Hierarchie bzw. Dimensionsstruktur bezeichnet, wobei zwischen den verschiedenen Typen der elementbestimmten und der ebenenbestimmten zu unterscheiden ist. Bei der ersteren können die recht wenigen Elemente einzeln aufgelistet werden, so wie beispielsweise bei einer Dimension für verschiedene Planungsszenarien. Die andere Variante identifiziert ganze Gruppen von Elementen über ihre Konsolidierungsebene, so etwa die Ebenen einer Produkthierarchie.

Die nun näher zu beschreibenden verschiedenen Dimensionsstrukturvarianten sind grundlegend für mehrdimensionale Datenstrukturen deren Abbildungsmöglichkeiten in einem Data Warehouse-System eine wesentliche Forderung darstellen.

### **Flache Strukturen**

Im einfachsten Fall besteht eine Dimension nur aus gleichgewichtig nebeneinander stehenden Elementen, die diese ohne jegliche hierarchische Struktur formen. Für das Beispiel der Szenariodimension könnte dies etwa die Menge aller einzelnen Datenarten, wie in Abbildung 9 dargestellt, sein.

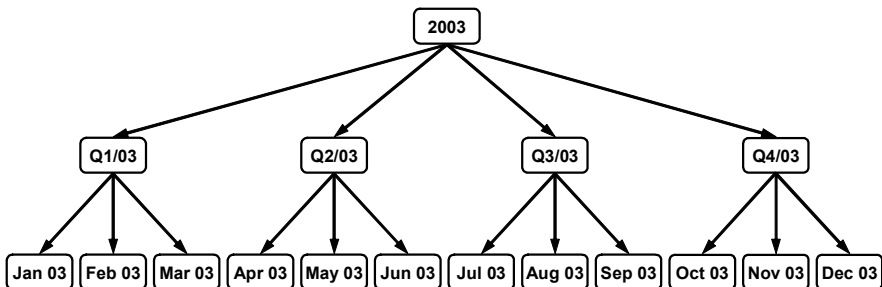


**Abb. 9.** Datenart-Dimension

Diese Struktur ist im Sinne der Operationen auf mehrdimensionalen Datenstrukturen wenig spektakulär und bildet nur eine Basis für weitere Strukturtypen. Der Vollständigkeit halber soll dieser Strukturtyp jedoch nicht unerwähnt bleiben.

### **Balancierte Baumstrukturen**

Typische Dimensionsstrukturen sind balancierte Baumstrukturen wie in Abbildung 10 dargestellt, die wie ein Baum aufgebaut und dadurch gekennzeichnet sind, dass die Basiselemente alle über gleich viele Stufen bis zu einem obersten Konsolidierungselement verbunden sind und dass es genau ein oberstes Wurzelement gibt. In der Baumdarstellung werden die Pfeile in der Richtung von der geringen zu der hohen Granularität gezeichnet, also in der Richtung des drill-down.



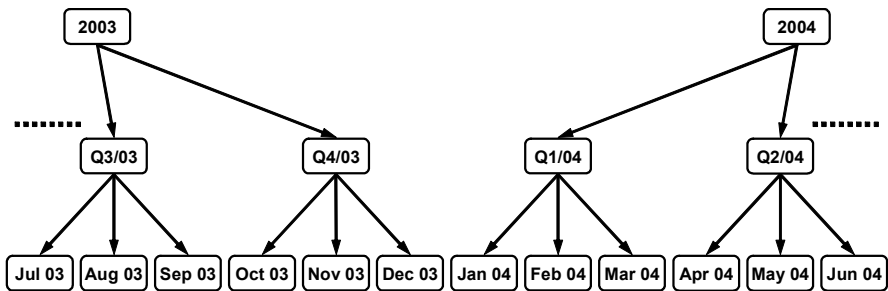
**Abb. 10.** Balancierte Baumstruktur als Dimensionshierarchie



Zu jedem Element auf der untersten Ebene, den Ausprägungen der Monatswerte, gibt es genau ein zugeordnetes Element in der Ebene darüber, gerade die Quartale. Jedem Quartal ist nun seinerseits genau ein übergeordnetes Element, im Beispiel genau das Jahr, zugeordnet.

### **Balancierte Waldstrukturen**

Eine Frage in dem Beispiel oben ist, wie sich die Hinzunahme eines weiteren Jahreselementes auf die Hierarchie auswirkt. Bei der Baumstruktur gab es nur ein oberstes Element, ein eindeutig festgelegtes Wurzelement. Eine Waldstruktur kann als Zusammenschluss von Baumstrukturen betrachtet werden, womit auch die Namensgebung verständlich wird. In dem Beispiel wird die Struktur der Zeitdimension gedoppelt, d. h. für jedes Jahr gibt es eine Baumstruktur. Dies ist in Abbildung 11 dargestellt.



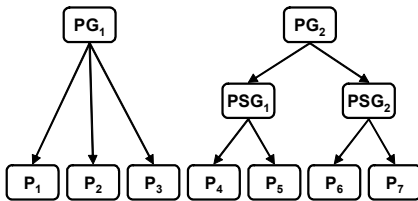
**Abb. 11.** Waldstruktur als Dimensionshierarchie

In diesem Fall gibt es einen relativ einfachen Übergang von der Waldstruktur zu einer Baumstruktur, in dem über die Knoten der Jahre eine weitere Ebene mit genau einem Element gehängt wird, welche die Verdichtung über alle Elemente repräsentiert (künstlicher Wurzelknoten).

### **Unbalancierte Baum- und Waldstrukturen**

Eine weitere Bedingung, die beim Aufbau von Dimensionsstrukturen fallen gelassen werden kann, ist die Eigenschaft der Ausgeglichenheit, die beschreibt, dass alle Dimensionselemente auf unterster Ebene über gleich lange Wege bis zu einem obersten Wurzelement führen und somit die Ebenen keine Lücken aufweisen. Strukturen der besprochenen Art mit Lücken in mindestens einer Ebene werden unter dem Begriff der unbalancierter

ten Baum- und Waldstrukturen erfasst.<sup>7</sup> Ein typisches Beispiel für eine unbalancierte Struktur ist eine Kostenstellenhierarchie, bei der eine eindeutige Zuordnung der Dimensionselemente zu den Konsolidierungsgruppen anhand der Navigationsschritte von den Basiselementen bis zum Wurzelement nicht möglich ist. Eine weitere Dimension, die oftmals eine unbalancierte Konsolidierungshierarchie umfasst, ergibt sich aus Produkten (siehe Abbildung 12). Dort sind exemplarisch einige Produkte, Produktgruppen sowie Untergruppen dargestellt.



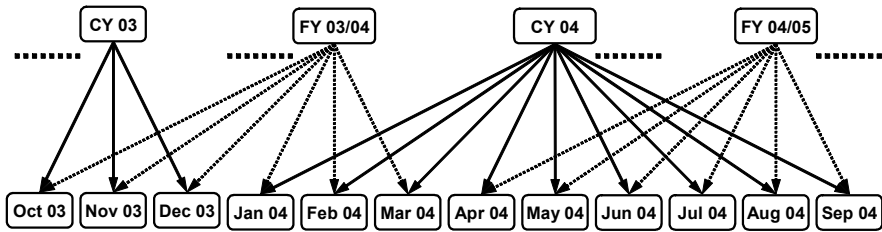
**Abb. 12.** Unbalancierte Dimensionshierarchie

### **Parallele Hierarchien**

Eine ebenfalls häufig anzutreffende Struktur bilden parallele Hierarchien, in denen Basiselemente auf unterschiedlichen Wegen konsolidierbar sind, beispielsweise eine Kundendimension, in der die Gruppierung von Kunden zum einen regional und zum anderen nach sachlogischen Aspekten erfolgen kann. Statt von parallelen Hierarchien wird oftmals auch von verschiedenen Hierarchien in einer Dimension gesprochen.

Vielfach wird in der Jahresbetrachtung zwischen der Kalenderjahres- und der Geschäftsjahresebene differenziert. Diese müssen nicht zwingend übereinstimmen. Dann ist die Verdichtung von den Monaten nicht eindeutig, da genauer spezifiziert werden muss, ob die Verdichtung zu Kalenderjahren oder zu Geschäftsjahren erfolgen soll. Dies ist ebenfalls eine parallele Hierarchie, da die Verdichtungsmöglichkeiten gleichgewichtig nebeneinander stehen. Ein exemplarischer Ausschnitt hierzu findet sich in Abbildung 13.

<sup>7</sup> Unbalancierte bzw. unausgeglichene hierarchische Strukturen werden auch unter dem Begriff der *Strukturanomalie* diskutiert.



**Abb. 13.** Beispiel einer parallelen Hierarchie

Gibt es einen gemeinsamen obersten Knoten entlang der beiden Hierarchien, so führen verschiedene Verdichtungswege in dem mehrdimensionalen Modell zu dem gleichen Element und eine konsistente Verdichtung ist zu gewährleisten.

### ***Hierarchien mit m:n-Beziehungen (Heterarchien)***

Nachdem zuvor schon die Bedingung der Balanciertheit aufgebrochen wurde, kann nun eine weitere Eigenschaft als bindende Restriktion weggelassen. In den bisherigen Strukturen ist jedem Element in jeder Hierarchie eindeutig genau ein übergeordnetes Element zugeordnet. Wird diese Forderung weggelassen, treten im Allgemeinen m:n-Beziehungen zwischen den Ebenen in einer Hierarchie auf. Diese Strukturen sollen als *Heterarchien* bezeichnet werden. Diese Strukturen stellen eine andere Komplexität dar, die besonderer Berücksichtigung bedarf. Die Zuordnung von Kalenderwochen zu Monaten ist ein Beispiel hierfür, da nicht alle Kalenderwochen eindeutig genau einem Monat zugeordnet sind. Die Wochen an den Monatsübergängen werden anteilig in die Monate verdichtet. Alternativ zu dieser Sichtweise können Kalenderwochen auch als eine parallele Hierarchie betrachtet werden, sofern die unterste Ebene die Tagesebene ist. In diesen Fällen ist ein besonderes Augenmerk auf die Verdichtungsregeln zu legen.

In einer Dimensionsstruktur, die z. B. Zuordnungen von Tochtergesellschaften zu Muttergesellschaften innerhalb einer Holding abbildet (siehe Abbildung 14), sind die Anteilsverhältnisse bei der Konsolidierung zu berücksichtigen. Diese Formen von Hierarchien mit m:n-Beziehungen zwischen einzelnen Dimensionsebenen stehen auch unter dem Begriff der anteiligen Verrechnung in der Diskussion.

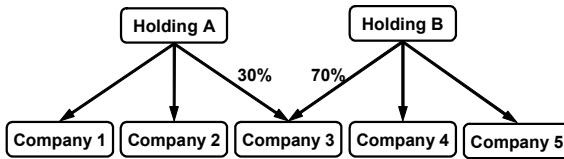


Abb. 14. Beispiel einer Heterarchie (Ausschnitt)

### 3.2.3 Hierarchie-Attribute

In allen Strukturvarianten von Hierarchien sind die Elemente nicht nur über ihren Namen identifiziert, denn insbesondere für Selektionen dient eine Vielzahl von Hierarchieattributen. Im Wesentlichen sind dabei technische und fachliche Attribute zu differenzieren.

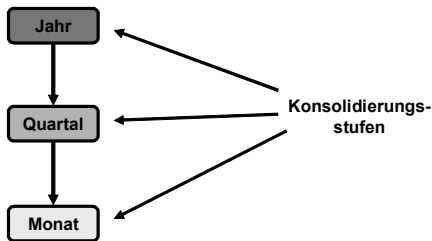
Eine Dimensionsstruktur ist formal gesehen ein gerichteter Graph, wobei die Pfeile von der obersten zur untersten Verdichtungsstufe gerichtet sind. Diejenigen Elemente, zu denen nur Pfeile hinführen, aber keine von ihnen weg, werden als *Blattelemente* bezeichnet und repräsentieren die Ebene der höchsten Granularität. Demgegenüber heißen die Elemente, von denen lediglich Pfeile wegführen, *Wurzelement* und stellen die höchste Verdichtung dar. Elemente, die nicht Wurzel und auch nicht Blatt sind, sind die *inneren Knoten* des Dimensionsgraphen.

Nun kann jeder Knoten, also jedes Element einer Dimensionshierarchie, von einem Wurzelknoten aus entlang der Pfeile erreicht werden. Die Anzahl der Pfeile auf einem kürzesten Weg von einem Wurzelement zu einem Knoten wird als dessen *Tiefe* bezeichnet. Analog ist eine *Zählweise* von einem Blattelement aus denkbar und führt zu dem Begriff der *Höhe* eines Dimensionselementes. Im Kontext von OLAP Anwendungen sind aber die Begriffe *level* und *generation* (*Ebene* und *Generation*) gebräuchlicher. Die Höhe korrespondiert dabei mit der Ebene und die Tiefe entspricht der Generation. Anzumerken ist, dass in den meisten Systemprodukten diese Begriffe eine unterschiedliche Bedeutung haben.

Die angesprochenen Konsolidierungsebenen in Dimensionshierarchien sind insbesondere im Fall unbalancierter Strukturen nicht eindeutig mit den Begriffen Ebene und Generation zu fassen, da die Elemente gleicher Konsolidierungsebenen nicht notwendigerweise auch die gleiche Ebenen- und Generationenzuordnung haben müssen.

Neben den technischen Attributen spielen für OLAP Anwendungen insbesondere fachliche Attribute von Dimensionselementen eine Rolle. Zu diesen zählen alle näheren beschreibenden Informationen eines Elementes, die nicht selbst als Dimension modelliert sind.

Eine Art von Attribut ergibt sich aus dieser Anforderung, ganze Ebenen von Dimensionselementen zu einer Gruppe zusammenzufassen, die gerade den Konsolidierungsebenen entsprechen. So werden beispielsweise alle Monats-Elemente in der Zeitdimension unter dem Gruppennamen Monate zusammengefasst. Die Ebene darüber würde mit dem Namen Quartale angesprochen usw. Dies wird für die exemplarische Zeitdimension in der Abbildung 15 verdeutlicht. Auf Ebene dieser Gruppen können dann weitere Attribute festgelegt werden.



**Abb. 15.** Ebenenbezogene Gruppenbildung

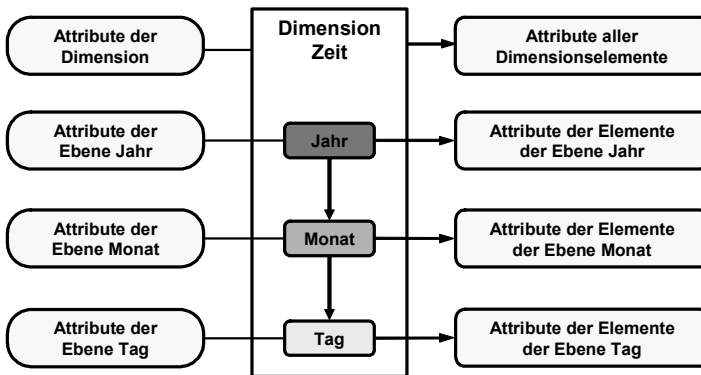
In der bildhaften Betrachtung von Würfeln und Würfelkanten entsprechen Dimensionen den Kantenbeschriftungen und sind in diesem Sinne ein eindimensionales Gebilde. Diese Darstellung reicht für das Verständnis mehrdimensionaler Strukturen nicht aus. Jedem Dimensionselement können weitere beschreibende Attribute zugeordnet sein, beispielsweise in der Zeitdimension mit Tagesgranularität weitere Attribute, die jedem Tag auch ein Wochentag sowie ein Kennzeichen, ob dieser Tag ein Feiertag ist, zuordnen. Dies geht aber nicht auf Monatsebene. Andere Attribute gelten etwa für alle Elemente, so z. B. ein englischer oder französischer Bezeichner für das Dimensionselement.<sup>8</sup> Allerdings sind unter diesem Attributsbegriff ebenfalls Aggregatknoten in der Hierarchie mit zu fassen. Es sind demzufolge Attribute für jede Ebene und die Elemente insgesamt zu unterscheiden. Gleichermäßen kann auch der Dimension als Gesamtheit ein Attribut zugeordnet werden.

Zusammenfassend ergeben sich folgende Arten von Attributen, die in mehrdimensionalen Datenstrukturen zu berücksichtigen sind:

<sup>8</sup> In der Darstellung mehrdimensionaler Datenstrukturen in der Literatur spielen Attribute häufig eine eher untergeordnete Rolle. In Darstellungen, die auf den Einsatz relationaler Datenbanktechnik abzielen, wird hingegen betont, dass vernünftige und zahlreiche Attribute das Wesen eines mehrdimensionalen Modells ausmachen.

- Jedem Dimensionselement können Attribute zugeordnet werden. Diese gelten für alle Elemente gleichermaßen, unabhängig von der Position innerhalb der Dimensionshierarchie. Beispielsweise sind dies die Elementbezeichnungen in verschiedenen Sprachen.<sup>9</sup>
- Der Dimension selbst können Attribute zugeordnet werden, so etwa Dimensionsbezeichnungen in verschiedenen Sprachen.
- Die Dimensionshierarchie kann in (semantisch sinnvolle) Konsolidierungsebenen aufgeteilt werden. Diesen Gruppen als solches können Attribute zugeordnet werden, etwa eine Ebenenbezeichnung in verschiedenen Sprachen.
- Den Elementen einer Konsolidierungsstufe können gruppenspezifische Attribute zugewiesen werden, so etwa das Attribut Wochentag auf Ebene der Partition Tage.

Dies wird zusammenfassend in der folgenden typisierten Darstellung in Abbildung 16 verdeutlicht.



**Abb. 16.** Typisierte Darstellung von Attributen

### 3.3 Kennzahlen und deren Berechnung

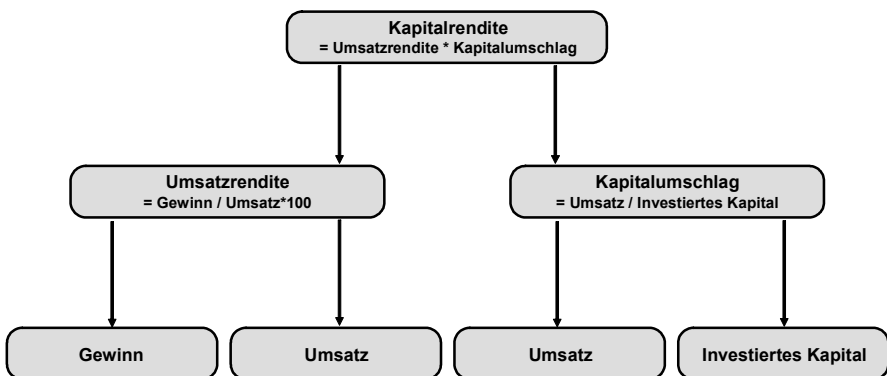
In mehrdimensionalen Datenmodellen werden die Wertausprägungen, die quantitativen Größen, in einem mehrdimensionalen Raum nach den Dimensionen, den qualitativen Größen, strukturiert betrachtet. Für den Anwendungsfall im Bereich analyseorientierter Informationssysteme beinhal-

<sup>9</sup> Das Attribut (als Schema) ist dann für alle Dimensionselemente definiert, die Attribut-Ausprägungen der einzelnen Elemente können unterschiedlich sein.

ten diese Datenwürfel Werte zu betriebswirtschaftlichen Kennzahlen, um damit Anwendern eine realitätsgetreue Sicht auf ihr Arbeitsumfeld zu ermöglichen. Die Berechnung von Kennzahlen entlang der Konsolidierungshierarchien ist dabei nicht immer eine reine Summierung und auch abhängig von den beteiligten Dimensionen. Auf Kennzahlen und Kennzahlensysteme wird in Abschnitt 3.3.1 eingegangen, deren mögliche Gestaltungsvarianten im mehrdimensionalen Modell sind Gegenstand von Abschnitt 3.3.2 und die Verdichtungseigenschaften stehen in Abschnitt 3.3.3 im Vordergrund.

### 3.3.1 Kennzahlen und Kennzahlensysteme

Unter Kennzahlen sollen im folgenden Zahlen, die Informationen über betriebswirtschaftliche Tatbestände beinhalten, verstanden werden. Eine einzelne Kennzahl kann demzufolge z. B. *Umsatz* sein. Aber auch *Deckungsbeitrag* und *Umsatzrendite* sind mögliche Kennzahlen. Aus gegebenen Kennzahlen können nach bestimmten fest vorgegebenen Rechenvorschriften neue Kennzahlen gebildet werden. Eine Menge von solchen Kennzahlen und abgeleiteten Kennzahlen bezeichnet man auch als Kennzahlenschema oder *Kennzahlensystem*. Ein bekanntes Beispiel für ein solches Kennzahlensystem ist das *Du Pont*-Kennzahlensystem, welches in der folgenden Abbildung 17 dargestellt ist. Für Informationssysteme im betriebswirtschaftlichen Umfeld spielen Kennzahlen und Kennzahlensysteme eine wichtige Rolle.



**Abb. 17.** Ausschnitt aus dem Du Pont-Kennzahlensystem

Kennzahlensysteme werden graphisch oftmals als gerichteter endlicher Baum oder Graph dargestellt. Die Knoten sind dabei mit den Namen der

Kennzahlen beschriftet. Nachfolger eines Knotens sind die Knoten, zu denen von diesem Knoten aus ein Pfeil führt.

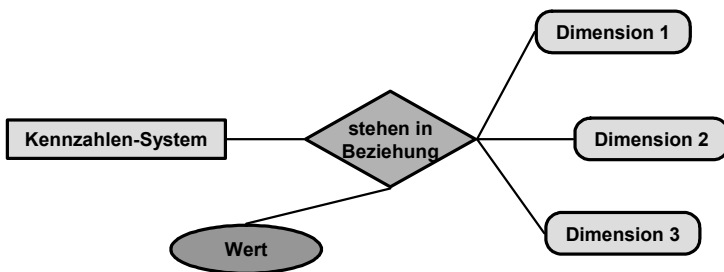
Bei der Struktur eines gerichteten endlichen Graphen müssen Zyklen ausgeschlossen werden, damit keine Rekursionen in der Berechnung von Kennzahlen auftreten. Um die Eindeutigkeit zu wahren, ist zu gewährleisten, dass verschiedene Berechnungswege zu einer Kennzahl zum gleichen Ergebnis führen.

In einem Kennzahlensystem bestehen die Knoten aus der Angabe der Kennzahl und der Angabe der Funktion, mit welcher der Wert der Kennzahl aus den Werten der Nachfolgerknoten berechnet wird. Die Kanten in der graphischen Darstellung sind dabei der Berechnungsrichtung entgegengesetzt orientiert.

### 3.3.2 Kennzahlen im mehrdimensionalen Modell

In mehrdimensionalen Datenmodellen werden Kennzahlen und Kennzahlensysteme mit den Beschreibungselementen Dimensionen und Würfel abgebildet. Für den Zusammenhang zwischen Dimensionen und Kennzahlen gibt es zwei verschiedene Ansätze, die im Folgenden näher erläutert werden sollen.

Zunächst wird davon ausgegangen, dass lediglich ein Kennzahlensystem zu berücksichtigen ist. Die Kennzahlen dieses Systems werden dann alle nach den gleichen Dimensionen aufgegliedert. Diese Beziehung kann wie in Abbildung 18 für exemplarische drei Dimensionen dargestellt werden, in der das Kennzahlensystem im mehrdimensionalen Modell dann durch eine eigene Dimension abgebildet wird.



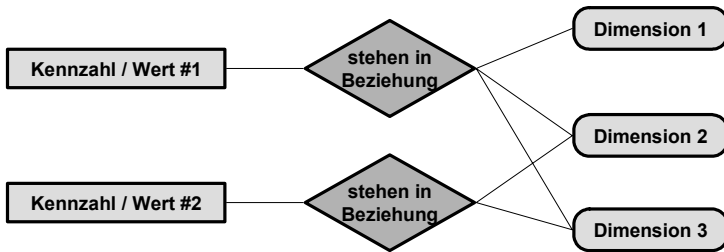
**Abb. 18.** Beziehung zwischen Dimensionen und Kennzahlensystem

Die Graphik drückt aus, dass zu jeder Kombination von Dimensionselementen, genau genommen aus jeder weiteren Dimension ein Element, eine Beziehung zu jeder Kennzahl besteht, deren Attribut *Wert* der eigentliche Wert der Kennzahl ist. Anders ausgedrückt lässt sich Wert als 4-fach



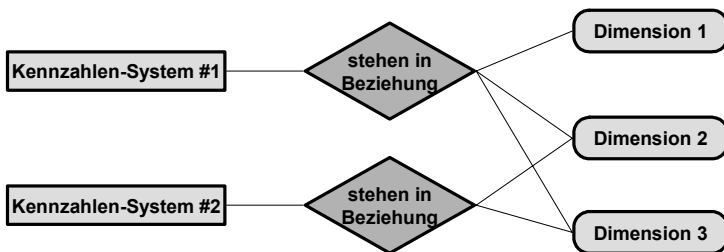
indizierte Variable interpretieren, deren Indexmengen die Knotenmengen der Dimensionen und des Kennzahlensystems sind. In diesem Sinne tritt das Kennzahlensystem als Dimension auf. Diese Betrachtungsweise impliziert, dass alle Kennzahlen des Kennzahlensystems nach genau den gleichen Dimensionen aufgegliedert betrachtet werden.

In der anderen Variante der Beziehungsmodellierung wird jede Kennzahl für sich einzeln betrachtet. In einem mehrdimensionalen Datenmodell wird dann für jede Kennzahl ein eigener Datenwürfel modelliert. Dann können die in dem Modell abgebildeten Kennzahlen auch unterschiedlich dimensioniert sein, was exemplarisch in der Abbildung 19 dargestellt ist.



**Abb. 19.** Modellierung von dedizierten Würfeln je Kennzahl

Die Einschränkung, dass nur ein Kennzahlensystem im mehrdimensionalen Datenmodell berücksichtigt sei, kann nun fallengelassen werden. Die darzustellende Beziehung verbindet verschiedene Kennzahlensysteme, welche ähnlich den verschiedenen Kennzahlen in Abbildung 19 auch unterschiedlich dimensioniert sein können. Die erhaltene Struktur ist eine hybride Form der beiden dargestellten Ansätze und ist in Abbildung 20 visualisiert.



**Abb. 20.** Mehrere Kennzahlensysteme in einem Datenmodell

Diese bisherige Betrachtungsweise ist insofern statisch, als die Werte in diesen modellierten Datenwürfeln als gegeben vorausgesetzt werden. Im Allgemeinen werden die Werte jedoch nur auf einer (Detail)-Ebene, meis-

tens der untersten Ebene, eingelesen, und die verbleibenden Ebenen werden dynamisch kalkuliert. Damit ist aber auch in den Dimensionen zu definieren, wie eine Berechnung verdichteter Werte aus den Detailwerten erfolgen soll. Neben die bisher betrachteten strukturellen Beziehungen in den Dimensionshierarchien tritt dann eine dynamische Beziehung, die diesen rechnerischen Zusammenhang modelliert.

Ein Berechnungsmodell für ein mehrdimensionales Datenmodell soll festlegen, auf welchen Ebenen bzw. für welche Kombinationen von Dimensionselementen Eingaben möglich sind und wie sich aus diesen Werten die anderen Werte berechnen.

### **3.3.3 Additivitätseigenschaften**

Die meisten Kennzahlen, so auch z. B. der Umsatz, haben die Eigenschaft, dass die Verdichtungsoperation entlang der Konsolidierungspfade die Addition ist. So ist beispielsweise der Umsatz einer Warengruppe gleich der Summe der Umsätze der Produkte in dieser Warengruppe. Diese Eigenschaft hat die Kennzahl Umsatz im Allgemeinen bezogen auf jede Dimension. Dies führt zur Definition der Additivität: Ist die Verdichtungsoperation einer Kennzahl in einer Dimension die Addition, so heißt diese Kennzahl additiv bezogen auf diese Dimension. Ist die Kennzahl darüber hinaus additiv bezogen auf jede verbundene Dimension, so wird allgemein von einer additiven Kennzahl gesprochen.

Der Lagerbestand als Kennzahl ist ein gutes Beispiel dafür, dass eine Kennzahl nicht bezogen auf jede Dimension additiv sein muss. Der Lagerbestand in Stück einer Warengruppe ist zum einen gleich der Summe der Bestände der zugehörigen einzelnen Artikel. Aber der Lagerbestand am Ende eines Jahres ist im Allgemeinen sehr verschieden von der Summe der Bestände an den jeweiligen Monatsenden. Die Kennzahl Lagerbestand ist somit bezogen auf die Zeitdimension nicht additiv, hingegen bezogen auf die anderen Dimensionen sehr wohl additiv. In diesem Fall wird von semi-additiven Kennzahlen gesprochen.

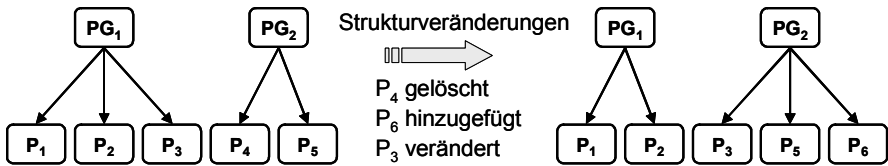
Kennzahlen, die Verhältniszahlen wie etwa Gewinnspannen darstellen, sind bezogen auf keine Dimension additiv und heißen nicht-additiv.

## **3.4 Temporale Aspekte**

Da alle Fakten in einem Data Warehouse schon einen Zeitbezug über die Zeitdimension haben, reicht es im Folgenden für die Historisierung im

konzeptionellen Modell aus, die Dimensionen und deren Hierarchien zu berücksichtigen.

Die strukturellen Veränderungen in Konsolidierungshierarchien entstehen durch Ändern, Löschen und Hinzufügen von Knoten, wie exemplarisch für den Fall einer zweistufigen Hierarchie mit Produktgruppen und Produkten in Abbildung 21 dargestellt ist. Im Fall der Änderung eines Hierarchieknotens erfolgt eine Veränderung der Zuordnung zu über- und untergeordneten Elementen in der Hierarchie. Die Aufnahme neuer Knoten und das Löschen bestehender Knoten verändert ebenfalls die Konsolidierungshierarchie.



**Abb. 21.** Strukturelle Veränderungen in Konsolidierungshierarchien

Um die Auswirkungen dieser strukturellen Veränderungen zu verdeutlichen, werden diese anhand eines einfachen Zahlenbeispiels mit den möglichen Anforderungen an Berichte betrachtet. Alle drei Veränderungen der Hierarchie wie in Abbildung 21 dargestellt fanden beim Übergang von einer fiktiven Zeitperiode 1 zu einer zweiten fiktiven Zeitperiode 2 statt. Die fiktiven Werte zu den korrespondierenden Buchungsperioden sind in Tabelle 1 zusammengefasst.

**Tabelle 1.** Kennzahlenwerte des Rechenbeispiels

Buchungsperiode	Produkt	Wert
Zeitperiode 1	P <sub>1</sub>	110
Zeitperiode 1	P <sub>2</sub>	120
Zeitperiode 1	P <sub>3</sub>	130
Zeitperiode 1	P <sub>4</sub>	140
Zeitperiode 1	P <sub>5</sub>	150
Zeitperiode 2	P <sub>1</sub>	110
Zeitperiode 2	P <sub>2</sub>	120
Zeitperiode 2	P <sub>3</sub>	130
Zeitperiode 2	P <sub>5</sub>	150
Zeitperiode 2	P <sub>6</sub>	160

Eine häufig anzutreffende Berichtsanforderung ist die Analyse aller Daten nach der jeweiligen aktuellen, nach einer historischen oder nach der jeweils gültigen Struktur („*historische Wahrheit*“). Eine andere Anforderung an das Berichtswesen kann sein, nur die Strukturbestandteile zu berücksichtigen, die im gesamten Berichtszeitraum gültig sind („*Äpfel nicht mit Birnen zu vergleichen*“).

Für die Darstellung der unterschiedlichen Berichtsergebnisse bei der Berücksichtigung der verschiedenen Anforderungen im Umgang mit strukturellen Veränderungen werden für die vier aufgelisteten Szenarien die Ergebnisse der Berechnung für die Produktgruppen im Folgenden aufgelistet.

Zunächst erfolgt die Berechnung anhand der Dimensionsstruktur mit Gültigkeit in der alten Zeitperiode 1. Dies führt u. a. dazu, dass Werte für P<sub>6</sub> nicht berücksichtigt werden. Das Ergebnis der Abfrage ergibt sich aus Tabelle 2.

**Tabelle 2.** Berichtsergebnis nach historischer Struktur

	PG <sub>1</sub>	PG <sub>2</sub>
Zeitperiode 1	360	290
Zeitperiode 2	360	150

Die in der zweiten betrachteten Periode gültige Struktur unterscheidet sich neben dem neuen Element P<sub>6</sub> dabei durch den fehlenden Knoten P<sub>4</sub>

und die Umordnung von  $P_3$ . Demnach erfolgt keine Berücksichtigung der Werte von  $P_4$  mit dem Ergebnis wie in Tabelle 3 dargestellt.

**Tabelle 3.** Berichtsergebnis nach aktueller Struktur

	$PG_1$	$PG_2$
Zeitperiode 1	230	280
Zeitperiode 2	230	440

Von diesen beiden Betrachtungsweisen unterscheidet sich die Berechnung nach der historischen Wahrheit deutlich, da hierbei die Bewegungsdatensicht eine Rolle spielt. Die Berechnung ergibt sich dann immer anhand der jeweils gültigen Struktur und liefert die Ergebnisse wie sie Tabelle 4 verdeutlicht.

**Tabelle 4.** Berichtsergebnis mit historischer Wahrheit

	$PG_1$	$PG_2$
Zeitperiode 1	360	290
Zeitperiode 2	230	440

Alle drei Varianten haben bei Zeitreihenanalysen den Nachteil, dass die Berechnung zu nicht unmittelbar vergleichbaren Werten führt. Daher wird auch oft gefordert, nur die Elemente in eine Abfrage mit aufzunehmen, die in dieser Form für den gesamten Betrachtungszeitraum Gültigkeit haben (siehe Tabelle 5).

**Tabelle 5.** Berechnungen nur für vergleichbare Elemente

	$PG_1$	$PG_2$
Zeitperiode 1	230	150
Zeitperiode 2	230	150

Zur systemtechnischen Umsetzung der verschiedenen Berichtsanforderungen stehen die folgenden vier Möglichkeiten zur Verfügung:

1. Anpassung des historischen Datenmaterials an neue Strukturen.
2. Separate Speicherung des historischen Datenbestandes zusätzlich zum Komplettestand mit neuen Strukturen.
3. Aufbau paralleler Hierarchien mit alten bzw. neuen Strukturen.

#### 4. Temporale Datenbanken und Gültigkeitsstempel.

Die Anpassung der historischen Daten hat den Vorteil, dass der Datenbestand nicht aufgebläht wird und die Datenstrukturen überschaubar bleiben. Da die alten Strukturen aber verloren sind, können nicht alle Berichtsanforderungen abgedeckt werden.

Bei der Vorgehensweise der separaten Speicherung können alte Auswertungen abgerufen werden, dies wird aber mit einem größeren Datenvolumen und einer aufwendigen Aktualisierung erkauft.

Die Lösungsalternative paralleler Hierarchien bietet den Vorteil, dass alle Zahlen mit beliebigen Strukturen angezeigt werden können. Die entstehenden Dimensionsstrukturen sind aber eher unübersichtlich.

Die Historisierung über Zeitstempel ermöglicht maximale Flexibilität für das Berichtswesen mit beliebigen historischen Varianten von Strukturen. Nachteilig sind aber die tendenziell schlechtere Performance und teilweise noch unausgereifte Konzepte.

### 3.5 Weiterführende Literatur

Zum Datenmodellbegriff wird auf (Luft 1990), (Brodie 1984), (Lockemann 1990) und (Vossen 1999) verwiesen.

Die Behandlung mehrdimensionale Datenstrukturen erfolgt eingehender bei (Holthuis 1997), (Schelp 2000) sowie bei (Hahne 2002).

Für eine detaillierte Auseinandersetzung mit Kennzahlen und Kennzahlensystemen wird auf (Meyer 1994), (Michel 1999) und (Küting 1983) verwiesen.

Zu temporalen Aspekten im Data Warehouse allgemein siehe (Chamoni u. Stock 1998) sowie (Stock u. Beekmann 2003), speziell die Zeitabhängigkeit im SAP Business Information Warehouse steht in (Hahne 2003) im Vordergrund.

Brodie ML (1984) On the developement of data models. In: Brodie ML, Mylopoulos J, Schmidt J (Hrsg) On conceptual modeling. Springer-Verlag, New York, S 19–47

Chamoni P, Stock S (1998) Temporale Daten in Management Support Systemen. Wirtschaftsinformatik 6:513–519

Hahne M (2002) Logische Modellierung mehrdimensionaler Datenbanksysteme. Deutscher Universitäts-Verlag, Wiesbaden

Hahne M (2003) Time aspects in SAP Business Information Warehouse. In: Jardim-Gonçalves R, Cha J, Steiger-Garção A (Hrsg) Concurrent Engineering. A.A. Balkema Publishers, Lisse, S 69–74

- Holthuis J (1997) Multidimensionale Datenstrukturen. In: Mucksch H, Behme W (Hrsg) Das Data Warehouse Konzept – Architektur – Datenmodelle – Anwendungen. Gabler, Wiesbaden
- Kütting K (1983) Kennzahlensysteme in der betrieblichen Praxis. WiSt
- Lockemann PC, Radermacher K (1990) Konzepte, Methoden und Modelle zur Datenmodellierung. HMD 152:3–16
- Luft AL (1990) Datenmodelle. In: Mertens P (Hrsg) Lexikon der Wirtschaftsinformatik, 2. Aufl. Berlin, S 132–134
- Meyer C (1994) Betriebswirtschaftliche Kennzahlen und Kennzahlensysteme, 2. Aufl. Poeschel, Stuttgart
- Michel R (1999) Komprimiertes Kennzahlen-Know-How: Analysemethoden, Frühwarnsysteme, PC-Anwendungen, Checklisten. Gabler, Wiesbaden
- Schelp J (2000) Modellierung mehrdimensionaler Datenstrukturen. Deutscher Universitäts-Verlag, Wiesbaden
- Stock S, Beekmann F (2003) Temporal aspects of data warehousing. In: Jardim-Gonçalves R, Cha J, Steiger-Garção (Hrsg) Concurrent Engineering. A.A. Balkema Publishers, Lisse, S 53–59
- Vossen G (1999) Datenmodelle, Datenbanksprachen und Datenbank-Management-Systeme, 3. Aufl. Addison-Wesley, Bonn

### 3.6 Zusammenfassung

Eine weit verbreitete Klassifizierung von Datenmodellen unterscheidet die Ebenen der semantischen, logischen und physischen Modellierung. Im Gegensatz zur semantischen Ebene, die der Realwelt am nächsten steht und den betrachteten Realitätsausschnitt auf fachlicher Ebene abstrahiert, ist die logische Ebene schon abhängig vom Zielsystem aber noch losgelöst von der physischen Repräsentation. In diesem Buch erfolgt die Diskussion mehrdimensionaler Datenstrukturen, die Kennzahlen nach verschiedenen Dimensionen aufgegliedert betrachten und bildlich die Kanten eines mehrdimensionalen Datenwürfels festlegen, auf allen drei Ebenen der Modellierung.

Mehrdimensionale Datenstrukturen bilden die wesentliche Grundlage analyseorientierter Informationssysteme mit den Basiskonzepten des Data Warehouse und OLAP. Die üblichen mehrdimensionalen Operationen betrachten insbesondere das Drehen und Selektieren von Würfeln und das Navigieren in Hierarchien auf Dimensionen, die durch Über- und Unterordnungsbeziehungen von Elementen einer Dimension entstehen. Diese sind die klassischen Operationen des slice, dice und drill.

Die verschiedenen Varianten von Dimensionsstrukturen, die damit vielfältige Konsolidierungspfade festlegen, bestimmen maßgeblich die analytischen Möglichkeiten von mehrdimensionalen Modellen. Die strukturellen

Veränderungen in Dimensionshierarchien werden unter dem Stichwort der Zeitabhängigkeit diskutiert und deren Abbildung in Modellen bestimmt ebenfalls die Möglichkeiten zur Analyse.



## 4 Architektur des SAP Business Information Warehouse

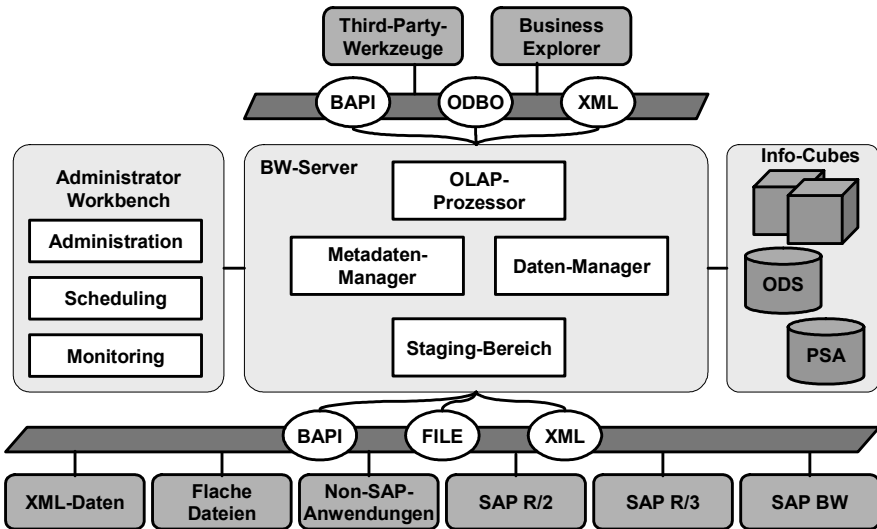
Die Architektur des Business Information Warehouse ist an die allgemeine Referenzarchitektur für ein Data Warehouse angelehnt. Ausgehend von verschiedenen möglichen Quellsystemen erfolgt ein im Allgemeinen periodischer Datenladevorgang in die zentrale Datenhaltung des Business Information Warehouse (BW). R/3-Systeme haben eine besondere Bedeutung als vorgelagertes OLTP System, aber auch die Anbindung von R/2-, Non-SAP- und anderen BW Systemen ist neben dem Import über flache Dateien möglich.

Die Verwaltung des BW-Systems erfolgt in der Administrator Workbench, einer Administrationskonsole, in der die Steuerung, Überwachung und Pflege des gesamten Extraktions- und Lademanagements von Quelldaten, die Benutzerverwaltung sowie die Modellierung der Datenstrukturen erfolgt. Im Mittelpunkt der Architektur steht der eigentliche Business Information Warehouse Server, der für die Ablage, Aufbereitung und Abfrage der Daten im BW verantwortlich ist. Die wesentlichen Objekte zur Datenablage sind dem Datenfluss folgend zunächst die Persistent Staging Area (PSA), der Operational Data Store (ODS) und die Info-Cubes. Während die PSA temporären Charakter hat, sind die Daten im ODS und in den Info-Cubes für die dauerhafte Ablage gedacht. In den relational strukturierten ODS-Objekten erfolgt die Speicherung von belegnaher Information, die Info-Cubes bilden die mehrdimensionalen Strukturen als Grundlage für OLAP-Analysen ab.

Basis für Auswertungen und die Erstellung von Berichten mit verschiedenen Front-End-Werkzeugen sind die im Business Information Warehouse abgelegten Datenstrukturen. Mögliche Kategorien von Anwendungen basieren zum einen auf Werkzeugen zur Unterstützung des Managements, die von einfachen Berichts- und Abfragesystemen über leistungsstarke Werkzeuge zur Entscheidungsunterstützung bis zu den Führungsinformationssystemen reichen. Zum anderen handelt es sich um die unter den Stichworten Data Mining und OLAP diskutierten Anwendungsbereiche. Eine einheitliche Metadatenbasis ist im Allgemeinen für alle auf das Busi-

ness Information Warehouse zugreifenden Front-End-Werkzeugen essentiell.

Eine grobe Übersicht der Architektur und des Zusammenspiels der beteiligten Komponenten des Business Information Warehouse ergibt sich aus Abbildung 22.



**Abb. 22.** Architektur des Business Information Warehouse 3.1

Integraler Bestandteil des Business Information Warehouse ist der Business Explorer Analyzer als Excel-basiertes OLAP-Werkzeug. Der Bereich des Webreportings wird durch die neue Technologie auf Basis des Web Application Servers und dem Web Application Designer adressiert. Das zentrale Front-End für SAP BW ist der Business Explorer (BEx). Dieser besteht seit der Version 3.0 des BW aus den Komponenten BEx Analyzer, BEx Browser, BEx Query Designer und dem BEx Web Application Designer.

Basis zur Speicherung der Daten im Business Information Warehouse bildet die relationale Datenbanktechnik. Für die Speicherung mehrdimensionaler Datenstrukturen in relationalen Datenbanken hat sich das Star Schema als Standard durchgesetzt (siehe Kapitel 7).

Auf Grund ihrer Bedeutung für den Modellierungsprozess erfolgt in den folgenden Abschnitten eine detailliertere Betrachtung der Ebenen der Datenspeicherung sowie daran anknüpfend der Auswertung bzw. Analyse.

## 4.1 Komponenten der Datenspeicherung im BW

Die Speicherung von Daten erfolgt im Business Information Warehouse in Objekten entlang einer mehrstufigen Architektur. In der idealtypischen Konzeption von Systemlösungen auf dieser Basis erfüllt jede Ebene eine spezifische Funktion innerhalb eines Prozesses der Transformation und Aufbereitung.

Grundbausteine zur Speicherung sind die in Abschnitt 4.1.1 dargestellten Info-Objekte, die als kleinste Einheit angesehen werden können und auf denen alle anderen Speicherobjekte basieren. Dem Quellsystem am nächsten im Sinne eines periodischen Datenaktualisierungsprozesses ist die Persistent Staging Area, deren Darstellung Gegenstand von Abschnitt 4.1.2 ist. Die Möglichkeiten der Transformation stehen in Abschnitt 4.1.3 im Vordergrund. Durch diesen Prozess erfolgt die Aufbereitung der Quelldaten etwa aus dem PSA heraus für die nachfolgenden Objekte des Operational Data Store (ODS) und der eigentlichen Info-Cubes, die die wesentlichen Speicherobjekte im Business Information Warehouse darstellen und die in den Abschnitten 4.1.4 und 4.1.5 beschrieben werden. Aspekte des Datenflusses innerhalb des BW-Systems sind Gegenstand der in Abschnitt 4.1.6 diskutierten Integrationsarchitektur und der möglichen Ausgestaltungsvarianten hiervon.

### 4.1.1 Info-Objekte

Info-Objekte bilden die Basis aller Info-Provider wie ODS-Objekte oder Info-Cubes. Sie sind eindeutig gekennzeichnet durch ihren technischen Namen der durch eine Bezeichnung ergänzt wird. Für jedes Info-Objekt ist festgelegt, von welchem Datentyp und welcher Datenlänge es ist. Info-Objekte dienen der Speicherung von Merkmalen, Zeitangaben, Einheiten und Kennzahlen. Dieser Typ eines Info-Objektes definiert neben den allgemeinen weitere spezifische Eigenschaften.

Info-Objekte vom Typ eines Merkmals können zur weiteren Beschreibung Attribute besitzen, die als Stammdaten ebenfalls auf Basis von Info-Objekten definiert sind. Mit den datentechnischen Zusammenhängen zwischen Merkmalen und Attributen befasst sich Abschnitt 8.1 detailliert. Neben dieser Form der attributiven Beschreibung bietet das Business Information Warehouse die Möglichkeit der Speicherung von beschreibenden Texten zu Merkmalsausprägungen. Diese können in verschiedenen im System gepflegten Sprachen verfasst sein.

Die Texte eines Info-Objektes müssen nicht eindeutig sein, die Schlüssel hingegen schon. Diese können im BW auch aus mehreren Komponen-

ten zusammengesetzt sein. Dieses Konstrukt wird als Klammerung bezeichnet. Ein Beispiel zur Nutzung von Klammer-Objekten im BW ist die Verwendung des Sachkontos mit der Bedeutung wie im R/3, das nur in Verbindung mit der Angabe eines Kontenplans eindeutig festgelegt ist. Dazu wird das Info-Objekt des Sachkontos an den Kontenplan geklammert.

Es ist auch möglich, Info-Objekte als Referenz-Merkmal anzulegen, wodurch ein Verweis auf ein bereits bestehendes Info-Objekt angelegt wird. Sie benutzen beide die gleichen Stammdatentabellen, diese sind also immer für das zugrunde liegende Info-Objekt und alle davon abhängig definierten Referenz-Merkmale gleichermaßen gültig. Der klassische Anwendungsfall hierfür sind Partnerrollen wie etwa für die verschiedenen Typen eines Kunden, so dass neben dem Kunden selbst auch noch die Rollen Warenempfänger und Rechnungsempfänger von Bedeutung sind.

Info-Objekte zur Aufnahme von Zeitangaben haben die spezifische Eigenschaft, dass für diese zeitlichen Objekte im System vordefinierte Konvertierungsroutinen zur Verfügung stehen. Zeitmerkmale können jedoch nicht selbst im BW angelegt werden, es muss hierbei immer auf die vom Business Content zur Verfügung gestellten Objekte zurückgegriffen werden.

Zur Aufnahme von Kennzahlenwerten steht im Business Information Warehouse ein eigener Info-Objekt-Typ zur Verfügung. Für diesen stehen neben numerischen Datentypen auch Typen mit Einheit zur Verfügung und ermöglichen die Speicherung von Mengen und Betragsgrößen. Für die Angabe von Werten mit Währungen stellt das System dafür spezifische Umrechnungsfunktionalitäten zur Verfügung, so dass innerhalb eines Modells eine Kennzahl auch mit unterschiedlichen Währungen vorkommen kann, die vom System automatisch nach verschiedenen Umrechnungsvarianten ineinander überführt werden können. Bei den Mengeneinheiten steht diese automatische Umrechnungsfunktionalität nicht zur Verfügung.

Einheiten-Info-Objekte basieren auf den im System vorgegebenen Objekten zur Speicherung von Währungen und Mengeneinheiten und stellen selbst nur eine Referenz darauf dar.

#### **4.1.2 Persistent Staging Area**

Bewegungsdaten wie auch Stammdatattribute und Texte finden in der Persistent Staging Area (PSA) eine Eingangsablage, in der die Daten untransformiert und ohne Modifikation für die weitere Verarbeitung gespeichert werden. Die in transparenten Tabellen gespeicherten Quelldaten

können dort bei entsprechender Berechtigung einer Qualitätsprüfung unterzogen und ggf. auch verändert oder gelöscht werden.

Über die PSA wird im Business Information Warehouse auch eine Verbesserung der Ladeperformance erreicht, da der eigentliche Ladevorgang von der Weiterverarbeitung mit den notwendigen Transformationsschritten entkoppelt ist. Die PSA-Daten dienen ebenfalls einem möglichen Neuaufbau von Info-Cubes, die aus diesen Daten befüllt wurden.

### **4.1.3 Extraktion und Transformation**

Die Quelldaten, die in einem Business Information Warehouse als Informationsspeicher landen sollen, müssen zunächst einen Prozess der Extraktion, Transformation und des Ladens durchlaufen. Dieser als ETL-Prozess umschriebene Vorgang hat im BW zwei wesentliche Komponenten, in denen die Aufgaben der Transformation durchgeführt werden.

Ausgehend vom Quellsystem stellt dieses eine Datenlieferung dem BW auf Anforderung in einer so genannten Extraktstruktur zur Verfügung. Eine Teilsicht hiervon stellt die Transferstruktur da, die eine direkte Entsprechung im BW hat. Aus dem Quellsystem erfolgt dann die Weitergabe einer Menge von Datensätzen ohne Transformation über die Transferstruktur in die PSA. Die Kombination der Extraktstruktur einerseits und der Transferstruktur sowohl auf Quellsystemseite wie auch im Business Information Warehouse andererseits wird auch als Data-Source bezeichnet.

Die erste Transformationskomponente entlang des Datenflusses basiert auf dem Konstrukt der Info-Source, die Daten verschiedener Data Sources zusammenfasst, einer Modifikation in Form der sog. Übertragungsregeln unterzieht und für die weitere Verarbeitung zur Verfügung stellt. Info-Sources sind die zentralen Objekte, auf deren Basis die Befüllung von ODS-Objekten und Info-Cubes erfolgt. Die Kommunikationsstruktur spiegelt dabei die transformierte Struktur wieder.

Aus den Info-Sources erfolgt die Datenweitergabe an die Zielobjekte über die Fortschreibungsregeln, die den zweiten wesentlichen Teil der Transformation im Business Information Warehouse implementieren. In Abbildung 23 ist dieser Zusammenhang dargestellt und die Komponenten der Data-Source und Info-Source sind unmittelbar verdeutlicht.

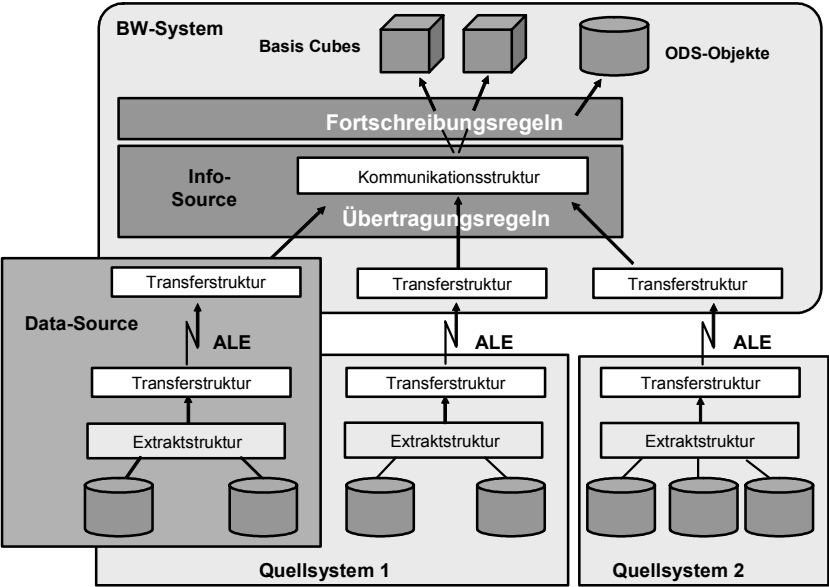


Abb. 23. Stufen im Transformationsprozess

Die erste Transformation entlang des gesamten Prozesses erfolgt durch die Festlegung von Übertragungsregeln. In Abbildung 24 ist ein exemplarisches Beispiel hierfür dargestellt. Aus der zur Verfügung stehenden Extraktstruktur des Quellsystems repräsentiert die Transferstruktur die selektierte Teilmenge. Einzelne Felder unterliegen nun durch die Übertragungsregeln einer weiteren Umwandlung, so etwa im Beispiel der Ableitung eines Feldes Umsatz aus den Quellfeldern für Menge und Preis.

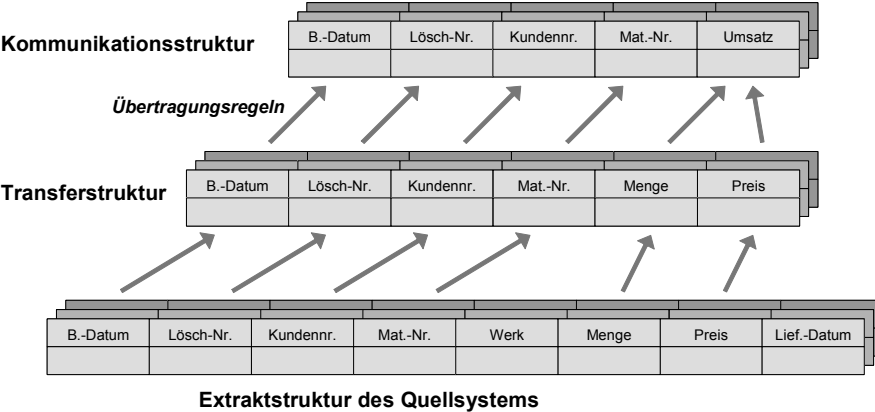


Abb. 24. Transformation mit Übertragungsregeln

Die Kommunikationsstruktur definiert damit den Satzaufbau der Daten, die in einem weiteren Schritt auf Basis der Info-Source an die ODS-Objekte bzw. an die Info-Cubes weitergereicht werden. Der hier ansetzende Schritt der Fortschreibung ermöglicht die Implementierung einer umfangreichen Business-Logik zur Datenaufbereitung und Transformation.

#### **4.1.4 Operational Data Store**

Zur Speicherung der konsolidierten und bereinigten Bewegungsdaten auf detaillierter Belegebene dient im Business Information Warehouse der Operational Data Store. Hier erfolgt die Ablage in relationalen Tabellen und es wird von einem ODS-Objekt gesprochen. Neben den identifizierenden Schlüsselfeldern gibt es die Datenfelder, in denen die Werte und beschreibenden Merkmale abgelegt sind.

Ein wesentlicher Vorteil von ODS-Objekten ist die Möglichkeit, Datensätze ändern zu können. Damit sind auch sehr volatile und zeitnahe Berichtsanforderungen möglich zu implementieren. Die Ebene des ODS ermöglicht die persistente Ablage bereits transformierter, integrierter und gleichwohl granularer Daten aus dem Extraktionsprozess.

Die erste wesentliche Aufgabe eines ODS-Objektes im gesamten Datenfluss ist die Transformation und Datenbereinigung beim Datentransfer von der Persistent Staging Area aus auf Basis der Übertragungsregeln und dient der physischen Speicherung der aufbereiteten Daten. Weiterhin dient es der Integration verschiedener Datenquellen, die den gleichen Prozess abbilden. Dieser auch als „inbound level“ bezeichnete Bereich im Business Information Warehouse ist eine wesentliche Grundlage eines Data Warehouse, denn es stellt einen themenorientierten, konsolidierten und integrierten Datenbestand bezogen auf einen Geschäftsprozess dar, der auch verschiedene Quellsysteme umfasst.

Für die weitere Integration über verschiedene Geschäftsprozesse hinweg ist ebenfalls der Aufbau entsprechender ODS-Objekte und/oder Info-Cubes im Business Information Warehouse vorgesehen. Auf Basis von ODS-Objekten können Benutzer auch direkt Auswertungen und Analysen durchführen, da dies aber auch eine Frage der Performance ist, sollte diese Möglichkeit nur dann genutzt werden, wenn diese detaillierte Ebene für Auswertungen benötigt wird und die Nutzung eines Info-Cubes nicht möglich oder sinnvoll ist. Die im Gegensatz zu den ODS-Objekten nicht relational sondern mehrdimensional strukturierten Speicherobjekte im BW stellen die im Allgemeinen bessere Wahl als Basis für Analysen und Berichtssysteme dar.

### 4.1.5 Info-Cubes

Die wichtigsten Objekte für verschiedenste Arten von Analysen und Berichten im Business Information Warehouse sind die Info-Cubes, die aus fachlicher Sicht einen in sich geschlossenen Datenbestand repräsentieren. Die Datenorganisation erfolgt mehrdimensional und ist damit klar auf OLAP-Anforderungen ausgerichtet. Im BW gibt es verschiedene Typen dieser zentralen mehrdimensionalen Datenspeicher. Die wichtigsten Info-Cubes sind die Basis-Cubes, denn nur in diesen erfolgt eine physische Speicherung von Daten. Im Rahmen eines meist zyklischen Datenladevorganges erfolgt deren Befüllung und Aktualisierung.

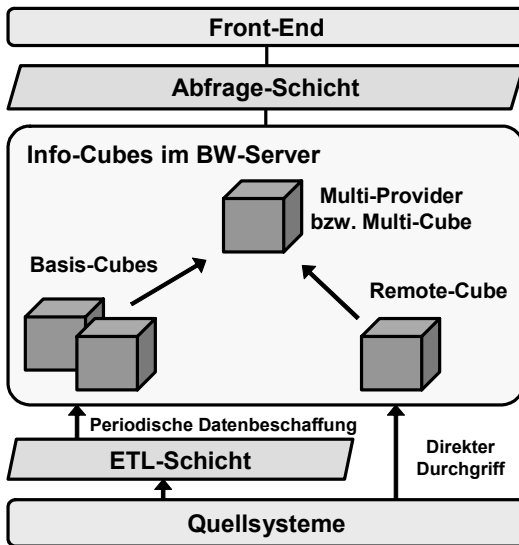
Der Informationsgehalt eines Basis-Cubes bezieht sich auf einen klar definierten betriebswirtschaftlichen Kontext und bildet die Grundlage zur Unterstützung von Fach- und Führungskräften. Bei stark durch strategische Fragestellungen bestimmten Aufgabengebieten werden übergreifende Sichtweisen benötigt. Hierzu stellt das BW die Möglichkeit der Berechnung abgeleiteter Kennzahlen auf Basis eines Info-Cubes bereit. Oftmals sind Fragen auf strategischer Ebene nicht auf Basis eines einzelnen Info-Cubes mit abgegrenztem betriebswirtschaftlichem Fokus zu beantworten, sondern benötigen eine Verbindung verschiedener Sichtweisen und Aspekte. Um dazu keine Datenredundanzen in einem eigenständigen übergeordneten Basis-Cube zu erzeugen, stellt das BW das Konstrukt eines Multi-Providers bereit, dessen Festlegung nur eine Art Layer darstellt und verschiedene auswertbare Objekte im BW miteinander virtuell verknüpft. Ein Multi-Provider beinhaltet somit keine physische Datenablage, sondern die Abfrage der Daten wird zur Laufzeit umgesetzt in Abfragen an die verknüpften Basisobjekte.

Auf Basis eines Multi-Providers können verschiedene Basis-Cubes zu einer übergeordneten betriebswirtschaftlichen Sicht verknüpft werden. Daneben besteht aber auch noch die Möglichkeit, Remote-Cubes mit in die Definition aufzunehmen. Diese werden genutzt, um kleine Datenbestände direkt aus den operativen Vorsystemen wie beispielsweise R/3 für Reportingzwecke zur Verfügung zu stellen. Die möglichen Formen von Info-Cubes sind in Abbildung 25 zusammengefasst.<sup>10</sup>

---

<sup>10</sup> Für den in der Abbildung aufgeführten Multi-Provider war im BW2 die korrekte Bezeichnung Multi-Cube. In der Definition von BW3 Multi-Providern können hingegen auch Stammdaten und ODS-Objekte integriert werden.





**Abb. 25.** Verschiedene Typen von Info-Cubes

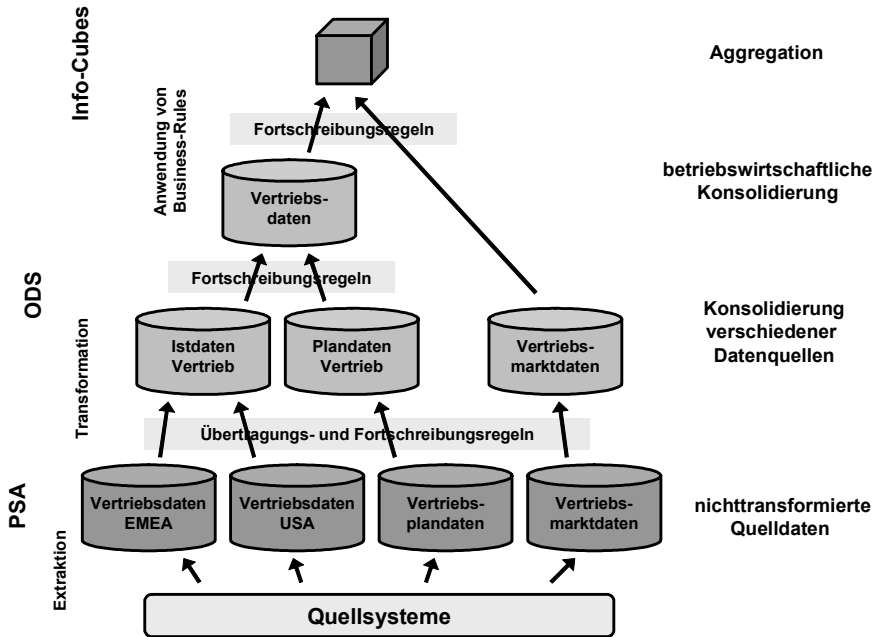
Die als auswertbare Datenlieferanten in Frage kommenden Objekte werden als Info-Provider bezeichnet. Diese bilden die Grundlage für Reporting und Analyse, die auf Basis von Abfragen (Queries) implementiert werden. Aus Sicht der Abfragedefinition spielt der Typ des zugrunde liegenden Info-Providers keine Rolle, da diese uniform behandelt werden, allerdings ergeben sich erhebliche Unterschiede hinsichtlich der Performance.

#### 4.1.6 Varianten der Datenbereitstellung

Entlang des Datenflusses ausgehend vom Quellsystem ist zunächst die Persistent Staging Area zu nennen. Diese fungiert als Sammler der von Quellsystemen gelieferten Daten. Der so genannte Operational Data Store ist im Kern eine relationale Speicherkomponente, die in einer Referenzarchitektur auf der Ebene der Integration und Transformation angesiedelt wird. Im Business Information Warehouse sind diese Objekte zwar ebenfalls als Informationsspeicher für Auswertungen und Analysen optional verfügbar, hierfür sind aber die mehrdimensional strukturierten Info-Cubes dediziert konzipiert. Diese sind die wichtigsten Objekte für Analysen und Auswertungen.

Eine zusammenfassende Darstellung des Datenflusses findet sich in Abbildung 26. Ausgehend von den Quellsystemen dient die Ebene der Persistent Staging Area der Speicherung der nicht aufbereiteten und nicht mo-

difizierten Daten aus den vorgelagerten Quellsystemen. Eine erste Transformation ermöglicht die Datenweitergabe an die ODS-Objekte, die in der Darstellung der Konsolidierung verschiedener Datenquellen dienen. Nicht dargestellt ist in der Abbildung die ebenfalls bestehende Möglichkeit der direkten Datenbefüllung der Info-Cubes aus der PSA unter Umgehung zwischengeschalteter ODS-Objekte. In einer weiteren Stufe erfolgt die betriebswirtschaftliche Konsolidierung und Aggregation u. a. mit dem Aufbau von Info-Cubes.



**Abb. 26.** Datenfluss und Integrationsarchitektur im SAP BW

Diese skizzierte Architektur hat einen gewissen Referenzcharakter, da dies eine optimale Ausgestaltung repräsentiert. In der Praxis sind natürlich verschiedenste Varianten hiervon denkbar, gerade in Fällen geringer Transformationsanforderungen ist beispielsweise die Konsolidierungsebene der ODS-Objekte obsolet.

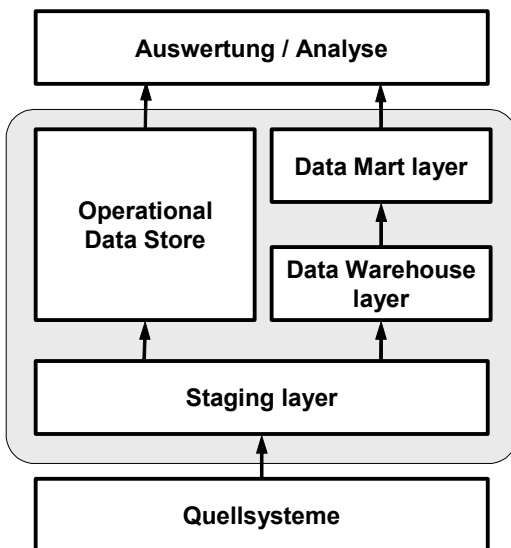
Die Ebene der ODS-Objekte erfüllt im Business Information Warehouse die zentrale Funktion der Speicherung auch historischer transformierter und aufbereiteter Daten zur weiteren Verwendung in Info-Cubes. Diese können dann unabhängig vom gesamten Datenaktualisierungsprozess neuen Anforderungen flexibel angepasst und neu aufgebaut werden. Der Aufbau von Info-Cubes ist im Wesentlichen determiniert von den Berichts-

und Analyseanforderungen, die ODS-Objekte stellen die Basisdaten für verschiedene Szenarien und Reporting-Anforderung zu Verfügung.

Die Diskussion der verschiedenen Varianten der Ausgestaltung des Datenbereitstellungsprozesses im Business Information Warehouse erfolgt unter dem Stichwort der Staging-Szenarien. Die dargestellte mehrstufige Lösung mit den Ebenen der Persistent Staging Area, den ODS-Objekten und den Info-Cubes ist die idealtypische Lösung und wird im Allgemeinen empfohlen. In einem weiteren Szenario erfolgt die Analyse bereits auf der ODS-Ebene, diese Vorgehensweise empfiehlt sich insbesondere bei Berichtswesen auf Belegebene, da die ODS-Objekte das Überschreiben von Daten unterstützen. Ebenfalls ist es denkbar, dass keine ODS-Objekte im Datenfluss benötigt werden, da die Quellsysteme bereits integrierte Daten bereitstellen. Dies ist im Falle eines R/3-Quellsystems beispielsweise für die Extraktion aus der Ergebnisrechnung im CO-PA sinnvoll.

Weitere Varianten der Kombination mit verschiedenen Info-Cubes sind im Data Mart Konzept wie in Abschnitt 10.3 dargestellt berücksichtigt.

Einen weiteren Aspekt bringt der Ansatz des unternehmensweiten Data Warehouses und Forderungen nach mehr Real-Time-Nähe in die Diskussion. Ein unternehmensweites Data Warehouse hat noch stärker die Aspekte der Integration im Fokus. Etwas verallgemeinert wird von einer konzeptionellen Data Warehouse Schichten-Architektur gesprochen, die in Abbildung 27 dargestellt ist.



**Abb. 27.** Konzeptionelle Schichten-Architektur im Data Warehousing

Dabei sind die folgenden Schichten (*layer*) zu differenzieren:

- Staging layer
- Data Warehouse layer
- Data Mart layer
- Operational Data Store layer

Die erste Schicht deckt den Bereich des Staging ab und dient der Ablage der Quelldatenextrakte auf Belegebene (*row level*). Ziel der Data Warehouse Schicht ist der Aufbau des integrierten Datenmaterials auf granularer Ebene. Diese stellt damit die Historie zur Verfügung und ist gleichermaßen unabhängig von der (betriebswirtschaftlichen) Anwendung. Die Data Mart Schicht bietet nun ihrerseits genau diese Sicht eines spezifischen Anwendungsbereiches und stellt dafür aggregierte integrierte Daten zur Verfügung. Ebenfalls Anwendungsspezifisch ist der Operational Data Store, in diesem sind die Daten aber granular abgelegt. Diese Schicht bedient die Anforderungen an ein Real-Time-nahes Data Warehousing, da hier kurze Aktualisierungszeiten möglich sind.

Zu beachten ist, dass diese Schichten-Architektur rein konzeptionell ist, die physische Ebene und die Strukturierung verteilter Systeme zu einer gesamten Data Warehouse Landschaft ist eine andere Fragestellung, die unter dem Begriff der *Topologie* getrennt zu diskutieren ist.

## 4.2 Auswertung und Analyse auf Basis vom BW

Die gespeicherten Inhalte einer Data Warehouse-Lösung bestimmen maßgeblich deren mögliches Einsatzspektrum. Für die Endanwender sind neben den Inhalten jedoch auch die verfügbaren Front-End-Tools zur Auswertung und Analyse entscheidend und bilden neben dem allgemeinen Antwortzeitverhalten ein wesentliches Kriterium zur Akzeptanz des analytischen Systems, denn gerade durch die Endbenutzerwerkzeuge sind die analytischen Funktionalitäten festgelegt, die den Anwender in seinen fachlichen Aufgabenstellung unterstützen. Zu berücksichtigen ist dabei, dass unterschiedliche Benutzertypen durchaus verschiedene Anforderungen und Präferenzen an das analytische Front-End haben können.

Grundlage jeder Abfrage und Auswertung gespeicherter Daten im Business Information Warehouse sind die in Abschnitt 4.2.1 behandelten Queries. Darauf setzen die verschiedenen Front-End Werkzeuge auf, von denen explizit der Business Explorer Analyzer in Abschnitt 4.2.2 sowie der Bereich der Web-Anwendungen mit BEx Web Applications in Abschnitt 4.2.3 näher betrachtet werden.

### 4.2.1 Queries als mehrdimensionale Schemata

Queries formen die Grundlage jeder Abfrage und Auswertung gespeicherter Daten im Business Information Warehouse und stellen bereits spezifische Sichten auf die vorhandenen Daten dar. Auf der Ebene der Queries erfolgt auch die Festlegung von berechneten und eingeschränkten Kennzahlen, die nicht physisch gespeichert sind sondern erst zur Laufzeit kalkuliert werden.

Zur Gestaltung benutzerspezifischer Sichten lassen sich überdies so genannte Strukturen nutzen, die im Sinne einer virtuellen Dimension vorhandene Kennzahlen und Merkmale zu einer anforderungsspezifischen Teilsicht bündeln und kombinieren. Ebenso sind Ausnahmen bzw. Exceptions definierbar, über die sich Zahlenwerte farblich markieren und darüber geschäftliche Tendenzen besser hervorheben und erfassen lassen. Dagegen dienen Bedingungen bzw. Conditions dazu, anzuzeigende Datensätze nach bestimmten Filterkriterien zu selektieren und dadurch das angezeigte Datenvolumen auf die relevanten Inhalte zu reduzieren.

Zu Beginn der Querydefinition ist der eindeutige Info-Provider zu selektieren, auf den sich die Query bezieht. Dies kann beispielsweise ein Info-Cube oder ODS-Objekt sein. Die Erstellung von Queries erfolgt im Query Designer und diese sind im Business Information Warehouse mit einem eindeutigen technischen Namen zentral gespeichert. In Abbildung 28 ist der Weg der Daten bei der Abfrageausführung von einem Info-Cube bis zur Präsentationsebene dargestellt.

Die wesentliche Komponente zur Datenaufbereitung ist der OLAP-Prozessor des BW-Systems, dieser führt auch die Navigationsschritte auf Basis der Benutzerinteraktionen durch. Auf der Präsentationsebene ist dann der jeweilige Navigationszustand abgebildet. Dem Benutzer bietet sich hier auch die Möglichkeit der Speicherung eines Navigationszustandes in Form eines Bookmarks. Hierbei sind die Varianten der Speicherung inklusive der angezeigten Daten sowie eine dynamische Definition zu differenzieren.

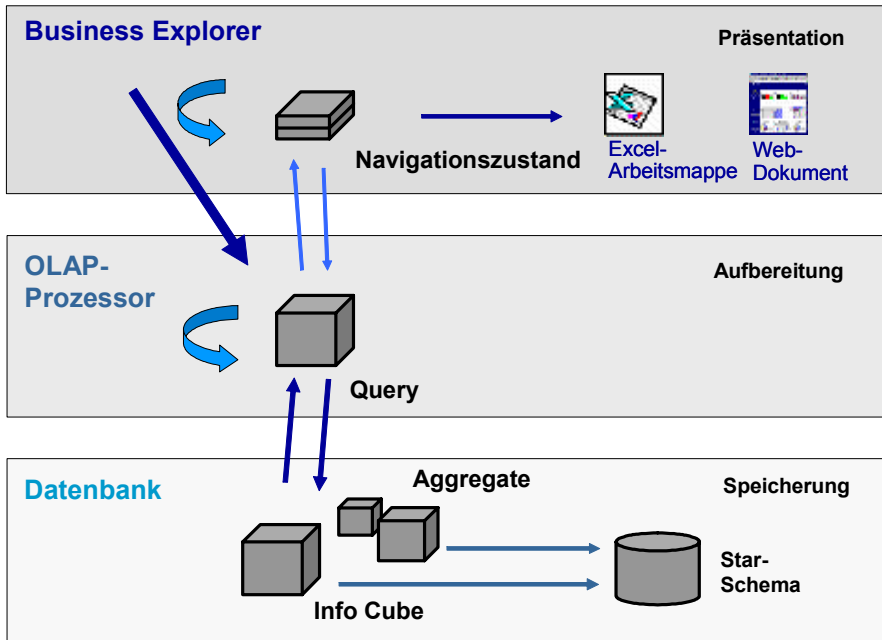


Abb. 28. Ebenen der Query-Ausführung

### Eingeschränkte und berechnete Kennzahlen

Für die Ablage von betriebswirtschaftlichen Kennzahlen gibt es im Business Information Warehouse eine spezifische Form von Info-Objekten, die auch Einheiten wie beispielsweise Währungen berücksichtigen können und als Kennzahl in die Definition der Info-Provider aufgenommen werden. Dabei sind diese Kennzahlen bei der Fortschreibung in die Info-Provider zu berechnen bzw. direkt aus den Vorsystemen zu übernehmen. Darüber hinaus besteht die Möglichkeit, auf der Ebene der Query abgeleitete Kennzahlen festzulegen. Zu diesen gehören im Business Information Warehouse die *berechneten Kennzahlen* und die *eingeschränkten Kennzahlen*.

Sofern sich eine Kennzahl aus den vorhandenen Kennzahlen eines Info-Providers durch Berechnung ergibt, kann diese Formel fest hinterlegt werden. Diese so genannte *berechnete Kennzahl* steht dann global auf Ebene des Info-Providers zur Verfügung. Die Definition der Berechnungslogik kann dabei auf verschiedene, auch komplexe mathematische Operatoren zurückgreifen, wobei das System den Benutzer bei der Erstellung berechneter Kennzahlen u. a. durch eine Syntaxprüfung unterstützt.

Über die Definition von ebenfalls global auf der Ebene des Info-Providers gespeicherter *eingeschränkter Kennzahlen* ist es möglich, für de-

ren Berechnung eine Beschränkung auf relevante Inhalte vorzunehmen. Hierzu können Restriktionen für einzelne Werte, Intervalle und Hierarchieknoten festgelegt werden. Diese Form einer abgeleiteten Kennzahl findet häufig bei der Berechnung von Referenzgrößen Anwendung, auf deren Basis beispielsweise Indexberechnungen möglich sind.

Nach der globalen Speicherung einer berechneten oder eingeschränkten Kennzahl steht diese für alle weiteren Queries zur Verfügung, die sich auf den betreffenden Info-Provider beziehen.

### ***Erweiterte Query-Bestandteile***

Die Parametrisierung von Abfragen erfolgt im Business Information Warehouse über so genannte Variablen, durch die eine Selektion für Merkmale, Hierarchien und Hierarchieknoten möglich ist. Des Weiteren können sie auch Texte und Formelelemente aufnehmen. Durch den Einsatz von Variablen können Abfragen bereits vor der erstmaligen Ausführung benutzer-spezifisch eingegrenzt werden und erst zur Laufzeit erfolgt die Belegung der Variablen mit konkreten Ausprägungen. Somit sind Queries sehr flexibel für verschiedene Auswertungen auf Basis des gleichen grundsätzlichen Aufbaus verwendbar. Die Speicherung von Variablen erfolgt global auf der Ebene des Info-Providers und diese stehen somit für alle Queries des entsprechenden Info-Providers zur Verfügung.

Eine Klassifizierung von Variablen kann nach der Art der Verarbeitung und Wertzuweisung erfolgen. Hierbei sind die folgenden fünf unterschiedlichen Varianten zu unterscheiden:

- Bei der manuellen Wertzuweisung gibt der Benutzer vor der Ausführung einer Query einen Wert für die Variable ein, der auch mit einem Vorschlagswert (Default-Value) vorbelegt sein kann.
- Über den Ersetzungspfad ergibt sich der Wert der Variablen automatisch aus dem Kontext und wird vom System gefüllt.
- Über einen Customer Exit ist es möglich, eine spezifische Logik zur Ermittlung des Variablenwertes individuell zu programmieren.
- Die Verarbeitungsart Berechtigung dient dem Befüllen von Variablen mit Werten auf Basis von Benutzerberechtigungen.
- Für die Variablen des Business Content steht die Verarbeitungsart eines SAP-Exits zur Verfügung.

Eine Variable ist durch den Typ, den Namen, einen beschreibenden Text und die Verarbeitungsart charakterisiert. In Abhängigkeit vom Variablentyp sind unterschiedliche weitere Einstellungen vorzunehmen, die u. a. das zugrunde liegende Merkmal und die Eingabeoptionen festlegen. Hier wird auch festgelegt, ob die Werteingabe verpflichtend oder optional ist.

Die Pflege von Variablen erfolgt assistentengestützt im Query Designer. Über den Business Content steht ein umfangreicher Fundus an vordefinierten SAP-Exit-Variablen zur Verfügung, die u. a. spezifische Zeitreihensichten ermöglichen und automatische Kumulierungen auf Basis von Zeitmerkmalen vornehmen.

In Verbindung mit der Definition von Strukturen bilden Variablen ein sehr mächtiges Instrumentarium zur flexiblen Analyse von Info-Providern. In diesem Sinne stellen die Queries eine besondere Form eines mehrdimensionalen Schemas dar. In Abbildung 29 ist eine exemplarische Strukturdefinition einer Zeitreihe visualisiert.

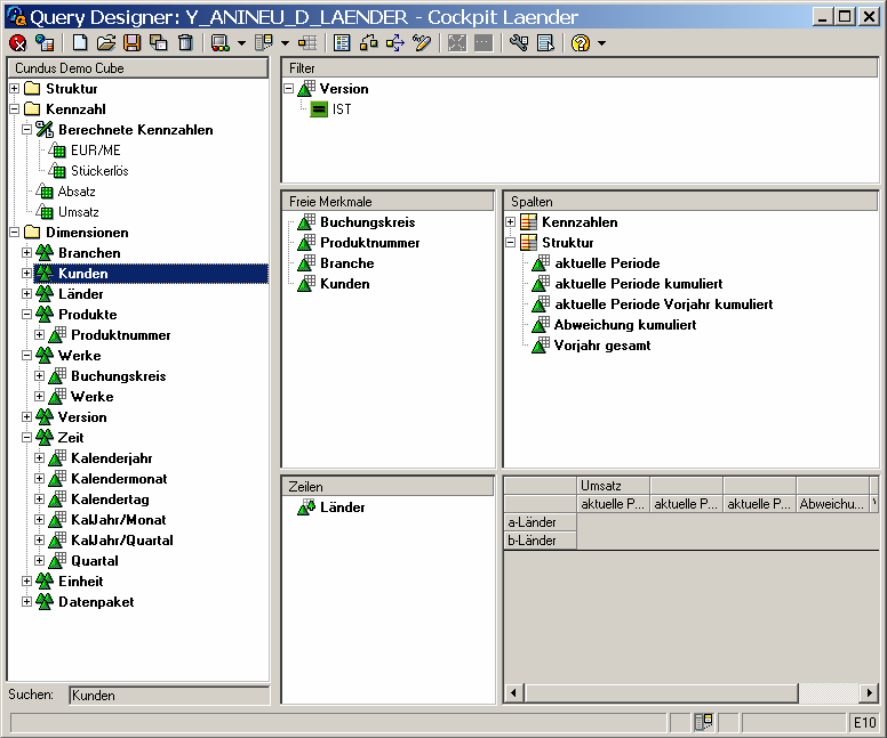


Abb. 29. Strukturdefinition im Query Designer

Mit dem in einer Query festgelegten Aufbau in den Spalten, Zeilen und den freien Merkmalen ist die Sichtsicht der Abfrage nach dem Aufruf festgelegt. Diese dient aber oftmals nur als Ausgangspunkt für weitere Analysen durch Navigation. Interessante Anknüpfungspunkte für weitere Navigationsoperationen, die sich etwa durch Über- oder Unterschreiten von definierten Schwellwerten ergeben, können durch die Definition von Ex-



*ceptions* (Ausnahmen) in der Datensicht farblich hervorgehoben werden. Dies kann auch als Ausnahmeberichtswesen umschrieben werden. Durch *Conditions* (Bedingungen) sind die dargestellten Inhalte für spezifische Analysezwecke weiter einschränkbar. Bedingungen und Ausnahmen lassen sich im Query Designer anlegen, können aber auch vom Benutzer zur Zeit der Ausführung einer Query angepasst bzw. verändert werden.

#### 4.2.2 Business Explorer Analyzer

Eine zentrale Komponente des Business Information Warehouse ist der Business Explorer Analyzer, der als Plug-In für Microsoft Excel dieses um die OLAP-spezifischen Funktionalitäten und den Datenzugriff auf das BW erweitert. Zu den möglichen Operation gehören neben den klassischen Funktionalitäten der Navigation durch drill-down und roll-up auch die völlig flexible Veränderung des Aufbaus und auch der drill across, der Sprünge in andere Abfragen bzw. Berichte implementiert. Die Bearbeitung der Navigationsanforderungen erfolgt dabei im OLAP-Prozessor auf dem BW-Server und nur das Ergebnis einer Abfrage wird zum Excel-Client übertragen. Die mehrdimensionalen Funktionen stehen in Excel über eine Symbolleiste und spezifische Kontextmenüeinträge zur Verfügung.

Nach dem Aufruf einer Query in Excel ergibt sich eine Datensicht auf Basis des initial festgelegten Aufrisses (siehe Abbildung 30). Die Darstellung der möglichen Merkmale zur Selektion und Navigation erfolgt oberhalb des Datenbereiches. Innerhalb dieses aufgespannten Datenraumes können nun die folgenden Navigationsoperationen durchgeführt werden:

- Auf- oder Zuklappen von Hierarchieknoten (*drill down* bzw. *roll up*)
- Setzen von Filterwerten für Merkmale,
- Veränderung des Aufbaus von Führungsspalten,
- Veränderung des Aufbaus der Kopfzeile
- Vorgabe von Sortierungen.

Zur optischen Anreicherung können in die Excel-Arbeitsmappen zusätzlich Logos oder spezielle Hintergrundbilder eingefügt werden. Zur Speicherung der Arbeitsmappen stehen die gleichen Optionen zur Verfügung, wie bei der Ablage der Queries. Diese lassen sich entweder in den Favoriten des Benutzers oder in einer Benutzerrolle speichern.

The screenshot shows a Microsoft Excel window titled 'Mappe2'. A context menu is open over the 'Werke' column, showing options like 'Zurück', 'Zurück zum Anfang', 'Hinzufügen Aufruf nach', 'Austauschen Werke mit', 'Hierarchie expandieren', 'Hierarchie aktiv', 'Sortieren', 'Springen', 'Exceptions', 'Währungsumrechnung', 'Werke', 'Alle Merkmale', and 'Eigenschaften ...'. The table data is as follows:

	Umsatz 1999	Umsatz 2000	Abw Umsatz	Bruttoerlös 1999	Bruttoerlös 2000
WERKE GESAMT	250.493.286,59 DM	354.261.075,87 DM	103.767.789,28 DM	265,23 DM/ST	318,50 DM/ST
▷ BOSSETTO GMBH	6.452.477,80 DM	6.922.132,37 DM	469.654,57 DM	339,46 DM/ST	350,00 DM/ST
▷ NORMBAU AG	34.621.712,60 DM	93.270.639,05 DM	58.648.926,45 DM	244,45 DM/ST	327,00 DM/ST
▷ MODELLBAU GMBH	22.004.612,86 DM	20.899.304,29 DM	-1.105.308,57 DM	1.105,20 DM/ST	1.133,00 DM/ST
▷ HORNBAU AG	187.414.483,33 DM	233.169.000,16 DM	45.754.516,83 DM	245,34 DM/ST	298,00 DM/ST

Abb. 30. Darstellung des Query-Ergebnisses im BEx Analyzer

### 4.2.3 BEx Web Applications

Gerade für unternehmensweite Berichtssysteme bietet sich der Einsatz von Webtechnologien an. Über die so genannten *Web Applications* gibt es die Möglichkeit der Auswertung und Analyse von BW-Daten über einen üblichen Browser.

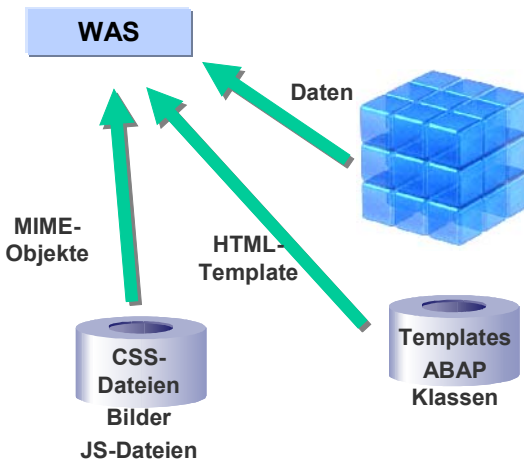
Seit dem Release 3 des Business Information Warehouse wird die hierfür notwendige Funktionalität direkt vom Web Application Server, dem Kern des BW-Systems, wahrgenommen und ein Rückgriff auf den Internet Transaction Server (ITS) ist nicht mehr notwendig. Auf dieser technologischen Basis ist neben der schnellen Erstellung einfacher Web-Anwendungen auch die Implementierung selbst komplexer Funktionalitäten möglich.

Die Erstellung dieser Web-Applikationen erfolgt im Wesentlichen mit dem *BEx Web Application Designer* in dem die so genannten *Web Templates* erstellt werden, für anspruchsvollere Designanforderungen kann der Entwickler jedoch auch auf dedizierte Werkzeuge zur Web-Entwicklung

zurückgreifen. Dabei kommen HTML, Java Script und ABAP Objects als Programmiersprachen zum Einsatz.

Die Ausführung der Web-Templates zur Laufzeit obliegt direkt dem Web Application Server, durch den die Aufbereitung der HTML-Seiten auf Grundlage der Templates erfolgt. Für die Erzeugung von Grafiken ist hingegen eine weitere Komponente, der *Internet Graphic Server*, zuständig. Das Zusammenspiel der Komponenten einer Web Application ist in Abbildung 31 skizziert.

An vielen Stellen erlangt ein derart gestaltetes Informationssystem weitere Funktionalitäten durch einzelne Anwendungselemente auf Basis der Business Server Page Technologie (BSP), mit der die Erstellung kompletter Anwendungen auf Basis des BW-Systems (genauer dem Web Application Server) als Entwicklungs- und Laufzeitumgebung möglich ist.



**Abb. 31.** Komponenten der Web Applications

Für den Aufbau von webbasierten Berichtssystemen kann neben den Query-basierten Standard-Berichten auf individuelle analytische Applikationen, deren Entwicklung mit dem Web Application Designer erfolgt, zurückgegriffen werden.

### **Query-basierte Standard-Berichte**

Die einfachste Form einer BEx Web Application besteht in der direkten Darstellung einer Query über die Option „Query im Web anzeigen“ des Query Designers. Die dabei generierte Standardsicht mit den verfügbaren Funktionalitäten basiert auf einer Schablone, die u. a. die üblichen Möglichkeiten der Navigation und Filterung vorsieht. Der Aufbau des generier-

ten Webberichtes ergibt sich aus der exemplarischen Darstellung in Abbildung 32 und umfasst eine Symbolleiste für allgemeine Funktionen, einen Navigationsblock für Funktionen der Filterung und Aufrissveränderung sowie den eigentlichen Datenbereich.

Neben den Möglichkeiten der grafischen Visualisierung in Form verschiedener Darstellungsvarianten wie Kreis-, Säulen- und Liniendiagrammen umfasst die Standardfunktionalität auch Möglichkeiten des Exports in Excel- oder CSV-Dateien sowie der Definition von Ausnahmen (Exceptions) und Bedingungen (Conditions).

Länder	Absatz	Umsatz	Stückzahl
<b>Gesamtergebnis</b>	<b>1.111.026 ST</b>	<b>354.261.076 EUR</b>	<b>318,86</b>
Welt Gesamt	1.110.301 ST	353.519.084 EUR	318,40
Europa	956.278 ST	298.355.675 EUR	312,00
Westeuropa	923.121 ST	286.245.381 EUR	310,08
SAN MARINO	11 ST	13.030 EUR	1.184,57
SCHWEDEN	13.599 ST	3.541.061 EUR	260,39
PORTUGAL	333 ST	319.366 EUR	960,14
NORWEGEN	1.066 ST	396.240 EUR	371,71
NIEDERLANDE	10.334 ST	4.626.263 EUR	447,66
MALTA	3 ST	2.255 EUR	751,59
LICHTENSTEIN	23.684 ST	3.967.845 EUR	167,53
ITALIEN	17.832 ST	7.717.042 EUR	432,78 EUR/ST
IRLAND	31 ST	42.634 EUR	1.397,83 EUR/ST
GRIECHENLAND	7.903 ST	1.986.477 EUR	251,34 EUR/ST
GROSSBRITANNIEN	9.771 ST	11.609.193 EUR	1.188,10 EUR/ST
FRANKREICH	50.536 ST	13.107.294 EUR	259,36 EUR/ST

Abb. 32. Web-Standardbericht auf Basis einer Query-Definition

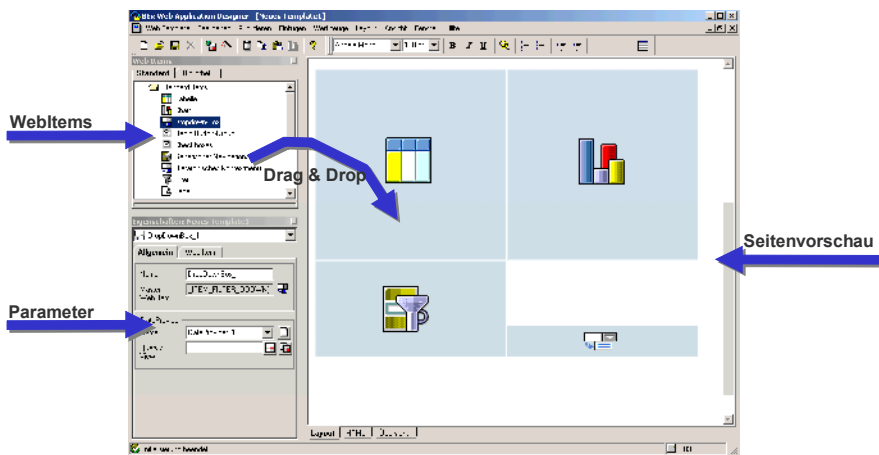
Durch die Speicherung eines Bookmarks kann die gerade durch Navigation erdrillte Sicht in Form einer URL gespeichert werden, die z. B. über die Favoritenliste einen jederzeitigen erneuten Aufruf ermöglicht. Auch eine Weiterleitung dieses Bookmarks per E-Mail ist damit möglich.

Die Definition eines Bookmarks ist dynamisch, d. h. ein Aufruf generiert stets eine Abfrage auf die aktuellen Daten im Business Information Warehouse. Alternativ kann ein Bookmark auch mit den angezeigten Daten gespeichert werden, so dass dieser die statische Sicht repräsentiert.

## Web Application Designer

Gerade für unternehmensweite Berichtslösungen auf Basis der Web-Technologie ist die Anpassung an unternehmensspezifische Layoutvorgaben notwendig und die skizzierten Möglichkeiten der standardmäßigen Darstellung einer Query im Web sind hierfür nicht ausreichend. Die Gestaltung individueller *BEx Web Applications* lässt sich mit dem Werkzeug *Web Application Designer* (WAD) durchführen, der die Erstellung der *Web Templates* unterstützt.

Für die individuelle Gestaltung dieser Web Templates stellt das System vordefinierte Objekte für den Zugriff auf BW-Daten sowie die Navigation darin zur Verfügung. Diese so genannten *Web Items* sind umfangreich parametrisierbar. Die Kombination der verschiedenen Web Items zusammen mit allgemeinen HTML-Elementen erfolgt in dem Werkzeug in der Layout- und HTML-Sicht. Darüber hinaus sind auch externe Editoren zur Webseitengestaltung in die Entwicklung integrierbar. Die verschiedenen Bereiche im Web Application Designer sind in Abbildung 33 dargestellt.



**Abb. 33.** Gestaltung eines Vertriebsberichtes im WAD

Die zur Verfügung gestellten Web Items befinden sich in einer Liste oben links und können direkt in die Layout-Ansicht auf der rechten Seite gezogen werden. Sie sind über den Bereich unten links parametrisierbar. Für die Navigation in umfangreichen Berichtssystemen bietet sich die Nutzung des spezifischen Web Items „Rollenmenü“ an, das die benutzerabhängig zur Verfügung stehenden Favoriten und Rollen auslistet und darüber eine hierarchische Navigation aufbaut. Dieses Verfahren basiert auf dem Berechtigungssystem des Business Information Warehouse. Der Einsatz

einer Linksammlung kann in einfachen Fällen hingegen auch schon ausreichend sein.

### 4.3 Weiterführende Literatur

Zur vertiefenden Betrachtung der allgemeinen Architektur des SAP Business Information Warehouse sind (Chamoni et al. 2004) und (Mehrwald 2003) geeignet. Für die Aspekte der Datenauswertung wird insbesondere auf (Egger 2004) verwiesen. Die technologischen Aspekte des Web Application Servers stehen bei (Wolf 2003), (Färber und Kirchner 2002) und (Heinemann und Rau 2003) im Vordergrund. Der Aufbau eines unternehmensweiten Data Warehouse wird ausführlich in (Haupt 2003) diskutiert.

- Chamoni P, Gluchowski P, Hahne M (2004) Business Information Warehouse. Springer-Verlag, Heidelberg
- Egger N (2004) Praxishandbuch SAP BW 3.1. Galileo Press, Bonn
- Färber G, Kirchner J (2002) mySAP Technology – Einführung in die neue Technologie-Plattform der SAP. Galileo Press, Bonn
- Haupt J (2003) Enterprise Data Warehousing with SAP BW – An Overview. SAP
- Heinemann R, Rau C (2003) SAP Web Application Server. Galileo Press, Bonn
- Mehrwald C (2003) SAP Business Information Warehouse 3 – Architektur, Konzeption, Implementierung. dpunkt-Verlag, Heidelberg
- Wolf F (2003) SAP Web Application Server – Internet-Anwendungen entwickeln mit ABAP, HTML und JavaScript. dpunkt-Verlag, Heidelberg

### 4.4 Zusammenfassung

Die Architektur des Business Information Warehouse ist angelehnt an die allgemeine Referenzarchitektur für ein Data Warehouse. Ausgehend von verschiedenen möglichen Quellsystemen erfolgt ein im Allgemeinen periodischer Datenladevorgang in die zentrale Datenhaltung des Business Information Warehouse (BW).

Im Mittelpunkt der Architektur steht der eigentliche Business Information Warehouse Server, der für die Ablage, Aufbereitung und Abfrage der Daten im BW verantwortlich ist. Die wesentlichen Objekte zur Datenablage sind dem Datenfluss folgend zunächst die Persistent Staging Area (PSA), der Operational Data Store (ODS) und die Info-Cubes. Während die PSA temporären Charakter hat, sind die Daten im ODS und in den Info-Cubes für die dauerhafte Ablage gedacht. In den relational strukturierten ODS-Objekten erfolgt die Speicherung von belegnaher Information,

die Info-Cubes bilden die mehrdimensionalen Strukturen als Grundlage für OLAP Analysen ab.

Ausgehend von den Quellsystemen dient die Ebene der Persistent Staging Area der Speicherung der nicht aufbereiteten und nicht modifizierten Daten aus den vorgelagerten Quellsystemen. Eine erste Transformation ermöglicht die Datenweitergabe an die ODS-Objekte, die in der Darstellung der Konsolidierung verschiedener Datenquellen dienen, wobei auch ein direktes Befüllen der Info-Cubes aus der PSA unter Umgehung zwischengeschalteter ODS-Objekte möglich ist. In einer weiteren Stufe erfolgt die betriebswirtschaftliche Konsolidierung mit dem Aufbau von Info-Cubes.

Grundlage jeder Abfrage und Auswertung gespeicherter Daten im Business Information Warehouse sind die Queries, die bereits spezifische Sichten auf die vorhandenen Daten darstellen. Auf der Ebene der Queries erfolgt auch die Festlegung von berechneten bzw. abgeleiteten Kennzahlen, die nicht physisch gespeichert sind sondern erst zur Laufzeit kalkuliert werden. Jede Query bezieht sich stets auf genau einen Info-Provider, z. B. Info-Cubes oder ODS-Objekte.

Integraler Bestandteil des Business Information Warehouse ist das Werkzeug Business Explorer Analyser als Excel-basiertes OLAP-Werkzeug. Der Bereich des Webreportings wird durch die neue Technologie auf Basis des Web Application Servers und dem Web Application Designer adressiert.

## 5 Anwendungsbeispiel

In diesem Abschnitt wird die dem Buch zugrunde liegende Fallstudie dargestellt. Die in den folgenden Kapiteln dargelegten Methoden und Techniken der mehrdimensionalen Datenmodellierung werden anhand dieser Fallstudie entwickelt. Der Bezug zu einem fiktiven Data Warehouse-Projekt, verbunden mit den daraus resultierenden Fragestellungen eines Projektes im Bereich Business Intelligence, ermöglicht die Darstellung von Modellierungstechniken und deren exemplarische Anwendung.

Im folgenden Abschnitt wird die betrachtete Modellunternehmung skizziert. Die Darstellung eines fiktiven Projektes in Form eines Projektzieles schließt sich daran an. Abgerundet wird dies durch eine Aufarbeitung des Fachmodells zu dieser Fallstudie.

### 5.1 Ausgangssituation

Gegenstand des Beispiels ist eine mittelständische Unternehmung der Elektronikbranche, die sich mit der Produktion und dem Vertrieb von elektronischen Geräten und Komponenten beschäftigt. Zum Programm gehören neben allen relevanten elektronischen Bauteilen wie etwa Widerstände, Transistoren, Kondensatoren etc. auch Laborgeräte und Produkte aus dem High Fidelity Bereich.

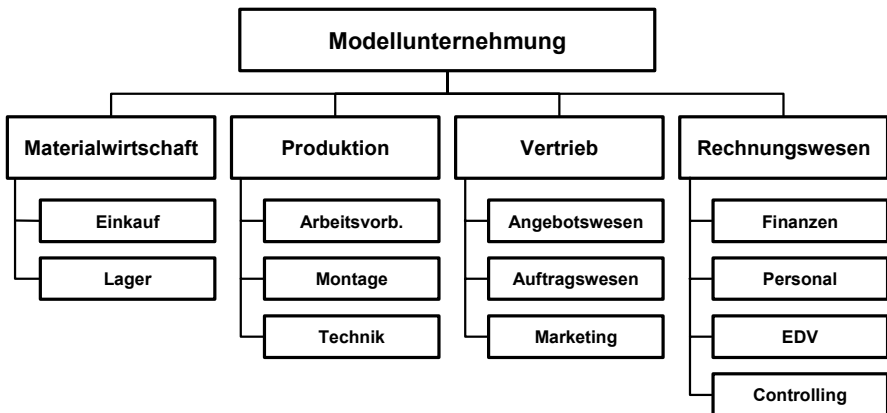
Der Fokus des Unternehmens liegt nicht auf der Produktion einzelner Komponenten, sondern auf der Auswahl geeigneter Lieferanten und Vorprodukte, die zu einem qualitativ hochwertigen zielgruppengerechten Endprodukt zusammengesetzt werden. Zum Einsatz kommen Produkte verschiedener Lieferanten, wobei auf gleich bleibend hohe Qualitätsstandards sehr viel Wert gelegt wird.

Kunden der Unternehmung sind große Warenhausketten und viele kleine Einzelhandelsgeschäfte. Neben diesem klassischen Vertriebsweg erfolgt der Verkauf insbesondere von elektronischen Bauteilen über ihren Katalog. Dieser erscheint zweimal jährlich und wird regelmäßig verschickt. Ein weiterer Vertriebskanal erlangt zunehmende Bedeutung für das Unterneh-



men: Über eine E-Commerce-Lösung bietet das Unternehmen seine Produkte im Internet an.

Die anfallenden Aufgabenbereiche werden komplett intern abgewickelt. Die Materialwirtschaft erledigt die termingerechte Beschaffung benötigter Vormaterialien für die Montage sowie aller Artikel der Ersatzteil- und Zubehörproduktgruppen und deren Lagerung und Verwaltung. In der Produktion werden aus den Einsatzstoffen die Endprodukte gefertigt. Alle Aufgaben der Produktentwicklung und Qualitätssicherung sind ebenfalls in diesem Bereich angesiedelt. Der Vertrieb gestaltet den Katalog, verwaltet den Online-Shop im Internet und betreut die Partner. Die Aufbauorganisation der Beispielunternehmung geht aus Abbildung 34 hervor.



**Abb. 34.** Organigramm der Modellunternehmung

## 5.2 Projektdefinition und Projektziel

In der Unternehmung sind die operativen Geschäftsprozesse analysiert und optimiert worden. Die Prozessunterstützung durch die EDV kann als gut angesehen werden. Neben diesen operativen Bereich treten aber zunehmend Anforderungen auf dispositiver und strategischer Ebene. Eine EDV-seitige Unterstützung von Planungs- und Entscheidungsprozessen ist seitens der Geschäftsführung ausdrücklich erwünscht. Ein hierzu gegründetes Projektteam erhielt dazu den folgenden Projektauftrag:

*Für die Unterstützung der dispositiven und strategischen Prozesse ist eine EDV-Systemlösung zu konzipieren und zu realisieren. Zu unterstützen sind im Einzelnen die Prozesse der Planung und Budgetierung im betriebswirtschaftlichen Controlling sowie die Analyseprozesse im Marke-*

*ting und Vertrieb. Des Weiteren soll die Geschäftsführung durch ein adäquates Informationssystem unterstützt werden. Im Rahmen des Projektes ist eine Anforderungsdefinition zu erarbeiten und mit der Geschäftsführung abzustimmen.*

*Projektziel ist die Bereitstellung regelmäßig aktualisierter Datenbestände in einem zentralen Data Warehouse auf Basis des SAP Business Information Warehouse und die Implementierung der darauf zugreifenden Systemlösungen zur Unterstützung der analytischen Prozesse.*

Die Projektarbeit in der Fallstudie mündet in einem weiteren Schritt in einer Beschreibung der Anforderungen an das zu erstellende System. Dieses ist Gegenstand des folgenden Abschnittes.

## 5.3 Fachkonzept der Fallstudie

Viele Data Warehouse-Projekte beginnen mit der Erarbeitung einer Datenquellen- und Informationsbedarfsanalyse. In der Datenquellenanalyse werden die vorhandenen Datenquellen formal beschrieben. Dies sind neben den DV-Systemen auch nicht DV-mäßig gespeicherte Informationen, wie beispielsweise Papierdokumente. Ziel der Informationsbedarfsanalyse ist eine möglichst exakte Beschreibung der zu erwartenden Fragestellungen, die an das zu entwickelnde System gerichtet werden.

Die in das Data Warehouse implizit oder explizit hereingebrachten Daten sind die Grundlage für analytische Abfragen basierend auf diesem Datenbestand. Der Aufbau einer solchen integrierten Datenbasis ist somit die entscheidende Grundlage für eine erfolgreiche Nutzung eines Data Warehouse-Systems.

Für die betrachtete Fallstudie ist es ausreichend, ausschließlich die eingesetzten operativen DV-Systeme als Datenquelle zu betrachten. Aspekte externer Daten bleiben dabei ebenso wie nicht in elektronischer Form vorliegende Informationen außen vor.

Zunächst sind die Geschäftsprozesse Gegenstand der Betrachtung in Abschnitt 5.3.1. Daran anknüpfend werden die Datenquellen in Form von ER-Modellen im darauf folgenden Abschnitt 5.3.2 dieses Kapitels beschrieben. Anschließend erfolgt eine Beschreibung der von dem System erwarteten Berichte und Abfrage-Möglichkeiten in Abschnitt 5.3.3, wobei auch die zu betrachtenden Kennzahlen entwickelt und informal dargestellt werden.

### 5.3.1 Beschreibung der Geschäftsprozesse

Die Hauptprozesse sind analog der Hauptabteilungsstruktur, wie im Organigramm dargestellt, differenzierbar. Demzufolge sind die Hauptprozesse *Vertrieb*, *Materialwirtschaft*, *Produktion* und *Rechnungswesen* aufzuführen.

#### **Vertriebsprozess**

Zum Vertrieb gehören die Bearbeitung von Anfragen, Aufträgen und alle Aufgaben des Marketings. Darüber hinaus erfolgt auch die Kundenbetreuung im Vertrieb. Zu Kundenanfragen bzgl. Preis und Lieferbarkeit bestimmter Mengen von einzelnen Bauteilen oder anderen Produkten werden im Vertrieb Angebote erstellt. Eingehende Bestellungen werden durch eine Auftragsbestätigung beantwortet und die Lieferung über die Materialwirtschaft veranlasst. Grundlage der Kalkulation von Marketing-Aktivitäten sind die im Vertrieb auszuwertenden Monats- und Quartals-Statistiken mit den Auftragsbeständen und den kundenbezogenen Erlös- und Umsatz-Übersichten.

#### **Prozess in der Materialwirtschaft**

Die aus dem Bereich Vertrieb hereingereichten Auslieferungsanweisungen werden in der Materialwirtschaft gegen den Lagerbestand geprüft. Ist die Lagerkapazität zur Befriedigung des Auftrages ausreichend, werden die Versandpapiere und Lieferscheine erstellt, die Artikel werden ausgelagert und zum Versand gegeben und schließlich erfolgt die Mitteilung an das Rechnungswesen zur Rechnungserstellung. Bei einer nicht ausreichenden Lagerkapazität und auch bei Unterschreiten einer bestimmten Lagermindestmenge erfolgt ein Produktionsauftrag an die Produktionsabteilung zur Herstellung der benötigten Artikel bzw. wird die Beschaffung der fehlenden Artikel initiiert. In der Materialwirtschaft werden die Statistiken hinsichtlich der Lagerumschlagshäufigkeit, Lagerauslastung etc. geführt.

#### **Produktionsprozess**

In der Produktionsabteilung ist der gesamte Prozess der Entwicklung und Durchführung der Geräteherstellung sowie die Entwicklung von neuen Produkten angesiedelt. Einzelne Artikel werden auf Anforderung der Materialwirtschaft hin in gewissen Mindestlosgrößen produziert und sodann einer Qualitätsprüfung unterzogen. Die Beschaffung der benötigten Komponenten und Bauteile für die Produktion erfolgt über die Materialwirtschaft. Produzierte und geprüfte Artikel werden der Materialwirtschaft zur

Einlagerung übergeben. Im Produktionsbereich sind insbesondere perioden- und artikelbezogene Qualitätsstatistiken gefordert.

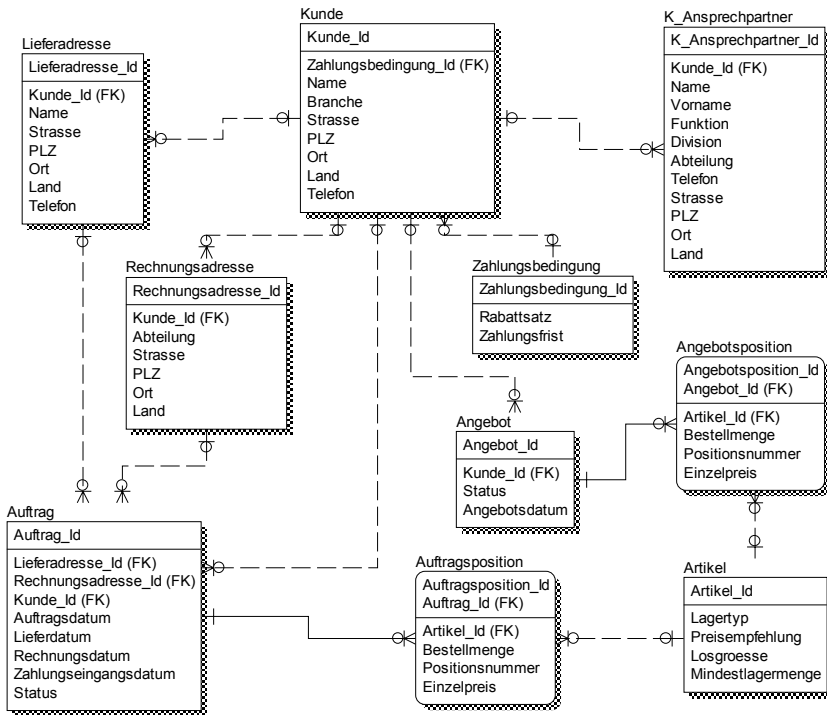
### ***Prozess des Rechnungswesens***

Im Rechnungswesen erfolgt die Rechnungserstellung für gelieferte Artikel. Des Weiteren erfolgt im Rechnungswesen die Kalkulation von Verkaufspreisen. Darüber hinaus ist die Abteilung Rechnungswesen auch für alle Controlling-Aktivitäten zuständig. Hierzu werden detaillierte Statistiken benötigt.

### **5.3.2 Datenquellenbeschreibung**

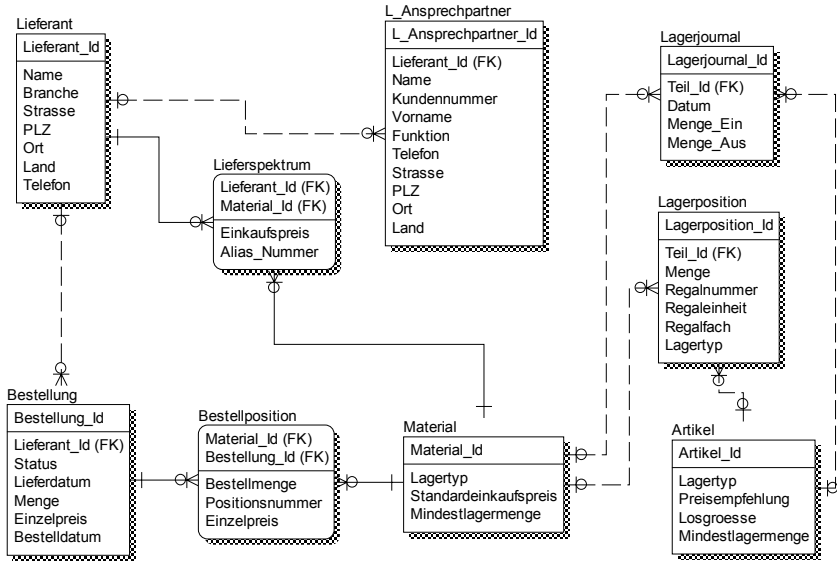
Im Allgemeinen kann weder davon ausgegangen werden, dass alle in operativen Systemen vorhandenen Daten in einem ERM modelliert vorliegen, noch, dass es eine adäquate Ausgangslage zur Modellierung eines Data Warehouse darstellt, wenn die operativen Systeme mit einem ERM dokumentiert vorliegen. In dieser Fallstudie soll lediglich zur Vereinfachung ein ERM als Ausgangsbasis dienen. Dabei sind die dargestellten Modelle der operativen Systemlösungen etwas ausgedünnt in dem Sinne, dass sie von realen Modellen etwas stärker abstrahieren wodurch die Komplexität reduziert ist.

Der zuerst dargestellte Prozess des Vertriebsbereiches spiegelt sich in einem Datenmodell des unterstützenden Systems wider. Ein exemplarisches ERM für diesen Bereich ist in Abbildung 35 dargestellt. Wesentliche Entitäten bilden dabei die Kunden und die Artikel, die über Angebote und Aufträge in Beziehung zueinander stehen.



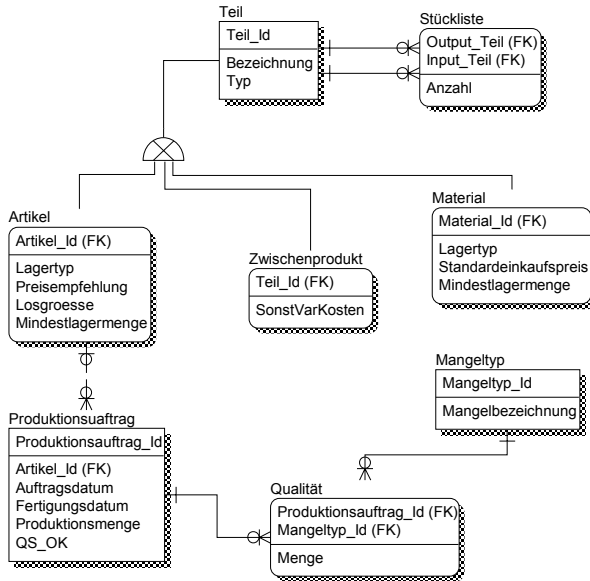
**Abb. 35.** ERM des Vertriebsbereiches

In der Materialwirtschaft erfolgt die Bestellung von notwendigen Materialien für die Produktion. Dieser Zusammenhang spiegelt sich in dem in Abbildung 36 visualisierten Datenmodell in der Beziehung zwischen Lieferanten und Materialien über die Bestellungen wider. Die Betrachtung der Lagerbestände findet sich in der Entität des Lagerjournals, die Grundlage für die Überwachung der artikelbezogenen Lagermindestmenge ist.



**Abb. 36.** ERM der Materialwirtschaft

Im Produktionsbereich findet sich das in Abbildung 37 dargestellte Modell als Grundlage für das zugrunde liegende Datenbanksystem. Wesentlich ist in diesem Modell die Spezialisierung der Teile in Artikel, Zwischenprodukte und Materialien sowie die Beziehung der Artikel zu den Produktionsaufträgen.



**Abb. 37.** ERM des Produktionsbereiches

Diese drei beschriebenen Bereiche bilden den Kern des operativen Basissystems, auf dem die Fallstudie aufsetzt. Sie sind nicht isoliert in getrennten Systemen zu betrachten, sondern als integrierte Komponenten eines einzigen Systems, dessen Gesamtmodell sich aus den dargestellten Teilsichten ableiten lässt.

### 5.3.3 Skizzierung des Informationsbedarfs

Im Folgenden werden die (exemplarischen) Ergebnisse einer Informationsbedarfsanalyse in den Bereichen Marketing und Controlling verkürzt dargestellt. Diese sind die Ausgangsbasis für die mehrdimensionalen Datenmodelle, die im weiteren Verlauf des Buches entwickelt und dargestellt werden.

#### **Controllingbereich**

Als Ausgangsbasis für die Anforderungen im Bereich Controlling wird zunächst nur eine Betrachtung bezogen auf Kostenarten und Kostenstellen gewünscht, auf der basierend später aber eine Ergebnisrechnung anknüpfen soll.

Neben den Ist-Daten sollen auch Plandaten angezeigt werden können und damit verbunden eine Abweichungsanalyse möglich sein. Die Aus-

wertungen werden für Monate, Quartale und Jahre benötigt. Für erhöhten Analysekomfort und zielgerichtete Auswertungen ist der Kostenstellenverantwortliche mit zu berücksichtigen.

### **Marketingbereich**

Die wesentliche Fragestellung im Marketingbereich ist, wann welche Produkte über welchen Vertriebskanal in welcher Größenordnung verkauft wurden sowie daraus ableitbare Prognosen und Abweichungsanalysen. Dabei stehen die Absatz- und Umsatzinformationen aufgegliedert nach den Vertriebskanälen auf Produktebene im Vordergrund. Aufgrund der verschiedenen Vertriebskanäle und der daraus resultierenden unterschiedlichen Informationen sind Auswertungen auf Ebene des Partner-Marketings wie auch auf Ebene des Kunden-Marketings zu berücksichtigen. Eine aggregierte Sicht verbindet die beiden Detailsichten zu einem Gesamt-Marketing, das lediglich die gemeinsamen Informationen enthält.

Für die Vertriebskanäle, in denen der Kunde namentlich bekannt ist, d.h. bei Verkäufen über den E-Shop und über den Katalog, sind kundenbezogene detaillierte Informationen zu berücksichtigen. Für den dritten Vertriebsweg sind detaillierte Informationen für partnerbezogene Auswertungen notwendig.

Bei Produktauswertungen sind neben der Grundinformation über das eigentliche Produkt auch Summierungen anhand von Warenhauptgruppen, Warengruppen und Warenuntergruppen zu berücksichtigen. Analog dazu ist eine Auswertung über die verschiedenen Hersteller der Produkte gewünscht.

Neben der eigentlichen Produktnummer sind für detaillierte und zuverlässige Auswertungen Zusatzinformationen wie Produktbezeichnung, Verpackungsart, -größe, Gewicht, Palettenanzahl sowie die Differenzierung von Eigenproduktion und Fremdbezug erforderlich.

Auswertungen nach zeitlichen Aspekten sollen neben den Standardsektionen nach Jahren, Quartalen und Monaten auch Informationen über die Geschäftsjahre anbieten können, da diese im Unternehmen von den Kalenderjahren abweichen. Es ist weiterhin relevant, ob ein Tag Werktag oder der letzte Tag im Monat ist. Ebenso sind Ferienzeiten, besondere Ereignisse und saisonale Aspekte mit zu betrachten.

Im Hinblick auf die speziellen Anforderungen bei der Auswertung des Partnergeschäftes liegt das Hauptaugenmerk auf der Auswertung nach Partnern und nach der Effizienz der einzelnen Werbemaßnahmen, um die Auswirkungen im Partnergeschäft nachvollziehen zu können.

Die Auswertung der Partner bzw. Filialen erfolgt einerseits nach der organisatorischen Zuordnung zum Konzern und zum anderen über regionale



Zuordnungen. Dabei sollen weiterhin zur Auswahl nicht nur die reine Partner- bzw. Filialnummer zur Verfügung stehen, sondern auch Informationen wie Name, Straße, PLZ, Eröffnungsdatum und Größe.

Bei den Werbemaßnahmen muss erkennbar sein, ob es sich um eine Preisreduktion und bzw. oder um eine Anzeige handelt, wobei für den Fall einer Anzeige zusätzlich das Medium selektierbar sein soll. Zusatzinformationen zu den einzelnen Werbemaßnahmen wie Kosten, Beginn und Ende sind für Auswertungen ebenfalls zu berücksichtigen.

Beim Direktkundengeschäft über den Katalog oder die On-Line-Aktivitäten wird das Hauptaugenmerk der Auswertungen auf den angesammelten Informationen über die Kunden liegen.

Aus marketingstrategischen Gesichtspunkten ist es gewünscht, die Kunden in einer regionalen Gruppierung auswerten zu können, wobei die Gruppierung den üblichen regionalen Anforderungen entsprechen soll. Weiterhin ist zu jedem Kunden eine Gruppierung nach Familienstand, Geschlecht und Altersgruppe zu berücksichtigen, wobei die relevanten Altersgruppen <30 Jahre, 30-40 Jahre und >40 Jahre sind. Zusätzlich zu den Gruppierungsinformationen sind zu jedem Kunden neben der Kundennummer Detailinfos wie Name, Vorname, etc. zur Verfügung zu stellen.

### **Vertriebsbereich**

Die Absatzplanung für alle Produkte über alle Vertriebswege erfolgt im Vertriebsbereich und wird quartalsweise aktualisiert. Die einzelnen Vertriebswege werden sowohl bezogen auf die Kunden als auch in der Granularität der Produktdimension auf unterschiedlichen Ebenen geplant. Die Absatzplanung erfolgt für die Vertriebswege E-Shop und Katalog auf Ebene der Warengruppen. Für die Einzelhandelskunden erfolgt eine detaillierte Planung auf Produktebene, da langfristige Rahmenverträge als Planungsgrundlage vorliegen.

## 6 Semantische mehrdimensionale Modellierung

Die Analyse auf fachkonzeptueller Ebene ist ein wesentlicher Schritt beim Aufbau einer Data Warehouse-Lösung. Die für OLAP-Analysen benötigten mehrdimensional strukturierten Datenbestände sind an den Anforderungen orientiert zu gestalten. Daher hat die semantische Modellierung mehrdimensionaler Datenstrukturen für die frühen konzeptgetriebenen Projektphasen einen hohen Stellenwert. Die hierzu geeigneten Methoden und Beschreibungsmittel der Modellierung sind Gegenstand dieses Kapitels. Zunächst erfolgt eine kurze Beschreibung einer Vielzahl von Methoden im Überblick in Abschnitt 6.1. Die Darstellung semantischer Modelle basiert in diesem Buch auf der spezifischen Methode ADAPT, die anschließend in dem eigenen Abschnitt 6.2 dargestellt wird. Die Berücksichtigung von temporalen Aspekten in der Modellierung mit ADAPT steht in Abschnitt 6.3 im Vordergrund, bevor die Modellierung mit ADAPT auf das in Kapitel 5 aufgebaute Beispiel angewendet wird..

### 6.1 Methoden mehrdimensionaler semantischer Datenmodellierung

Zur semantischen Modellierung mehrdimensionaler Informationssysteme wurden verschiedene Ansätze in der Literatur in die Diskussion gebracht. Neben vielen Vorschlägen, die eher der logischen Modellebene zuzurechnen sind, gibt es auch eine Reihe von Modellierungsmethoden, die für eine Diskussion auf Fachbereichs- bzw. Anwenderebene geeignet erscheinen.

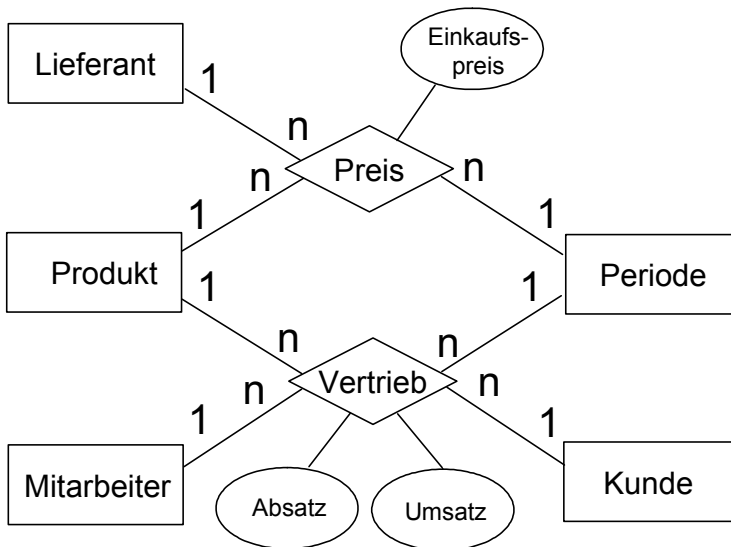
Neben den Ansätzen basierend auf dem Entity Relationship-Modell erfolgt in den nächsten Abschnitten die kurze Skizzierung einiger spezifischer Vorschläge.

#### 6.1.1 Entity Relationship-Modelle

Eines der bekanntesten semantischen Datenmodelle ist sicherlich das Entity Relationship-Modell (ERM) von Chen, in dem die Objekte der Modellwelt als *Entities* und die Beziehungen als *Relationship* bezeichnet werden.

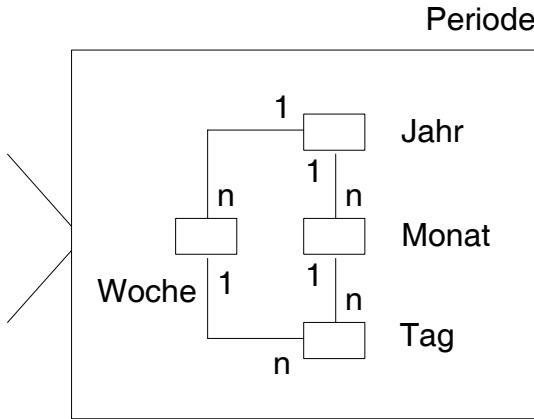
Beide können näher spezifizierende Attribute aufweisen. Üblich ist eine grafische Darstellung mit Rechtecken für die Entities und deren Verbindung über Rauten für die Beziehungen. Die als Oval dargestellten Attribute sind über eine Kante mit den Entities bzw. Relationships verbunden. Weiterhin erfolgt die Notation der Kardinalitäten an den Beziehungskanten. Darüber hinaus gibt es verschiedene Varianten des ER-Ansatzes mit unterschiedlicher Notation von Beziehungen, Attributen und Kardinalitäten.

Das ER-Modell ist von seinem Ausgangspunkt her ein methodischer Ansatz für die Modellierung operativer Datenstrukturen. Gleichwohl können mit dieser Menge an Beschreibungsmitteln auch mehrdimensionale Strukturen abgebildet werden, wie Abbildung 38 belegt. Die einzelnen Dimensionen finden dabei als Entities Eingang in die Darstellung deren Verbindung über ein Beziehungsobjekt mit den angehängten Attributen zur Darstellung von Kennzahlen erfolgt. Die Darstellung ist natürlich angelehnt an den Entwurf eines Star Schemas, dem üblichen Modell zur Abbildung mehrdimensionaler Strukturen in relationalen Datenbanken, auf das im folgenden Kapitel 7 detailliert eingegangen wird. Die ER-Methode ist zwar auf der semantischen Ebene angesiedelt, eine starke Nähe zum relationalen Datenmodell auf der logischen Modellebene kann jedoch nicht verhehlt werden.



**Abb. 38.** ER-Modellierung mehrdimensionaler Datenstrukturen

In einem ersten Schritt zur verbesserten Darstellung der mehrdimensionalen Spezifika ist es möglich, Dimensionshierarchien explizit als eine Kette von 1:n-Beziehungen abzubilden. Diese auch als *Aggregation* bezeichnete Erweiterung ist in Abbildung 39 am Beispiel einer Zeitdimension verdeutlicht.

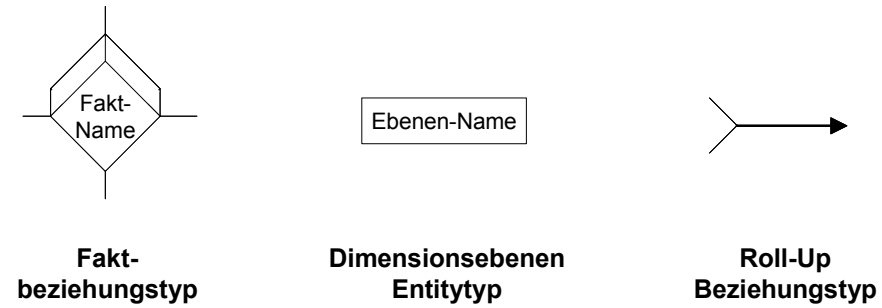


**Abb. 39.** Erweiterte ER-Darstellung von Dimensionshierarchien

Eine weitere Erweiterung des ER-Ansatzes um Konstrukte der Generalisierung und Spezialisierung, die zu einem neuen Entity-Typen der *is-a*-Beziehung zusammengefasst sind, ermöglichen eine noch detailliertere Darstellung, aber auch dieser Ansatz entstand vor dem Hintergrund der Gestaltung operativer Informationssysteme.

### 6.1.2 Mehrdimensionales ER-Modell (ME/R)

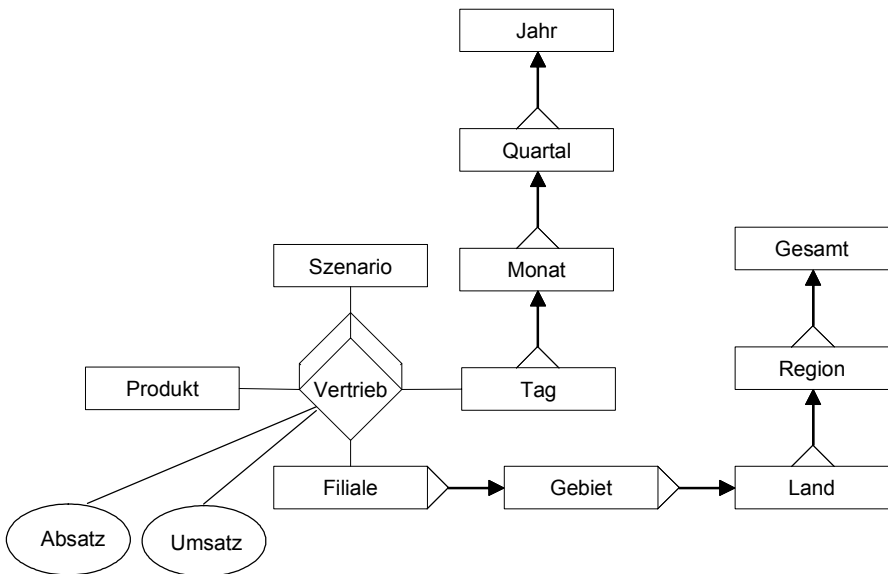
Das Mehrdimensionale ER-Modell (ME/R) ist eine explizite Erweiterung zur Modellierung mehrdimensionaler Datenstrukturen und ist 1998 von Sapia, Blaschka, Höfling und Dinter in die Diskussion gebracht worden. Der Ansatz sieht spezifische neue Darstellungsobjekte vor, um eine adäquate grafische Notation mehrdimensionaler Datenstrukturen zu ermöglichen. Diese drei neuen Elemente sind in Abbildung 40 in der Übersicht dargestellt.



**Abb. 40.** Darstellungsobjekte im ME/R-Modell

Zur Modellierung von Dimensionsstrukturen stehen der Entity-Typ *Dimensionsebene* und der Beziehungstyp *Roll-Up* zur Verfügung. Hiermit ist die Vielzahl ebenenbestimmter Dimensionsstrukturen durch einen Graph von Ebenen-Entities verbunden durch Roll-Up-Beziehungen darstellbar.

Der Fakt-Beziehungstyp, an den die betriebswirtschaftlichen Kennzahlen als Attribute angehängt sind, verknüpft die einzelnen Dimensionen miteinander. Dieses Zusammenspiel der einzelnen Komponenten wird am Beispiel in Abbildung 41 deutlich.



**Abb. 41.** Beispiel eines ME/R-Modells

Wie auch beim ER-Modell sieht diese Methode keine Berechnungsvorschriften für abgeleitete Werte entlang der Dimensionshierarchien oder für

Kennzahlensysteme vor. Auch die Modellierung von Heterarchien bzw. anteiligen Verrechnungen ist derzeit im ME/R-Modell nicht möglich.

### 6.1.3 Dimensional Fact Modeling

Das *Dimensional Fact Modeling* wurde 1998 von Golfarelli, Maio und Rizzi vorgestellt. Neben einer formalen und grafischen Beschreibung konzeptioneller mehrdimensionaler Datenmodelle umfasst der Vorschlag auch ein Vorgehensmodell zur Ableitung eines *dimensional fact model* aus bestehenden ER-Schemata.

Die Beschreibungselemente in diesem Ansatz sind Fakten, Dimensionen und Hierarchien. Dabei stellen die Fakten die relevanten betriebswirtschaftlichen Betrachtungsgegenstände dar, denen konkrete Kennzahlen in Form von Attributen zugeordnet sind. Die Aufgliederungsrichtungen, nach denen die Fakten betrachtet werden können, sind die Dimensionen die durch Hierarchien weiter strukturiert werden. Ein Beispiel zur grafischen Darstellung im *Dimensional Fact Modeling* findet sich in Abbildung 42.

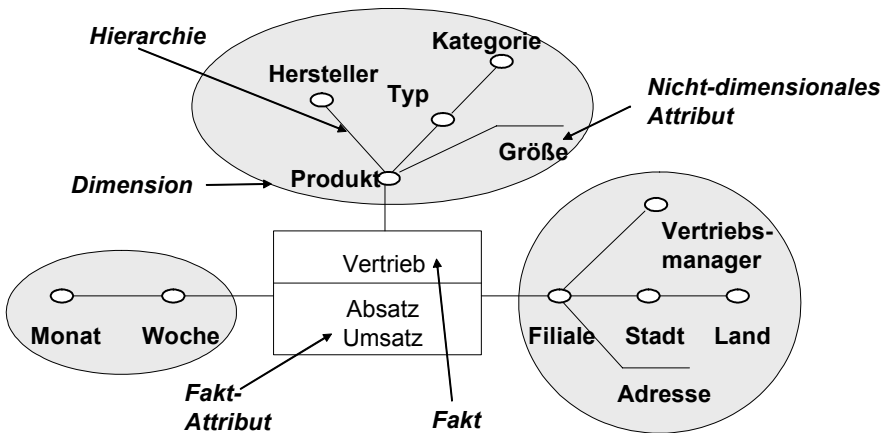
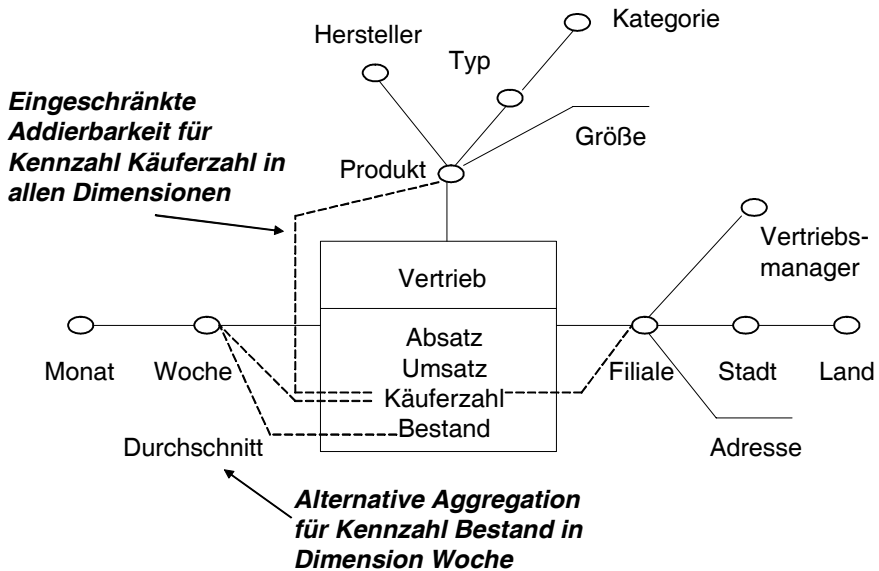


Abb. 42. Beispiel im Dimensional Fact Modeling

Der betriebswirtschaftliche Gegenstand der Betrachtung liegt in diesem Beispiel im Vertrieb mit den Kennzahlen Absatz und Umsatz und ist als zentrales Fakt dargestellt. Die einzelnen Dimensionen sind darum herum angeordnet und durch die Schattierung jeweils umschlossen. Innerhalb einer Dimension erfolgt die Darstellung von Hierarchien durch aneinander gehängte Kreisobjekte, die jeweils eine Konsolidierungsstufe repräsentieren und entsprechend beschriftet sind. Zusätzliche (nicht-dimensionale)

Attribute finden ihren Eingang in die Darstellung direkt in Form einer beschrifteten Linie ohne Kreisobjekt.

Für die Aggregation entlang der verschiedenen Hierarchiestufen ist die Addition als Standard vorgesehen. Für Fakt-Attribute, die sich nicht sinnvoll entlang aller Dimensionen addieren lassen, also für den semi-additiven und nicht-additiven Fall, sieht das Modell eine gesonderte Notation der Verdichtungsoperation vor. In Abbildung 43 ist exemplarisch dargestellt, wie die abweichende Aggregation eines Fakt-Attributes in Bezug auf eine oder mehrere Dimensionen durch gestrichelte Linien und Angabe der alternativen Operation zur Verdichtung im Modell visualisiert wird.



**Abb. 43.** Aggregation im Dimensional Fact Modeling

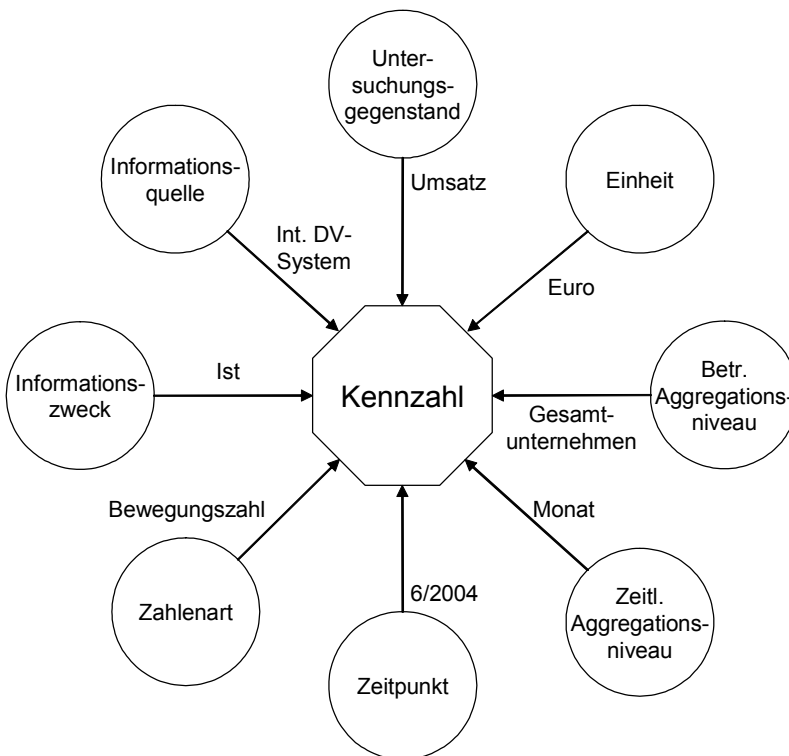
Auf die Beschreibung der weiteren Modellbestandteile des Dimensional Fact Modeling zur Darstellung von Abfragen sowie die vorgeschlagene Vorgehensweise zur Ableitung eines mehrdimensionalen Modells aus einem ER-Schema wird an dieser Stelle verzichtet.

Das Dimensional Fact Modeling ist mathematisch formal definiert. Problematisch ist, dass in den Hierarchien nur 1:n-Beziehungen vorgesehen sind. Für nicht-additive Kennzahlen besteht zumindest die Möglichkeit der Kennzeichnung, wenn auch konkrete Berechnungsvorschriften nicht Bestandteil des Modells sind. Somit ist auch die Darstellung von Kennzahlensystemen nicht möglich.

### 6.1.4 Kennzahlendatenmodell

Das 1992 von Groffmann vorgestellte Kennzahlendatenmodell ist unabhängig von der Data Warehouse Diskussion entstanden. Zentraler Ausgangspunkt ist die Positionierung von Kennzahlen und Kennzahlensystemen als ein wichtiges Informations- und Steuerungsinstrument für das Management. Aufgrund der heterogenen Informationsbedarfe der einzelnen Informationsempfänger sind unterschiedliche Sichten auf die Kennzahlen notwendig, damit diese sinnvoll eingesetzt werden können.

Aus der Analyse von Kennzahlen und Kennzahlensystemen einerseits und dem Informationsbedarf der Unternehmensführung andererseits leiten sich die Differenzierungsmerkmale von Kennzahlen ab. Diese auch als „Dimensionen“ bezeichneten Merkmale sind in Abbildung 44 exemplarisch für eine Kennzahl dargestellt.



**Abb. 44.** Acht Dimensionen zur Differenzierung von Kennzahlen

Der zu quantifizierende Sachverhalt einer Kennzahl wird als *Untersuchungsgegenstand* bezeichnet. Damit dieser messbar ist, bedarf es der An-



gabe einer *Einheit*. Mit dem *betrieblichen Aggregationsniveau* ist die Betrachtungstiefe bzw. Verdichtungsstufe umschrieben, die bzgl. des betrachteten Sachverhaltes und in zeitlicher Hinsicht (*zeitliches Aggregationsniveau*) zu spezifizieren ist. Weiterhin ist kenntlich zu machen, zu welchem *Zeitpunkt* die Quantifizierung des Sachverhaltes erfolgt. Die *Zahlenart* differenziert u. a. zwischen Bestands- und Bewegungsdaten. Zur Unterscheidung der Datenart bzw. des betrachteten Szenarios dient das Merkmal des *Informationszweckes*. Das achte Merkmal hat den Fokus auf der *Informationsquelle* auf deren Basis die Quantifizierung erfolgt.

In einem weiteren Schritt werden die dergestalt modellierten Kennzahlen zu Hierarchien bzw. Kennzahlensystemen zusammengestellt.

Aufgrund der starken isolierten Fokussierung auf die Kennzahlensysteme ist dieser Modellierungsansatz nicht geeignet, um eine umfassende Unterstützung bei dem Aufbau mehrdimensionaler Datenmodelle zu bieten. Es ist jedoch in der Phase der Kennzahlendefinition ein durchaus hilfreiches Werkzeug gerade beim Aufbau von Führungsinformationssystemen, da hier eine sehr differenzierte Sichtweise auf die betriebswirtschaftlichen Variablen notwendig ist.

### 6.1.5 Multidimensional Data Model ( $\mathcal{MD}$ )

Ein weiterer konzeptioneller Vorschlag zur Modellierung von mehrdimensionalen Datenstrukturen stammt von Cabibbo und Torlone aus dem Jahr 1997 und wird als Multidimensional Data Model ( $\mathcal{MD}$ ) bezeichnet. Dieser umfasst neben einer grafischen Repräsentation auch ein Vorgehensmodell, mit dem aus vorhandenen ER-Modellen ein entsprechendes mehrdimensionales Schema abgeleitet werden kann. Diese Vorgehensweise ist somit stark von den verfügbaren Quellstrukturen getrieben und geht nicht direkt von den Informationsbedarfen aus.

In der als *f-graph* bezeichneten Repräsentation wie in Abbildung 45 für unser Vertriebsbeispiel dargestellt befindet sich ein zentraler Knoten, der als *f-node* die Verbindung zwischen den Variablen und den verschiedenen Dimensionen des Würfels herstellt.

### Beschreibung eines Dimensionslevels

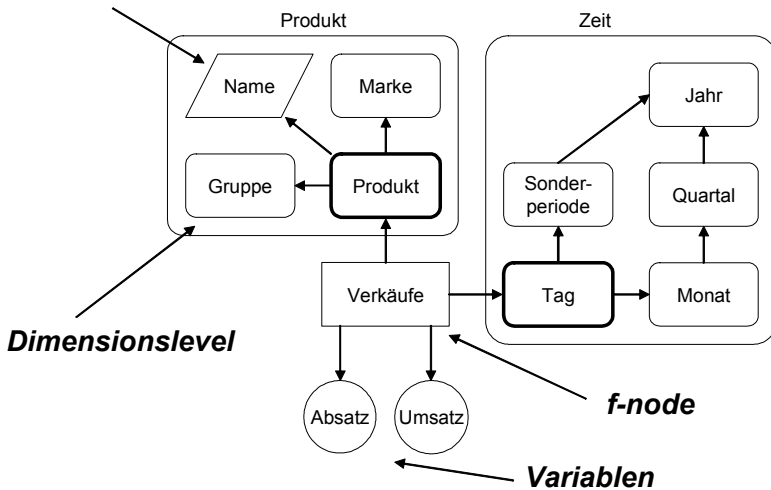


Abb. 45. Beispiel im Multidimensional Data Model ( $\mathcal{MD}$ )

Innerhalb der umrahmten Dimensionen finden sich die Objekte zur Abbildung der Ebenen, aus denen sich die Dimensionshierarchien zusammensetzen, sowie beschreibender Attribute auf Hierarchiestufenebene.

Bei dem Ansatz des Multidimensional Data Model ist nur die Abbildung von 1:n-Beziehungen in Dimensionshierarchien möglich, die Modellierung anteiliger Verrechnungen ist gar nicht vorgesehen. Die Darstellung von Kennzahlensystemen ist insofern nicht möglich, als betriebswirtschaftliche Variablen als Attribute des zentralen Fakts eines Graphen modelliert werden und nicht in einer eigenen Dimension abbildbar sind.

## 6.2 Semantische Modellierung mit ADAPT

Neben Modifikationen von ER-Modellen und objektorientierten Ansätzen treten auch Vorschläge mit einer eigenen Methodologie zur semantischen Modellierung mehrdimensionaler Datenstrukturen auf. Eine solche Methode stellt das in Amerika verbreitete *Application Design for Processing Technologies* (ADAPT) dar, das 1996 von Bulos vorgestellt wurde.

Es handelt sich bei ADAPT um einen Modellierungsansatz, in dessen Fokus explizit mehrdimensionale Datenstrukturen für OLAP-Anwendungen stehen. Zusätzlicher praktischer Nutzen der Methode liegt in der Ver-

fügbare der zugrunde liegenden Symbole für Microsoft Visio, wodurch die Modelle einfach zu generieren sind.<sup>11</sup>




In Deutschland ist ADAPT erst seit einigen Jahren in der Diskussion. Die in der ersten Veröffentlichung zahlreichen Objekttypen standen als wesentlicher Kritikpunkt im Raum und heute beinhaltet die Methode nur noch die wesentlichen Objekttypen und ermöglicht damit leicht lesbare verständliche konzeptionelle Modelle. Diese angepasste Modellierungsmethode wird im Folgenden detailliert beschrieben. Der Schwerpunkt der Modellierung mit ADAPT liegt in der Darstellung von Dimensionsstrukturen und hierarchischen Strukturen, deren Darstellung erfolgt im ersten Abschnitt. Auf die Aspekte der Modellierung von Cubes wird anschließend kurz eingegangen.

6.2.1 Dimensionsmodellierung





Als Modellier-Methode auf konzeptioneller Ebene betrachtet ADAPT eher den betriebswirtschaftlichen Kontext einer Dimension als den strukturellen Aufbau. Daher umfasste ADAPT in einer frühen Version sehr viele verschiedene Dimensionstypen als Objekttypen zur Klassifizierung von Dimensionen. In der aktuellen Version findet nur ein einziger Dimensionstyp Berücksichtigung, da die Objektvielfalt auch einer der wesentlichen Kritikpunkte an der Methode war.

Die Stärke liegt in der Darstellung vielfältiger Dimensionsstrukturen und deren betriebswirtschaftlichem Bezug. Die einzelnen Grundbestandteile zur Modellierung von Dimensionen und darauf definierten hierarchischen Strukturen sind in der folgenden Tabelle 6 zusammengefasst.

Tabelle 6. ADAPT Objekte zur Dimensionsmodellierung





Objekt	Bedeutung
 Dimension	Objekt zur Darstellung einer Dimension (Achse eines Cubes)
 Hierarchy	Objekt zur Darstellung von Hierarchien (Verdichtungsweg von Dimensionselementen)
 Level	Ebene in einer Hierarchie (Konsolidierungsstufe)

<sup>11</sup> Die Schablonendateien sowie unterschiedliche Vorlagen können unter [www.hahneonline.de](http://www.hahneonline.de) sowie [www.cundus.de](http://www.cundus.de) heruntergeladen werden.

 Attribute	Darstellungsobjekt für Attribute (Zusätzliche Information zu Dimensionselementen)
 Member	Objekt für ein Dimensionselement (Eindeutig identifizierender Wert)
 Scope	Darstellung eines Dimensionsausschnittes (Teilmenge einer Dimension)
 Model	Berechnungsvorschrift (Formeln zur Berechnung abgeleiteter Daten)

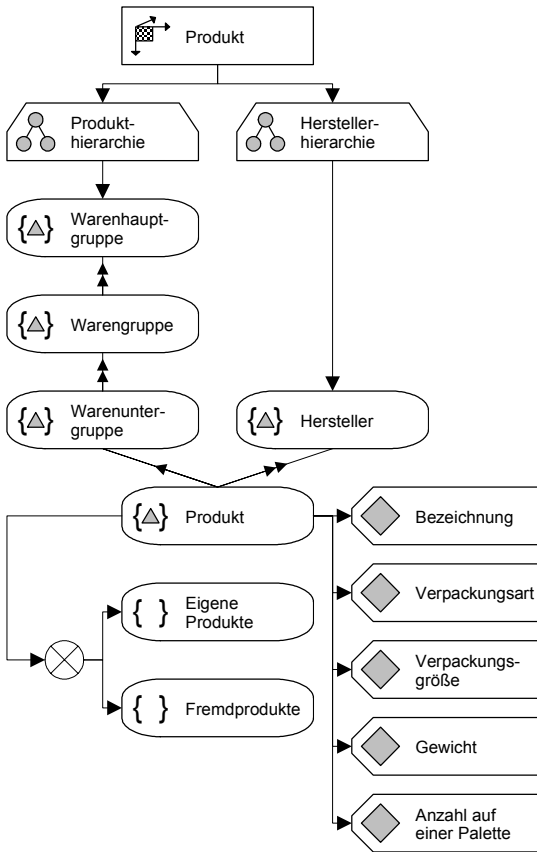
Neben den eigentlichen Objekten ist auch die Form der Verbindung untereinander wichtig. Eine Zusammenstellung der möglichen Verbindungen zur Dimensionsmodellierung findet sich in der folgenden Tabelle 7. Die Möglichkeit zur Darstellung von m:n-Beziehungen bzw. Heterarchien ist im Standard nicht vorgesehen, wurde jedoch vom Autor mit aufgenommen.<sup>12</sup>

**Tabelle 7.** Verbindungsobjekte für die Dimensionsmodellierung

Objekt	Bedeutung
	Lockere Beziehung zwischen Ebenen
	Strenge Beziehung
	Rekursive Beziehung
	m:n-Beziehung

Die Darstellung der wesentlichen Komponenten zur Abbildung von Dimensionen in ADAPT erfolgt anhand eines Beispiels, der exemplarisch aufgerissenen Produktdimension in Abbildung 46.

<sup>12</sup> Dieses Objekt ist in der unter [www.hahneonline.de](http://www.hahneonline.de) erhältlichen Visio-Schablonendatei ebenfalls enthalten.



**Abb. 46.** Exemplarische Produktdimension in ADAPT Notation

Dem mehrdimensionalen Grundverständnis folgend sind eine oder mehrere Hierarchien das Wesensmerkmal einer Dimension. Diese basieren auf Über- und Unterordnungsverhältnissen und bilden die Konsolidierungspfade in Dimensionen. Sie sind damit die Grundlage für Operationen des *drill-down* und *roll-up*. Wie der Darstellung zu entnehmen ist, sind für Produkte zwei Hierarchien definiert, die jeweils unterschiedlich aufgebaut sind.

Bei der Darstellung handelt es sich um eine ebenenbestimmte Dimension, so dass die Modellierung aller Hierarchien über Dimensionsebenen erfolgt. Diese verschiedenen Hierarchien werden oftmals auch als parallele Hierarchien bezeichnet. Die gemeinsame Basisebene wird durch das Objekt für die Produkt-Ebene repräsentiert, die damit die feinste Granularität in dieser Dimension festlegt.

Die Hersteller-Hierarchie zeigt zunächst einen einfachen Fall, da nur eine Verdichtung von Produkten zu Herstellern erfolgt. Dabei ist jedem Produkt genau ein Hersteller zugeordnet. Auch der Fall einer m:n-Beziehung zwischen Produkten und Herstellern ist denkbar. Für diesen Fall bedeutete dies, dass jedem Produkt auch mehrere Hersteller zugeordnet werden könnten und das many-many Verbindungsobjekt zur Darstellung heranzuziehen wäre.





Das verwendete Symbol mit dem Doppelpfeil zur Verbindung der Ebenen verdeutlicht die strenge Beziehung, demzufolge jedem Produkt genau ein übergeordnetes Element zugewiesen wird. Dies ist in der anderen Hierarchie nicht der Fall. Dort werden die Ebenen Warenuntergruppe, Warengruppe und Warenhauptgruppe differenziert, jedoch kann es auch Produkte geben, die keiner Warenuntergruppe zugeordnet sind, sondern direkt an einer Warengruppe hängen. Die Notation dieses Zusammenhangs erfolgt in ADAPT mit dem Verbindungspfeil mit nur einer Spitze.

Ein weiteres Grundkonstrukt zur Darstellung mehrdimensionaler Strukturen auf semantischer Ebene sind Attribute zu Dimensionen oder Dimensionsebenen. In dem dargestellten Beispiel befinden sich die Attribute ausschließlich auf der untersten Ebene aller Produkte und modellieren weitere Eigenschaften von Produkten wie etwa deren Gewicht oder weitere verpackungsrelevante Informationen.

In der Modellierungsphase ist oftmals für diverse betriebswirtschaftliche Objekte zu entscheiden, ob diese als Attribut oder als eigene Dimension zu modellieren sind.

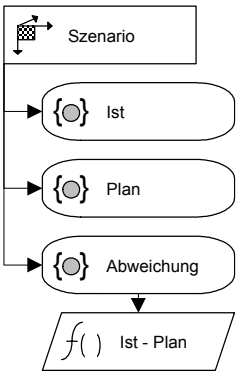
Eine Stärke von ADAPT ist die Modellierung von Dimensionsausschnitten, auch Dimensionssicht genannt, die eine logisch zusammenhängende Teilmenge einer Dimension darstellen. Für die Produkte sind beispielsweise Teilmengen für Eigenfertigung und Fremdbezug denkbar. Diese Dimensionsausschnitte stehen wiederum untereinander in einer Beziehung. So ist etwa ein Produkt entweder selbst produziert oder wird hinzugekauft. Dieser Beziehungstyp heißt umfassendes Exklusiv-Oder. Neben den Beziehungen zwischen Dimensionsausschnitten gibt es in ADAPT auch die Möglichkeit der Modellierung von Beziehungen zwischen Dimensionen. Eine Übersicht über die verschiedenen Beziehungstypen zwischen Dimensionsausschnitten gibt die folgende Tabelle 8.

**Tabelle 8.** Beziehungstypen in ADAPT

Objekt	Bedeutung
	Umfassendes Exklusiv-Oder
	Umfassendes Oder
	Partielles Exklusiv-Oder
	Partielles Oder

In der Bezeichnung der Beziehungen bedeutet *Exklusiv-Oder*, dass die beteiligten Teilmengen disjunkt sind, das nichtexklusive *Oder* hingegen erlaubt Überlappungen. Bildet die Vereinigung der Teilmengen die Gesamtheit, wird dies als *umfassend* bezeichnet. Andernfalls wird dies mit *partiell* umschrieben.

Neben den ebenenbestimmten Dimensionstypen sind auch Dimensionen mit ihren direkten Ausprägungen mit dem Objekt für Dimensionselemente abbildbar. Typischerweise sind dies die Dimensionen für Datenarten und Wertarten, die auch als *elementbestimmte Dimension* bezeichnet werden. In Abbildung 47 ist das Beispiel einer Szenariodimension aufgezeigt. An der Modellierung des berechneten Elementes für die Abweichung ist erkennbar, wie der Objekttyp für Berechnungsvorschriften zum Einsatz gelangt.

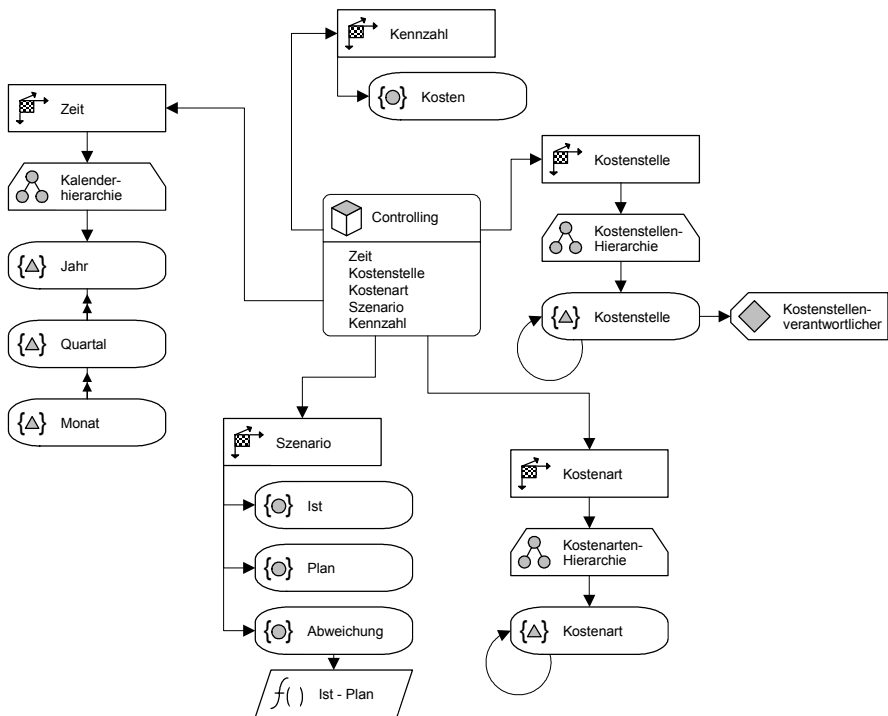


**Abb. 47.** Elementbestimmte Dimension

Die Dimensionen legen die Struktur der mehrdimensionalen Daten in Würfeln bzw. Cubes fest. Diese sind Gegenstand des nachfolgenden Abschnittes.

### 6.2.2 Modellierung von Cubes

Für die Darstellung eines Würfels steht in ADAPT das Cube-Objekt zur Verfügung. In diesem werden die Bezeichnung und die Dimensionierung manuell eingetragen. Über ein Verbindungspfeil-Objekt findet die Verknüpfung mit den beteiligten Dimensionen statt. Dabei reicht in der allgemeinen Darstellung die Modellierung mit dem Cube-Objekt und den angehängten Dimensionsobjekten, um die Komplexität der Darstellung zu reduzieren. Bei nicht so großen Modellen ist aber die vollständige Darstellung sinnvoller, da alle Modellbestandteile auf einen Blick sichtbar sind. Beispielhaft ist in Abbildung 48 ein einfach gehaltener Controlling-Cube modelliert, bei welchem auch die Dimensionen direkt mit dargestellt sind.

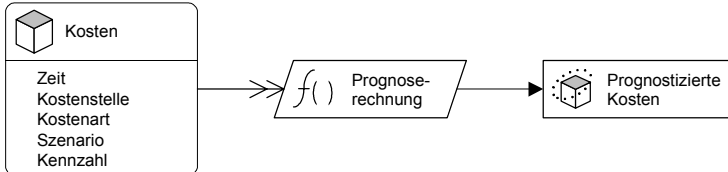


**Abb. 48.** Darstellung eines Cubes in ADAPT

Ähnlich der Modellierung von Dimensionsausschnitten gibt es auch für die Darstellung von Cubes zusätzliche Möglichkeiten in ADAPT, um noch mehr fachbezogenen Kontext abzubilden. Eine Teilmenge eines Cubes ergibt sich durch die Kombination verschiedener Dimensionsausschnitte, die jeweils eine Teilmenge in einer Dimension festlegen.



Neue Sichten kommen durch die Kombination mit dem Objekt für Berechnungsvorschriften hinzu. Geläufig ist hierfür auch die Bezeichnung Ableitungsregel. Mit diesem Konstrukt sind abgeleitete betriebswirtschaftliche Größen modellierbar. Ein recht abstrakt gehaltenes Beispiel in Abbildung 49 zeigt die Notation für die Ableitung der prognostizierten Kosten aus dem Kosten-Cube.



**Abb. 49.** Modellierung einer abgeleiteten Teilsicht eines Cubes

Neben den dargestellten Möglichkeiten der Verwendung des Objektes für Berechnungsvorschriften sind in der ADAPT-Spezifikation noch weitere Verwendungsmöglichkeiten referenziert, auf die aber an dieser Stelle nicht weiter eingegangen wird.

### 6.3 Erweiterung von ADAPT zu T-ADAPT

Durch die Modellierung soll es möglich sein, die in Abschnitt 3.4 aufgeführten Berichtsanforderungen abzubilden. Demnach muss die Unterscheidung verschiedener Versionen von Dimensionsstrukturen zum Ausdruck gebracht werden können. Es sind vier Szenarien zu berücksichtigen, wobei die Analyse auf Basis der jeweils aktuellen Dimensionsstruktur der bisher angesprochenen Hierarchie-Darstellung in ADAPT entsprechen soll.

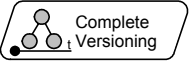
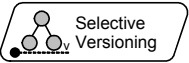
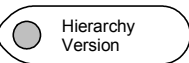
Die verbleibenden Möglichkeiten können danach klassifiziert werden, ob zu bestimmten Stichpunkten, im allgemeinen bestimmte Zeitpunkte, eine Hierarchieversion zusätzlich angelegt wird, oder ob alle Veränderungen kontinuierlich nachverfolgbar sind. Diese beiden Typen sind durch *selektive Versionierung* und *vollständige Versionierung* beschrieben.

In ADAPT können parallele Hierarchien direkt zum Aufbau einzelner Versionen von Dimensionsstrukturen genutzt werden. Dabei erfolgt aber die Modellierung der gesamten hierarchischen Struktur mit Ebenen, Elementen und Attributen mehrfach. Für die vollständige Versionierung steht in ADAPT nur die textliche Ergänzung um Kommentare zur Verfügung.

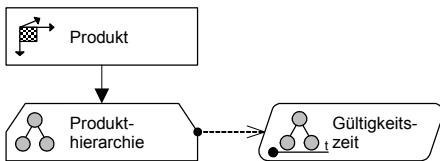
Um die Aspekte der Berücksichtigung struktureller Veränderungen in Dimensionen adäquat in ADAPT abbilden zu können, erfolgt in diesem Abschnitt die Darstellung eigener Objekte zur Abbildung temporaler Kon-

strukture, die in Tabelle 9 zusammengestellt sind. Die Erweiterung von ADAPT um diese Konstrukte wird mit T-ADAPT bezeichnet.

**Tabelle 9.** Objekte zur Historisierung von Hierarchien

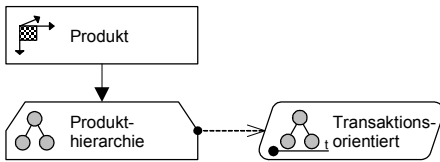
Objekt	Bedeutung
 Complete Versioning	Vollständige Historisierung
 Selective Versioning	Selektive Historisierung mit verschiedenen Versionen
 Hierarchy Version	Einzelne Hierarchie-Version

Mit diesen Objekten kann für jede Hierarchie einzeln eine Zeitabhängigkeit modelliert werden. Die herkömmliche Darstellung korrespondiert dann mit der Vernachlässigung jeglicher struktureller Veränderungen und setzt eine fixe hierarchische Struktur voraus. Die Historisierung kann entweder selektiver oder vollständiger Art sein, eine Mischung innerhalb einer Hierarchie ist ausgeschlossen. Bei der vollständigen Versionierung wird jede strukturelle Änderung berücksichtigt. Die Darstellung der stetigen Versionierung ist in Abbildung 50 für den Fall der Abhängigkeit von der Gültigkeitszeit dargestellt.



**Abb. 50.** Vollständige Historisierung über Gültigkeitszeit

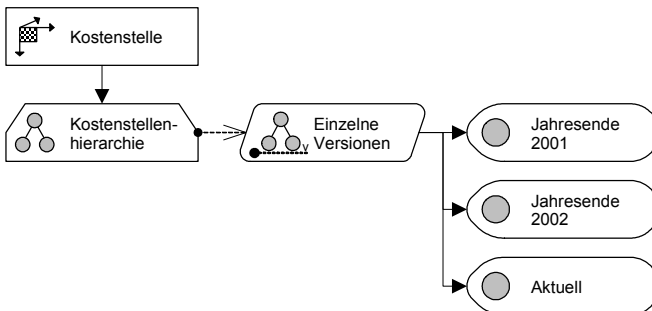
An das Hierarchieobjekt wird die temporale Eigenschaft angehängt und mit dem zentralen Einflussfaktor beschriftet. Neben der Gültigkeitszeit als maßgebendem Einflussfaktor für die Generierung einer neuen strukturellen Version ist es auch denkbar, dass für die Hierarchie die jeweils zum Belegdatum gültige Struktur betrachtet werden soll. Diese Sicht auf die Historisierung stellt einen Spezialfall der vollständigen Versionierung dar und kann als transaktionsorientiert charakterisiert werden. Sie ist in Abbildung 51 exemplarisch dargestellt.



**Abb. 51.** Transaktionsorientierte Sicht auf Hierarchien

Zusätzlich zu der transaktionsorientierten Historisierung ist es denkbar, einzelne herausragende Versionen gesondert hervorzuheben. Dies erfolgt nach dem gleichen Verfahren wie im Folgenden für die nicht-vollständige Versionierung dargestellt.

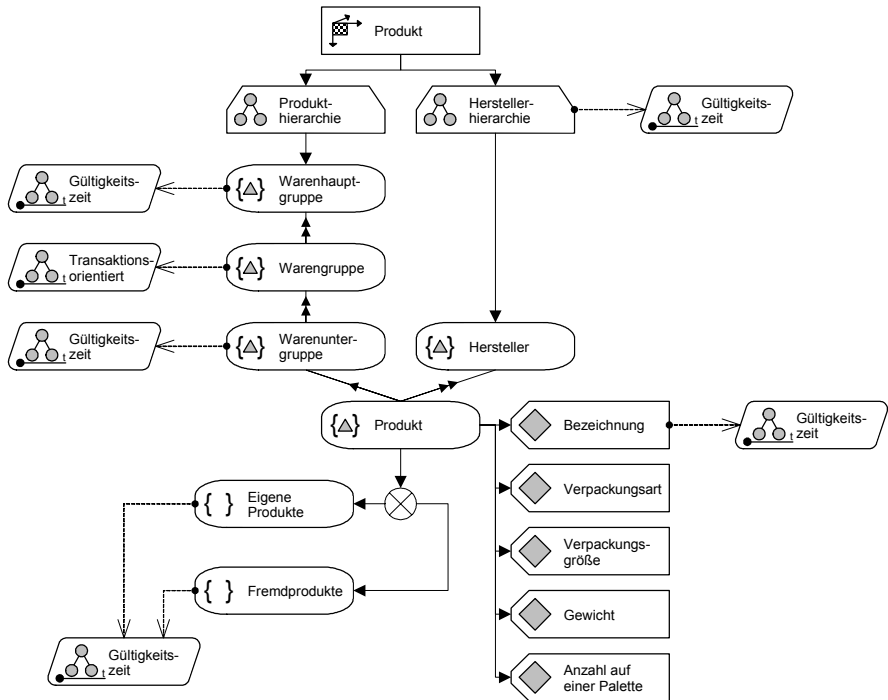
Im Fall der selektiven Historisierung erfolgt die Darstellung anhand des entsprechenden Objektes, das durch die gepunktete Linie im Symbol angedeutet, dass nur einzelne Versionen berücksichtigt werden. Zur Darstellung der Hierarchie-Varianten werden die einzelnen Hierarchie-Versionen als eigene Objekte einzeln aufgelistet und an das Objekt der temporalen Eigenschaft angehängt. Für den Fall einer Kostenstellendimension ist dies exemplarisch in Abbildung 52 dargestellt. Die selektive Historisierung im Beispiel basiert auf der Bildung einzelner Schnappschüsse, denn es werden gewisse Jahresend-Strukturen in eigenen Versionen zusätzlich zur aktuell gültigen Struktur fixiert.



**Abb. 52.** Historienverfolgung mit dedizierten Versionen

Mit den dargestellten Objekttypen der Modellierung in T-ADAPT ist es möglich, die dargestellten Formen der Historisierung von Hierarchien abzubilden. Oftmals tritt auch die Frage nach der Historisierung von Attributen und Dimensionsausschnitten hinzu. Die Darstellungsvarianten für die Zeitabhängigkeit sind direkt von den Hierarchien auf Attribute in ADAPT übertragbar. In Abbildung 53 ist dies exemplarisch für die Produktdimension dargestellt. Im Beispiel wird das Attribut zur Bezeichnung eines Pro-

duktes über die Gültigkeitszeit versioniert. Für die Dimensionsausschnitte, die den Fremdbezug bzw. die Eigenproduktion kennzeichnen, ist ebenfalls eine Abhängigkeit von der Gültigkeitszeit mit eingezeichnet um die Form der Modellierung von Zeitabhängigkeit für diese Objekte mit darzustellen.



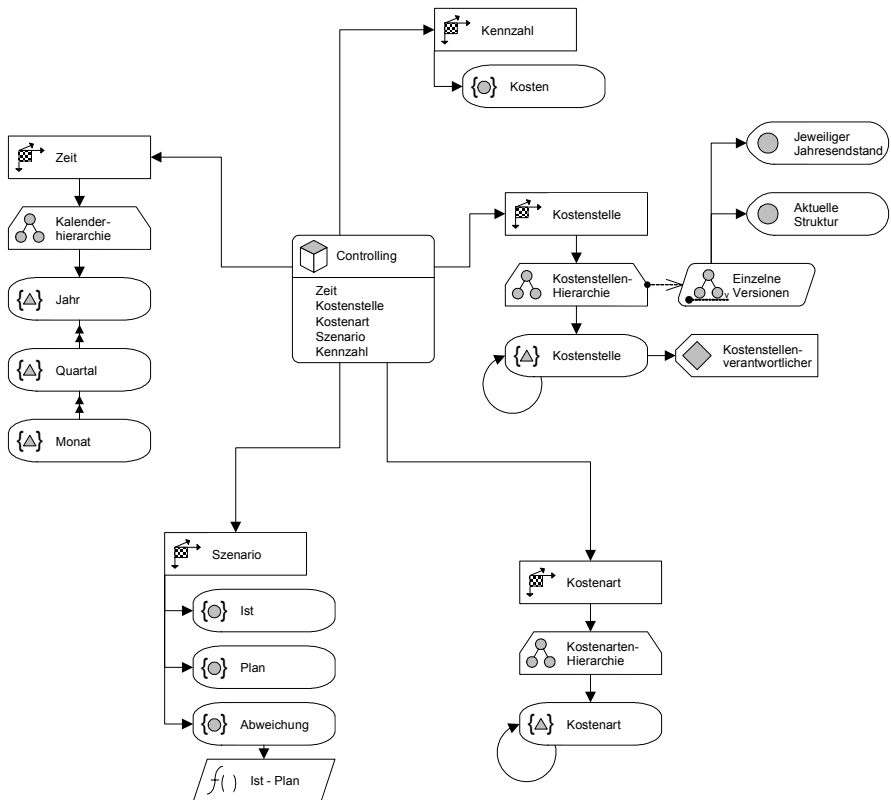
**Abb. 53.** Weitere Formen der Historisierung

In obiger Abbildung ist noch eine weitere Variante der Verwendung der Modellierungskonstrukte zur Zeitabhängigkeit dargestellt, die als hybride Versionierung von Dimensionshierarchien bezeichnet werden kann. In den bisherigen Ausführungen wurde ja die Randbedingung postuliert, dass für eine Hierarchie nur jeweils eine einzige Form der Berücksichtigung struktureller Veränderungen möglich ist. Diese Voraussetzung kann aber im Allgemeinen auch fallen gelassen werden, so dass hybride Varianten entstehen. In dem Beispiel ist in der Produkthierarchie für einzelne Ebenen eine transaktionsorientierte Sichtweise modelliert, so dass neben dem Produkt auch die Warengruppe eine Beleginformation ist, die über die Bewegungsdaten zur Verfügung gestellt wird. Hingegen erfolgt die Berücksichtigung struktureller Veränderungen in der Warenuntergruppe und der Warenhauptgruppe abhängig von der Gültigkeitszeit.

## 6.4 Semantisches Modell des Beispiels

In diesem Abschnitt erfolgt die Darstellung des T-ADAPT-Modells zu den in Kapitel 5 skizzierten Anforderungen.

Zunächst erfolgen die Aufarbeitung des Informationsbedarfes aus dem Controllingbereich und die Darstellung des damit verbundenen Modells. Als betriebswirtschaftlicher Untersuchungsgegenstand ist für diesen Bereich die Analyse von Kosten bezogen auf Kostenstellen und Kostenarten genannt worden. Die Anforderungen sind durch einen Datenwürfel abdeckbar, dessen T-ADAPT-Darstellung in Abbildung 54 visualisiert ist.

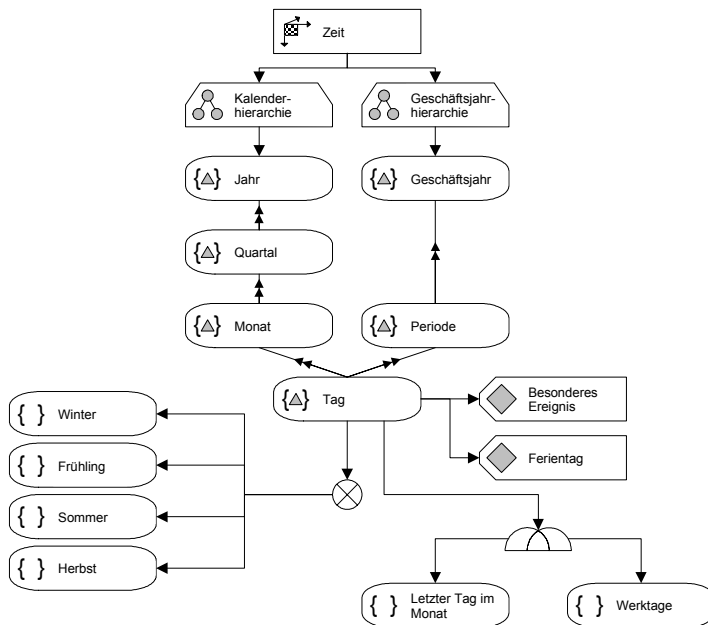


**Abb. 54.** Controlling Würfel in ADAPT-Notation

Im Anwendungsbereich Controlling ist die Kennzahl als eine eigene Dimension modelliert. Neben den bereits identifizierten Dimensionsbildenden Objekten Kostenstelle und Kostenart sind weiterhin die Zeit- und Szenariodimension zu nennen.

Die Dimensionshierarchien für die Kostenstellen und Kostenarten sind jeweils als rekursive Beziehung abgebildet. Bei den Kostenstellen ist eine selektive Zeitabhängigkeit in die Gestaltung eingeflossen, so dass neben der gerade aktuell gültigen Kostenstellenhierarchie auch auf die fixierten Versionen zum jeweiligen Jahresende zurückgegriffen werden kann. Der Kostenstellenverantwortliche findet seinen Eingang in das Modell als Attribut zur Kostenstelle. Die Zeitdimension ist als balancierte Waldstruktur mit den Ebenen Monat, Quartal und Jahr modelliert. Die Szenariodimension berücksichtigt die einzeln aufgelisteten Datenarten Ist, Plan sowie das berechnete Element der Abweichung.

Während für das Controlling eine Granularität auf Monatsebene ausreichend ist, benötigt der Marketingbereich eine Granularität auf Tagesebene. Dies drückt sich im Datenmodell auch durch eine differenziertere Zeitdimension, wie in Abbildung 55 erkennbar, aus.

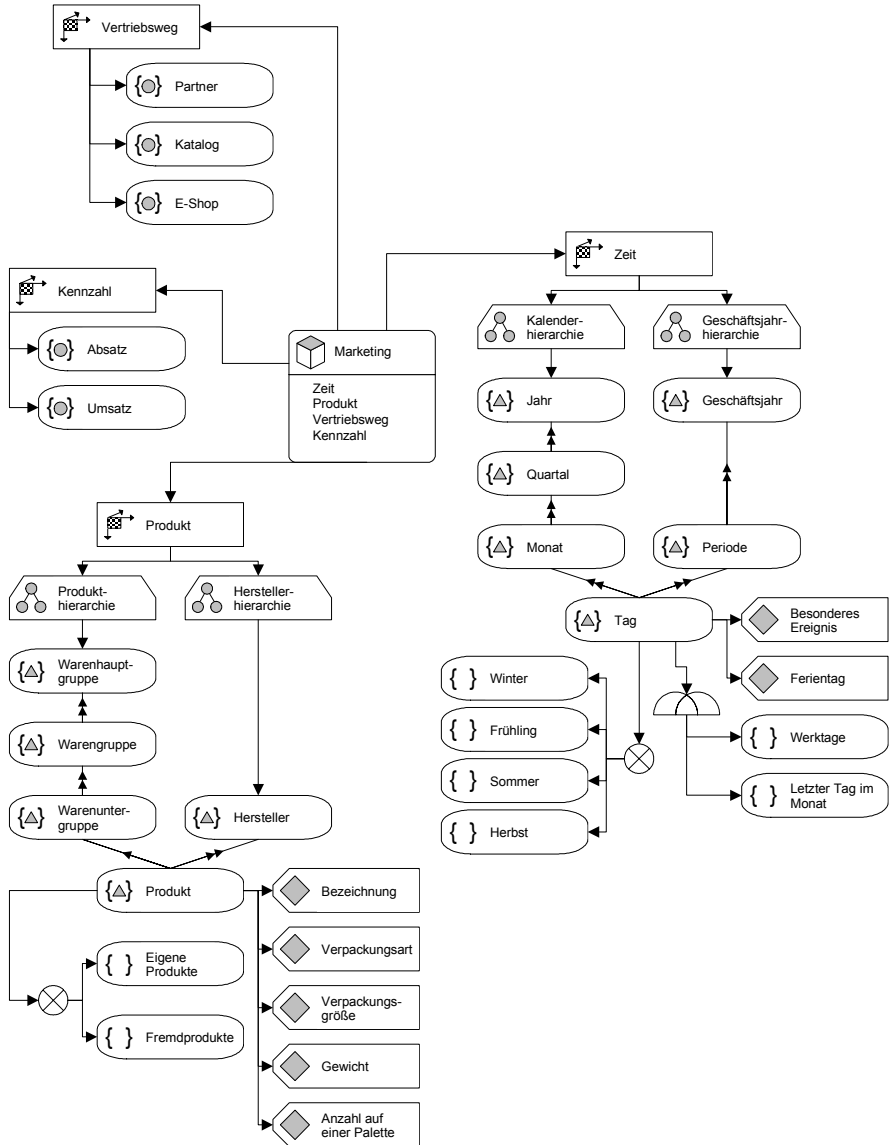


**Abb. 55.** Modellierung von Scopes in der Zeitdimension

Auswertungen nach zeitlichen Aspekten müssen neben den Standardselektionen nach Jahren, Quartalen und Monaten auch Informationen über die Geschäftsjahre anbieten können, da diese im Unternehmen von den Kalenderjahren abweichen. Die Darstellung in T-ADAPT erfolgt mit zwei getrennten Hierarchien.



Die aggregierte Sichtweise, die in den Anforderungen im Marketing artikuliert wurde, umfasst eine Betrachtung der Kennzahlen Absatz und Umsatz bezogen auf zeitliche Aspekte, betrachtete Produkte und den Vertriebsweg. Das korrespondierende Gesamtmodell in T-ADAPT-Notation ist in Abbildung 57 wiedergegeben.



**Abb. 57.** ADAPT-Cube für die aggregierte Marketingsicht



In dieser Darstellung des Gesamtmodells ist neben der Kennzahldimension nur die elementbestimmte Dimension für den Vertriebsweg noch nicht angesprochen worden. Auf der aggregierten Ebene ist keine Zeitabhängigkeit zu berücksichtigen, da nur die jeweils aktuell gültigen Strukturen zur Auswertung herangezogen werden.

Die Gesamtsicht für den Marketingbereich führt die spezielleren Sichtweisen des Partner- und des Kunden-Marketings zusammen. Zunächst sollen die Anforderungen bezogen auf Partner näher betrachtet werden.

Da hierbei ja das Hauptaugenmerk auf der Auswertung nach Partnern und nach der Effizienz der einzelnen Werbemaßnahmen liegt, um die Auswirkungen im Partnergeschäft nachvollziehbar zu machen, ist insbesondere der Bereich der Werbung näher zu betrachten. Dieser bildet die Dimension Werbung in unserem Beispiel-Modell, das in Abbildung 58 wiedergegeben ist.

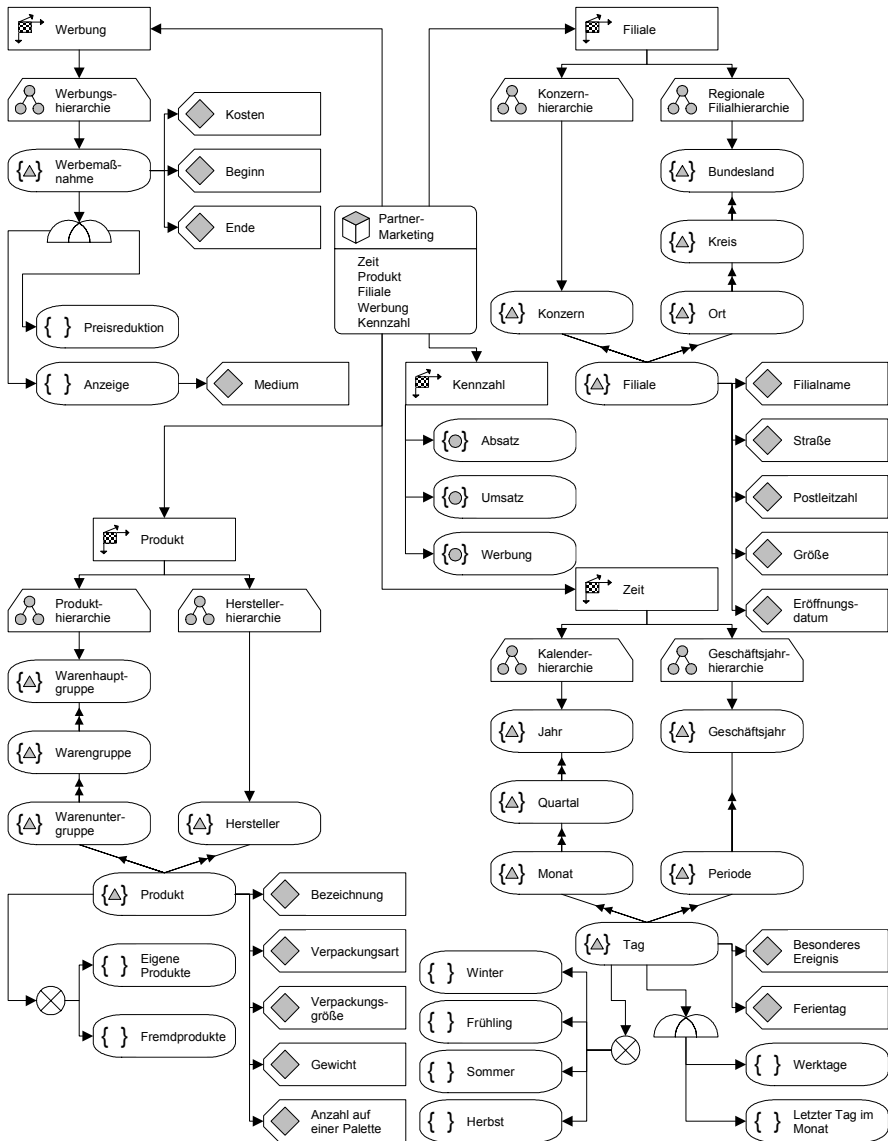
Die für die Auswertung von Werbemaßnahmen notwendigen Informationen, ob es sich um eine Preisreduktion und bzw. oder um eine Anzeige handelt, sind in Form einer Dimensionssicht an der Ebene der Werbemaßnahme festgemacht. Die Werbungshierarchie besteht allerdings nur aus dieser einen Ebene, es handelt sich also um eine flache Struktur. Die zusätzliche Information des Mediums bei Anzeigen ist als Attribut der korrespondierenden Dimensionssicht abgebildet. Die weiteren Zusatzinformationen zu den einzelnen Werbemaßnahmen wie Kosten, Beginn und Ende finden ihren Eingang in das Modell als Attribute zur Ebene der Werbemaßnahmen.

Die Auswertung der Partner bzw. Filialen muss einerseits nach organisatorischer Zuordnung zum Konzern und zum anderen über regionale Zuordnungen möglich sein, so dass hierfür die Modellierung von zwei Hierarchien in einer Dimension für die Filiale angemessen ist. Da die Zuordnung zu Konzernen optional ist und nicht jede Filiale einem Konzern zugeordnet werden kann, ist für die Verbindung der Ebenen innerhalb der Dimensionshierarchie der Pfeil für die lockere Beziehung zu wählen. Die regionale Hierarchie ist vollständig balanciert.

Die weiteren notwendigen Informationen wie Name, Straße, PLZ, Eröffnungsdatum und Größe einer Filiale sind klassischerweise Attribute der untersten Ebene der beiden Hierarchien, also direkt an die Filial-Ebene angebunden.

Auch für das Partner-Marketing wird auf die Berücksichtigung temporaler Aspekte in Form von strukturellen Änderungen in den Hierarchien oder der Änderung von Attributwerten verzichtet.

Für den dargestellten Bereich des Partner-Marketings kommt nur der Vertriebsweg über Partner in Frage, so dass auf die Modellierung einer eigenen Dimension hierfür verzichtet wird.



**Abb. 58.** Partner-Marketing-Cube in der Gesamtdarstellung

Der zweite Bereich, der seinen Eingang in die Gesamtsicht des Marketings findet, betrachtet die kundenbezogenen Aktivitäten.

Beim Direktkundengeschäft über den Katalog oder die On-Line-Aktivitäten liegt das Hauptaugenmerk der Auswertungen auf den angesammelten Informationen über die Kunden, die eine zentrale Rolle in der Gestaltung des Gesamtmodells wie in Abbildung 59 dargestellt ausmachen. Hier ist

die dargestellte Kunden-Dimension im Modell näher zu betrachten, denn die anderen Dimensionen entsprechen den bisherigen Darstellungen. Zu beachten ist, dass hier in der Dimension Vertriebsweg nur zwei Ausprägungen berücksichtigt sind, denn nur über den E-Shop und den Katalog liegen die Informationen kundenbezogen vor.

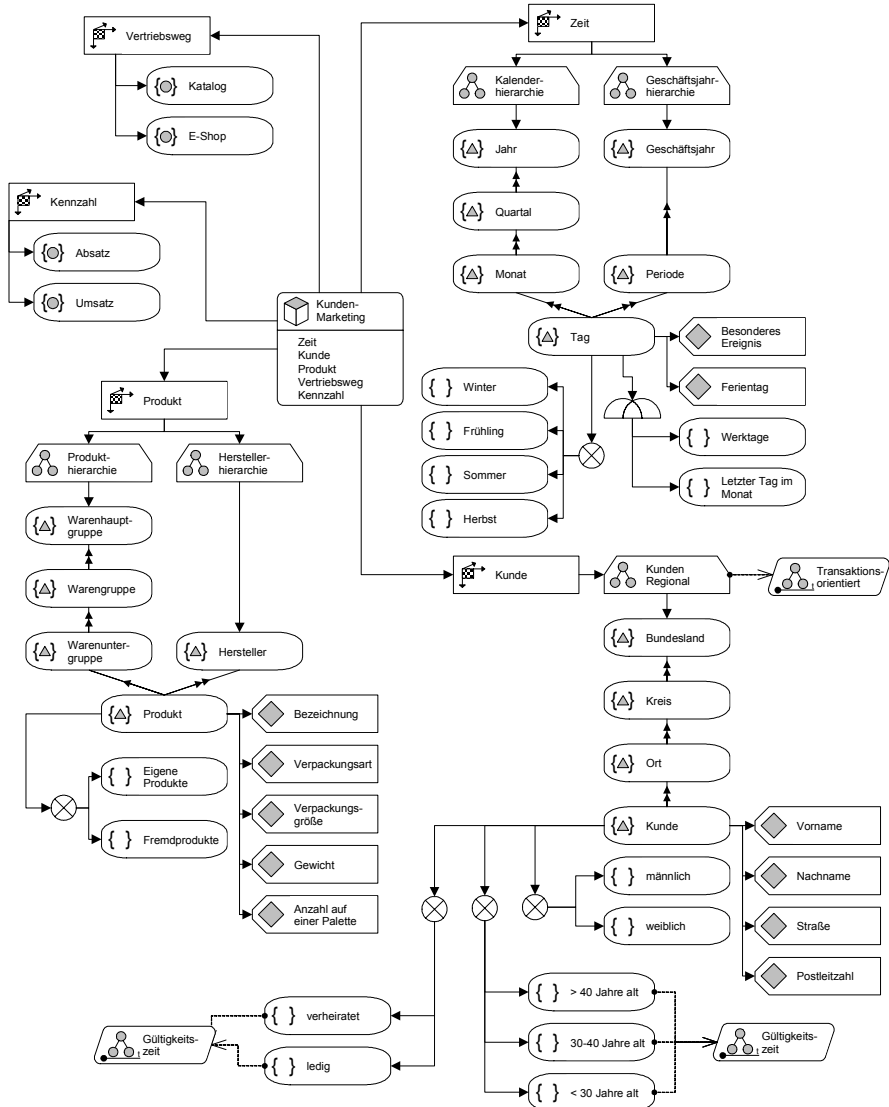


Abb. 59. Gesamtmodell für das Kunden-Marketing

Aus marketingstrategischen Gesichtspunkten ist es notwendig, Kunden in einer regionalen Gruppierung auswerten zu können, wobei die Gruppierung den üblichen regionalen Anforderungen entsprechen soll. Für diese Hierarchie ist eine transaktionsorientierte Versionierung abgebildet. Weiterhin kann zu jedem Kunden eine Gruppierung nach Familienstand, Geschlecht und Altersgruppe erfolgen. Zusätzlich sind zu jedem Kunden neben der Kundennummer Detailinfos wie Name, Vorname, Regionalinfos etc. zur Verfügung zu stellen. Bis auf das Geschlecht sind die demographischen Attribute als zeitabhängige Dimensionsausschnitte deklariert. Namens- und Adressinformationen sind als Attribute modelliert.

## 6.5 Weiterführende Literatur

Eine Übersicht zu den verschiedenen Ansätzen der semantischen Modellierung mehrdimensionaler Datenstrukturen für analyseorientierte Systeme bieten (Gabriel u. Gluchowski 1998) sowie (Schelp 2000).

Zum Multidimensional ER-Modell wird auf (Sapia et al. 1998) verwiesen. Näheres zum Dimensional Fact Modeling ist in (Golfarelli et al. 1998a) sowie (Golfarelli et al. 1998b) nachzulesen. Das Kennzahlendatenmodell ist in (Groffmann 1992) ausführlich beschrieben. Das Multidimensional Data Model ist in (Cabibbo u. Torlone 1997a) und (Cabibbo u. Torlone 1997b) beschrieben.

Zur ADAPT-Spezifikation speziell siehe (Bulos 1996), in (Bulos 1998) findet sich ein Nachdruck hiervon. Eingehender diskutiert wird ADAPT in (Totok u. Jaworski 1998), insbesondere die Bereiche der Dimensionsausschnitte und die Ableitung von Teilen eines Würfels werden dort intensiv diskutiert. Die dargestellten Erweiterungen von ADAPT insbesondere bezogen auf die Berücksichtigung temporaler Aspekte stammen vom Autor.

Bulos D (1996) A New Dimension. Database Programming & Design 9:33–38

Bulos D (1998) OLAP Database Design – A New Dimension. In: Chamoni P, Gluchowski P (Hrsg) Analytische Informationssysteme: Data Warehouse, On-Line Analytical Processing, Data Mining. Springer, Berlin, S 251–261

Cabibbo L, Torlone R (1997a) Querying Multidimensional Databases. In: Sixth International Workshop on Database Programming Languages (DBPL-97), S 319–335

Cabibbo L, Torlone R (1997b) A Systematic Approach to Multidimensional Databases. In: Schreiber F (Hrsg) Quinto Convegno Nazionale su Sistemi Evoluti per Basi di Dati (SEBD-97)

- Gabriel R, Gluchowski P (1998) Grafische Notation für die semantische Modellierung multidimensionaler Datenstrukturen in Management Support Systemen. *Wirtschaftsinformatik* 6:493–502
- Golfarelli M, Maio D, Rizzi S (1998a) The Dimensional Fact Model: a Conceptual Model for Data Warehouses. In: *International Journal of Cooperative Information Systems* 2-3:215-247
- Golfarelli M, Maio D, Rizzi S (1998b) Conceptual design of data warehouses from E/R schemes. In: El-Rewini H (Hrsg) *Proceedings of the 31<sup>st</sup> Hawaii International Conference on System Sciences (HICCS-31)*, S 334-343
- Groffmann HD (1992) Kennzahlendatenmodell (1992) als Grundlage aktiver Führungsinformationssysteme. In: Rau KH, Stickel E (Hrsg) *Daten- und Funktionsmodellierung – Erfahrungen – Konzepte – Perspektiven*. Gabler, Wiesbaden
- Sapia C, Blaschka M, Höfling G, Dinter B (1998) Extending the E/R Model for the Multidimensional Paradigm. In: Kambayashi Y, Lee DL, Lim E, Mohania M, Masunaga Y (Hrsg) *Advances in Database Technologies, ER '98 Workshops on Data Warehousing and Data Mining, Mobile Data Access and Collaborative Work Support and Spatio-Temporal Data Management*. *Proceeding Nr. 1552 der Reihe "Lecture Notes in Computer Science (LNCS)"*, Springer, Berlin, S 105–116
- Schelp J (2000) *Modellierung mehrdimensionaler Datenstrukturen*. Deutscher Universitäts-Verlag, Wiesbaden
- Totok A, Jaworski R (1998) *Modellierung von multidimensionalen Datenstrukturen mit ADAPT*. Arbeitsbericht 98-11, Technische Universität Braunschweig

## 6.6 Zusammenfassung

Die Analyse auf fachkonzeptueller Ebene ist ein wesentlicher Schritt beim Aufbau einer Data Warehouse-Lösung. Die für OLAP-Analysen benötigten mehrdimensional strukturierten Datenbestände sind an den Anforderungen orientiert zu gestalten. Daher hat die semantische Modellierung mehrdimensionaler Datenstrukturen für die frühen konzeptgetriebenen Projektphasen einen hohen Stellenwert.

Neben vielen Vorschlägen zur konzeptionellen Gestaltung mehrdimensionaler Informationssysteme, die eher der logischen Modellebene zuzurechnen sind, gibt es auch eine Reihe von Modellierungsmethoden, die für eine Diskussion auf Fachbereichs- bzw. Anwenderebene geeignet erscheinen.

Eines der bekanntesten semantischen Datenmodelle ist sicherlich das Entity Relationship-Modell (ERM) von Chen, in dem die Objekte der Modellwelt als *Entities* und die Beziehungen als *Relationship* bezeichnet werden. Das ER-Modell ist von seinem Ausgangspunkt her ein methodischer Ansatz für die Modellierung operativer Datenstrukturen. Gleichwohl kön-

nen mit dieser Menge an Beschreibungsmitteln auch mehrdimensionale Strukturen abgebildet werden.

Das Mehrdimensionale ER-Modell (ME/R) ist eine explizite Erweiterung zur Modellierung mehrdimensionaler Datenstrukturen und sieht spezifische neue Darstellungsobjekte vor, um eine adäquate grafische Notation mehrdimensionaler Datenstrukturen zu ermöglichen. Hierbei handelt es sich um einen Entity-Typen für die Darstellung von Dimensionsebenen sowie die Beziehungs-Typen für die hierarchischen Anordnungen und die Fakten. Berechnungsvorschriften sind in den ER-Ansätzen ebenso wenig berücksichtigt wie Kennzahlensysteme.

Im *Dimensional Fact Modeling* sind die Beschreibungselemente Fakten, Fakt-Attribute, Dimensionen und Hierarchien sowie nicht-dimensionale Attribute für die grafische Beschreibung eingeführt. Daneben umfasst der Vorschlag auch ein Vorgehensmodell. Das Modell sieht eine gesonderte Notation der Verdichtungsoperation für semi-additive und nicht-additive Fakten bzw. Fakt-Attribute vor. Eine explizite Darstellung von Kennzahlensystemen ist in dem Modell nicht vorgesehen.

Das *Kennzahlendatenmodell* ist unabhängig von der Data Warehouse Diskussion entstanden. Zentraler Ausgangspunkt ist die Positionierung von Kennzahlen und Kennzahlensystemen als ein wichtiges Informations- und Steuerungsinstrument für das Management. Es werden die folgenden acht Sichten auf eine Kennzahl unterschieden: Untersuchungsgegenstand, Einheit, betriebliches Aggregationsniveau, zeitliches Aggregationsniveau, Zeitpunkt, Zahlenart, Informationszweck und Informationsquelle.

Dieses Modell ist in der Phase der Kennzahlendefinition ein durchaus hilfreiches Werkzeug gerade beim Aufbau von Führungsinformationssystemen, da hier eine sehr differenzierte Sichtweise auf die betriebswirtschaftlichen Variablen notwendig ist.

Das *Multidimensional Data Model* umfasst neben einer grafischen Repräsentation auch ein Vorgehensmodell, mit dem aus vorhandenen ER-Modellen ein entsprechendes mehrdimensionales Schema abgeleitet werden kann. In der als *f-graph* bezeichneten Repräsentation befindet sich ein zentraler Knoten, der als *f-node* die Verbindung zwischen den Variablen und den verschiedenen Dimensionen des Würfels herstellt. Bei diesem Ansatz ist nur die Abbildung von 1:n-Beziehungen in Dimensionshierarchien möglich, die Modellierung anteiliger Verrechnungen ist gar nicht vorgesehen. Die Darstellung von Kennzahlensystemen ist insofern nicht möglich, als betriebswirtschaftliche Variablen als Attribute des zentralen Fakts eines Graphen modelliert werden und nicht in einer eigenen Dimension abbildbar sind.

Bei *Application Design for Processing Technologies (ADAPT)* handelt es sich um einen Modellierungsansatz, in dessen Fokus explizit mehrdi-

mensionale Datenstrukturen für OLAP-Anwendungen stehen. Zusätzlicher praktischer Nutzen der Methode liegt in der Verfügbarkeit der zugrunde liegenden Symbole für Microsoft Visio, wodurch die Modelle einfach zu generieren sind. Diese stehen unter [www.hahneonline.de](http://www.hahneonline.de) zum Download zur Verfügung. Die Stärke von ADAPT liegt in der Darstellung vielfältiger Dimensionsstrukturen und deren betriebswirtschaftlichem Bezug. In der hier dargestellten erweiterten Form sind neben den klassischen Dimensionsstrukturen auch rekursive Beziehungen und m:n-Beziehungen darstellbar. Auch eine explizite Differenzierung von ebenenbestimmten und elementbestimmten Dimensionen ist berücksichtigt. Ein weiteres Grundkonstrukt zur Darstellung mehrdimensionaler Strukturen auf semantischer Ebene sind Attribute zu Dimensionen oder Dimensionsebenen, die beide in ADAPT darstellbar sind. Eine Stärke von ADAPT ist die Modellierung von Dimensionsausschnitten bzw. Dimensionssichten die eine logisch zusammenhängende Teilmenge einer Dimension darstellen. Neben den Objekten zur Modellierung von Cubes stehen in ADAPT mit den Modellen bzw. Ableitungsregeln auch Möglichkeiten der Modellierung abgeleiteter Teilsichten von Würfeln zur Verfügung.

Die Berücksichtigung von Aspekten struktureller Veränderungen und von Zeitaspekten ist mit der Erweiterung T-ADAPT möglich, die zusätzliche Objekte zur Modellierung von Zeitabhängigkeit für Hierarchien und Attribute bereitstellt. Diese Objekte wurden vom Autor eingeführt und sind ebenfalls in den unter [www.hahneonline.de](http://www.hahneonline.de) erhältlichen Schablonen enthalten.

## 7 Allgemeines Star Schema

Zur Abbildung mehrdimensionaler Datenstrukturen in relationalen Systemen hat sich mittlerweile ein Standard entwickelt, der unter dem Sammelbegriff Star Schema bekannt ist. Hierunter ist eine facettenreiche Vielzahl von Varianten einer relationalen Modellklasse zu verstehen, deren Ursprung in der mehrdimensionalen Betrachtungsweise liegt.

Die Grundbestandteile des Star Schema-Ansatzes sind Gegenstand des Abschnittes 7.1. Dort steht die Modellierung von Dimensionshierarchien und Kennzahlensystemen im Vordergrund. Auf die Aspekte der Vorberechnung konsolidierter Werte und der Speicherung von Aggregaten geht Abschnitt 7.2 ein. In diesem Zusammenhang sind auch die Konzepte des Fact Constellation Schemas und des Snow Flaking zu nennen. Weitere besondere Modellierungsvarianten sind Gegenstand des Abschnittes 7.3.

### 7.1 Grundform des Star Schemas

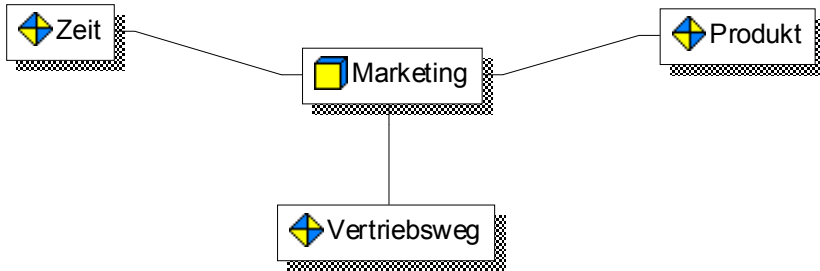
In dem grundsätzlichen Ansatz des Star Schemas werden die quantifizierenden Informationen in einer zentralen Tabelle gehalten, die Faktentabelle genannt wird. Die Ablage der qualifizierenden Informationen erfolgt in Form von Satellitentabellen, die sternförmig um die Faktentabelle herum angeordnet sind und Dimensionstabellen genannt werden. Der identifizierende Schlüssel in der Faktentabelle ist dabei der zusammengesetzte Schlüssel bestehend aus den Fremdschlüsseln, welche die Primärschlüssel aller Dimensionstabellen referenzieren.

Während in der Faktentabelle die Bewegungsdaten enthalten sind, beinhalten die Dimensionstabellen die Stammdaten und beschreiben die Bewegungssätze. Die Dimensionstabellen haben drei wesentliche Aufgaben:

- Sie beschreiben die Fakten, um daraus sinnvolle Aussagen entstehen zu lassen.
- In ihnen sind die Suchkriterien festgelegt, nach denen Fakten sinnvoll auswertbar sind.
- Sie definieren die Hierarchien, entlang derer die Verdichtungsstufen für die Auswertungen festgelegt werden können.



In der Grundform des Star Schemas sind alle Kennzahlen als Spalten in einer großen Faktentabelle abgelegt. Als einfaches Beispiel für ein Modell in Form eines Star Schemas soll das Beispiel der Anwendung im Marketingbereich dienen. Die abstrakte Sicht ohne detaillierte Tabellendarstellung ist in Abbildung 60 dargestellt. Dieses stellt die Kennzahlen aufgliedert nach Produkten und Vertriebswegen für einzelne Zeitperioden dar.



**Abb. 60.** Einfaches Star Schema

Mit diesem Modell, das in den weiteren Ausführungen detaillierter erörtert wird, ist es möglich, Fragen der folgenden Art zu beantworten:

- Gesamtumsatz für ein Produkt?
- Anteil am Gesamtumsatz für einen Vertriebsweg?
- Welche Produkte ergeben wann und über welchen Vertriebsweg den größten Anteil am Gesamtumsatz?
- Welche Artikel einer Warengruppe verkaufen sich unterproportional?

Die Granularität der Auswertungen und die Möglichkeiten der Verdichtung ergeben sich aus der Modellierung der Dimensionstabellen die nachfolgend dargestellt wird.

### 7.1.1 Modellierung von Dimensionshierarchien

Die verschiedenen Formen von hierarchischen Strukturen in Dimensionen auf semantischer Ebene standen in Abschnitt 3.2.2 im Vordergrund. Nun erfolgt die Darstellung der Abbildungsmöglichkeiten dieser Strukturen im Star Schema. Als Grundlage zur Veranschaulichung dient der Marketingbereich des Anwendungsbeispiels.

### **Flache Strukturen**

Die Dimension *Vertriebsweg* hat als Dimensionselemente ausschließlich die Knoten *Partner*, *Katalog* und *E-Shop*, die in keiner hierarchischen Beziehung zueinander stehen. Für die Verknüpfung der Dimensionstabellen mit der Faktentabelle dienen im Allgemeinen künstliche Primärschlüssel. Für den Vertriebsweg ergibt sich die folgende Dimensionstabelle:

**Tabelle 10.** Dimensionstabelle Vertriebsweg

Vertriebsweg_Id	Vertriebsweg
1	Partner
2	Katalog
3	E-Shop

Die in diesem Beispiel verwendeten Primärschlüssel sind lediglich eine fortlaufende Nummer, die die Eindeutigkeit der Datensätze gewährleistet und damit keine semantische Information trägt.

### **Balancierte Baum- und Waldstrukturen**

Die Zeitdimension hat in fast allen Data Warehouse-Modellen eine herausragende Bedeutung, da die zeitliche Qualifizierung des betrachteten Zahlenmaterials essentiell ist. Die kalendarische Sicht ist eine klassische Form einer balancierten Struktur. Für die Abbildung in einem Star Schema gibt es verschiedene Varianten. Zunächst erfolgt die Abbildung in einer Dimensionstabelle. Dann entspricht jede Konsolidierungsebene der Hierarchie einer eigenen Spalte in der Dimensionstabelle. Die Hierarchiestufen der Verdichtungswege in der Dimension sind in dieser Form die Spalten *Monat*, *Quartal* und *Jahr* in der Tabelle, wobei von einer Granularität auf Monatsebene ausgegangen wird.

**Tabelle 11.** Dimensionstabelle Zeit

Zeit_Id	Monat	Quartal	Jahr
1	Januar	Q1	2004
2	Februar	Q1	2004
3	März	Q1	2004
4	April	Q2	2004
5	Mai	Q2	2004
...	...	...	...
12	Dezember	Q4	2004
13	Januar	Q1	2005
14	Februar	Q1	2005
...	...	...	...
24	Dezember	Q4	2005

In der skizzierten Form der Dimensionstabelle für die Zeit ist der Primärschlüssel wieder ein einfacher künstlicher numerischer Schlüssel. In einer anderen Variante der Modellierung wird ein zusammengesetzter Primärschlüssel bestehend aus Komponenten für jede Konsolidierungsstufe verwendet. Dann sind automatisch auch Schlüsselinformationen zu den Verdichtungsebenen bereits in der Faktentabelle abgelegt und in vielen Fällen erspart dies in der Abfrage einen Join mit der Dimensionstabelle.

***Unbalancierte Strukturen***

Eine weitere Klasse von Baum- und Waldstrukturen sind die mit unterschiedlich langen Wegen von der Wurzel zu den Blättern, die unter dem Begriff unbalanciert oder unausgeglichen eingeführt wurden. Die dritte Dimension in dem betrachteten Beispiel, die Produktdimension, ist ein guter Kandidat für eine solche Struktur, denn teilweise werden die Produkte neben der Strukturierung nach Produkt- bzw. Warengruppen auch in Untergruppen einsortiert. Eine exemplarische Ausprägung der Dimensionstabelle könnte die in der folgenden Tabelle dargestellte Form haben.

**Tabelle 12.** Ausschnitt der Dimensionstabelle Produkt

Produkt_Id	Produkt	Untergruppe	Gruppe	Hauptgruppe
1	100	Widerstände	Bauteile	Elektronik
2	1 k	Widerstände	Bauteile	Elektronik
...	...	...	...	...
51	250 F	Kondensatoren	Bauteile	Elektronik
52	500 F	Kondensatoren	Bauteile	Elektronik
...	...	...	...	...
101	PA600	NULL	Verstärker	High Fidelity
102	PAX300	NULL	Verstärker	High Fidelity
103	PAX450	NULL	Verstärker	High Fidelity
...	...	...	...	...

An den NULL-Einträgen in der Spalte der Untergruppe ist die Eigenschaft der Unbalanciertheit erkennbar. Die dargestellte Struktur stellt eine unausgeglichene Waldstruktur da. Zu einer Baumstruktur wird dies durch Hinzunahme einer weiteren Spalte, die allerdings für jede Zeile der Tabelle die Ausprägung *Alle Produkte* hätte und damit nicht wirklich sinnvoll ist.

### **Parallele Hierarchien**

In Abschnitt 3.2.2 diente die Zeitdimension der Verdeutlichung paralleler Hierarchien. Dieses Anwendungsbeispiel soll nun auf den Fall des Star Schemas Anwendung finden. Wie auch bei den anderen diskutierten Strukturformen wird im Fall der parallelen Hierarchie jede Konsolidierungsebene durch eine eigene Spalte der Dimensionstabelle repräsentiert, und zwar dies für alle parallelen Hierarchien. Demzufolge kann die Zeitdimension im Star Schema wie in der folgenden Tabelle 13 dargestellt implementiert werden.

**Tabelle 13.** Zeitdimension mit Kalender- und Geschäftsjahressicht

Zeit_Id	Monat	Quartal	Jahr	Geschäftsjahr
1	Januar	Q1	2004	2003/2004
2	Februar	Q1	2004	2003/2004
3	März	Q1	2004	2003/2004
4	April	Q2	2004	2004/2005
5	Mai	Q2	2004	2004/2005
...	...	...	...	...
12	Dezember	Q4	2004	2004/2005
13	Januar	Q1	2005	2004/2005
14	Februar	Q1	2005	2004/2005
...	...	...	...	...
24	Dezember	Q4	2005	2005/2006

An der Tabelle selbst ist dann aber nicht mehr ablesbar, welche Spalten zu welcher Hierarchie gehören. Wie auch bei den anderen Strukturen ist die Information über die Konsolidierungsebenen ebenfalls nicht mehr direkt aus einer Tabelle ableitbar.

### ***Anteilige Verrechnung und Heterarchien***

Bei den Strukturformen, in denen die Bedingung der 1:n-Beziehungen zwischen den Konsolidierungsebenen fallen gelassen wird, sind zwei Varianten zu differenzieren. Im einfachsten Fall basiert die Verdichtung auf der üblichen Annahme der Summation, die in der zweiten Form ebenso fallen gelassen wird. Für letzteres wurde in Abschnitt 3.2.2 das Beispiel der Konsolidierung entlang einer Hierarchie von Tochtergesellschaften in eine Holding aufgeführt.

In der Implementierung im Star Schema ist die Bedingung einer 1:n-Beziehung impliziert fixiert, da für ein Tupel jeder Attributwert eindeutig ist. Üblicherweise erfolgt die Realisierung dieser Beziehung zwischen Ebenen im Star Schema in Form zweier eigenständiger Dimensionen.

Für die Realisierung von Heterarchien der aufgeführten Form mit einer nicht standardmäßigen Verdichtung, die ebenfalls über zwei eigenständige Dimensionen erfolgen muss, ist eine Abfrage verdichteter Werte mit normalen Abfragen nicht möglich und es sind andere Formen der Berechnung

zu berücksichtigen, etwa durch vorberechnete Aggregatwerte. Dies wird in Abschnitt 7.2 wieder aufgegriffen.

### 7.1.2 Attribute in Dimensionen

Wie in Abschnitt 3.2.3 dargestellt sind zwei Arten von Attributen zu unterscheiden, die eher technisch orientierten für die Navigation entlang der Ebenen von Konsolidierungshierarchien und die fachlich bestimmten Attribute, die einen Großteil der analytischen Möglichkeiten eines Modells ausmachen. Für die Implementierung im Star Schema spielt die Art des abzubildenden Attributes keine Rolle, da jedes durch eine eigene Spalte in der Dimensionstabelle repräsentiert wird. Die Abbildung von Attributen in der Dimensionstabelle ist exemplarisch in der Abbildung 61 für den Fall der Zeitdimension dargestellt.

DTZeit

Zeit_Id
Tag
Wochentag
Monat
Quartal
Jahr
Geschäftsjahr
Werktag_flag
Ferien_flag
Jahreszeit
Ereignis
Letzter_Tag_im_Monat

**Abb. 61.** Attribute in der Zeitdimension

Neben den im Beispiel aufgeführten fachlichen Attributen erfolgt auch die Darstellung der Attribute für Ebene und Generation sowie die Ebenenbezeichnung in Form von eigenen Spalten in der Tabelle.

Im Star Schema ist der Unterschied zwischen den Spalten für Hierarchiestufen und denen für Attribute, egal welcher Form, nicht erkennbar, da alle Spalten gleichberechtigt nebeneinander stehen. Die strenge Differenzierung zwischen diesen Komponenten mehrdimensionaler Modelle, wie sie auf semantischer Modellebene propagiert werden, ist im Star Schema nicht ohne zusätzliche Informationen möglich.

### 7.1.3 Normalisierung von Dimensionen

Die bisher dargestellten Star Schema Modelle haben die Eigenschaft, dass ihre Dimensionstabellen in erster und zweiter Normalform sind. In der Theorie und Praxis relationaler Datenbanksysteme hat aber die dritte Normalform eine besondere Bedeutung, die dadurch gekennzeichnet ist, dass neben den Eigenschaften der ersten und zweiten Normalform kein Nichtschlüsselattribut transitiv vom Primärschlüssel abhängt.

Die Überführung einer Dimensionstabelle in die dritte Normalform soll im Folgenden am Beispiel der Produktdimension dargestellt werden. Dazu sei die in Abbildung 62 dargestellte, um zahlreiche Attribute erweiterte Dimensionstabelle zugrunde gelegt.

 DTProdukt

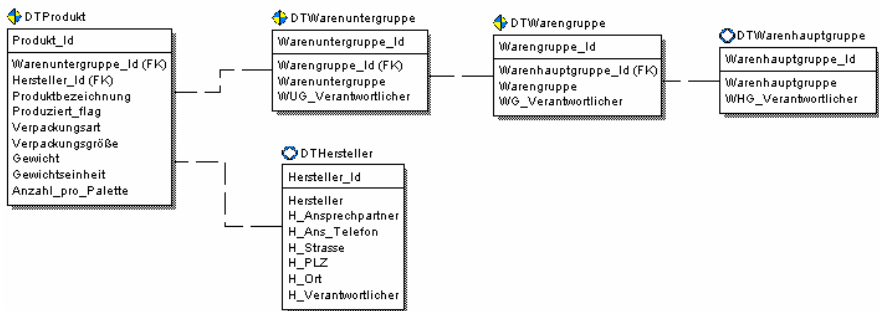
Produkt_Id
Hersteller_Id
Hersteller
H_Ansprechpartner
H_Ans_Telefon
H_Strasse
H_PLZ
H_Ort
H_Verantwortlicher
Warenuntergruppe_Id
Warenuntergruppe
WUG_Verantwortlicher
Warengruppe_Id
Warengruppe
WG_Verantwortlicher
Warenhauptgruppe_Id
Warenhauptgruppe
WHG_Verantwortlicher
Produktbezeichnung
Produziert_flag
Verpackungsart
Verpackungsgröße
Gewicht
Gewichtseinheit
Anzahl_pro_Palette

**Abb. 62.** Herkömmliche Produktdimension

Die vielen Attribute in der Dimension zeigen auch die vielfältigen Möglichkeiten von Attributen auf unterschiedlichen Konsolidierungsebenen. So ist etwa der Warengruppen-Verantwortliche ein Attribut der Ebene der Warengruppe, nicht jedoch der anderen Hierarchieebenen. Die Spalte für

den Hersteller eines Produktes repräsentiert dabei eine parallele Hierarchie, zu der auch die Attribute mit „H\_“ beginnend gehören.

Aus dieser Dimensionstabelle werden nun bis auf die Attribute auf der untersten Ebene und den Schlüsseln zu den Ebenen der nächsthöheren Verdichtungsstufen alle Attribute herausgebrochen. Konkret sind alle weiteren Attribute bezogen auf die Hersteller und die Warenuntergruppe sowie alle Attribute der Warengruppe bzw. Warenhauptgruppe aus der eigentlichen Dimensionstabelle entfernt. Durch diesen Schritt der Normalisierung, dessen Ergebnis in Abbildung 63 zusammengefasst ist, wird die Tabelle in mehrere über Beziehungen verbundene Tabellen umgeformt.



**Abb. 63.** Normalisierte Produktdimension

An dieser Darstellung ist sehr gut die Parallelität zur typisierten Darstellung der Dimension erkennbar, in der die Knoten für die Konsolidierungsebenen den normalisierten Dimensionstabellen entsprechen. Dies ist allerdings nicht der Regelfall, denn die Entscheidung Attribut vs. Dimensionsebene ist einer der Freiheitsgrade in der Modellierung.

Wichtig ist bei diesem Modellierungsansatz, dass bei Abfragen, in denen Attribute abgefragt werden, die nicht in der eigentlichen Dimensionstabelle gespeichert sind, die Einbeziehung der weiteren Tabellen notwendig ist. Trotz der gewonnen Übersichtlichkeit der Dimensionsstruktur leidet die Performance bei dieser Modellierungsvariante zumindest bei entsprechend großen Datenmengen.

Für diesen dargestellten Fall der Abfrageanforderungen bietet sich der Ansatz der Partitionierung an, in dem die Attribute der höheren Ebenen in die zu den unteren Ebenen gehörigen Tabellen dupliziert werden. Hierdurch entsteht eine bewusste Redundanz, die aber Vorteile in der Abfrageperformance impliziert. Die Ableitung eines partitionierten Modells aus der Darstellung der Produktdimension ist in Abbildung 64 wiedergegeben



und unterscheidet sich nur durch die redundanten Attributspalten von dem normalisierten Modell.

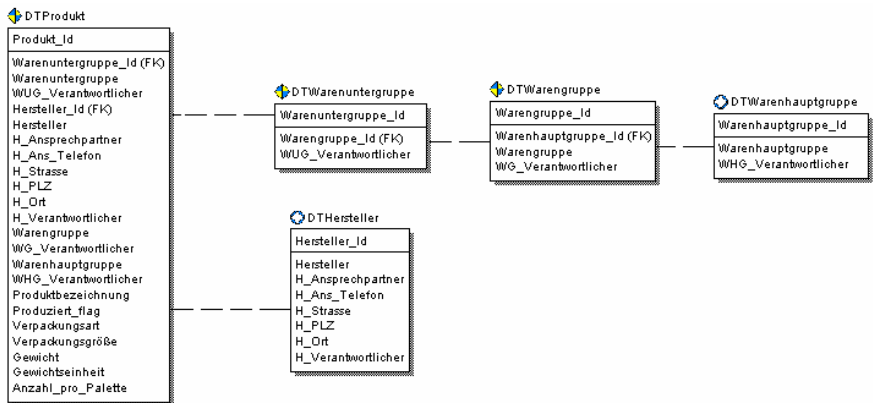


Abb. 64. Partitionierte Dimensionstabelle

Die Verwendung von normalisierten oder partitionierten Dimensionstabellen führt in Verbindung mit der Verwendung von Aggregattabellen zu einer weiteren Designvariante des Star Schemas und wird in Abschnitt 7.2 aufgegriffen.

### 7.1.4 Abbildung von Kennzahlen und Kennzahlensystemen

Im Ansatz des klassischen Star Schemas und der bisher diskutierten Varianten erfolgt die Speicherung der betrachteten Kennzahlen, die als eigentliche qualifizierte Wertgrößen für das Modell eine herausragende Bedeutung haben, in der Faktentabelle. Das Kennzahlensystem des Marketingbereiches ist im Folgenden die Grundlage für die Analyse der Möglichkeiten zur Abbildung von Kennzahlensystemen im Star Schema. Die Abbildung des exemplarischen Kennzahlensystems führt zu der in Abbildung 65 dargestellten Faktentabelle.



Abb. 65. Kennzahlen in der Faktentabelle

In dieser Darstellung tritt das wesentliche Defizit dieser Modellierungsvariante hervor: Jede Kennzahl steht für sich und die Abhängigkeiten, die sich aus der Semantik des Kennzahlensystems ergeben, gehen dabei verloren. Die rechnerischen Abhängigkeiten sind somit ebenfalls nicht darstellbar, sondern sind Teil des Datenbeschaffungsprozesses.

Ein alternativer Ansatz zur Modellierung von Kennzahlensystemen ergibt sich aus deren Struktur eines Graphen der zufolge dieses System als Dimension des mehrdimensionalen Modells aufgefasst werden kann. Das impliziert, dass in der Faktentabelle nur eine künstliche Faktenspalte für den Wert einer über die Dimensionsausprägung in der Kennzahlendimension näher bestimmten Kennzahl vorhanden ist. Diese Form der Darstellung hat den Vorteil, dass die hierarchische Information zur Navigation im Kennzahlensystem erhalten bleibt.

In den Darstellungen der Abbildung von Kennzahlensystemen wird im Allgemeinen davon ausgegangen, dass die Beziehungen der Kennzahlen untereinander, die sich aus den Abhängigkeiten in der Berechnung ergeben, auch gleichzeitig die Beziehung zur Navigation widerspiegeln. In der dargestellten Form einer Kennzahlendimension ist das implizit ebenso. Da jedoch die Berechnung der Kennzahlen ohnehin Teil des Prozesses der Datenbeschaffung und kein direkter Modellbestandteil ist, kann diese Annahme fallen gelassen werden und die Dimensionshierarchie repräsentiert die Navigationspfade. Da sich aber aus der Hierarchie demzufolge keine Berechnungsvorschriften ableiten lassen, handelt es sich bei dieser Form der Modellierung automatisch um ein Star Schema mit gespeicherten Aggregaten, auf die im folgenden Abschnitt näher eingegangen wird.

## 7.2 Aggregate

In der bisherigen Betrachtung des Star Schemas wurden nur Detailwerte auf unterster Ebene jeder Dimension in der Faktentabelle gespeichert. Die analytischen Fähigkeiten basieren aber gerade auf Abfragen auf unterschiedlichste verdichtete Werte. Diese können zum einen zur Laufzeit berechnet werden oder aber zum anderen vorberechnet vorliegen. Sind dabei nicht alle kombinatorischen Möglichkeiten der Verdichtung bereits vorberechnet, muss die Kenntnis über vorhandene Aggregate mit in die Abfrageerstellung eingehen.

In Abschnitt 7.2.1 erfolgt zunächst die Darstellung des einfachen Falles der dynamischen Berechnung verdichteter Werte zur Laufzeit. Anschließend wird der Fall der Speicherung von Aggregatwerten in der Faktentabelle in Abschnitt 7.2.2 diskutiert. Die Auslagerung der Aggregate in eige-

ne Faktentabellen ist der Grundgedanke des in Abschnitt 7.2.3 dargestellten Fact Constellation Schemas. Die Einbeziehung der Aspekte der Normalisierung von Dimensionen führt in Abschnitt 7.2.4 zu der Modellierungsvariante des Snow Flake Schemas.

### 7.2.1 Dynamische Berechnung

Unabhängig vom Werkzeug, mit dem ein Star Schema ausgewertet wird, kommt dabei SQL (*Structured Query Language*) zum Einsatz. In der Abfrage erfolgt der Join der Faktentabelle mit den beteiligten Dimensionstabellen. Über die Gruppierungsoperatoren erfolgt die Verdichtung, so dass nicht nur Detailwerte abgefragt werden. Typisch für diese Art von Abfragen ist, dass normalerweise sehr viele Datensätze der Faktentabelle abgefragt werden, dass alle betroffenen Tabellen mit einem Join verbunden werden und dass die Ergebnismenge im Vergleich zu den beteiligten Basisdatensätzen recht klein ist. Dies impliziert auch den hohen Stellenwert der Performance-Optimierung in diesem Zusammenhang.

Aufgrund der Gleichberechtigung aller Nichtschlüsselspalten in den Dimensionstabellen geht der semantische Unterschied zwischen Attributen und Konsolidierungsebenen verloren, jedes Attribut kann demzufolge als eine Verdichtungsoption betrachtet werden.

Ein grundsätzliches Problem bei dieser Vorgehensweise der Berechnung verdichteter Werte ist, dass keine Werte auf aggregiertem Niveau erfasst werden können. Dies lässt sich aber durch eine Speicherung von Aggregaten in der Faktentabelle umgehen und wird im folgenden Abschnitt dargestellt.

### 7.2.2 Vorberechnete Aggregate in der Faktentabelle

Grundgedanke bei der Bildung von Aggregaten ist die Erkenntnis, dass ein bereits vorberechnet vorliegender Wert zur Laufzeit nur abgefragt und nicht mehr berechnet werden muss. Dieser Vorteil ist mit dem Nachteil abzuwägen, dass diese Werte trotzdem irgendwann, zumeist in der Batchverarbeitung, zu berechnen sind.

Um auch Aggregate in der Faktentabelle speichern zu können, müssen diese Datensätze über entsprechende Dimensionselemente, d. h. Datensätze in den Dimensionstabellen, qualifiziert werden. Ausgangspunkt der Betrachtungen ist wieder der Marketing Data Mart und die zugrunde liegende Zeitdimension.

Veränderungen ergeben sich für Dimensionen, in denen eine Hierarchie für Konsolidierungswege vorgegeben ist, wie beispielsweise die Kalender-

jahressicht in der Zeitdimension. Die Tabellendefinition selbst bleibt unverändert, jedoch treten weitere Tupel für die Qualifizierung der Verdichtungsebenen hinzu, wie in Tabelle 14 für die Dimension des Beispiels zu erkennen ist.

**Tabelle 14.** Dimensionstabelle Zeit mit Aggregattupeln

Zeit_Id	Monat	Quartal	Jahr
1	Januar	Q1	2004
2	Februar	Q1	2004
3	März	Q1	2004
4	April	Q2	2004
...	...	...	...
12	Dezember	Q4	2004
13	Januar	Q1	2005
14	Februar	Q1	2005
...	...	...	...
24	Dezember	Q4	2005
101	NULL	Q1	2004
...	...	...	...
104	NULL	Q4	2004
105	NULL	Q1	2005
...	...	...	...
108	NULL	Q4	2005
1001	NULL	NULL	2004
1002	NULL	NULL	2005

Für eine Speicherung der kompletten Verdichtung über alle Elemente, was in dem Fall der Jahre meistens nicht sinnvoll ist, müsste ein weiterer Eintrag, der auch in der Jahresspalte den Wert NULL trägt, hinzukommen. Aus den Einträgen mit NULL in den Tupeln ergibt sich, welche Aggregation dieser Knoten repräsentiert.

Somit gibt es zu jeder Kombination von Elementen in Aggregationsebenen in allen Dimensionen genau eine Schlüsselkombination, zu der ein passender Datensatz in der Faktentabelle eingetragen werden kann und so-

gar muss, denn für Abfragen ist es nicht unterscheidbar, für welche Kombinationen vorberechnete Werte in der Faktentabelle vorliegen.

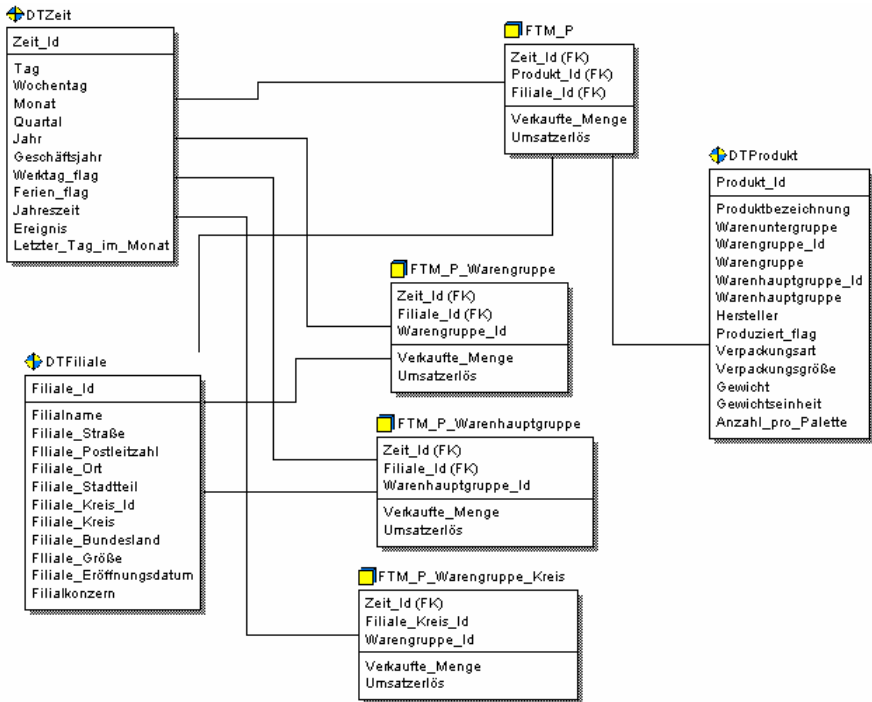
Deutlich vereinfacht werden die Abfragen an solche Modelle, wenn in den Dimensionen zusätzlich noch ein Level-Attribut mit gespeichert wird, das die Ebene genauer bezeichnet. Damit sind Selektionen ganzer Verdichtungsebenen wesentlich einfacher zu formulieren.

Die Abfragen an ein Star Schema mit vorberechneten Aggregaten sind sehr performant. Dieser Vorteil wird aber durch den Nachteil erkauft, dass sowohl das Ebenen-Attribut mit gepflegt werden muss als auch alle Kombinationen der Aggregatbildung im Rahmen des ETL-Prozesses berechnet werden müssen. Letzteres ist sehr komplex und zeitaufwendig. Eine Möglichkeit, die Fülle der Aggregate in den Griff zu bekommen, bietet der im folgenden Abschnitt diskutierte Ansatz.

### **7.2.3 Fact Constellation Schema**

Der konzeptionelle Grundgedanke des Fact Constellation Schemas als ein alternativer Designansatz ist die Auslagerung von Aggregaten aus der eigentlichen Faktentabelle in eigenständige Aggregattabellen. Dieses ergibt sich aus dem normalen Star Schema, in dem für jede mögliche Kombination von Hierarchieebenen aller beteiligten Dimensionen eine Aggregatfaktentabelle eingeführt wird, deren Tupel genau die Werte dieser Ebenenkombination bilden.

Da in einem Fact Constellation Schema die Aggregate jeder Kombination von Verdichtungsstufen in einer eigenen Faktentabelle abgespeichert werden, ist die Angabe eines Level-Attributes obsolet. Die Anzahl der hinzukommenden Aggregattabellen steigt jedoch extrem mit der Anzahl der Dimensionen und Konsolidierungsebenen. Die gestiegene Komplexität führt somit auch zu einer schlechten Wartbarkeit. In Abbildung 66 erfolgt daher auch nur die Darstellung eines Ausschnittes, um das Funktionsprinzip dieses Modellierungsansatzes zu verdeutlichen.

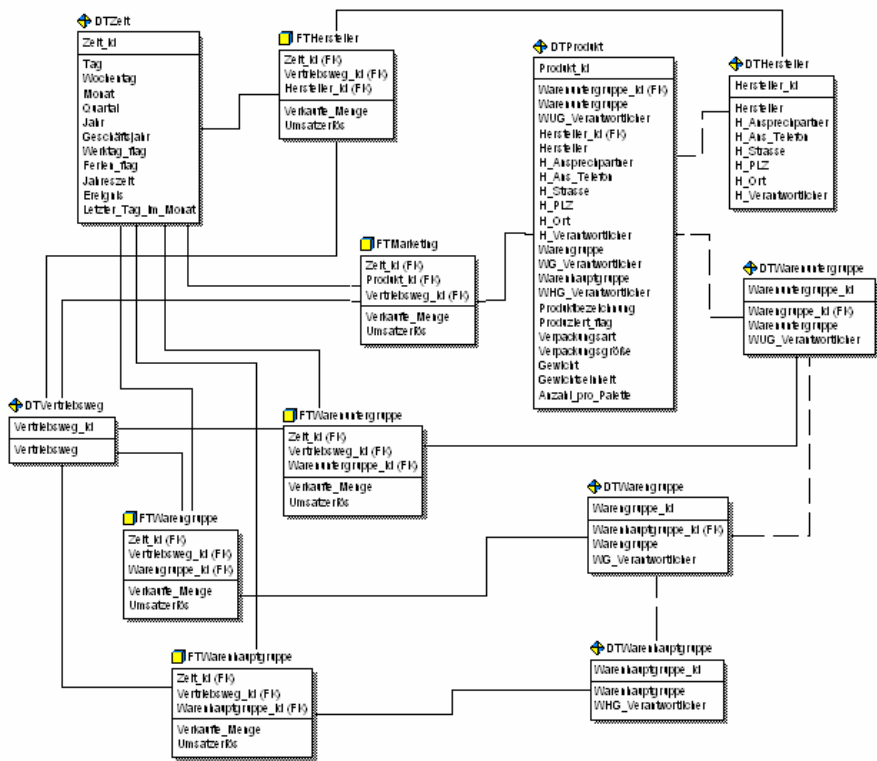


**Abb. 66.** Ausschnitt aus einem Fact Constellation Schema

In der dargestellten Form sind in den Aggregattabellen die Primärschlüssel nicht mehr aus lauter Primärschlüsseln von Dimensionstabellen zusammengesetzt, vielmehr tauchen die Attribute für die Ebenen im Schlüssel wieder auf. Im Beispiel ist etwa das Attribut Warenhauptgruppe\_Id Bestandteil des Schlüssels der Aggregattabelle FTM\_P\_Warenhauptgruppe, dieses Feld ist jedoch in keiner Dimensionstabelle Teil des Primärschlüssels. Dies liegt daran, dass die Dimensionstabellen nicht normalisiert sind. Dies führt somit zum Ansatz des Snow Flake Schemas und ist Gegenstand des folgenden Abschnittes.

### 7.2.4 Snow Flake Schema

Die Kombination der Partitionierung als besondere Form der Normalisierung von Dimensionen mit der Ausgestaltung von Aggregattabellen wie im Fact Constellation Schema führt zu einer Gestaltungsalternative, die als Snow Flake Schema bekannt ist. Ein Auszug eines solchen Schemas ist für die Anwendung im Bereich Kunden-Marketing des Beispiels in Abbildung 67 dargestellt.



**Abb. 67.** Snow Flake Schema für den Bereich Kunden-Marketing

Die Faktentabellen ergeben sich bei dieser Form der Modellierung analog zum Fact Constellation Schema, wodurch sich eine ähnliche Komplexität ergibt, da wieder für jede Kombinationen von Verdichtungen in unterschiedlichen Dimensionen eine eigene Aggregattabelle implementiert ist. Eine zusätzliche Komplexitätssteigerung ergibt sich durch die zusätzlichen Dimensionstabellen aufgrund des Normalisierungsschrittes.

Im allgemeinen Ansatz des Snow Flake Schemas werden alle kombinatorischen Möglichkeiten der Aggregation berücksichtigt. Um einen guten Kompromiss zwischen Performance einerseits sowie Speichernutzung und Übersichtlichkeit andererseits zu ermöglichen, sind zusätzliche Werkzeuge notwendig, damit nur ein Teil der Aggregattabellen tatsächlich vorkalkuliert benötigt wird und verbleibende Verdichtungen aus anderen vorhandenen Werten zur Laufzeit berechnet werden können.

## 7.3 Weitere Modellierungsvarianten

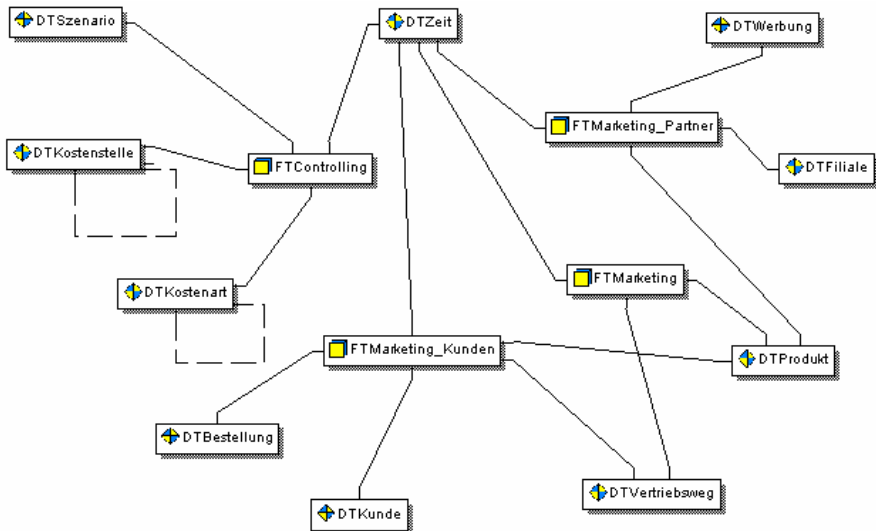
Aus dem praktischen Einsatz heraus sind verschiedene Varianten des Star Schemas eingeführt worden, die sich grob in Varianten der Modellierung von Faktentabellen und von Dimensionstabellen aufteilen lassen.

### 7.3.1 Galaxien

Im bisher dargestellten Star Schema erfolgt die Speicherung aller Fakten in einer Tabelle. Gerade wenn viele Fakten sehr unterschiedlicher Dimensionierung zu berücksichtigen sind ist dies sehr nachteilig.

Eine naheliegende Verbesserung liegt in der Trennung in Faktentabellen, in denen jeweils nur Fakten gleicher Dimensionierung gespeichert werden. Darüber hinaus ist durch diese Trennung auch der semantische Zusammenhang besser abbildbar. Im Extremfall ist jedes Fakt in einer eigenen Tabelle abgebildet, was den Speicherbedarf allerdings signifikant erhöhen würde.

Die in der folgenden Abbildung 68 dargestellte Galaxie bildet die Anwendungsbereiche Marketing und Controlling des Beispiels ab.



**Abb. 68.** Modell der Galaxie des Beispiels

An diesem Beispiel lässt sich aber noch ein weiteres Modellierungsproblem der Galaxie darstellen: Alle dargestellten Faktentabellen sind an die gleiche Dimensionstabelle Zeit gebunden, obwohl ihnen eine zum Teil



unterschiedliche zeitliche Granularität zugrunde liegt. Im Unterschied zu den anderen Faktentabellen ist für den Controllingbereich eine Darstellung auf Monatsebene ausreichend und es wird demzufolge nur ein Teil der Dimensionselemente benötigt.

### 7.3.2 Faktenlose Faktentabellen

In allen bisher dargestellten Star Schema-Modellen sind die Kennzahlen als Bewegungsdaten in der Faktentabelle gespeichert. Eine ganz andere Variante von Faktentabellen stellen diejenigen dar, die gar keine abzuspeichernde Kennzahl beinhalten und deswegen Faktenlose Faktentabelle (*Factless Fact Table*) heißen.

Ein typisches Anwendungsbeispiel dient der Verfolgung von Ereignissen. Dann bedeutet das Vorkommen einer Schlüsselkombination in der Faktentabelle, dass dieses Ereignis eingetreten ist. Dies wird in der Faktentabelle oftmals durch ein künstliches Zähl-Faktum ergänzt.

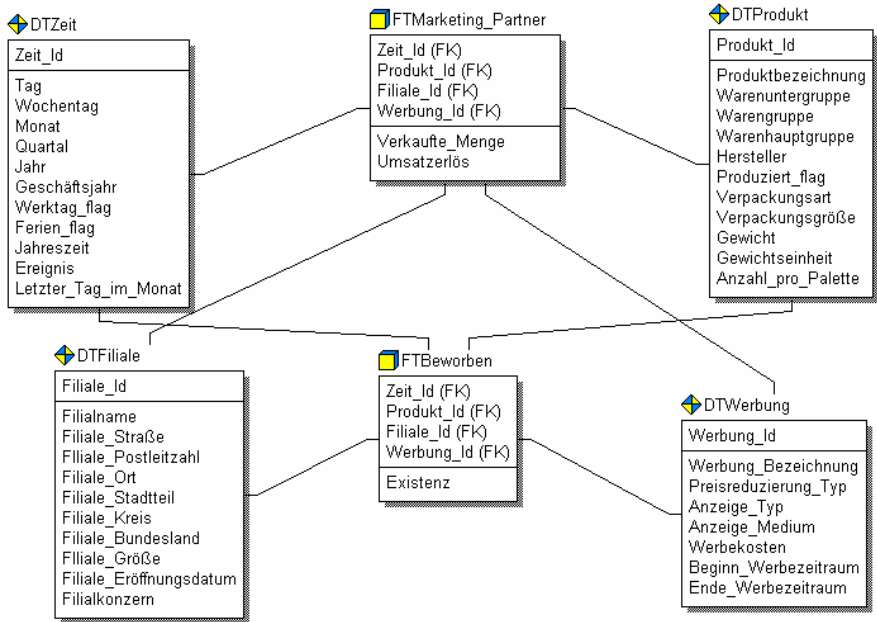
Neben den Ereignissen, die eingetreten sind, gibt es auch Anwendungsfälle, in denen die Fragestellung nach nicht eingetretenen Ereignissen auftritt. Die Bewerbung von Produkten liefert eine Fragestellung nach den trotz Werbemaßnahme nicht verkauften Produkten. Neben die Faktentabelle mit den Fakten verkaufter Produkte tritt dann noch eine Überlagerungstabelle, in welcher die beworbenen Produkte nachgehalten werden. Über eine Differenzmengenbildung ergeben sich die beworbenen nicht verkauften Produkte.

Auch in dem betrachteten Beispiel gibt es einen Anwendungsbereich für Faktenlose Faktentabellen. Im Partner-Marketing treten Fragestellungen bezüglich der Wirksamkeit von Werbemaßnahmen auf. Für Marketing-Strategen sind dabei die folgenden Fragen vorrangig von Interesse:

- Ist der Umsatz beworbener Produkte höher als üblich? (*lift*)
- Fallen die Umsätze der beworbenen Produkte nach Ablauf der Werbemaßnahme wieder auf das vorherige Niveau zurück? (*time shifting*)
- Verdrängt der Mehrumsatz eines beworbenen Produktes den Umsatz eines anderen Produktes? (*cannibalization*)
- Hat die Werbemaßnahme dauerhaften Einfluss in Form größerer Marktanteile?
- Wie ist die Profitabilität?

Für die Analysen bezogen auf die Werbemaßnahmen wird dem Modell eine Dimension Werbung hinzugezogen, die durch die Tabelle DTWerbung repräsentiert wird und die Faktentabelle im Partner-Marketing ist dann zusätzlich mit dieser Tabelle verbunden.

Aber die Information der nicht verkauften jedoch beworbenen Produkte ist damit allein noch nicht ableitbar. Hierzu wird noch die Überlagerungstabelle benötigt, die genauso wie die andere Faktentabelle dimensioniert ist. Das Faktum dieser Tabelle ist ein Artefakt, ein künstliches Faktum, dass nur die Existenz der Schlüsselkombination der Dimensionselemente bezeichnet. Das Modell zusammen mit der normalen Faktentabelle und der Überlagerungstabelle ist in Abbildung 69 dargestellt.



**Abb. 69.** Überlagerungstabelle für das Partner-Marketing

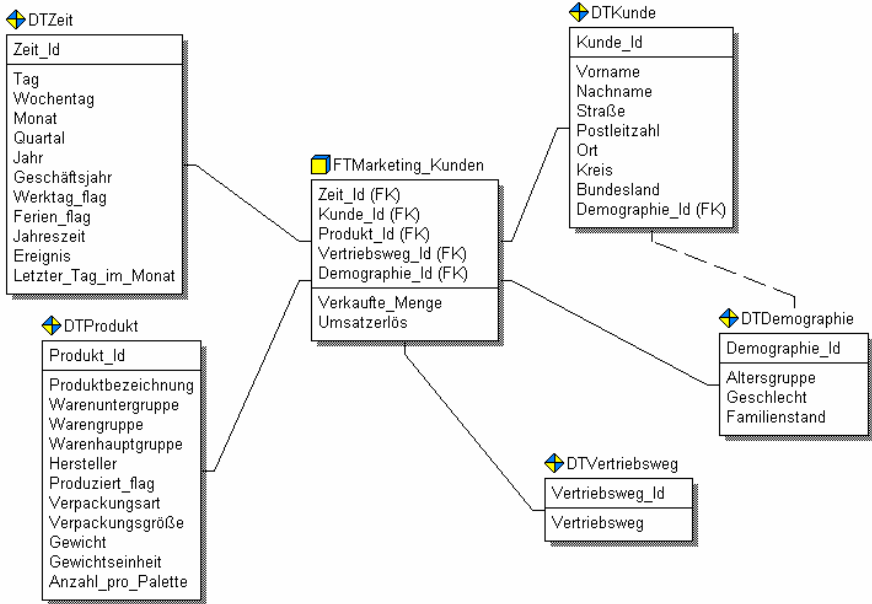
Die meisten der eingangs aufgeführten Fragestellungen können mit diesem Modell beantwortet werden. Aspekte des Marktanteils bedürfen allerdings einer breiteren Datenbasis, in der auch Markt- und Wettbewerberdaten mit einfließen.

### 7.3.3 Minidimensionen

Abfragen an besonders große Dimensionstabellen sind trotz der weitreichenden Optimierungspotentiale auf physischer Modellebene potentieller Engpass für das Antwortzeitverhalten. Eine Designvariante des Star Schemas vermag hier Abhilfe zu schaffen: Attribute in großen Dimensionen mit einer diskreten Wertemenge formen eine weitere Dimension, die Mi-

nidimension genannt wird. Nicht diskrete Attribute können über Bandbreiten bzw. Intervalle ebenfalls in Minidimensionen eingehen.

Insbesondere demographische Attribute in einer Kundendimension sind hervorragende Kandidaten für eine Minidimension. Dies trifft auch für den Bereich des Kunden-Marketings im Anwendungsbeispiel zu und das resultierende Modell für diese Teilsicht ist in Abbildung 70 wiedergegeben.



**Abb. 70.** Modell mit demographischer Minidimension

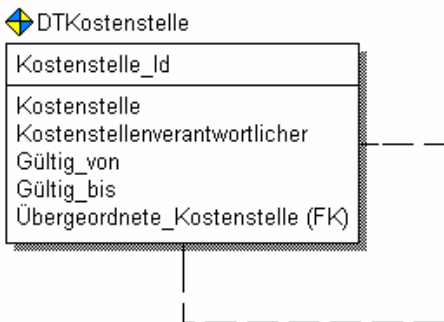
Deutlich erkennbar ist in der Abbildung, dass im Unterschied zu einem herkömmlichen Modell die demographischen Attribute in der Dimensionstabelle durch eine Referenz auf die Minidimension ersetzt werden und diese zusätzlich an die Faktentabelle gebunden ist. Dieser Verweis in der Faktentabelle auf die Minidimension ist somit redundant und könnte auch ausbleiben.

Die sich im Zeitablauf verändernden Attribute in großen Dimensionen sind oftmals die gleichen wie die Kandidaten für die Bildung von Minidimensionen. Für die Historisierung von Strukturbrüchen in diesen Dimensionen können demzufolge die Minidimensionen mit herangezogen werden.

### 7.3.4 Hierarchien über rekursive Beziehungen

Die bisher diskutierte Form der Modellierung von Hierarchien basiert auf dem Grundprinzip, dass jede Hierarchiestufe wie auch jedes Attribut einer Spalte in der Dimensionstabelle entspricht. Einerseits sind mit dieser Vorgehensweise vielfältige Hierarchien modellierbar, aber andererseits geht dies einher mit einer gewissen Unflexibilität etwa beim Einfügen neuer Hierarchiestufen. Alternativ besteht auch die Möglichkeit, die Beziehungen der Dimensionselemente über eine Parent-Child-Relation (rekursive Beziehung) zu modellieren. Sinnvoll ist bei dieser Variante die Nutzung einer zusätzlichen Spalte für die Ebenenzuordnung eines Elementes (*level*). Die Navigation in derart abgebildeten Dimensionen gestaltet sich im Allgemeinen komplexer als in der Implementierung über normale Dimensionstabellen.

Für das entwickelte Anwendungsbeispiel aus Kapitel 5 erfolgte in Kapitel 6 die Modellierung eines semantischen Modells mit ADAPT. In diesem ist in den beiden Hierarchien für die Kostenstelle und die Kostenarten jeweils eine rekursive Beziehung abgebildet. Das Beispiel der Kostenstellenhierarchie dient im Folgenden der Verdeutlichung, denn gerade Kostenstellenhierarchien weisen oftmals eine stark unbalancierte Struktur auf und können daher sehr gut durch rekursive Beziehungen wie in Abbildung 71 skizziert abgebildet werden. Gerade Änderungen in der Dimensionsstruktur, die sich auf die Anzahl der Ebenen oder die Ebenenzugehörigkeit einzelner Elemente auswirken, sind in dieser Form sehr flexibel handhabbar.



**Abb. 71.** Rekursive Beziehung einer Kostenstellenhierarchie

Im skizzierten einfachen Fall sind neben den Attributen wie dem Kostenstellenverantwortlichen und der eigentlichen Kostenstelle noch der Fremdschlüssel für die Beziehung modelliert, der Primärschlüssel ist wieder ein künstlicher Schlüssel. Exemplarisch könnte diese Dimensionstabelle wie in der folgenden Tabelle dargestellt gefüllt sein.

**Tabelle 15.** Dimensionstabelle Kostenstelle

Kostenstelle_Id	Kostenstelle	Übergeordnete_Kostenstelle_Id
1	40	NULL
2	41	1
3	45	1
4	451	3
5	452	3

Die Abfrage auf bestimmte Ebenen innerhalb der Hierarchie oder die Selektion aller Blattelemente ist recht komplex und bedarf im Allgemeinen einer rekursiven Prozedur. Daher erleichtert die Ergänzung der Tabelle um eine Spalte für die Ebene (*level*-Attribut) die Abfragen deutlich.

Auch diese Form der Hierarchie basiert noch auf einer Baumstruktur, jedoch ist dies keine konzeptionelle Bedingung. Vielmehr sind mit dieser Variante auch Heterarchien (m:n-Beziehungen) abbildbar, wie an der beispielhaften Ausprägung in Tabelle 16 deutlich wird.

**Tabelle 16.** Heterarchie in der Dimensionstabelle Kostenstelle

Kostenstelle_Id	Kostenstelle	Übergeordnete_Kostenstelle_Id
1	40	NULL
2	41	1
3	45	1
4	46	1
5	451	3
6	452	3
7	452	4

Zu der Kostenstelle 452 gibt es somit zwei Einträge in der Dimensionstabelle mit der Folge, dass Abfragen und Aggregationen einer besonderen Sorgfalt bedürfen.<sup>13</sup>

<sup>13</sup> Eine eingehende Diskussion der damit verbundenen Problematiken erfolgt in (Hahne 2002).

### 7.3.5 Temporale Aspekte im Star Schema

Bereits in Abschnitt 3.4 wurde die besondere Bedeutung des Zeitbezugs herausgearbeitet. Änderungen der Dimensionsstrukturen im Zeitablauf müssen im Modell gesondert berücksichtigt werden, sofern die Anforderungen des Benutzers an Auswerte- und Analysemöglichkeiten dies erforderlich machen.

Für die Historisierung im Star Schema hat Kimball drei Vorschläge gemacht (Kimball 1996 S. 100ff.):

1. keine Historisierung
2. einzelne Versionen
3. erste und aktuelle Ausprägungen

Fällt die Historisierung weg, werden die einzelnen Attributwerte in der Dimensionstabelle überschrieben, wodurch historische Daten nur noch nach den aktuellen Strukturen auswertbar sind.

Bei der Verfolgung kompletter Historiensätze ist eine Erweiterung der Dimensionstabelle um eine weitere Spalte für die Version notwendig. Da der bisherige Primärschlüssel diese Funktion nicht mehr wahrnehmen kann, ist ein künstlicher Primärschlüssel zu definieren. Tabelle 17 verdeutlicht den Aufbau der Dimensionstabelle für das Beispiel der Produktdimension.

**Tabelle 17.** Dimensionstabelle mit kompletten Versionen

Produkt_AId	Produkt_Id	Produkt	Version	Gruppe
...	...	...	...	...
67	101	PA600	1	High Fidelity
68	102	PAX300	1	High Fidelity
314	102	PAX300	2	Consumer Electronic
69	103	PAX450	1	High Fidelity
...	...	...	...	...

Der Eintrag mit dem künstlichen Schlüsselwert 314 entsteht nach der Änderung des Attributes Gruppe von High Fidelity auf Consumer Electronic. Der alte Satz bleibt in der Tabelle mit der Versionsnummer 1 enthalten, der neue Satz bekommt die Versionsnummer 2 zugewiesen und ist damit als aktuellerer erkennbar. In Ergänzung zu diesem Vorschlag nach Kimball bietet sich die Einführung weiterer Attribute, die den Zeitpunkt der Änderung berücksichtigen, an. Statt einer Versionsnummer kann dann

auch mit Attributen für die Gültigkeitszeit gearbeitet werden, da dies eine größere Flexibilität in der Nachverfolgbarkeit von strukturellen Veränderungen mit sich bringt.

Die angeführte dritte Variante berücksichtigt nicht alle Veränderungen, sondern lediglich den ersten Attributswert sowie den jeweils aktuell gültigen. Die Dimensionstabelle erhält dann die Gestalt wie in Tabelle 18 dargestellt.

**Tabelle 18.** Dimensionstabelle mit zusätzlichem Aktuell-Feld

Produkt_Id	Produkt	Gruppe	Aktuelle_Gruppe
...	...	...	...
101	PA600	High Fidelity	High Fidelity
102	PAX300	High Fidelity	Consumer Electronic
103	PAX450	High Fidelity	High Fidelity
...	...	...	...

Welche der dargestellten Varianten die für eine konkrete Aufgabenstellung angemessene ist, hängt vom jeweiligen Analyse- und Berichtsbedarf der Benutzer ab. Die für diese Entscheidung zugrunde liegenden Berichtsszenarien sind Gegenstand des Abschnittes 3.4 und hier entsprechend zu berücksichtigen. Die größtmögliche Flexibilität offeriert dabei die komplette Versionierung mit Zeitstempelfeldern für die Gültigkeit einer Attributsausprägung.

### 7.4 Weiterführende Literatur

Einer der Klassiker der Data Warehouse Literatur ist (Kimball 1996), in dem die Aspekte der Star Schema Modellierung sehr anschaulich dargestellt sind. Sehr intensiv werden die verschiedenen Modellierungsvarianten in (Hahne 2002) diskutiert. Ebenfalls lesenswert zum Thema sind (Hahne 1999), (Gluchowski 1997), (Nußdorfer 1998a) und (Nußdorfer 1998b). Die Zeitaspekte stehen bei (Stock 2001) im Vordergrund.

Gluchowski P (1997) Data Warehouse-Datenmodellierung – Weg von der starren Normalform. Datenbank-Fokus 11;62-66

Hahne M (1999) Logische Modellierung für das Data Warehouse – Bestandteile und Varianten des Star Schemas. In: Chamoni P, Gluchowski P (Hrsg) Analy-

- tische Informationssysteme – Data Warehouse, On-Line Analytical Processing, Data Mining. Springer-Verlag, Berlin, S 145-170
- Hahne M (2002) Logische Modellierung mehrdimensionaler Datenbanksysteme. Deutscher Universitäts-Verlag, Wiesbaden
- Kimball R (1996) The Data Warehouse Toolkit – Practical Techniques for Building Dimensional Data Warehouses. John Wiley & Sons, New York
- Nußdorfer R (1998a) Star Schema, das E/R-Modell steht Kopf – Teil 1: Modellieren von Faktentabellen. Datenbank-Fokus 10:22-28
- Nußdorfer R (1998b) Star Schema – Teil 1: Modellierung von Dimensionstabellen. Datenbank-Fokus 11:16-23
- Stock S (2001) Modellierung zeitbezogener Daten im Data Warehouse. Deutscher Universitäts-Verlag, Wiesbaden

## 7.5 Zusammenfassung

Hinter dem Begriff Star Schema verbirgt sich eine facettenreiche Vielzahl von Varianten zur Abbildung mehrdimensionaler Datenstrukturen in relationalen Systemen, die sich mittlerweile zu einem Standard entwickelt haben.

In dem grundsätzlichen Ansatz des Star Schemas werden die quantifizierenden Informationen in einer zentralen Tabelle, der Faktentabelle, gehalten. Die Ablage der qualifizierenden Informationen erfolgt in Form von sternförmig angeordneten Dimensionstabellen. Der identifizierende Schlüssel in der Faktentabelle ist dabei der zusammengesetzte Schlüssel bestehend aus den Fremdschlüsseln, welche die Primärschlüssel aller Dimensionstabellen referenzieren.

Während in der Faktentabelle die Bewegungsdaten enthalten sind, beinhalten die Dimensionstabellen die Stammdaten und beschreiben die Bewegungssätze. In ihnen entspricht jede Konsolidierungsebene der Dimensionshierarchie einer eigenen Spalte. Daneben gibt es auch die Möglichkeit der Modellierung von hierarchischen Strukturen über rekursive Beziehungen in der Dimensionstabelle.

Es sind zwei Arten von Attributen zu unterscheiden, die eher technisch orientierten für die Navigation entlang der Ebenen von Konsolidierungshierarchien und die fachlich bestimmten Attribute, die einen Großteil der analytischen Möglichkeiten eines Modells ausmachen. Im Star Schema wird jedes Attribut durch eine eigene Spalte in der Dimensionstabelle repräsentiert und eine Unterscheidung zwischen Hierarchiestufen und Attributen ist nicht ohne ergänzende Information gegeben.

Erfolgt eine Normalisierung der Dimensionstabellen, so führt dies zu der Modellierungsvariante des Snow Flake Schemas. Dieses ist eine Kom-



bination der Normalisierung mit zusätzlich aufgebauten Aggregattabellen zur Speicherung verdichteter Teilsichten. Ein Star Schema mit zusätzlichen Aggregattabellen wird auch als Fact Constellation Schema bezeichnet.

Eine nahe liegende Verallgemeinerung des Star Schemas liegt in der Trennung in Faktentabellen, in denen jeweils nur Fakten gleicher Dimensionierung gespeichert werden. Darüber hinaus ist durch diese Trennung auch der semantische Zusammenhang besser abbildbar. Im Extremfall ist jedes Fakt in einer eigenen Tabelle abgebildet. Diese Form ist mit dem Begriff der Galaxie umschrieben.

Faktentabellen, die gar keine abzuspeichernde Kennzahl beinhalten, heißen auch Faktenlose Faktentabelle (*Factless Fact Table*) und dienen der Verfolgung von Ereignissen. Neben den Ereignissen, die eingetreten sind, gibt es auch Anwendungsbereiche, in denen die Fragestellung nach nicht eingetretenen Ereignissen auftaucht.

Für die Berücksichtigung der Zeitabhängigkeit im Star Schema gibt es die drei Möglichkeiten der Bildung einzelner Versionen zur kompletten Historienbildung, die Schaffung zusätzlicher Attributspalten für die jeweils aktuellen neben den originalen Ausprägungen sowie die dritte Möglichkeit des Verzichts auf jegliche Historisierung.

## 8 Logisches Datenmodell des BW

Die Datenhaltung des Business Information Warehouse basiert auf der relationalen Datenbanktechnik und die Speicherung mehrdimensional strukturierter Informationen folgt dem Ansatz eines Star Schemas, dessen Darstellung im vorigen Kapitel erfolgte. Das erweiterte Star Schema im BW ist durch eine große Anzahl von Tabellen gekennzeichnet, über welche die Informationen verteilt abgelegt sind, und ist gegenüber klassischen Ansätzen aufgrund dieser Verteilung deutlich komplexer.

Im folgenden Abschnitt 8.1 steht der grundsätzliche Aufbau des BW-Modells im Vordergrund. Die variantenreiche Abbildung von Dimensionsstrukturen ist Gegenstand von Abschnitt 8.2 an dessen Darstellung sich die Abbildung zeitabhängiger Berichtsszenarien anschließt. Nach der Diskussion der Modellierung von Kennzahlen und der Möglichkeiten der grafischen Repräsentation von BW-Modellen steht die Ableitung eines BW-Modells aus einem semantischen T-ADAPT-Modell im Vordergrund.

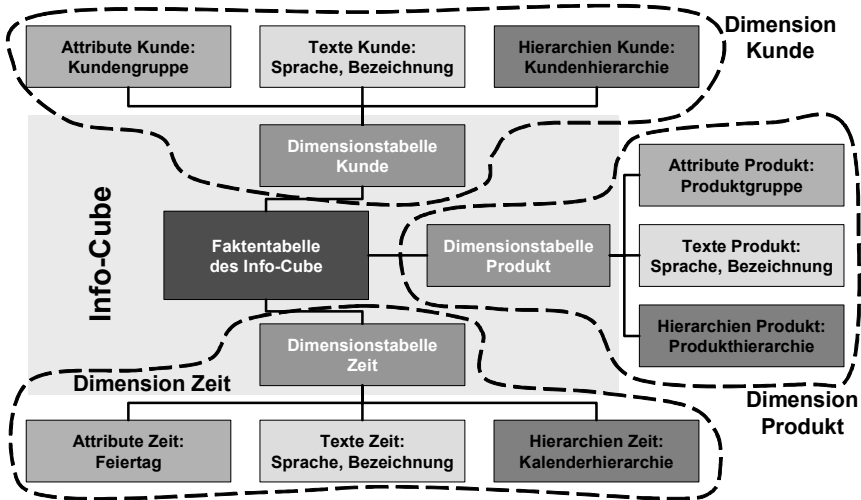
### 8.1 Erweitertes Star Schema der SAP

Der Info-Cube ist das zentrale Objekt zur Speicherung mehrdimensional strukturierter Daten im BW. Dass die Stammdaten übergreifend gültig sind und über alle Info-Cubes hinweg zur Verfügung stehen, ist dabei ein wesentliches Grundprinzip.

Die Faktentabelle, deren Wertspalten im BW auch als Key Figures (z.B. Umsatz oder Absatz) bezeichnet werden, bildet auch im erweiterten Star Schema den Mittelpunkt des Modells. Sie ist von Dimensionen umgeben, die jeweils aus der Dimensionstabelle und den Stammdatentabellen bestehen. Exemplarisch ist dies in Abbildung 72 für ein Modell mit den drei Dimensionen Kunde, Produkt und Zeit dargestellt. Dabei fungiert die Dimensionstabelle als Verbindungstabelle von der Faktentabelle zu den Stammdatentabellen, in denen sich die eigentlichen Dimensionswerte mit beschreibenden Texten sowie Attributen befinden.

Im Business Information Warehouse sind pro Info-Cube maximal 16 Dimensionen definierbar, wobei bereits drei vom System vorgegeben sind.

Neben der Zeitdimension sind dies die eher technisch ausgerichteten Dimensionen für Pakete und Einheiten, in deren Fokus die Berücksichtigung von einzelnen Ladeprozessschritten einerseits und die Implementierung von Währungen bzw. Einheiten andererseits liegen.



**Abb. 72.** Erweitertes Star Schema im SAP BW

Die elementaren Objekte im Business Information Warehouse sind die global definierten und über ihren technischen Namen eindeutig identifizierbaren Info-Objekte. Die im BW-Modell differenzierten betriebswirtschaftlichen Auswertungsobjekte sind zum einen die Merkmale, die den allgemeinen Dimensionselementen entsprechen, und die Kennzahlen, mit denen die Fakten beschrieben werden. Beide sind technisch gesehen Info-Objekte.

Merkmale als Dimensionselemente sind oftmals als numerischer Schlüssel definiert, beispielsweise die Artikelnummer aus einem vorgelagerten ERP-System wie z. B. R/3, zu deren weiterer Beschreibung ein Bezeichner gehört, beispielsweise der Artikelname. Die Ablage dieser beschreibenden Texte von Merkmalen erfolgt im Business Information Warehouse in einer separaten Texttabelle (siehe Abbildung 73), wobei die Texte in verschiedenen im System definierten Sprachen verfasst sein können. Beispielsweise sind zu einem Info-Objekt, das einen Länderschlüssel darstellt, mehrsprachige Landesbezeichnungen sinnvoll. Das Business Information Warehouse unterscheidet kurze, mittellange und lange Texte zur Beschreibung von Info-Objekt-Werten.

Land	Sprache	TXTSH	TXTMD	TXTLG
GER	DE	Deutschland	Deutschland	Deutschland
GER	EN	Germany	Germany	Germany

Abb. 73. Ablage von Texten im erweiterten Star Schema

Die Berücksichtigung von Dimensionsattributen ist im erweiterten Star Schema über Stammdatentribute eines Merkmals gewährleistet. Diese werden in einer separaten Stammdatentabelle eines Merkmals gespeichert. Im BW-Sprachgebrauch wird allgemein von Attributen (z. B. Materialgruppe) gesprochen, wobei zwischen reinen Anzeigeattributen und Navigationsattributen zu differenzieren ist. Bei der Definition eines Info-Objektes wird festgelegt, ob überhaupt Attribute definierbar sind. Ist dies der Fall, wird von einem stammdatentragenden Merkmal gesprochen. Sowohl Anzeige- als auch Navigationsattribute beziehen sich auf ein stammdatentragendes Merkmal, aber nur Navigationsattribute sind losgelöst von diesem in Berichten darstellbar. Anzeigeattribute als ergänzende Information zu einem Merkmal können nur zusammen mit diesem in Berichten verwendet werden.

Die Stammdatentabellen sind nicht direkt an die Info-Cubes gebunden, denn die Verbindung erfolgt über die auf künstlichen Schlüsseln basierenden SID-Tabellen (siehe Abbildung 74). Hierdurch können auch Merkmale auf Basis einer m:n-Beziehung wie z. B. Material und Farbe in einer Dimension abgebildet werden. Die Modellierung von m:n-Beziehungen in einer Dimension ist zwar technisch möglich, doch oftmals wenig sinnvoll.

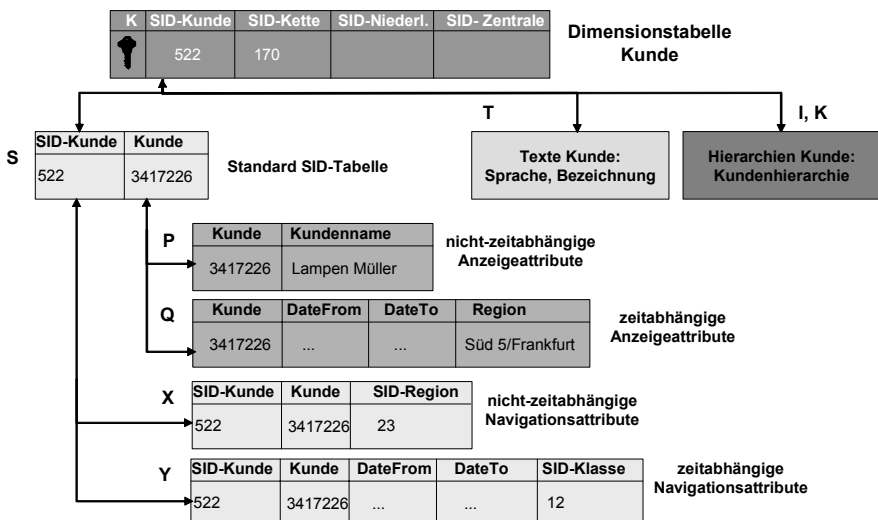
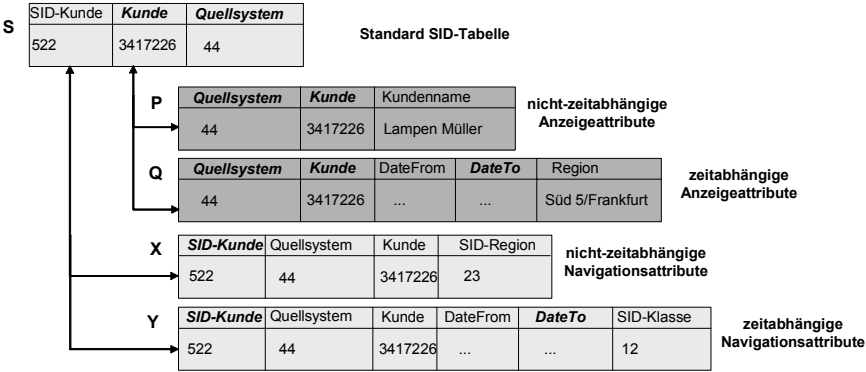


Abb. 74. Verschiedene Stammdatentabellen im BW

Im Business Information Warehouse ist es möglich, alle Stammdaten als zeitabhängig zu deklarieren. Dadurch erhalten Attribute, Texte oder Hierarchien die zusätzliche Information, von wann bis wann sie genau gelten. Für Abfragen an Modelle mit zeitabhängigen Objekten ergibt sich die Besonderheit, dass in der Abfrage festgelegt werden muss, welcher Stichtag für die Selektion herangezogen werden soll. Dieser wird als Query Key Date bezeichnet und ist innerhalb einer Abfrage für alle zeitabhängigen Objekte gleichermaßen gültig. Diese Zuordnung der Gültigkeitszeit erfolgt vom System eindeutig während der Ladevorgänge zum Befüllen des Data Warehouse. Die Verknüpfung von Info-Cubes mit zeitabhängigen und nicht zeitabhängigen Stammdaten erfolgt über getrennte SID-Tabellen.

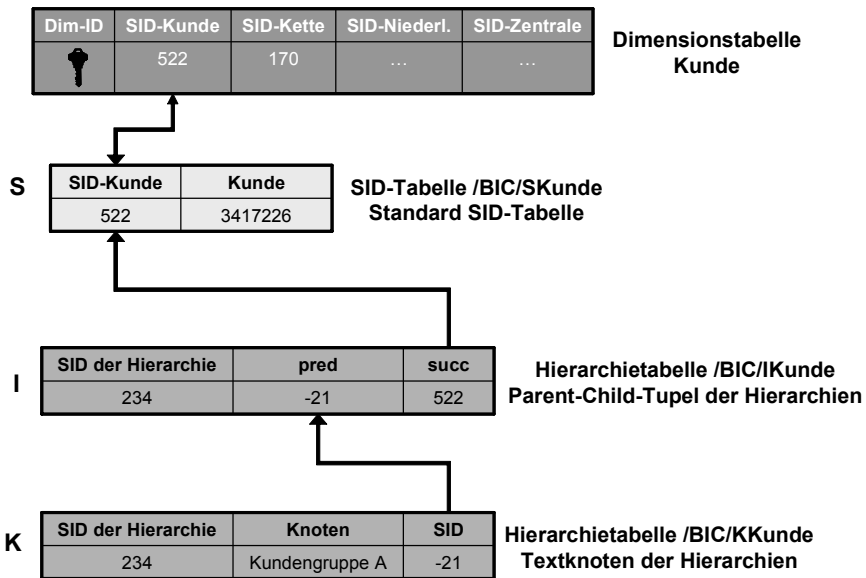
Bei geklammerten Info-Objekten besteht der Primärschlüssel aus mehreren Komponenten. Übertragen auf das Beispiel des Kunden ist es möglich, zur Unterscheidung der Stammdaten verschiedener Systeme das liefernde Quellsystem in Form der Klammerung zur eindeutigen Kennzeichnung mit aufzunehmen. Dieser zusammengesetzte Schlüssel überträgt sich dann entsprechend auf die einzelnen Stammdatentabellen wie in Abbildung 75 in der Übersicht dargestellt. Dort sind ebenfalls die Primärschlüssel der einzelnen Tabellen gesondert hervorgehoben.



**Abb. 75.** Stammdatentabellen bei geklammerten Info-Objekten

Hierarchien eines Merkmals wie beispielsweise eine Länderhierarchie können in separaten Hierarchietabellen auf Basis der Parent-Child-Informationen gespeichert werden. Diese Hierarchien werden *Externe Hierarchien* genannt und stellen vordefinierte Konsolidierungspfade dar. Die dazu verwendeten Tabellen im BW und deren Verknüpfung ergibt sich aus Abbildung 76. Die Benutzung des Begriffes Hierarchie ist im BW potenziell missverständlich, denn im normalen Verständnis ist eine Hierarchie eine Sequenz von 1:n-Beziehungen zwischen Merkmalen. Demzufolge

gibt es Hierarchien sowohl in der Dimensionstabelle über Merkmale, in der Stammdatentabelle über Attribute als auch in der eigentlichen Hierarchietabelle.



**Abb. 76.** Hierarchietabellen

Der Begriff der Dimensionstabelle ist im erweiterten Star Schema mit einer besonderen Bedeutung versehen, denn sie bildet den Verknüpfungspunkt von der Faktentabelle zu den abgekoppelten Stammdatentabellen, in denen die eigentlichen Dimensionswerte und deren beschreibende Attribute abgelegt sind. Die Ablage der Informationen, die in Dimensionstabellen eines allgemeinen Star Schemas abgelegt sind, erfolgt im BW in den so genannten Stammdatentabellen, die eigentliche Dimensionstabelle im BW bildet nur die Schnittstelle von der Faktentabelle zu den verschiedenen Tabellen, in denen die eigentlichen Dimensionsinformationen abgelegt sind. Die Line-Item Dimension stellt eine spezielle in Abbildung 77 dargestellte Variante einer Dimension dar, bei der die Verknüpfungstabelle entfällt und die Stammdatentabelle eines einzelnen Merkmals direkt an die Faktentabelle gebunden ist.

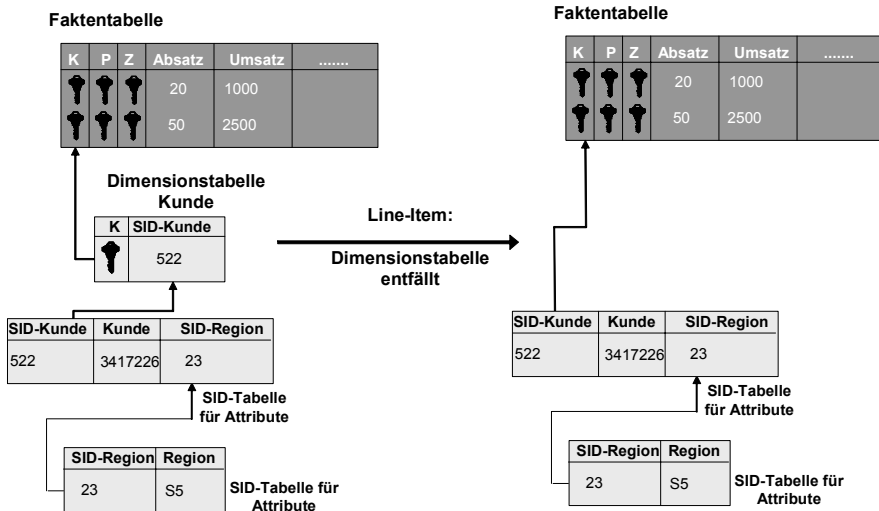


Abb. 77. Line-Item Dimension

Der Zugriff auf das Modell bei konkreten Abfragen erfolgt immer von den Dimensionstabellen ausgehend und je nach konkreter Fragestellung dienen unterschiedliche Stammdatentabellen zur näheren Eingrenzung und Bestimmung der qualifizierten Tupel aus der eigentlichen Dimensionstabelle, über die der Join mit der Faktentabelle durchgeführt wird. Bei der Analyse der Abbildungsmöglichkeiten der verschiedenen zeitabhängigen Berichtsszenarien in Abschnitt 8.3 wird auf die unterschiedlichen Abfragewege etwas detaillierter eingegangen. Doch zunächst erfolgt eine ausführliche Darstellung der Varianten zur Gestaltung von Dimensionsstrukturen.

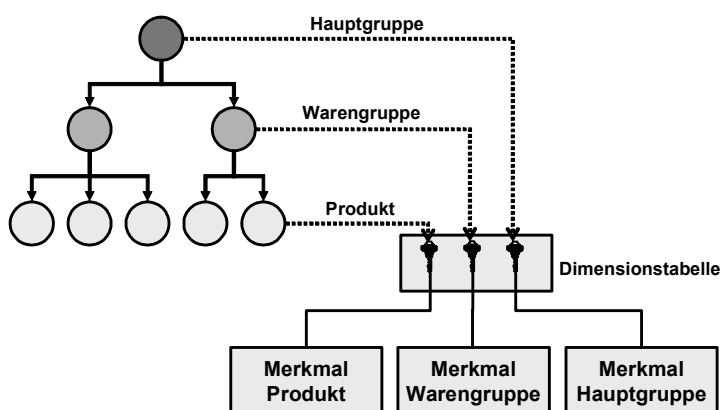
## 8.2 Modellierungsvarianten hierarchischer Dimensionsstrukturen

In mehrdimensionalen Modellen definieren die abgebildeten Dimensionsstrukturen die Grundlage für analytische Operationen des On-Line Analytical Processing. Neben verschiedenen Aspekten der Darstellung statischer Dimensionsstrukturen spielt für die Berücksichtigung von Berichtsanforderungen insbesondere der Umgang mit Veränderungen innerhalb von hierarchischen Strukturen eine Rolle. Die Bedeutung dieser temporalen Aspekte wurde schon in den vorangegangenen Darstellungen in den Abschnitten 3.4 und 6.3 herausgearbeitet und diese sind daher schon bei der Modellierung besonders mit zu berücksichtigen.

### 8.2.1 Hierarchische Beziehungen zwischen Merkmalen

Im Business Information Warehouse sind die Daten konsequent in die zwei Bereiche der Bewegungsdaten und der Stammdaten getrennt, die über jeweils unterschiedliche Ladeprozesse in das BW transportiert werden. Die Bewegungsdaten beziehen sich immer auf einen Info-Provider bzw. Info-Cube, die geladenen Stammdaten stehen jedoch Info-Cube übergreifend zur Verfügung.

Bei der Modellierung einer Dimensionshierarchie über Merkmale in einer Dimension wird jede Konsolidierungsebene der Hierarchie auf ein Merkmal abgebildet, die Hierarchie ist somit in den Bewegungsdaten abgebildet. Ein typisches Beispiel hierfür im BW ist eine Zeithierarchie mit den Merkmalen Tag, Monat und Jahr oder eine Hierarchie über die Merkmale Produkt, Warengruppe und Hauptgruppe (siehe Abbildung 78). Die Merkmale in einer Dimension qualifizieren eine Wertinformation in einem Bewegungsdatensatz.



**Abb. 78.** Hierarchische Beziehungen über Bewegungsdaten

Bei dieser Art der Hierarchiemodellierung werden nicht alle möglichen Strukturtypen unterstützt. Da jede Konsolidierungsstufe mit einem Merkmal korrespondiert, sind balancierte Strukturen eine Voraussetzung, die nur mit besonderer Kenntnis beim Anwender umgangen werden kann. Da die hierarchische Information in den Bewegungsdaten abgebildet ist und in den Dimensionen ein künstlicher Schlüssel die Merkmale einer Dimension zusammenfasst, ist es sogar möglich, die Daten auf einer Konsolidierungsebene statt für ein Basiselement zu laden. Diese Beziehung z. B. zwischen Produkt, Warengruppe und Hauptgruppe ist im BW technisch nicht gepflegt, alle Merkmale stehen gleichberechtigt nebeneinander, und es gibt



daher auch keine vordefinierten Navigationspfade. Bei einem Drill können somit einzelne Ebenen auch übersprungen werden. Dies ist seit der Version 3.0 des Business Information Warehouse zumindest in der Darstellung im Business Explorer Analyzer, dem Excel-basierten OLAP Analysewerkzeug, das mit dem BW ausgeliefert wird, anders darstellbar, da die Möglichkeit der Gruppierung von Merkmalen zu einer Anzeigehierarchie mittlerweile möglich ist.

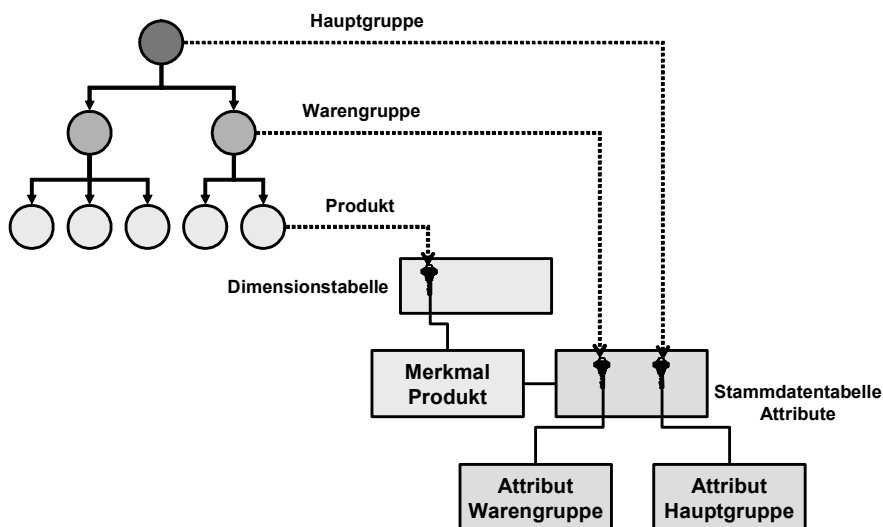
Strukturelle Änderungen in der hierarchischen Beziehung zwischen Merkmalen sind bei dieser Abbildungsform ohne Datenreorganisation nicht abbildbar, das Berichtswesen kann nur auf Basis der tatsächlich gebuchten Zuordnung, d. h. der historischen Wahrheit, erfolgen.

### **8.2.2 Navigationsattribute als Basis für hierarchische Strukturen**

Ein nicht unerheblicher Aspekt beim Aufbau eines Data Warehouse ist die Frage nach den Veränderungen in Dimensionshierarchien im Zeitablauf. Zum einen gibt es die Anforderung, verschiedene Strukturvarianten berücksichtigen zu können. Diese wird im Zusammenhang des Business Information Warehouse unter dem Stichwort *Zeitabhängigkeit* diskutiert. Zum anderen kann minimal gefordert werden, dass eine Reorganisation des Datenbestandes zur Anpassung an veränderte Strukturen erfolgt. Sind die einzelnen Konsolidierungsebenen Merkmale in den Bewegungsdaten, ist eine Reorganisation nur mit einem Neuaufbau des Info-Cubes abbildbar und daher sehr aufwendig. Das ist ein Grund, warum die Modellierung in den Stammdaten einen Vorteil darstellt. Dabei korrespondiert jede Konsolidierungsebene außer der untersten mit einem Navigationsattribut zu dem Merkmal der Basiselemente (siehe Abbildung 79). Im Extremfall besteht eine Dimension dann nur aus einem einzigen Merkmal und alle hierarchischen Informationen sind in den Stammdaten dieses Merkmals abgebildet. In diesem Fall ist der Vorteil einer Line-Item Dimension nutzbar, d. h. die Dimensionstabelle als Transfertabelle zwischen einem künstlichen Dimensionsschlüssel und den SIDs der Merkmale entfällt und die Anbindung an die Faktentabelle erfolgt direkt über die SID-Tabelle.

Die Ablage der Hierarchieinformation in den Stammdaten hat zur Konsequenz, dass diese allen Info-Cubes gleichermaßen zur Verfügung steht und eine Veränderung leicht möglich ist, da nur Stammdatenänderungen durchgeführt werden müssen, diese aber auch für alle Info-Cubes greifen. Eine Analyse auf Konsolidierungsebenen entsprechend dieser Navigationsattribute erfolgt dann nach der geänderten Struktur für alle Werte in den betroffenen Info-Cubes. Navigationsattribute bieten ebenso wie hie-

rarchische Beziehungen über Merkmale keine vordefinierten Konsolidierungspfade, so dass bei der Navigation Ebenen übersprungen werden können und der Anwender von den Attributen, welche die Hierarchie formen, Kenntnis haben muss.



**Abb. 79.** Navigationsattribute als Grundlage für Dimensionsstrukturen

Mit der Modellierungsalternative der Hierarchieabbildung über Navigationsattribute steht in dieser Form ein Berichtswesen auf Basis der jeweils aktuellen Dimensionsstruktur zur Verfügung, eine Auswertung auf Basis historischer Zuordnungen kann erst mit einer zeitabhängigen Ausgestaltung der Navigationsattribute abgebildet werden. Bei zeitabhängigen Attributen ist zu beachten, dass die Möglichkeit der Bildung von Aggregaten zur Performancesteigerung auf diesen Attributen nur sehr eingeschränkt möglich ist, denn es bedarf der Angabe eines Stichtages, für welches das Aggregat gültig ist.

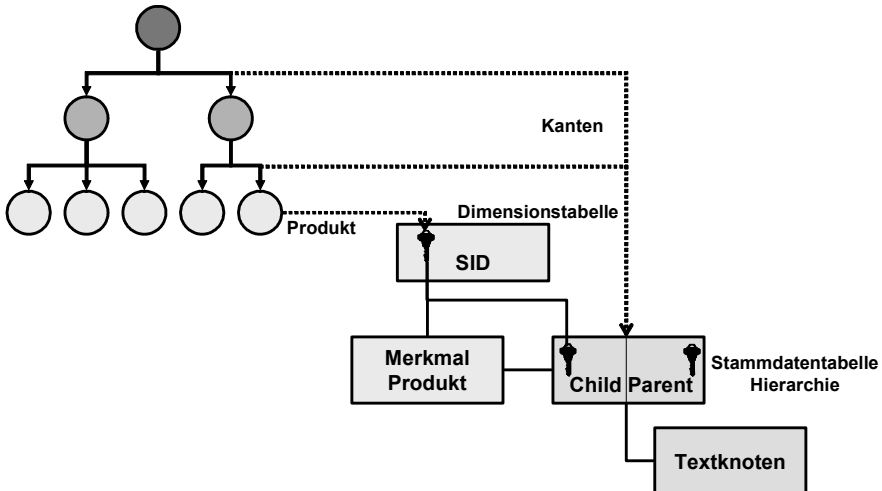
### 8.2.3 Externe Hierarchien in den Stammdaten

Die bzgl. der Dimensionsstruktur flexibelste Art der Modellierung von hierarchischen Strukturen in Dimensionen im BW stellen die Externen Hierarchien dar, da sie auf einer Darstellung in Form von rekursiven Beziehungen basieren. Sie bietet sich daher insbesondere bei unbalancierten Dimensionsstrukturen an. Externe Hierarchien sind in den Stammdaten abgelegt und somit für alle Info-Cubes, die ein spezielles Merkmal ver-

wenden, übergreifend gültig. Neben der Möglichkeit, verschiedene Hierarchien für ein Merkmal zu definieren, können einzelne Hierarchien zusätzlich in verschiedenen Versionen gepflegt sein.

Die Elemente Externer Hierarchien sind die Knoten, die ggf. Vorgänger und Nachfolger haben können. Elemente ohne Vorgänger sind Wurzelemente, die ohne Nachfolger werden Blattelemente genannt. Die Bausteine einer Externen Hierarchie sind zum einen die Merkmalsknoten und zum anderen die Textknoten, wobei nur zu Merkmalsknoten Daten in das BW geladen werden können. Merkmalsknoten werden aus den Ausprägungen eines Info-Objektes aufgebaut, Textknoten sind frei definierbare Elemente einer Hierarchie. Für die Anzeige werden die Texte der Merkmalsknoten aus den zugrunde liegenden Stammdaten des Merkmals gezogen, zu Textknoten wird der Anzeigetext direkt in der Hierarchiepflege gespeichert.

Die Externen Hierarchien eines Merkmals sind Bestandteil der Stammdaten und die Ablage erfolgt in mehreren Tabellen. Zentral ist die Tabelle, in der die rekursiven Beziehungen gespeichert sind. In dieser *Inclusion Table* genannten Tabelle werden die Beziehungen paarweise über Schlüssel abgelegt. Die Merkmalsknoten werden über den SID-Schlüssel des zugrunde liegenden Merkmals identifiziert, Textknoten bekommen einen identifizierenden Schlüssel mit negativem Vorzeichen, deren Texte in einer eigenen Tabelle gespeichert sind (siehe Abbildung 80).



**Abb. 80.** Dimensionsstrukturen auf Basis Externer Hierarchien

Die Identifikation von Konsolidierungsebenen in dieser Form einer Hierarchie erfolgt im BW nur auf Basis der Ebene, d. h. der Anzahl der Kanten ausgehend von der Wurzel bis zum Dimensionselement. Die Abbil-

dung von Dimensionsstrukturen, in denen die Bedingung einer 1:n-Beziehung zwischen den einzelnen Konsolidierungsstufen fallen gelassen wird, ist in der Form möglich, dass ein Knoten zu mehreren übergeordneten Elementen zugeordnet werden kann und die Gewährleistung der konsistenten Verdichtung im BW automatisch erfolgt.

Die Knoten einer Externen Hierarchie, die nicht Blattelement sind, werden im Allgemeinen als Textknoten definiert und nur die Blattelemente sind auf Basis eines Merkmals festgelegt. Für die Blattelemente ergeben sich damit die Texte aus den Stammdaten des zugrunde liegenden Info-Objektes. Aber auch für die übergeordneten Ebenen in einer Externen Hierarchie können Info-Objekte Verwendung finden, indem über fremde Merkmalsknoten einzelne Hierarchiestufen ebenfalls auf Basis von Info-Objekten aufgebaut werden. Dies hat zur Folge, dass die beschreibenden Texte dieser Knoten aus den entsprechenden Stammdaten ausgelesen werden und somit unabhängig von dem tatsächlichen Hierarchieaufbau immer den aktuellen Stammdaten entsprechen.

In Externen Hierarchien ist zwischen bebuchbaren und nicht bebuchbaren Knoten zu differenzieren. Erstere werden definiert durch die Knoten, die auf dem zugrunde liegenden Merkmal basieren. Die nicht bebuchbaren Knoten sind die Textknoten und fremde Merkmalsknoten. Die Anordnung beider Formen von Knoten innerhalb einer Externen Hierarchie ist völlig frei, so dass es auch möglich ist, nicht bebuchbare Knoten als Kindelement zu bebuchbaren Knoten festzulegen.

Eine weitere Besonderheit stellen die bebuchbaren Knoten basierend auf Hierarchieintervallen des Basismerkmals dar. Durch diese Festlegung werden alle Knoten, auch neu hinzukommende, automatisch zu den korrespondierenden Intervallknoten aggregiert.

Externe Hierarchien werden oft auch als Hierarchien im eigentlichen Sinne bezeichnet und über verschiedene Versionen von Hierarchien lassen sich schon Berichtsanforderungen abdecken, die unterschiedliche definierte Dimensionsstrukturen abbilden. Durch die zeitabhängige Modellierung von Hierarchien lässt sich dies auch direkt an einem Datum, dem so genannten Schlüsseldatum oder Query Key Date, festmachen. Dieses ist dann in einer Abfrage für alle zeitabhängigen Strukturbestandteile gültig. Für die zeitabhängige Modellierung von Hierarchien stehen die zwei Formen der Zeitstempelung der gesamten Hierarchie wie auch einzelner Hierarchieknoten zur Verfügung.

### 8.3 Abbildung zeitabhängiger Berichtsszenarien

Die Berücksichtigung temporaler Aspekte in analytischen Systemen ist gerade aus Anwendersicht essentiell. An dieser Stelle wird das bereits in Abschnitt 3.4 ausgiebig diskutierte Beispiel wieder aufgegriffen und auf das SAP Business Information Warehouse übertragen.

Dieses betrachtet den in Abbildung 21 dargestellten Fall einer strukturellen Änderung in einer Produktdimension. Für die Darstellung der Abfragewege im Business Information Warehouse und die Implementierung der verschiedenen Berichtsszenarien dienen von nun an die zwei abstrahiert dargestellten Dimensionen wie in Abbildung 81 aufgeführt, aus der ebenfalls die zugrunde liegende vereinfacht dargestellte Faktentabelle abzulesen ist.

**Produktdimension in 2003-04**

Produkt	Produktgruppe
P1	PG1
P2	PG1
P3	PG1
P4	PG2
P5	PG2

**Produktdimension in 2003-05**

Produkt	Produktgruppe
P1	PG1
P2	PG1
P3	PG2 (geändert)
P4	gelöscht
P5	PG2
P6	PG2 (neu)

**Faktentabelle**

Produkt	Periode	Umsatz
P1	2003-04	110
P2	2003-04	120
P3	2003-04	130
P4	2003-04	140
P5	2003-04	150
P1	2003-05	110
P2	2003-05	120
P3	2003-05	130
P5	2003-05	150
P6	2003-05	160

**Abb. 81.** Beispiel zur Analyse der Berichtsszenarien

Auf dieses Beispiel sind nun die vier Szenarien der Analyse nach der aktuellen Struktur, gemäß der alten Struktur, analog der historischen Wahrheit und der vergleichbaren Resultate anzuwenden.

#### 8.3.1 Aktuelle Struktur

Das erste auf das BW zu übertragende Szenario berücksichtigt für die Aggregation von Produkten zu Produktgruppen die jeweils gerade aktuelle Struktur. In dem Beispiel wurden nur zwei Zeitperioden betrachtet. Diese

stehen in den folgenden Ausführungen für die zwei Perioden 2003-04 und 2003-05 und somit ist letztere die aktuelle Struktur. Das in Abbildung 82 aufgeführte Berichtsergebnis bezogen auf die aktuelle Struktur entspricht dem Inhalt von Tabelle 3 in Abschnitt 3.4.

Produktdimension in 2003-05		Faktentabelle		
Produkt	Produktgruppe	Produkt	Periode	Umsatz
P1	PG1	P1	2003-04	110
P2	PG1	P2	2003-04	120
P3	PG2 (geändert)	P3	2003-04	130
P4	gelöscht	P4	2003-04	140
P5	PG2	P5	2003-04	150
P6	PG2 (neu)	P1	2003-05	110
		P2	2003-05	120
		P3	2003-05	130
		P5	2003-05	150
		P6	2003-05	160

Produktgruppe	Umsatz 2003-04	Umsatz 2003-05
PG1	230	230
PG2	280	440

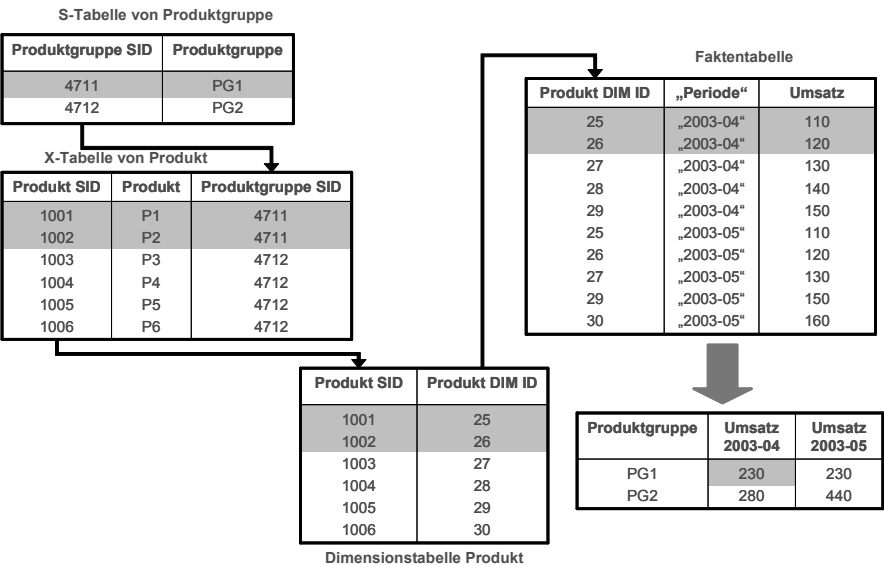
**Abb. 82.** Berichtsszenario „aktuelle Struktur“

Für die Implementierung dieses Szenarios im Business Information Warehouse stehen die beiden Varianten der Modellierung von Navigationsattributen und der Gestaltung von Externen Hierarchien zu Verfügung.

Bei dem ersten Implementierungsweg, wie in Abbildung 83 dargestellt, beginnt die Abfrage bei der S-Tabelle der Produktgruppe und geht mit dem resultierenden SID-Schlüssel der Produktgruppe in die nicht-zeitabhängige Attributtabelle des Produktes. Dies ist in der Darstellung exemplarisch für die Produktgruppe PG1 hervorgehoben gekennzeichnet. Diese hat den SID-Schlüssel 4711.

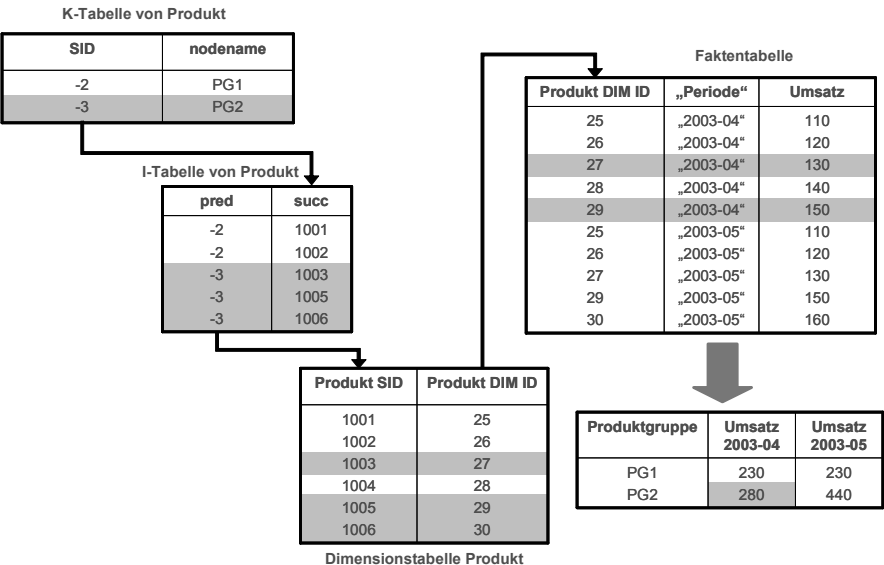
Mit den passenden SID-Schlüsseln des Produktes, im Beispiel sind dies 1001 und 1002, ergibt sich aus der Dimensionstabelle, welche Dimensions-IDs für die Abfrage an die Faktentabelle zu berücksichtigen sind. Die Perioden-Spalte in der Faktentabelle ist dabei aus Gründen der Komplexitäts-Reduktion abstrahiert dargestellt, da diese natürlich auch nur SID-Schlüsselwerte beinhaltet.

So ergibt sich wie ausgeführt der Wert 230 für die Produktgruppe PG1 in Periode 2003-04 als die Summe der Werte der beiden markierten Einträge in der Faktentabelle. Nach der aktuellen Struktur gehören ja nur die beiden Produkte P1 und P2 zu der Produktgruppe PG1.



**Abb. 83.** Aktuelle Struktur über Navigationsattribute

Bei der zweiten Implementierungsalternative in Form einer Externen Hierarchie ergibt sich die Situation wie in Abbildung 84 dargestellt.



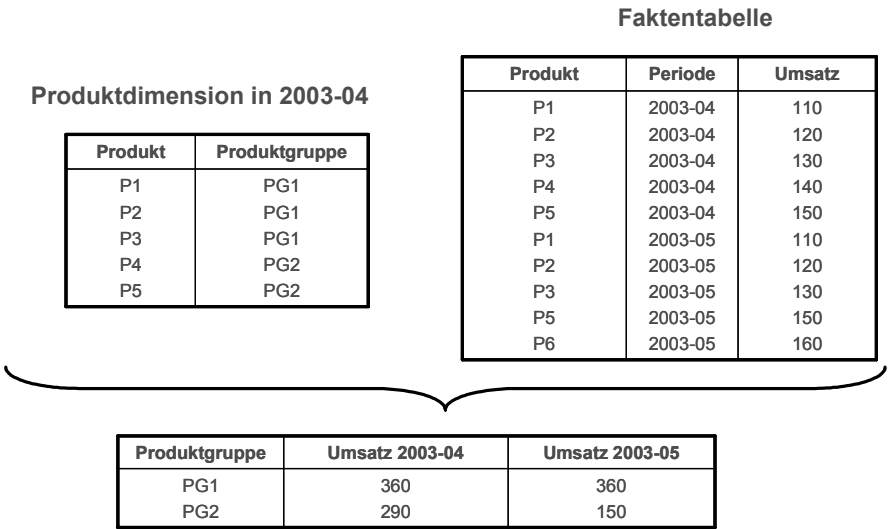
**Abb. 84.** Aktuelle Struktur über Externe Hierarchie

Die beiden Produktgruppen sind dabei als Textknoten gepflegt und stehen in der K-Tabelle der Hierarchie mit negativen SID-Schlüsseln. In der Parent-Child-Tabelle bilden diese zusammen mit den SID-Schlüsseln der Produkte die Über- bzw. Unterordnungsbeziehungen. Für das in der Abbildung besonders gekennzeichnete Beispiel der Abfrage der Produktgruppe PG2 ist der SID-Schlüssel dieses Textknotens -3, der gemäß der I-Tabelle die untergeordneten Elemente 1003, 1005 und 1006 besitzt. Mit diesen ergeben sich die für die Abfrage an die Faktentabelle zu berücksichtigenden Dimensions-IDs aus der Dimensionstabelle.

Bei dieser diskutierten Berichtsdarstellung findet die aktuell gültige Dimensionsstruktur auf die Daten aller Perioden Anwendung. Bei dem zweiten zu behandelnden Szenario soll hingegen eine historische bzw. alte Struktur auf die neuen Daten angewendet werden.

### 8.3.2 Historische Struktur

Dieses Szenario ist in Abbildung 85 zusammengefasst und korrespondiert inhaltlich mit der in Tabelle 2 in Abschnitt 3.4 dargestellten Situation. In unserem Beispiel erfolgt die Analyse auf Ebene der Produktgruppen anhand der Zuordnungen aus der Periode 2003-04, die sich dadurch unterscheidet, dass P3 noch zu der Produktgruppe PG1 zugeordnet ist, das Produkt P4 noch nicht gelöscht ist und Produkt P6 noch nicht vorhanden ist.

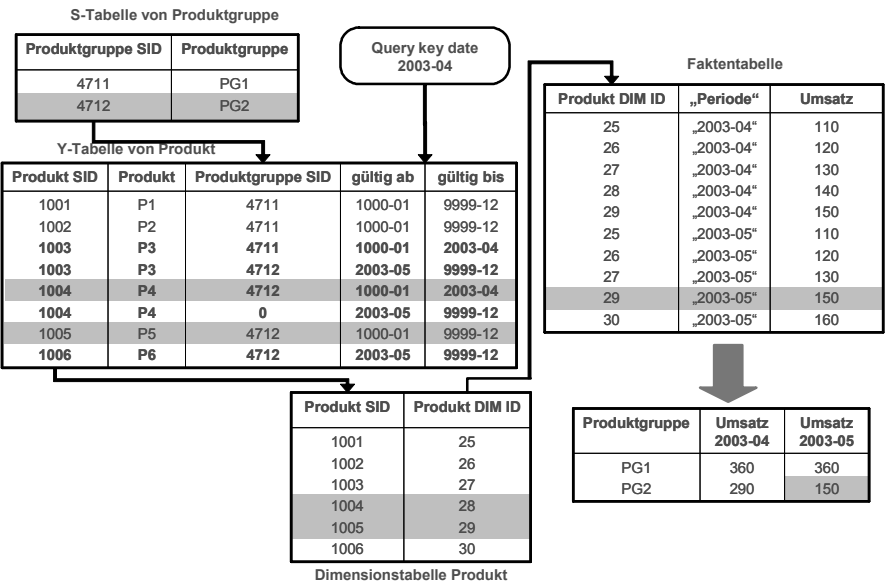


**Abb. 85.** Berichtsszenario „historische Struktur“



Wie auch bei dem ersten Szenario handelt es sich bei dieser Analyse für das Business Information Warehouse um eine stammdatenbezogene Sichtweise. Demzufolge kommen hier wieder die zwei Varianten der Implementierung von Navigationsattributen und von Externen Hierarchien zum Tragen, die aber zeitabhängig zu definieren sind, damit auf eine historische Struktur zurückgegriffen werden kann. Bei Abfragen, die Zeitabhängigkeit zu berücksichtigen haben, kommt das so genannte Query Key Date oder Schlüsseldatum zum Tragen. Dieses definiert den betrachteten Zeitpunkt der Gültigkeit und ist innerhalb einer Query eindeutig festzulegen.

Der Fall der Realisierung über zeitabhängige Navigationsattribute ist in Abbildung 86 wiedergegeben. Als Beispiel zur Verdeutlichung der Verknüpfungen der Tabellen dient hierbei die Abfrage des Umsatzwertes der Produktgruppe PG2 für die Periode 2003-05, das in der Abbildung wieder hervorgehoben dargestellt ist.



**Abb. 86.** Historische Struktur über zeitabhängige Navigationsattribute

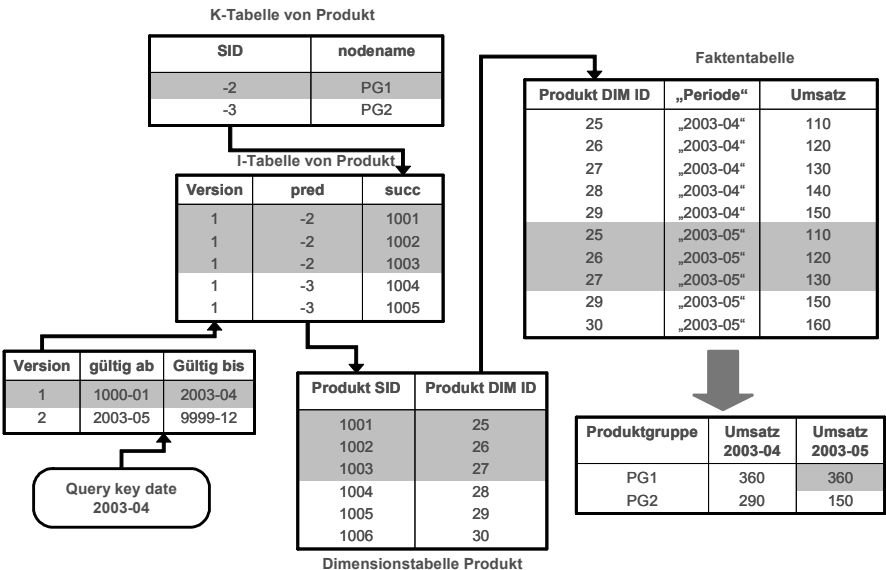
Für die Ablage der zeitabhängigen Attribute dient in den Stammdaten die Y-Tabelle. Diese hat, wie am Beispiel des Produktes ersichtlich, zusätzliche Spalten für die Gültigkeit, die mit dem Query Key Date eine zusätzliche Bedingung formen. Es finden nur die Datensätze der Attributtable Berücksichtigung, für die das Schlüsseldatum in dem (geschlossenen) Intervall definiert durch die Felder gültig ab (DATEFROM) und gültig bis

(DATETO) liegt. Für das Beispiel der Produktgruppe PG2 mit der SID 4712 sind dies lediglich die zwei markierten Zeilen, denn die Produkte P3 und P6 entfallen genau aufgrund der Schlüsseldatum-Restriktion. Der weitere Abfrageweg korrespondiert mit der bisherigen Darstellung und ergibt sich auch aus Abbildung 86.

Eine Besonderheit stellt der SID-Schlüssel 0 der Produktgruppe in der Y-Tabelle des Produktes dar. Durch das Löschen der Zuordnung von Produkt P4 zur Produktgruppe PG2 wird der Attributwert auf nicht zugeordnet gesetzt, was gerade durch die SID 0 ausgedrückt wird.

Für den Fall der Realisierung über eine Externe Hierarchie gibt es verschiedene Formen der Versionierung, die zeitabhängige Hierarchiestruktur (einzelne Knoten) und die zeitabhängige Gesamthierarchie, die folgende Darstellung bezieht sich dabei auf die zweite Variante.<sup>14</sup>

Technisch gesehen werden die einzelnen Gültigkeitsvarianten der Hierarchie als Versionen gespeichert, deren Nummer sich über das Schlüsseldatum eindeutig ergibt. Zu dieser sind dann wieder die Einträge in der Parent-Child-Tabelle abzufragen. In Abbildung 87 ist dies exemplarisch für die Abfrage der Produktgruppe PG1 dargestellt. Die Einschränkung auf der I-Tabelle kommt somit einerseits von der K-Tabelle, die die Produktgruppe festlegt, und andererseits von der Tabelle der Hierarchieversionen.



**Abb. 87.** Historische Struktur über zeitabhängige Externe Hierarchie

<sup>14</sup> Versionsabhängigkeit von Hierarchien ist hierfür nicht geeignet.

Die Einträge für die Version 2 sind, damit die Darstellung nicht zu unübersichtlich gerät, in der I-Tabelle weggelassen. Die Versionsnummer ist in dieser I-Tabelle natürlich auch für den Fall der nicht zeitabhängig implementierten Hierarchie vorhanden, wurde aber in der Darstellung bisher ausgelassen, da sie für das Verständnis der Zusammenhänge nicht notwendig war.

Der weitere Weg der Datenabfrage korrespondiert mit der bisherigen Darstellung, da die Verknüpfung über die Dimensionstabelle zur Faktentabelle natürlich immer die gleiche ist. Somit ergibt sich der dargestellte Abfragewert von 360 aus den drei qualifizierten Tupeln in der Faktentabelle.

Über die Selektion des Schlüsseldatums als aktuelles Datum, dies ist zugleich die Standardeinstellung für Abfragen an das Business Information Warehouse, ist auch das erste Szenario der aktuellen Struktur inkludiert. Ist aber neben dieser Anforderung keine Berücksichtigung von historischen Strukturen explizit gefordert, besteht keine Notwendigkeit, die mit der Verwendung zeitabhängiger Strukturen einhergehende gestiegene Komplexität in Kauf zu nehmen.

### **8.3.3 Transaktionsorientierte Sichtweise**

Von der Sichtweise der aktuellen und der historischen Struktur grenzt sich die historische Wahrheit, deren Szenario in Abbildung 88 noch mal zusammengefasst ist, in der Hinsicht grundsätzlich ab, als es sich dabei nicht um eine stammdatenbezogene Sicht, sondern um die transaktionsorientierte Bewegungsdatensicht handelt. Die einzige Möglichkeit, dies im Business Information Warehouse abzubilden, besteht ja in der Modellierung über Merkmale.

Bei der Implementierung über Merkmale sind die SID-Schlüssel für die Produktgruppe ebenso wie diejenigen für die Produkte direkt in den Dimensionstabellen abgelegt. Dies muss nicht notwendigerweise die gleiche Tabelle bzw. Dimension sein, ist aber für das dargestellte Beispiel zur Vereinfachung unterstellt.

Produktdimension in 2003-04

Produkt	Produktgruppe
P1	PG1
P2	PG1
P3	PG1
P4	PG2
P5	PG2

Produktdimension in 2003-05

Produkt	Produktgruppe
P1	PG1
P2	PG1
P3	PG2 (geändert)
P4	gelöscht
P5	PG2
P6	PG2 (neu)

Faktentabelle

Produkt	Periode	Umsatz
P1	2003-04	110
P2	2003-04	120
P3	2003-04	130
P4	2003-04	140
P5	2003-04	150
P1	2003-05	110
P2	2003-05	120
P3	2003-05	130
P5	2003-05	150
P6	2003-05	160

Produktgruppe	Umsatz 2003-04	Umsatz 2003-05
PG1	360	230
PG2	290	440

Abb. 88. Berichtsszenario „historische Wahrheit“

In Abbildung 89 ist die Abfrage für den Umsatz der Produktgruppe PG2 in der Periode 2003-05 besonders gekennzeichnet. Diese geht von der S-Tabelle der Produktgruppe direkt in die Dimensionstabelle und diese werden zu Dimensionsschlüsseln aufgelöst.

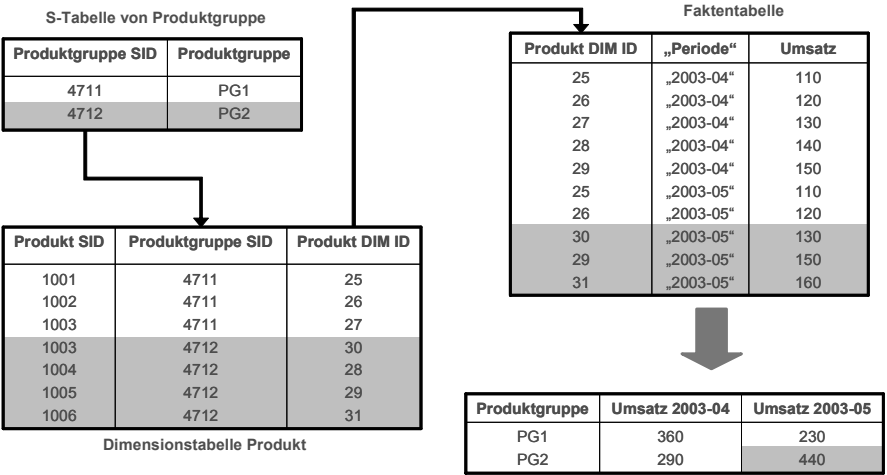


Abb. 89. Historische Wahrheit über Merkmale

Mit den Kombinationen, die gemäß der Struktur aus der aktuellen Periode in die Dimensionstabelle eingehen, ergeben sich neue Dimensions-IDs für noch nicht vorkommende Kombinationen. So tritt mit dem Produkt P6 ein neuer Eintrag mit dem Dimensionsschlüssel 31 hinzu und für die Umstrukturierung von Produkt P3 gibt es den weiteren Eintrag mit dem Dimensionsschlüssel 30, der diese Kombination abbildet. Im Rahmen der auftretenden Kombinationen aus Produkten und Produktgruppen in den Bewegungsdaten ergibt sich u. U. eine m:n-Beziehung, die mit dieser Modellierungsvariante ebenfalls abdeckbar ist.

8.3.4 Vergleichbare Resultate

In Abschnitt 3.4 ist ein weiteres Berichtsszenario Gegenstand der Betrachtungen gewesen. Dieses ist in Abbildung 90 zusammengefasst und berücksichtigt die Darstellung vergleichbarer Elemente bzw. Resultate.

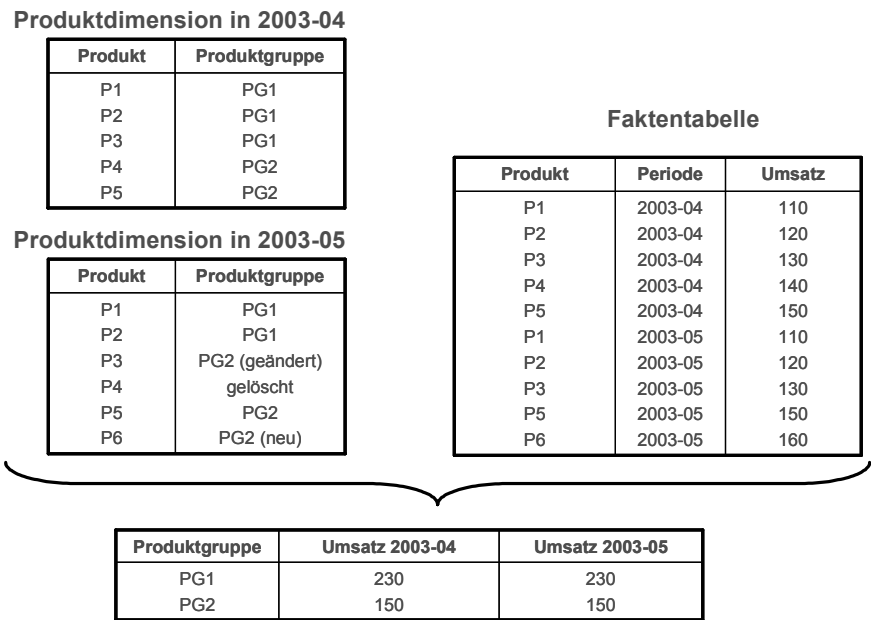


Abb. 90. Berichtsszenario „vergleichbare Elemente“

Zur Implementierung dieses eher seltenen Szenarios bedarf es neben der Verwendung zeitabhängiger Navigationsattribute zusätzlicher Felder zur benutzerdefinierten Abfrage auf die Gültigkeitszeit, da dies bei den Systemfeldern nicht möglich ist.

Um die Anforderungen des Szenarios zu erfüllen, sind nur diejenigen Produkte zu selektieren, deren Produktgruppenzuordnung sich über alle Berichtsperioden hinweg nicht verändert hat. Dazu ist einerseits das Schlüsseldatum auf eine der Berichtsperioden zu setzen und andererseits muss gewährleistet werden, dass die Gültigkeit spätestens mit der ersten Berichtsperiode beginnt und weiterhin frühestens mit der letzten endet.

Für die Realisierung sind zusätzliche Attribute des Produktes notwendig, die ebenfalls die Werte für Beginn und Ende der Gültigkeit repräsentieren (*bg ab* und *bg bis*). Diese werden allerdings nicht automatisch beim Ladevorgang gefüllt, sondern müssen im ETL-Prozess entsprechend abgeleitet werden. Die Eingrenzung der Tupel in der Y-Tabelle erfolgt zum einen wie gehabt über die S-Tabelle für die Produktgruppe, die im Beispiel aus Abbildung 91 die Eingrenzung auf PG2 vornimmt. Zum anderen sorgt die Eingrenzung auf das Schlüsseldatum, im Beispiel eine der beiden Perioden 2003-04 bzw. 2003-05, für die Berücksichtigung der gültigen Produktgruppenzuordnungen. Dies kann aber nur eine Periode berücksichtigen, damit die Gültigkeit über beide Perioden des Beispiels Berücksichtigung findet, ist die weitere Eingrenzung über die benutzerdefinierten Attribute notwendig. Dies führt somit zu den relevanten Einträgen in der Dimensionstabelle und dem damit korrekten Ergebnis auf Basis der Faktentabelle.

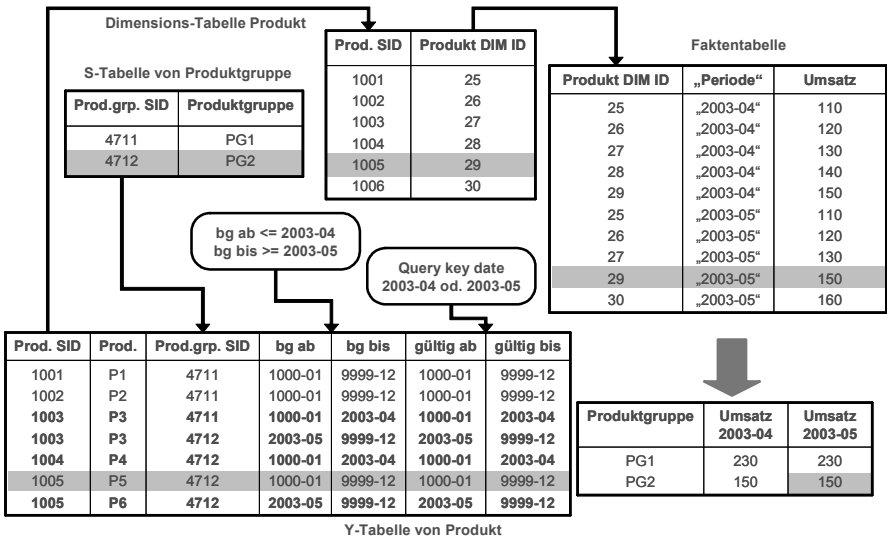


Abb. 91. Realisierung vergleichbarer Resultate

Die Ausführungen belegen, dass es grundsätzlich möglich ist die verschiedenen Berichtsszenarien im Business Information Warehouse abzubilden. Dabei sind auch Kombinationen der Ansätze denkbar. Ausgangspunkt für die Datenmodellierung ist dabei immer die Anforderung an die Möglichkeiten der Analyse. Die zu beantwortenden Fragen sind, ob historische Auswertungen gefragt sind, ob eine aktuelle Sichtweise genügt oder ob eine transaktionsorientierte Sicht den Anforderungen gerecht wird.

## 8.4 Modellierung von Kennzahlen

In betriebswirtschaftlichen Analysen ist eine Vielzahl von Kennzahlen relevant, die in verschiedenartigen zum Teil hierarchischen Systemen geordnet sind. Diese so genannten Kennzahlensysteme sind für die individuellen unternehmensspezifischen Analyseprozesse von Bedeutung.

Im erweiterten Star Schema des BW sind die physisch abgelegten Kennzahlen gerade die Wertspalten in der Faktentabelle, die somit keiner hierarchischen Strukturierung unterliegen. Die Berechnung dieser Kennzahlen erfolgt im Rahmen des ETL-Prozesses während der zumeist periodischen Aktualisierungsvorgänge des BW-Datenbestandes. Für die Aggregation entlang der Konsolidierungspfade wird standardmäßig die Addition herangezogen, aber durch die Angabe einer Ausnahmeaggregation sind hier auch alternative Möglichkeiten der Verdichtung realisierbar. Dies ist insbesondere für nicht additive Kennzahlen wie Bestandsgrößen oder Verhältniszahlen notwendig.<sup>15</sup> Für die Ausnahmeaggregation muss in der Kennzahlpflege die Bezugsgröße festgelegt werden. Diese definiert das Merkmal, auf das bezogen die Kennzahl nicht additiv ist. Eine Besonderheit stellen die Bestandskennzahlen dar, denn deren Modellierung erfolgt im Business Information Warehouse im Allgemeinen über eine Kennzahl für Bestandsveränderungen, aus welcher der Bestand berechnet wird. Eine weitere Form berücksichtigt getrennte Kennzahlen für Zu- und Abgänge.

Die Abbildung umfassender Kennzahlensysteme erfolgt im Business Information Warehouse nicht auf der Ebene der physischen Datenstrukturen, sondern erst auf der Ebene der Abfragen. Diese Queries stellen im BW eine mehrdimensionale Sichtweise auf den in Info-Providern abgelegten Datenbestand dar und ermöglichen die Implementierung komplexer Berechnungen und Operationen.

Auf der Auswertungsebene stehen den Benutzern die Möglichkeiten global definierter abgeleiteter Kennzahlen zur Verfügung. Im BW sind da-

---

<sup>15</sup> Zur Additivität siehe auch die Ausführungen in Abschnitt 3.3.3.

bei berechnete und eingeschränkte Kennzahlen zu differenzieren. Die Definition berechneter Kennzahlen erfolgt mithilfe vielfältiger mathematischer Funktionen auf Basis von physisch abgelegten wie auch von anderen abgeleiteten Kennzahlen. Eingeschränkte Kennzahlen bieten die Möglichkeit, Werte einer anderen Kennzahl auf spezielle Merkmalsausprägungen einzugrenzen. So sind etwa Ist- und Plandaten auf Basis von Kennzahlen eingeschränkt über diese Wertart abgrenzbar. Für Kennzahlen auf Ebene der Query gilt, dass die Berechnung nur auf dem von der Query selektierten Datenbereich erfolgt. Die Ausnahmeaggregation kann demgegenüber auf das gesamte Datenmodell zur Berechnung zugreifen.

In Abschnitt 3.3.2 wurden verschiedene grundsätzliche Modellierungsansätze für Kennzahlensysteme in die Diskussion gebracht. Das wesentliche Differenzierungsmerkmal ist dabei, ob das Kennzahlensystem als eine Dimension aufgefasst wird, oder ob jede Kennzahl durch einen Würfel repräsentiert wird. In der Diskussion des allgemeinen Star Schemas in Abschnitt 7.1.4 führte diese Unterscheidung zu den Formen einer Kennzahlendimension einerseits und der Abbildung von Kennzahlen als Spalten in der Faktentabelle andererseits. Dies ist in dieser Form auch auf das Datenmodell des Business Information Warehouse übertragbar. Soll die Darstellung des Kennzahlensystems in einer Dimension erfolgen, ist eine künstliche Kennzahl für die Aufnahme des Wertes zu definieren und die Bedeutung dieses Wertes ergibt sich dann aus einem Merkmal in einer Dimension, das die betriebswirtschaftliche Kennzahl repräsentiert. Während die Modellierung von Kennzahlen in der Faktentabelle zu breiten Tabellen führt, geht mit der anderen Variante eine zwar schmalere Tabelle einher, die aber wesentlich mehr Zeilen umfasst. Die Abbildung von Kennzahlen in einer eigenen Dimension führt darüber hinaus auch zu mehr Flexibilität im Falle hinzukommender Kennzahlen.

Weitere zu berücksichtigende Aspekte bei der Festlegung von Info-Objekten vom Typ Kennzahl im BW sind die zu berücksichtigenden Einheiten und die Möglichkeiten der Währungsumrechnung.

## 8.5 Grafische Repräsentation logischer BW-Modelle

Für die mehrdimensionale Modellierung auf semantischer Ebene stehen diverse grafische Methoden zur Verfügung, die eine übersichtliche Darstellung von Würfeln, Dimensionen und Hierarchien erlauben. Diese wurden ausführlich in Kapitel 6 diskutiert. Auf der logischen Modellebene sind für die mehrdimensionale Modellierung für das Business Information Warehouse keine standardisierten Verfahren verfügbar. In der Praxis wird



diese Phase der Modellierung oftmals auf Basis von tabellarischen Darstellungen in Tabellenkalkulationsprogrammen vorgenommen. Hierunter leidet insbesondere die Übersichtlichkeit, die Kommunikation mit dem Anwender und es erschwert die Arbeit des Datenmodellierers. Einen Ansatz bietet Microsoft Visio als grafisches Werkzeug zum Aufbau verschiedenster Modelle.

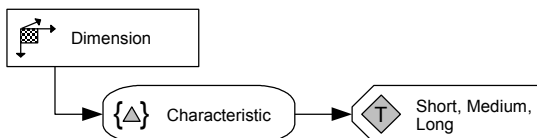
In diesem Abschnitt erfolgt der Entwurf einer eigenen Methode zur grafischen Beschreibung von logischen Datenmodellen des Business Information Warehouse, deren Objekte ebenfalls als Schablonen für Visio zur Verfügung stehen.<sup>16</sup>

### 8.5.1 Darstellung von Dimensionen

Zu den Aspekten der Dimensionsmodellierung gehört zum einen die Abbildung von Stammdaten wie Texten und Attributen, zum anderen aber auch die Frage der Ausgestaltung von hierarchischen Dimensionsstrukturen. Zunächst erfolgt die Darstellung der Grundbestandteile einer Dimension, die Darstellung der verschiedenen Varianten der Modellierung von Hierarchien schließt sich daran an.

Zur Abbildung einer Dimension gibt es das Darstellungsobjekt *Dimension*, in dessen Mitte der Dimensionsname steht. Eine Dimension besteht aus mindestens einem Merkmal (engl. *characteristic*). Die Darstellung von Merkmalen erfolgt durch das Objekt *Characteristic*. Das dreieckige Symbol deutet an, dass ein Merkmal auf der Definition eines Info-Objektes basiert. Dass über dieses Merkmal eine Ebene in einer Dimensionshierarchie repräsentiert wird ist an der geschweiften Klammerung ersichtlich. In dem Objekt steht der Name dieser Konsolidierungsebene.

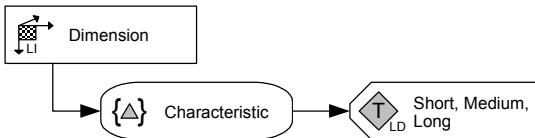
Die Beschreibung einer Merkmalsausprägung (*caption*) basiert im BW-Modell auf den Texten. Merkmale, die Texte haben, bekommen als weiteres anhängendes Darstellungsobjekt *Texts*, in dem dann die vorhandenen Texte direkt stehen. Dies können Kurztexte, mittellange Texte und Langtexte sein. Dieses Grundkonstrukt zur Modellierung von Dimensionen ist in Abbildung 92 dargestellt.



**Abb. 92.** Dimension mit einem Merkmal

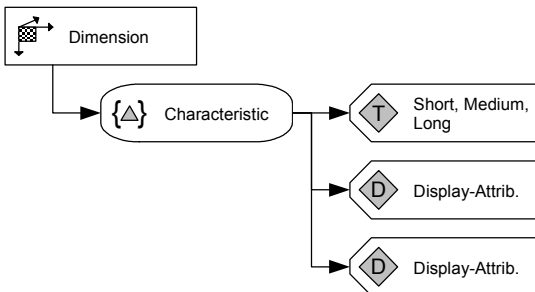
<sup>16</sup> Die Shape-Dateien können unter [www.hahneonline.de](http://www.hahneonline.de) heruntergeladen werden.

Texte sind im BW auch sprachabhängig modellierbar. Dann werden diese durch ein eigenes Objekt *Texts Language Dependent* repräsentiert. Eine weitere Modellierungsvariante entsteht gerade bei Dimensionen mit nur einem Merkmal, die dann aus Optimierungsgründen als Line-Item definiert werden sollten. Dieser Dimensionstyp wird durch das eigene Objekt *Line-Item Dimension* repräsentiert. Beide Objekttypen sind in Abbildung 93 dargestellt.



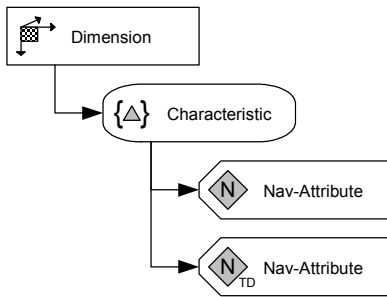
**Abb. 93.** Line-Item Dimension

Neben Texten als weiteres Beschreibungsmittel von Dimensionsausprägungen sind Attribute von Merkmalen darstellbar. Anzeigeattribute haben einen ähnlichen Charakter wie Texte, da sie lediglich als ergänzende Information zum Merkmal dienen, und werden mit dem Objekttyp *Display Attribute* notiert (siehe Abbildung 94).



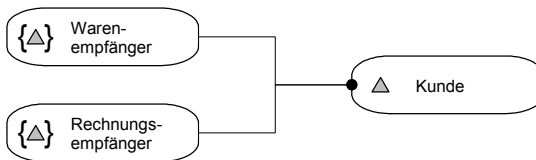
**Abb. 94.** Anzeigeattribute eines Merkmals

Neben den beschreibenden Attributen kennt das BW-Modell auch noch Navigationsattribute, die sich von den Anzeigeattributen dadurch unterscheiden, dass sie in Abfragen ohne das zugrunde liegende Merkmal angezeigt und für Navigationsschritte genutzt werden können. Navigationsattribute sind auch zeitabhängig definierbar. Die beiden Formen der Navigationsattribute sind in der Abbildung 95 mit den entsprechenden Objekten *Navigation Attribute* und *Navigation Attribute Time-Dep.* dargestellt.



**Abb. 95.** Navigationsattribute eines Merkmals

Damit verschiedene Merkmale in einem Info-Cube auf die gleichen Stammdaten zugreifen können, ist es möglich, diese als Referenz auf ein gemeinsames Info-Objekt zu definieren. Dann sind alle Einstellungen des referenzierten Info-Objektes auch für die darauf verweisenden gültig. In Abbildung 96 ist dies für Modellierung der verschiedenen Rollen des Warenempfängers und des Rechnungsempfängers eines Kunden dargestellt.



**Abb. 96.** Rollen über Referenzen auf Info-Objekte

Für Merkmale, die erst in Abhängigkeit von einem anderen Objekt eindeutig gekennzeichnet sind, steht im Business Information Warehouse der Mechanismus der Klammerung bereit. In Abbildung 97 ist hierzu das Beispiel der grafischen Visualisierung des Merkmals Kostenstelle aufgeführt, welches erst durch die Klammerung mit dem Info-Objekt Kostenrechnungskreis eindeutig ist.

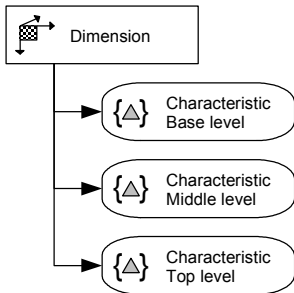


**Abb. 97.** Klammerung von Info-Objekten

Die dargestellten Objekte sind der Grundstock, mit dem auch hierarchische Dimensionsstrukturen abbildbar sind. Dies ist Gegenstand der folgenden Ausführungen.

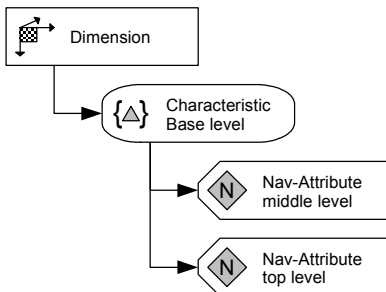
### 8.5.2 Abbildung von hierarchischen Strukturen

Im Business Information Warehouse werden hierarchische Dimensionsstrukturen auf vielfältige Weise dargestellt. Die Bewegungsdatensicht ergibt sich aus mehreren Merkmalen in einer Dimension, wobei jedes Merkmal eine Ebene einer im Allgemeinen balancierten Struktur repräsentiert. Dies ist in Abbildung 98 visualisiert.



**Abb. 98.** Hierarchie über Merkmale

Eine stammdatenbasierte Sicht ergibt sich aus der Modellierung mit Navigationsattributen. Diese können sowohl zeitabhängig als auch nicht zeitabhängig ausgestaltet sein. Die unterste Ebene der hierarchischen Struktur ist durch das Merkmal selbst repräsentiert, jede weitere Ebene korrespondiert mit einem eigenen Navigationsattribut. Der Fall der nicht zeitabhängigen Navigationsattribute ergibt sich aus Abbildung 99.



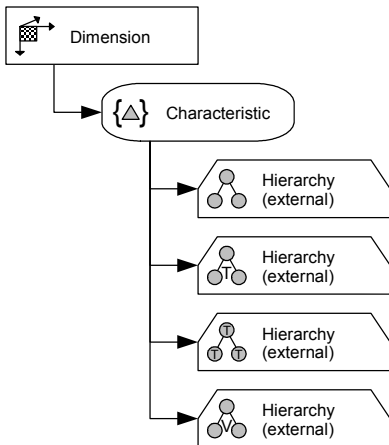
**Abb. 99.** Hierarchien über Navigationsattribute

Nicht alle hierarchischen Strukturen in mehrdimensionalen Modellen sind balanciert. Gerade für diesen Fall bietet sich die Modellierung über Externe Hierarchien im BW an. Diese sind ebenfalls stammdatenbezogen und bieten die Möglichkeit, hierarchische Strukturen auf Basis einer Parent-Child-Relationship darzustellen.

Zeitabhängigkeit ist bei Externen Hierarchien für die gesamte Struktur wie auch für einzelne Knoten aktivierbar. Des Weiteren gibt es auch die Möglichkeit der Speicherung einzelner Versionen einer Hierarchie. Alle vier Varianten von Externen Hierarchien sind in Abbildung 100 zusammengefasst.

In der Reihenfolge von oben nach unten stehen die Hierarchieobjekte für die Formen ohne Versionierung, diejenige mit Zeitabhängigkeit für die gesamte Struktur sowie für die Knoten und die Variante der versionsabhängigen Hierarchie.

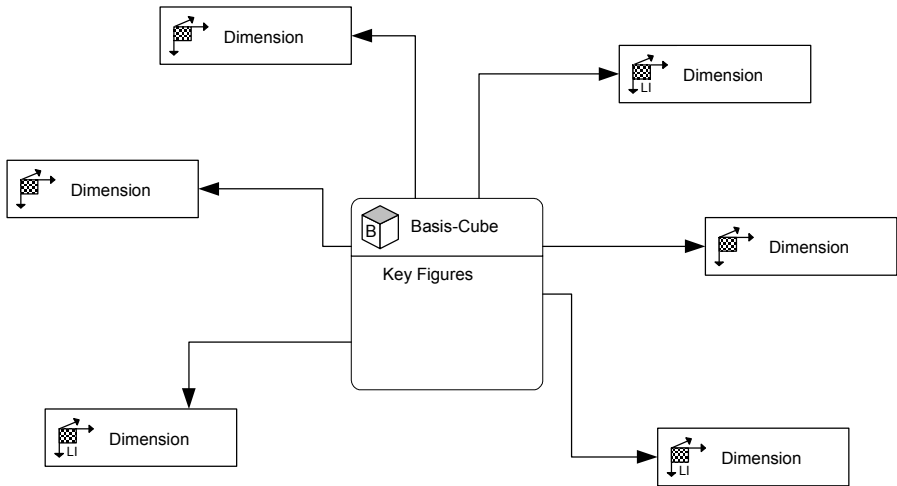
Alle dargestellten Bestandteile zur Modellierung von Dimensionen sind miteinander kombinierbar. So können innerhalb einer Dimension Externe Hierarchien gemeinsam mit zusätzlichen Navigationsattributen Verwendung finden und zugleich auch verschiedene Merkmale die transaktionsorientierte Sichtweise der Bewegungsdaten implementieren.



**Abb. 100.** Externe Hierarchien

### 8.5.3 Modellierung von Info-Cubes

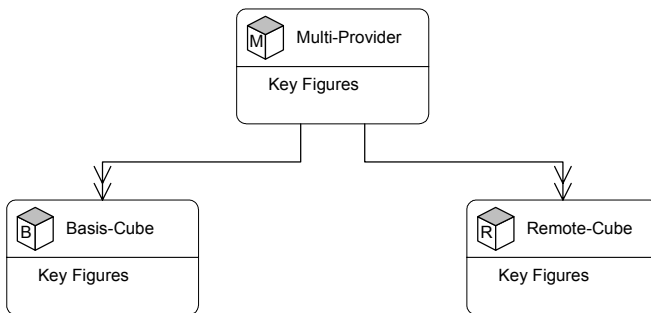
Wichtigster Typ eines Info-Cubes ist der Basis-Cube, dessen Darstellung auf dem Objekt *Basis Cube* basiert. Um den Cube herum werden die einzelnen Dimensionen angeordnet (siehe Abbildung 101). Bis zu dreizehn sind dabei zuzüglich der obligatorischen Zeit-, Einheiten- und Paket-Dimension möglich. Im Cube-Objekt erfolgt ebenfalls die Auflistung der abgebildeten Kennzahlen.



**Abb. 101.** Darstellung eines Basis-Cubes

In der Darstellung des Cubes kann die Dimension direkt auch detailliert modelliert werden. Ist das Ergebnis der gesamten Darstellung zu komplex und unübersichtlich erfolgt die Dimensionsmodellierung getrennt, der dargestellte Cube referenziert dann nur die einzelnen Dimensionen.<sup>17</sup>

Neben den Basis-Cubes gibt es noch Remote-Cubes und Multi-Provider, jeweils repräsentiert durch die Objekte *Remote Cube* und *Multi Provider*. Die Verbindung eines Multi-Providers mit den zugrunde liegenden Cubes erfolgt über den *based upon* Verbindungspfeil. Dieser Zusammenhang ist in Abbildung 102 wiedergegeben.



**Abb. 102.** Repräsentation eines Multi-Providers

<sup>17</sup> In Microsoft Visio kann beispielsweise per Doppelklick von dem Dimensionssymbol direkt in die Detaildarstellung verzweigt werden.

Da auch Multi-Cubes eine eigene Dimensionierung haben, ist mit dieser Darstellung die Modellierung des Multi-Providers noch nicht vollständig, es muss vielmehr auch noch eine eigene Cube-Modellierung analog wie bei den Basis-Cubes dazu stattfinden.

## **8.6 Ableitung eines BW-Modells für das Beispiel**

Entsprechend der vorgeschlagenen Vorgehensweise dient das semantische Datenmodell als Ausgangslage für die Ableitung des Modells auf der logischen Ebene. In einem späteren Schritt leitet sich daraus dann auch das physische Modell mit den spezifischen systembedingten Einstellungen ab.

Die Anforderungen der Benutzer bezogen auf Analyse und Reporting stellen den wesentlichen Punkt beim Aufbau des semantischen Modells dar. Die zentrale Bedeutung des Umgangs mit Zeitabhängigkeit und der Berücksichtigung verschiedener Zeitszenarien wurde bereits in den vorangegangenen Kapiteln herausgearbeitet. Daher dient auch T-ADAPT, als Erweiterung von ADAPT um temporale Konstrukte, zur Beschreibung bzw. grafischen Repräsentation auf der semantischen Gestaltungsebene in diesem Buch. Die Aufgabe der Modellierung ist es nun, aus dem vorhandenen T-ADAPT-Modell, das in Abschnitt 6.4 entwickelt wurde, ein adäquates Modell für das Business Information Warehouse abzuleiten. Dies geschieht in den folgenden Abschnitten sukzessive für die einzelnen thematischen Bereiche unseres Beispiels.

### **8.6.1 Modell für den Controllingbereich**

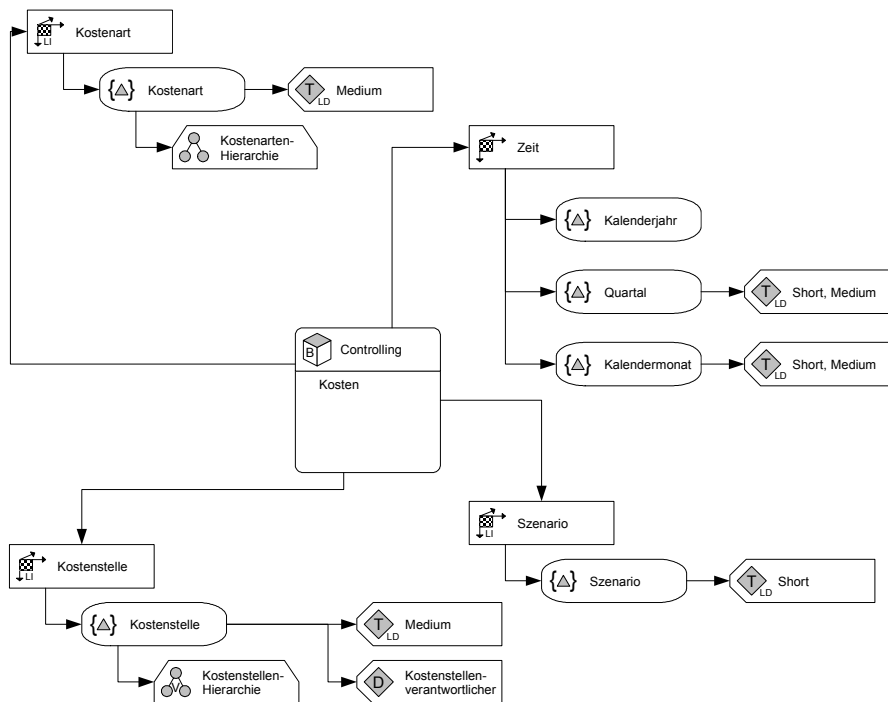
Zunächst ist auf der logischen Modellebene die Unterscheidung in Multi-Provider und Basis-Cubes noch nicht relevant, denn dieser Aspekt steht bei dem Aufbau des physischen Modells im Vordergrund. Somit korrespondieren die einzelnen Cubes in der T-ADAPT-Darstellung mit den Info-Cubes im Business Information Warehouse. Für das entworfene Modell, das die Belange des Controllingbereiches abbildet, handelt es sich daher um einen Controlling Basis-Cube im BW-Modell. Dies ist in der grafischen Repräsentation auf Basis der entwickelten Darstellungsobjekte, gestaltet mit den korrespondierenden Visio-Shapes, Gegenstand der Abbildung 103, in deren Zentrum der Basis-Cube steht.

Die Modellierung von Kennzahlen erfolgt in T-ADAPT entweder direkt im Würfel oder in einer Kennzahlendimension, wobei der letztere Weg in diesem Buch durchgängig bevorzugt wird. Diese sind dann in der Darstel-

lung des BW-Modells in das Cube-Objekt aufzunehmen, im Beispiel ist das die einzige Kennzahl Kosten.

Ein weiterer wesentlicher Transformationsschritt ist die Ableitung der Dimensionen, Merkmale, Hierarchien und Attribute aus den Dimensionen des T-ADAPT-Modells.

Eine herausragende Rolle spielt dabei die Zeitdimension im Business Information Warehouse, da diese als eine der drei fest vorgegebenen Dimensionen in jedem Cube enthalten ist und sich ausschließlich aus den vorgegebenen Zeitmerkmalen zusammensetzt. Im Beispiel basiert die Zeitdimension auf einer Kalenderjahreshierarchie mit den Ebenen Jahr, Quartal und Monat. Da in dem T-ADAPT-Modell für die Zeitdimension keine Versionierung für die Hierarchie zu berücksichtigen ist, erfolgt immer die Betrachtung gemäß der aktuellen Struktur.



**Abb. 103.** Info-Cube für den Controllingbereich

Im Business Information Warehouse erfolgt die Modellierung der Zeitdimension im Standard über Merkmale in der vorgegebenen Dimension. Hierzu dienen die Standard-Info-Objekte des Systems für die Abbildung von Jahren, Quartalen und Monaten (Eine Liste aller Zeitmerkmale im Business Content findet sich in Tabelle 19). Somit ist dies im BW normaler-



weise eine transaktionsorientierte Sicht, wobei allerdings die Zuordnung zwischen den zeitlichen Ebenen konstant ist und damit in der Konsequenz gerade der jeweils aktuellen Sichtweise entspricht.

Bei dem Einsatz benutzerdefinierter Info-Objekte ist auch die Implementierung der Zeithierarchien über Navigationsattribute möglich. Aufgrund der systemseitig vorgegebenen Möglichkeiten der Transformation und der speziellen Zeitarithmetik ist allerdings die Nutzung der Zeitmerkmale des Business Content im Allgemeinen zu empfehlen.

**Tabelle 19.** Zeitmerkmale des Business Content

Info-Objekt	Technischer Name	Beispiel
Kalendertag	0CALDAY	20040723
Kalenderjahr/Monat	0CALMONTH	200407
Kalendermonat	0CALMONTH2	07
Quartal	0CALQUART1	3
Kalenderjahr/Quartal	0CAL_QUARTER	20043
Kalenderjahr/Woche	0CALWEEK	200430
Kalenderjahr	0CALYEAR	2004
Geschäftsjahr/Geschäftsmonat	0FISCPER	2004007
Geschäftsmonat	0FISCPER3	007
Geschäftsjahresvariante	0FISCVARNT	K4
Geschäftsjahr	0FISCYEAR	2004
Halbjahr	0HALFYEAR1	2
Wochentag	0WEEKDAY1	6

Der Fall benutzerspezifischer Zeitmerkmale wird später noch aufgegriffen. Ergänzend sind die Bezeichnungen für Quartale und Monate als sprachabhängige Texte angehängt. Deren Darstellung ist auf der semantischen Ebene nicht Modellbestandteil gewesen. T-ADAPT bietet hierzu zwar Darstellungsmöglichkeiten, jedoch liegt der Fokus in dieser Phase der Modellierung auf einer allgemeinen konzeptuellen Beschreibung.

Für die elementbestimmte Szenariodimension gibt es keine direkte Entsprechung in der grafischen Darstellung des BW-Modells, da dort nur Ebenen berücksichtigt werden können. Somit besteht die Szenariodimension im BW nur aus dem einen entsprechenden Merkmal und kann demzufolge auch als Line-Item modelliert werden.

Sowohl für die Kostenstellen als auch für die Kostenarten kommt in dem T-ADAPT-Modell das Konstrukt der rekursiven Beziehung zum Tragen. Diese ist im Business Information Warehouse am geeignetsten durch eine Externe Hierarchie abzubilden. Für die Kostenartendimension beschränkt sich die Berücksichtigung von Zeitabhängigkeit auf die jeweils aktuelle Struktur, so dass hier keine zeitabhängige Hierarchie verwendet werden muss. Für die Kostenstellenhierarchie hingegen sind einzelne Versionen für die Jahresendstände neben der aktuellen Struktur zu berücksichtigen. Dies ist im BW über versionsabhängige Hierarchien und einzelne Externe Hierarchien gut abbildbar, zeitabhängige Hierarchien sind dazu weniger geeignet. Im Beispiel ist die Gestaltung von Versionen einer Externen Hierarchie bevorzugt.

Weitere Objekte, die ins logische Modell überführt werden müssen, sind die Attribute. Für das Controlling-Beispiel ist nur der Kostenstellenverantwortliche mit zu berücksichtigen. Diese Attribute können im Business Information Warehouse grundsätzlich durch Navigations- und auch Anzeigeattribute abgebildet werden. Tendenziell haben die Attribute in T-ADAPT jedoch den Charakter ergänzender Information, die in den Anzeigeattributen ihre Entsprechung haben. Navigationsattribute spielen später bei der Transformation von Dimensionsausschnitten (Scopes) im folgenden Abschnitt wieder eine Rolle.

### 8.6.2 Aggregierte Marketingsicht

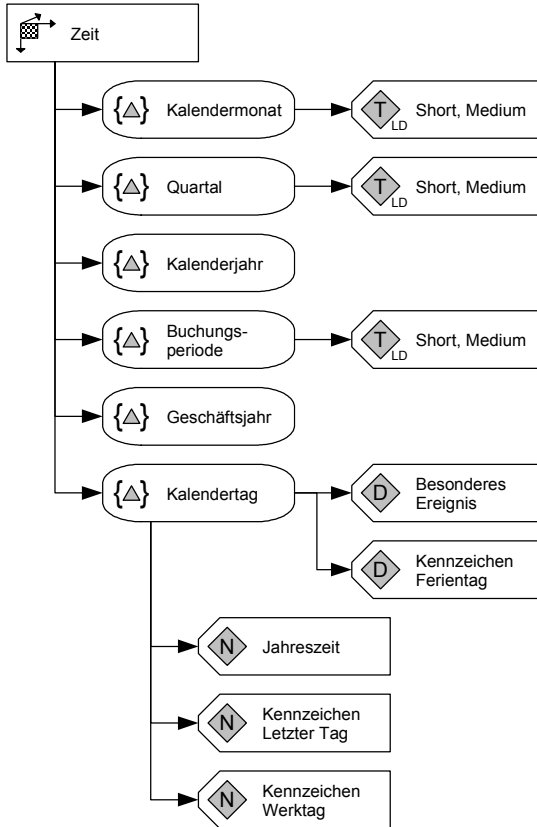
Die Modelle für den Marketingbereich unterscheiden sich von dem bereits dargestellten Controlling-Modell insbesondere durch eine veränderte Zeitdimension, da hier neben den beiden Hierarchien für die Kalender- wie auch die Geschäftsjahressicht noch eine andere Granularität gefordert ist.

Grundsätzlich kommt bei der Gestaltung der Zeitdimension im Business Information Warehouse die Standarddimension basierend auf den Business Content Zeitmerkmalen zum Tragen.

Besteht die Notwendigkeit, auf eigene Info-Objekte zurückgreifen zu müssen, erfolgt deren Ablage in einer eigenen Dimension und nicht in der Standarddimension. Es ist natürlich möglich, die zur Verfügung gestellten Info-Objekte den spezifischen Bedürfnissen anzupassen.

Auch für den Marketingbereich kann auf die Business Content-Objekte zurückgegriffen werden, so dass auch hier die normale Zeitdimension mit allen Ebenen als Merkmal zum Tragen kommt. Diese ist in Abbildung 104 visualisiert. Die Attribute des Kalendertages finden als Anzeigeattribute Eingang in die Darstellung. Die Dimensionsausschnitte haben eher den Charakter, dass sie Grundlage für Analysen und Navigationsoperationen

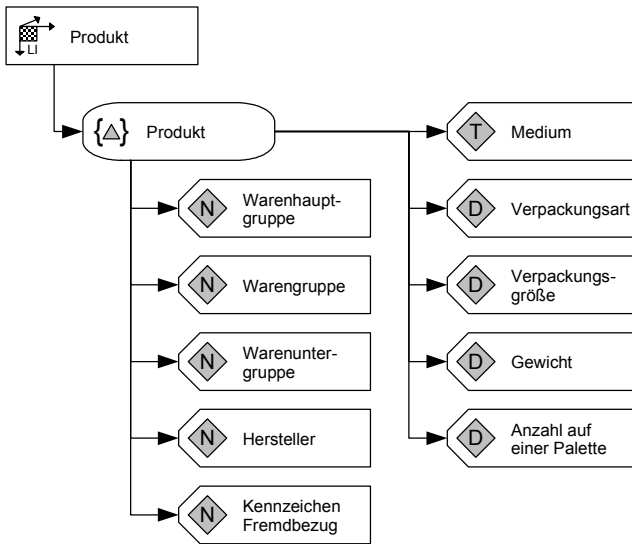
sind, so dass deren Modellierung im Business Information Warehouse am Besten über Navigationsattribute erfolgt.



**Abb. 104.** Zeitdimension der aggregierten Marketingsicht

Die nächste wichtige zu transformierende Dimension des Modells der aggregierten Marketingsicht berücksichtigt die Produkte mit deren vielfältigen Möglichkeiten auf Basis von Hierarchien, Attributen und Dimensionsausschnitten.

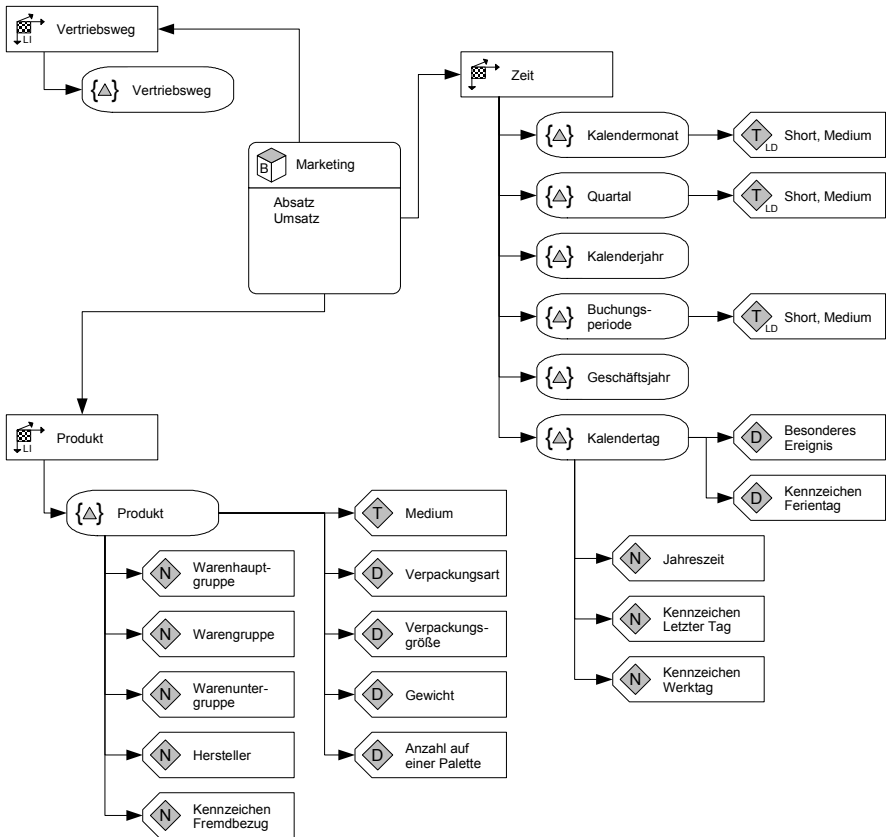
Erster Ansatzpunkt der Umsetzung ist die Berücksichtigung von Zeitabhängigkeit. Für die hier benötigte Sichtweise ist immer die aktuelle Struktur relevant, so dass eine Modellierung über Navigationsattribute angemessen ist. Dies trifft auch auf die Dimensionsausschnitte zur Differenzierung von Eigenproduktion und Fremdbezug zu. Die weiteren Attribute des T-ADAPT-Modells dieser Dimension können wieder als Anzeigeattribute abgebildet werden. Die grafische Repräsentation der Produktdimension ist in Abbildung 105 zusammengefasst.



**Abb. 105.** Dimension Produkt der aggregierten Marketingsicht

Um nun das gesamte Modell auf die logische Ebene des Business Information Warehouse umsetzen zu können, brauchen wir nur noch die Kennzahlen und die Dimension für den Vertriebsweg zu betrachten. Die Notation der Kennzahlen erfolgt wieder direkt in dem Info-Cube. Da es sich bei diesem Würfel um zwei zu berücksichtigende Kennzahlen handelt, könnte auch die Modellierung einer eigenen Kennzahlendimension erfolgen. Im Basis-Cube ist dann eine künstliche Kennzahl für den Wert zu implementieren. Dies macht natürlich nur dann Sinn, wenn alle Kennzahlen den gleichen Datentyp haben und sich auch bzgl. der Berücksichtigung und Umrechnung von Einheiten und Währungen gleich verhalten. Dies ist bei den im Beispiel zu berücksichtigenden Kennzahlen Absatz und Umsatz jedenfalls nicht gegeben und eine Modellierung physisch gespeicherter Kennzahlen im Basis-Cube ist daher zu bevorzugen.

Für die Dimension Vertriebsweg kann die Modellierung der Szenariodimension direkt übertragen werden, denn es handelt sich ebenfalls um eine kleine elementbestimmte Dimension. Diese Dimension basiert also auch auf einem Merkmal Vertriebsweg. Da die Stammdaten für alle Info-Cubes (in einem System) übergreifend gültig sind, ist zunächst an dem BW-Modell nicht erkennbar, dass für diesen Info-Cube nur der Partner-Vertriebsweg relevant ist. Auf diesen Umstand wird in Kapitel 10 bei der Diskussion der Anwendung strukturspezifischer Eigenschaften im Kontext der Partitionierung detaillierter eingegangen. Das resultierende Gesamtmodell ist in Abbildung 106 dargestellt.



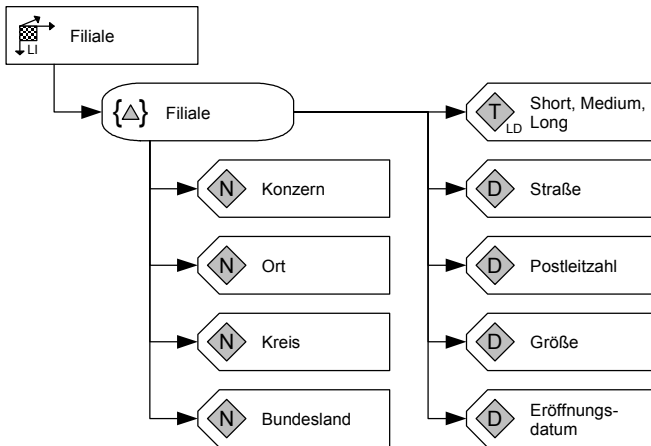
**Abb. 106.** Modell für die aggregierte Marketingsicht

Neben der übergeordneten Sicht des Bereiches Marketing sind je nach Vertriebsweg unterschiedliche detailliertere Sichtweisen zu berücksichtigen, deren Darstellung sich jetzt anschließt.

### 8.6.3 Modell für das Partner-Marketing

Der erste Aspekt, der nur den Partner-Vertriebsweg betrifft, ist die Analyse bezogen auf die Partner, die Filialen in denen die Produkte verkauft werden. In dem T-ADAPT-Modell existieren zu der Filiale eine regionale und eine konzernbezogene Hierarchie, die beide keine Versionierung berücksichtigen und daher immer in der aktuellen Struktur vorliegen. Somit ist die Gestaltung im BW-Modell mit Navigationsattributen und Externen Hierarchien möglich, wobei für diese Dimension im Beispiel die Navigationsattribute zum Einsatz kommen.

Die Attribute der Filialen wie beispielsweise die Adressinformationen entsprechen im logischen Modell des Business Information Warehouse den Navigationsattributen und sind ebenfalls in der Abbildung mit dargestellt.

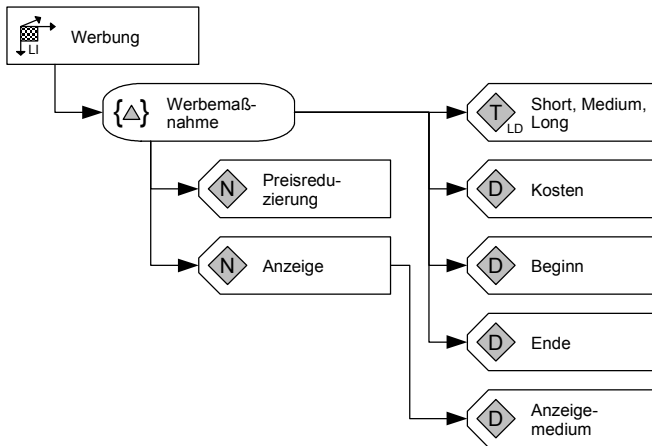


**Abb. 107.** Filialdimension im Partner-Marketing

Der zweite besondere Aspekt bei den partnerbezogenen Auswertungen betrifft die Analyse von Werbemaßnahmen, die auf der semantischen Ebene in Form der Werbungsdimension abgebildet sind. Auf der logischen Modellebene ist dazu ein Merkmal für die Werbemaßnahme in einer Dimension Werbung notwendig.

Die Attribute einer Werbemaßnahme werden im BW-Modell durch Anzeigeattribute abgebildet. Über Dimensionsausschnitte erfolgte im T-ADAPT-Modell die Unterscheidung verschiedener Arten von Werbemaßnahmen wie etwa Preisreduzierungen. Diese werden durch zwei Navigationsattribute abgebildet, denn da die beiden Dimensionsausschnitte überlappend sind, ist ein Navigationsattribut nicht ausreichend. Für den besonderen Fall einer Anzeige ist noch das Anzeigemedium als Attribut des Dimensionsausschnittes in das logische Modell zu transformieren. Navigationsattribute sind selbst auch Info-Objekte und können daher auch eigene Navigations- wie auch Anzeigeattribute haben. Da im Business Information Warehouse alle Anzeige- wie auch Navigationsattribute eines Info-Objektes, das selbst als ein Navigationsattribut verwendet wird, ebenfalls mit in die Anzeige in einer Query übernommen werden können, ist es möglich, das Medium als Attribut des Navigationsattributs Anzeige zu modellieren. Im Beispiel, wie in Abbildung 108 zusammengefasst, wurde

dabei ein reines Anzeigeattribut gewählt.<sup>18</sup> Navigationsattribute eines Navigationsattributes können dabei nicht für Navigationsoperationen verwendet werden, sondern verhalten sich wie Anzeigeattribute.

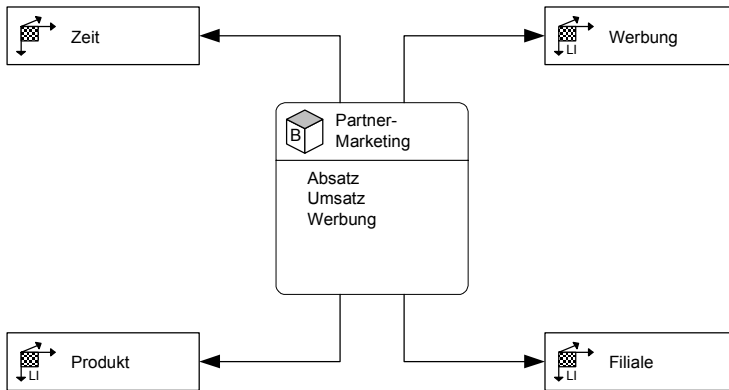


**Abb. 108.** Dimension Werbung im Partner-Marketing

Die in diesem Bereich zu berücksichtigenden Daten betreffen den Partner-Vertriebsweg. Da dies nur ein einziger ist, taucht diese Dimension im T-ADAPT-Modell gar nicht auf. In der aggregierten Sicht tritt der Vertriebsweg hingegen auf und ist dort als Dimension mit einem Merkmal abgebildet.

Als weiteren Bestandteil des T-ADAPT-Modells ist wieder die Kennzahlendimension in das logische Modell zu überführen. Diese können wie auch in der übergeordneten Marketingsicht als Spalten der Faktentabelle modelliert werden und finden ihren Eingang in die grafische Repräsentation direkt in dem Objekt für den Basis-Cube, dessen komprimierte Darstellung in Abbildung 109 zusammengefasst ist. Die weiteren Dimensionen für die Zeit, das Produkt und den Vertriebsweg entsprechen den bisherigen Darstellungen und brauchen nicht erneut aufgeführt zu werden.

<sup>18</sup> Info-Objekte, die nur als Anzeigeattribut verwendet werden, können im BW hierfür als ausschließliches Attribut definiert werden. Dies bewirkt eine Vereinfachung des Info-Objektes u. a. dadurch, dass dessen SID-Tabelle nicht gefüllt wird.



**Abb. 109.** Verdichtet dargestellter Basis-Cube des Partner-Marketings

Die Kennzahl Werbung hat hier in diesem Modell die gleiche Bedeutung wie in der Darstellung der Faktenlosen Faktentabellen des allgemeinen Star Schemas in Abschnitt 7.3.2 und in Abbildung 69 dargestellt. Aus Effizienzgründen könnte hier auch die Modellierung von zwei Basis-Cubes vorgezogen werden, dies ist aber ein Aspekt des physischen Modells.

#### 8.6.4 Modell des kundenbezogenen Marketings

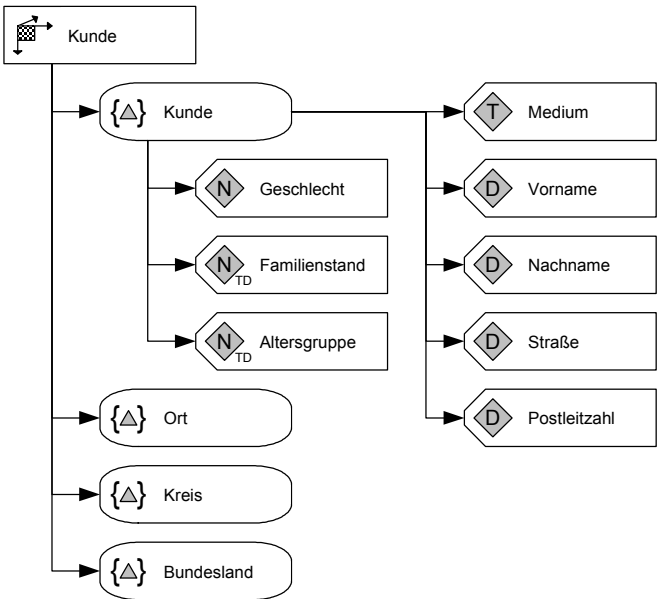
Für die Vertriebswege über Katalog und über den E-Shop stehen die Informationen kundenbezogen zur Verfügung. Für das Modell dieses Marketingbereiches verbleibt nur noch die Transformation der Dimension Kunde.

Kunden können gemäß dem semantischen Modell regional gruppiert werden. Diese Hierarchie ist transaktionsorientiert zu versionieren. Demzufolge sind die Ebenen dieser Hierarchie in Form von Merkmalen im BW-Modell abzubilden.

Die Attribute wie Name und Straße korrespondieren mit den Anzeigeattributen des Kunden im Business Information Warehouse. Die weiteren demografischen Informationen wurden im T-ADAPT-Modell in Form von Dimensionsausschnitten berücksichtigt und können als Navigationsattribute abgebildet werden. Da das Geschlecht eines Kunden als zeitkonstant vorausgesetzt wird bzw. nur das jeweils aktuell gültige zu berücksichtigen ist, kann für dieses Navigationsattribut auf Zeitabhängigkeit verzichtet werden. Der Familienstand und die Altersgruppe sind jedoch üblicherweise Veränderungen im Zeitablauf ausgesetzt, was durch die temporale Abhängigkeit von der Gültigkeitszeit in T-ADAPT und durch zeitabhängige

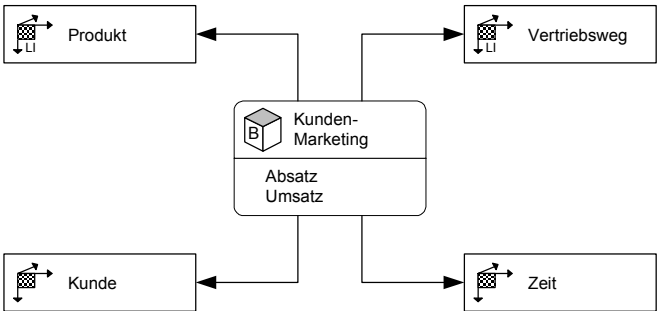


Navigationsattribute im logischen Modell des Business Information Warehouse ausgedrückt wird. Die grafische Repräsentation der Kundendimension findet sich in Abbildung 110.



**Abb. 110.** Kundendimension im BW

Die Modellierung der Kennzahlen erfolgt wieder direkt im Basis-Cube und die drei Dimensionen Produkt, Vertriebsweg und Zeit entsprechen den bisherigen Ausführungen. Das resultierende Modell für das Kunden-Marketing ist in Abbildung 111 zusammenfassend komprimiert dargestellt.



**Abb. 111.** Verdichtet dargestellter Basis-Cube des Kunden-Marketings

## 8.7 Weiterführende Literatur

Mit dem erweiterten Star Schema befassen sich zahlreiche Publikationen des Autors, so etwa (Hahne 2002a), (Hahne 2002b) und (Hahne 2003a).

Ein reichhaltiger Fundus an detaillierten Informationen zum Thema SAP und Business Information Warehouse hält der SAP Marktplatz bereit, den zugelassene Benutzer unter [service.sap.com](http://service.sap.com) erreichen. Dort gibt es u. a. auch die White-Paper (SAP 2000a) und (SAP 2000b) zur Modellierung im BW. Speziell mit der Zeitabhängigkeit befassen sich (Hahne 2003b) und (Stock u. Beekmann 2003). Die Diskussion der Aspekte der Transformation über die Ebenen der Modellierung hinweg erfolgt eingehender in (Hahne 2002c).

Allgemeiner wird das Thema Modellierung und SAP Business Information Warehouse in (Mehrwald 2003) und (Chamoni et al. 2004) behandelt.

Die Veröffentlichungen des Autors können zum Großteil auch unter [www.hahneonline.de](http://www.hahneonline.de) abgerufen werden.

- Chamoni P, Gluchowski P, Hahne M (2004) Business Information Warehouse. Springer-Verlag, Heidelberg
- Hahne M (2002a) Datenmodellierung für SAP Business Information Warehouse: Aufbau mehrdimensionaler Datenstrukturen im BW3.0 (Teil 1). itFokus 11:12-18
- Hahne M (2002b) Datenmodellierung für SAP Business Information Warehouse: Hierarchische Dimensionsstrukturen im BW3.0 (Teil 2). itFokus 12:56-59
- Hahne M (2002c) Transformation mehrdimensionaler Datenmodelle. In: von Maur E, Winter R (Hrsg) Vom Data Warehouse zum Corporate Knowledge Center. Physica-Verlag, Heidelberg, S 399-420
- Hahne M (2003a) Logische Datenmodellierung zur Abbildung mehrdimensionaler Datenstrukturen im SAP Business Information Warehouse. In: Weikum G, Schöning H, Rahm E (Hrsg) BTW 2003 Datenbanksysteme für Business, Technologie und Web, Tagungsband der 10. BTW-Konferenz, 26.-28. Februar 2003, Leipzig, Köllen Druck + Verlag, Bonn, S 630-647
- Hahne M (2003b) Time aspects in SAP Business Information Warehouse. In: Jardim-Gonçalves R, Cha J, Steiger-Garção A (Hrsg) Concurrent Engineering. A.A. Balkema Publishers, Lisse, S 69-74
- Mehrwald C (2003) SAP Business Information Warehouse 3 – Architektur, Konzeption, Implementierung. dpunkt-Verlag, Heidelberg
- SAP (2000a) Modeling with BW. ASAP for BW Accelerator
- SAP (2000b) Hierarchies in SAP BW. ASAP for BW Accelerator
- Stock S, Beekmann F (2003) Temporal aspects of data warehousing. In Jardim-Gonçalves R, Cha J, Steiger-Garção A (Hrsg) Concurrent Engineering. A.A. Balkema Publishers, Lisse, S 53-59

## 8.8 Zusammenfassung

Die Datenhaltung des Business Information Warehouse basiert auf der relationalen Datenbanktechnik und die Speicherung mehrdimensional strukturierter Informationen folgt dem Ansatz eines Star Schemas, das von der SAP AG um eigene Konstrukte erweitert wurde. Dieses so genannte erweiterte Star Schema im BW ist gekennzeichnet durch eine große Anzahl von Tabellen, über welche die Informationen verteilt abgelegt sind

Die elementaren Objekte im Business Information Warehouse sind die global definierten und über ihren technischen Namen eindeutig identifizierbaren Info-Objekte. Die im BW-Modell differenzierten betriebswirtschaftlichen Auswertungsobjekte sind zum einen die Merkmale, die den allgemeinen Dimensionselementen entsprechen, und die Kennzahlen, mit denen die Fakten beschrieben werden. Beide sind technisch gesehen Info-Objekte.

Der Info-Cube ist das zentrale Objekt zur Speicherung mehrdimensional strukturierter Daten im BW. Dass die Stammdaten übergreifend gültig sind und über alle Info-Cubes hinweg zur Verfügung stehen ist dabei ein wesentliches Grundprinzip. Die Faktentabelle bildet im erweiterten Star Schema den Mittelpunkt des Modells und ist von Dimensionen umgeben, die jeweils aus der Dimensionstabelle und den Stammdatentabellen bestehen. Dabei fungiert die Dimensionstabelle als Verbindungstabelle von der Faktentabelle zu den Stammdatentabellen, in denen sich die eigentlichen Dimensionswerte mit beschreibenden Texten sowie deren Attribute befinden. Im Falle einer als Line-Item deklarierten Dimension, die nur aus einem einzigen Merkmal besteht, fällt diese Verknüpfungstabelle weg.

Im Business Information Warehouse sind pro Info-Cube maximal 16 Dimensionen definierbar, wobei die drei Dimensionen für die Zeit, Pakete (eines Ladeprozesses) und für Einheiten (von Kennzahlen) bereits vom System vorgegeben sind. Die Ablage dieser beschreibenden Texte von Merkmalen erfolgt im Business Information Warehouse in einer separaten Texttabelle, wobei die Texte in verschiedenen im System definierten Sprachen verfasst sein können.

Die Berücksichtigung von Dimensionsattributen ist im erweiterten Star Schema über Stammdatenattribute eines Merkmals gewährleistet. Diese werden in einer separaten Stammdatentabelle eines Merkmals gespeichert. Im BW-Sprachgebrauch wird allgemein von Attributen gesprochen, wobei zwischen reinen Anzeigeattributen und Navigationsattributen zu differenzieren ist. Letztere können auch ohne das zugrunde liegende stammdatenträgende Merkmal in Abfragen berücksichtigt werden.

Für die Modellierung hierarchischer Dimensionsstrukturen gibt es im Business Information Warehouse drei verschiedene Varianten. Neben den auf Stammdaten basierenden Varianten der Abbildung über Externe Hierarchien und über Navigationsattribute gibt es noch die Bewegungsdatensicht der Gestaltung über Merkmale.

Bei der Modellierung einer Dimensionshierarchie über Merkmale in einer Dimension wird jede Konsolidierungsebene der Hierarchie auf ein Merkmal abgebildet, die Hierarchie ist somit in den Bewegungsdaten abgebildet. Daher sind im Wesentlichen nur balancierte Strukturen hierfür geeignet. Da alle Merkmale gleichberechtigt nebeneinander stehen, gibt es keine vordefinierten Navigationspfade. Strukturelle Änderungen in der hierarchischen Beziehung zwischen Merkmalen sind bei dieser Abbildungsform ohne Datenreorganisation nicht abbildbar, das Berichtswesen kann nur auf Basis der tatsächlich gebuchten Zuordnung, d. h. der historischen Wahrheit, erfolgen.

Mit der Modellierungsalternative der Hierarchieabbildung über Navigationsattribute steht ein Berichtswesen auf Basis der jeweils aktuellen Dimensionsstruktur zur Verfügung und eine Auswertung auf Basis historischer Zuordnungen kann erst mit einer zeitabhängigen Ausgestaltung der Navigationsattribute abgebildet werden. Bei dieser Variante korrespondiert jede Konsolidierungsebene außer der untersten mit einem Attribut zu dem Merkmal des Basiselementes. Strukturveränderungen sind bei dieser Stammdatensicht leicht durchführbar und greifen direkt für alle Info-Cubes. Vordefinierte Navigationspfade gibt es auch bei dieser Modellierungsform nicht.

Die bzgl. der Dimensionsstruktur flexibelste Art der Modellierung von hierarchischen Strukturen in Dimensionen im BW stellen die Externen Hierarchien dar, da sie auf einer Darstellung in Form von rekursiven Beziehungen basieren. Sie bietet sich daher insbesondere bei unbalancierten Dimensionsstrukturen an. Externe Hierarchien sind in den Stammdaten abgelegt und somit für alle Info-Cubes, die ein spezielles Merkmal verwenden, übergreifend gültig. Neben der Möglichkeit, verschiedene Hierarchien für ein Merkmal zu definieren, können einzelne Hierarchien zusätzlich in verschiedenen Versionen gepflegt sein.

Welche Gestaltungsalternative in einem konkreten Fall zu bevorzugen ist, ergibt sich im Wesentlichen aus den Berichtsanforderungen und der Berücksichtigung von Zeitabhängigkeit. Für transaktionsorientierte Berichte der so genannten historischen Wahrheit sind nur Merkmale geeignet. Die Darstellung bezogen auf die jeweils aktuell gültige Struktur ist über Navigationsattribute und über Externe Hierarchien abbildbar. Für die Realisierung von Berichten nach historischen Strukturen sind zeitabhängige

Navigationsattribute genauso wie auch zeitabhängige Externe Hierarchien geeignet.

Im erweiterten Star Schema des BW sind die physisch abgelegten Kennzahlen gerade die Wertspalten in der Faktentabelle, die somit keiner hierarchischen Strukturierung unterliegen. Die Berechnung dieser Kennzahlen erfolgt im Rahmen des ETL-Prozesses während der zumeist periodischen Aktualisierungsvorgänge des BW-Datenbestandes.

Die Abbildung umfassender Kennzahlensysteme erfolgt im Business Information Warehouse nicht auf der Ebene der physischen Datenstrukturen, sondern erst auf der Ebene der Abfragen. Diese so genannten Queries stellen im BW eine mehrdimensionale Sichtweise auf den in Info-Providern abgelegten Datenbestand dar und ermöglichen die Implementierung komplexer Berechnungen und Operationen mit berechneten und eingeschränkten Kennzahlen. Darüber hinaus gibt es auch die alternative Möglichkeit der Modellierung einer Kennzahlendimension und der Implementierung einer künstlichen Wertkennzahl.

Zur grafischen Beschreibung von mehrdimensionalen Datenmodellen des Business Information Warehouse steht eine Menge an Darstellungsobjekten zur Verfügung, die auch als Schablonen für Visio zur Verfügung stehen und damit eine leichte Visualisierung ermöglichen. Die hierzu benötigten Visio-Schablonen stehen unter [www.hahneonline.de](http://www.hahneonline.de) zum Download zur Verfügung.

## 9 Aspekte des physischen Modells

Im Rahmen des Modellierungsprozesses erfolgt die Ableitung eines logischen Schemas aus dem semantischen Modell. Dies wurde für den Fall eines T-ADAPT-Modells auf der semantischen Ebene für den Aufbau eines logischen Datenmodells des Business Information Warehouse in den bisherigen Kapiteln ausführlich beschrieben.

Mit dem nächsten Schritt erfolgt die Umsetzung des konzeptuellen Schemas in das interne Schema, also gerade die Gestaltung des physischen Modells. Neben den diversen Parametern für die Erstellung von Info-Objekten, die an dieser Stelle nicht weiter behandelt werden, sind für die physische Ebene einige Einflussfaktoren relevant, die insbesondere das Abfragezeitverhalten betreffen.

In diesem Sinne stehen die Prinzipien der physischen Modellierung im Business Information Warehouse, die einen signifikanten Einfluss auf das gesamte Zeitverhalten eines Datenbankmodells haben, in diesem Kapitel im Vordergrund. Die in diesem Kontext zu diskutierenden Konzepte sind die Multi-Provider, die Komprimierung und Partitionierung von Faktentabellen sowie der Aufbau von Aggregaten.

### 9.1 Multi-Provider

Die grundsätzliche Bedeutung von Multi-Cubes bzw. Multi-Providern wurde ja bereits in den Kapiteln 4 und 8 herausgearbeitet. Diese bilden neben dem Aufbau kombinierter neuer betriebswirtschaftlicher Sichtweisen auch eine wesentliche Säule für verschiedene Aspekte der physischen Modellebene.

Ein Multi-Provider stellt eine virtuelle Schicht ohne persistente Datenablage zur Verfügung. In ihm sind verschiedene andere Info-Provider in einem neuen Kontext zusammengeführt. Dabei können neben Basis-Cubes und Remote-Cubes auch ODS-Objekte, Info-Objekte und Info-Sets mit in die Definition eines Multi-Providers aufgenommen werden. Die Festlegung der Struktur verläuft analog zur Definition eines Basis-Cubes, so dass

insbesondere die Dimensionen und deren Merkmale festzulegen und zu berücksichtigende Navigationsattribute anzugeben sind.

Da mit der Definition eines Multi-Providers nur ein Layer festgelegt wird und keine physische Speicherung der Daten erfolgt, ist dieses Konzept sehr flexibel einsetzbar. So können die beteiligten Info-Provider eines Multi-Providers im Allgemeinen problemlos ausgetauscht werden, ohne dass dies Auswirkungen auf vorhandene Abfragen auf der Ebene des Multi-Providers hat.

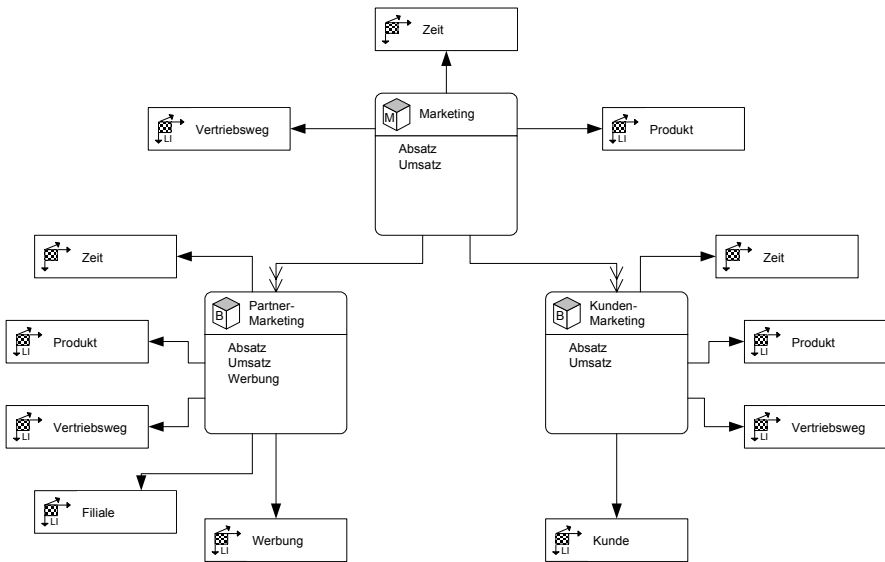
Mit Multi-Providern ist es möglich, die Daten verschiedener Info-Provider zu kombinieren. Eine selektive Eingrenzung der Daten einzelner beteiligter Info-Provider beispielsweise in einem Merkmal ist hingegen nicht möglich. Die Definition eines Multi-Providers basiert grundlegend auf der Festlegung von korrespondierenden Info-Objekten der eingeschlossenen Komponenten. Dabei können sich nur die Info-Objekte entsprechen, die gleich sind oder zumindest auf das gleiche Info-Objekt referenzieren, denn dies gewährleistet konsistente Stammdaten.

Für identische Kennzahlen der beteiligten Info-Provider erfolgt eine Summation im Multi-Provider, unterschiedliche Kennzahlen können hingegen nicht addiert werden. Eine weitere Transformation oder Berechnung ist im Konzept der Multi-Provider ebenfalls nicht vorgesehen, aber diese Funktion ist im Allgemeinen auch auf der Ebene der Query im Business Information Warehouse anzusiedeln.

Für Multi-Provider gibt es verschiedene Anwendungsbereiche:

- Verschiedene Info-Cubes, die jeweils unterschiedliche in sich abgeschlossene betriebswirtschaftliche Bereiche repräsentieren, verfügen über einen gemeinsamen Kern, der für Abfragen zur Verfügung gestellt werden soll.
- Technisch bedingt in unterschiedlichen Formen von Info-Providern (ODS-Objekte, Info-Cubes etc.) gespeicherte Daten sollen zusammengeführt werden.
- Abkopplung der physischen Struktur vom fachlich getriebenen Aufbau eines Info-Cubes und Performance-Optimierungsmaßnahmen (vgl. hierzu das 2-Layer-Konzept in Abschnitt 10.3).

Der erste Anwendungsbereich kann auch in unserem Beispiel identifiziert werden. Dort ist ja neben den detaillierten kunden- und partnerbezogenen Marketingsichten auch eine aggregierte Sicht zu berücksichtigen, die in dem bisher dargestellten Modell auch in Form eines Basis-Cubes abgedeckt ist. Genau hier ist das Prinzip des Multi-Providers anwendbar, dessen Ergebnis in Abbildung 112 in komprimierter Form dargestellt ist.



**Abb. 112.** Multi-Provider für die aggregierte Marketingsicht

Wie aus dieser Darstellung ersichtlich ist, wurde für diesen Würfel das Konstrukt eines Multi-Providers gewählt, denn die aggregierte Darstellung im Marketingbereich ergibt sich ja als übergeordnete Sichtweise aus den kundenbezogenen und partnerbezogenen Modellen.

Zunächst ist zu beachten, dass in dem Multi-Provider nur die Informationen, die in den beteiligten Basis-Cubes, im Beispiel also den beiden Würfeln für das Partner- und Kunden-Marketing, bereits vorhanden sind. Hier sind somit verschiedene Vertriebswege in einzelnen Cubes abgelegt, die über das Multi-Provider-Konzept zusammen geführt werden sollen. In der Modellierung für die partnerbezogenen Auswertungen fehlte die Vertriebsweg-Dimension bisher, für die aggregierte Sichtweise ist diese hier mit aufzunehmen, damit die einzelnen Info-Objekte korrekt zugeordnet werden können.

Die Gestaltung der Dimensionen, d. h. die Kombination der Merkmale, ist auf der Ebene des Multi-Providers isoliert von der Gestaltung der Basis-Cubes vorzunehmen. Auch für den Multi-Provider gibt es die Möglichkeit, einzelne Dimensionen als Line-Item zu deklarieren. Da in unserem Beispiel die Dimensionen identisch sind und lediglich einige Detail-Dimensionen der Basis-Cubes wegfallen, kann hier die Zusammenfassung von Merkmalen in Dimensionen direkt übernommen werden. Die einzelnen Merkmale sind bei der Definition des Multi-Providers zuzuordnen. Da es sich im Beispiel um die gleichen handelt, wird dies vom System automatisch korrekt vorgeschlagen.



Die übergeordnete Sicht im Marketing umschließt nur die beiden Kennzahlen Absatz und Umsatz, die in der Festlegung des Multi-Providers zu selektieren sind. Dabei erfolgt eine Aggregation identischer Kennzahlen in den verschiedenen beteiligten Basis-Info-Providern. Da in unserem Beispiel die Basis-Cubes bezogen auf den Vertriebsweg disjunkte Bereiche abbilden, greift hier die Info-Cube übergreifende Aggregation im Falle einer Selektion ohne Vertriebsweg. Unterschiedliche Kennzahlen können auf der Ebene des Multi-Providers nicht zusammengefasst werden, eine Addition beispielsweise ist hierüber nicht möglich.

## 9.2 Komprimierung von Faktentabellen

Dass es im erweiterten Star Schema des Business Information Warehouse eine vordefinierte Dimension für Pakete für die Berücksichtigung einzelner Ladeprozesse gibt, wurde bereits in Kapitel 8 kurz erwähnt. Die Berücksichtigung der so genannten Requests (Pakete) im Ladeprozess für Info-Cubes geschieht automatisch.

Durch die Berücksichtigung der Ladeprozesse in der Datenablage ergibt sich in der Faktentabelle ein größeres Datenvolumen als notwendig, da die betriebswirtschaftlichen Anforderungen die Berücksichtigung von Requests (Paketen) nicht erfordern. Dadurch ist es aber im Business Information Warehouse möglich, auf der Ebene einzelner Ladevorgänge Aktionen rückgängig zu machen, einzelne Ladeprozessschritte zu wiederholen und effizient Datenschiefstände zu korrigieren.

Gerade bei großen Datenvolumina und häufigen Datenänderungen kann dieses Konzept zu einer unnötig großen Faktentabelle mit den entsprechenden Implikationen auf das Abfragezeitverhalten führen. Daher bietet das System die Möglichkeit der Aggregation über die Datenpakete hinweg an. Dieses Prinzip ist mit Komprimierung bezeichnet, wobei allerdings nicht alle Requests mit komprimiert werden müssen.

Die Realisierung basiert auf dem Grundprinzip zweier Faktentabellen für jeden Info-Cube. Die so genannte F-Tabelle des Info-Cubes berücksichtigt die Requests und wird immer verwendet. Die zweite Faktentabelle ist die E-Tabelle, die ohne Bezug zu Requests aufgebaut ist und nur optional verwendet wird. Der Zusammenhang im gesamten Datenfluss ist in Abbildung 113 zusammengefasst.

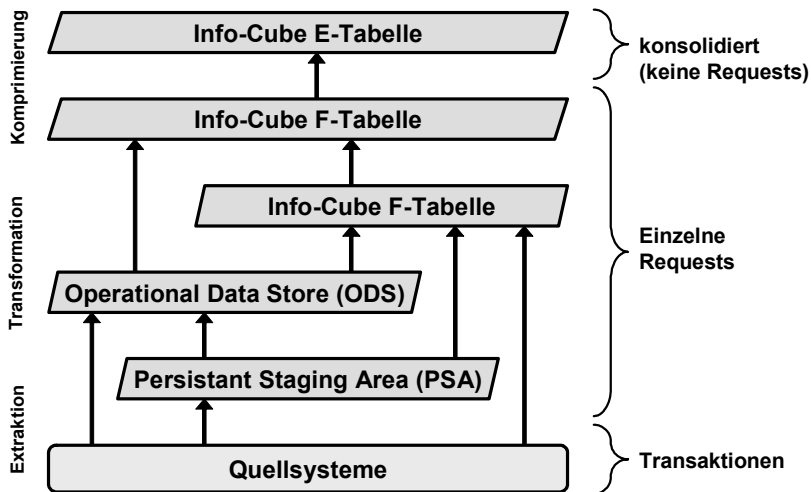


Abb. 113. Datenfluss zur Komprimierung einer Faktentabelle

Für die Komprimierung ist selektierbar, ob alle Requests oder nur diejenigen bis zu einem festgelegten Request in chronologischer Reihenfolge zu berücksichtigen sind. In Abbildung 114 ist exemplarisch die teilweise Komprimierung einer Faktentabelle dargestellt. Die Requests eins und zwei ergeben im Beispiel dann die aggregierte Teilsicht in der E-Faktentabelle, der Request drei verbleibt in der normalen F-Faktentabelle.

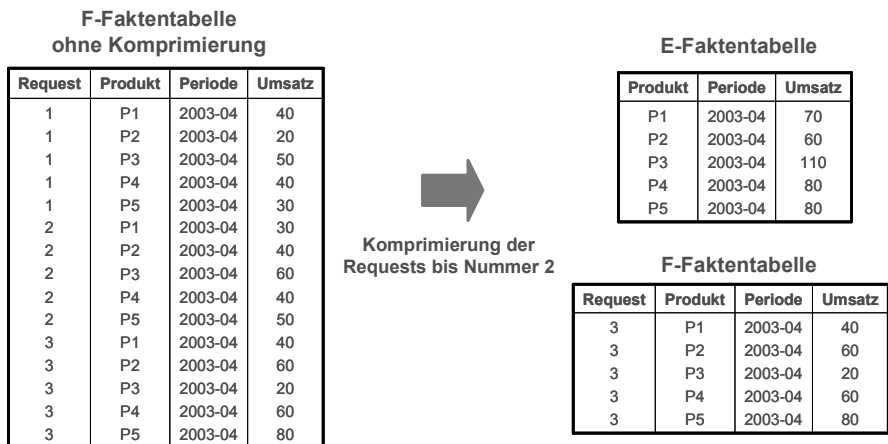


Abb. 114. Komprimierung einzelner Requests

Für den Benutzer ist diese Zweiteilung der Faktentabelle völlig transparent, so dass die Komprimierung aus Benutzersicht nicht relevant ist. Der

Vorteil der Komprimierung liegt in der reduzierten Datenmenge, die bei Abfragen zu berücksichtigen ist. Diese geht aber mit einer schlechteren Performance beim Laden von Daten einher und ist überdies nicht umkehrbar, so dass die üblichen Möglichkeiten der Qualitätssicherung und der Bereinigung von Schiefständen auf der Ebene eines Requests nicht verfügbar sind. Der aufwendige Prozess der Komprimierung hat des Weiteren erheblichen Einfluss auf das Abfragezeitverhalten und sollte in transaktionsarme Zeitfenster fallen.

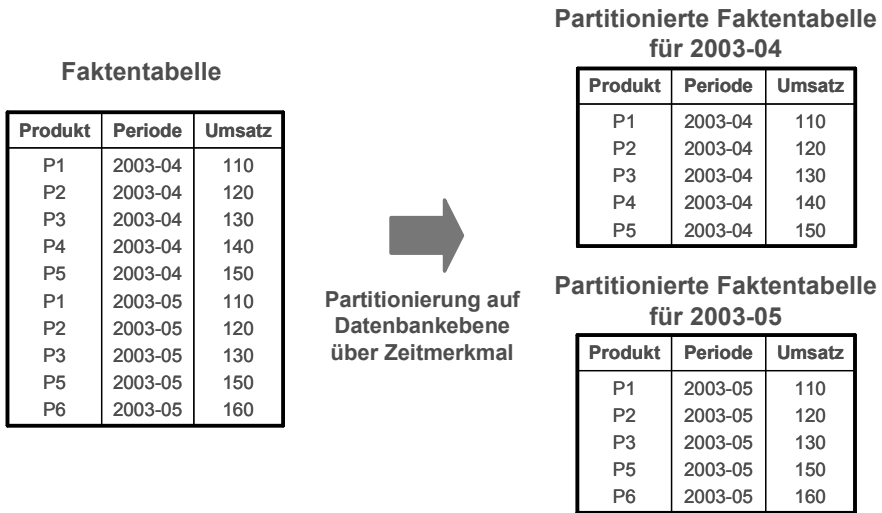
### **9.3 Partitionierung von Faktentabellen**

Die Partitionierung von Tabellen ist in relationalen Datenbanken ein mittlerweile üblicher Mechanismus zur Performance-Steigerung durch Parallelisierung. Dabei erfolgt die Aufteilung einer Tabelle in mehrere physisch getrennte Tabellen anhand der Ausprägungen einer Spalte. Auch für das Business Information Warehouse bietet sich dieser grundlegende Mechanismus zur Optimierung des Zugriffs auf besonders große Faktentabellen an. Dabei ist allerdings zu berücksichtigen, dass der Vorteil einer Partitionierung im Allgemeinen mit einem erhöhten administrativen Aufwand einhergeht.

Für das Business Information Warehouse gibt es zwei grundlegende Varianten der Partitionierung. Die erste nutzt genau die datenbankseitig zur Verfügung stehenden Möglichkeiten und eine weitere bietet diese Funktionalität auf der Anwendungsebene.

Der Mechanismus der Datenbank-Partitionierung ist nur für die Aufteilung der komprimierten Faktentabelle und dies nur anhand eines Zeitmerkmals nutzbar. Er steht für fast alle Datenbanksysteme des Business Information Warehouse zur Verfügung. Die Verteilung der Daten auf die Partitionen wird vom Datenbanksystem automatisch vorgenommen, daher kann diese Eigenschaft auch nur beim initialen Erstellen eines Basis-Cubes festgelegt werden. Die Partitionierung ist in Abbildung 115 für das Beispiel der Faktentabelle aus Kapitel 8, anhand derer schon die verschiedenen Zeitszenarien verdeutlicht wurden, dargestellt.

Die Verteilung auf verschiedene Tabellen führt durch die Möglichkeit parallelen Lesens zu einer verbesserten Abfrageperformance. Sind nur Daten einer Partition betroffen, wirkt sich die geringere Tabellengröße positiv auf das Zeitverhalten aus. Da mit diesem Konzept auch ein größerer Wartungsaufwand einhergeht, ist der Mechanismus gezielt dosiert anzuwenden.

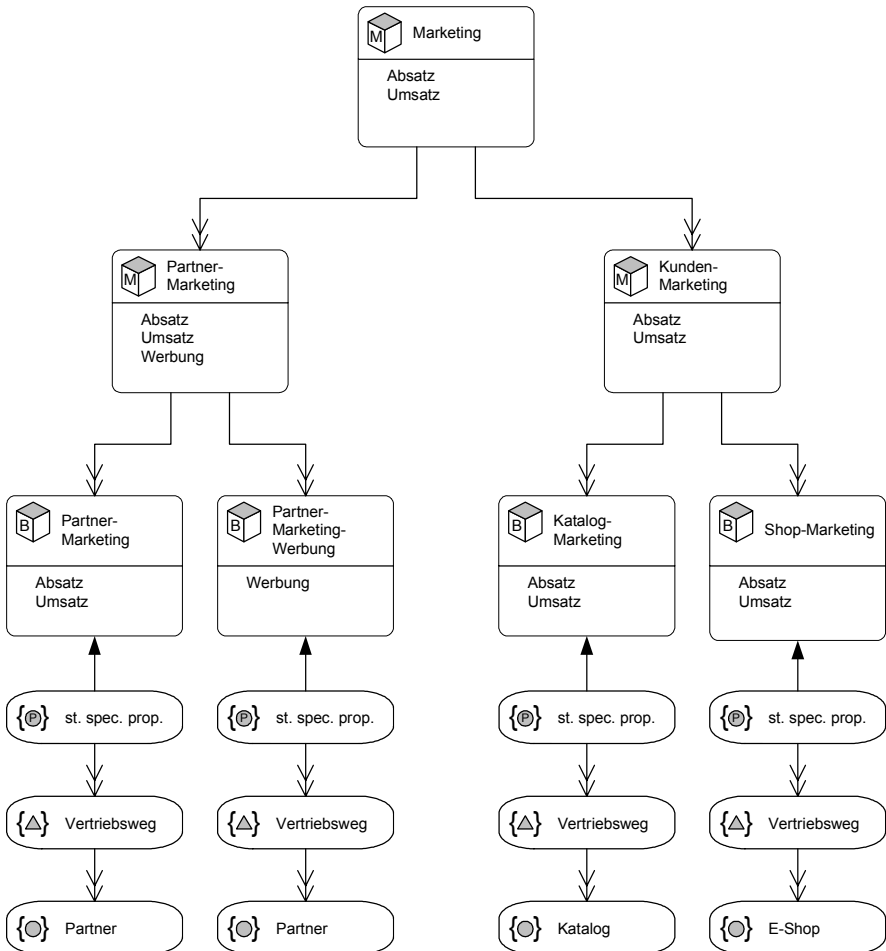


**Abb. 115.** Datenbank-Partitionierung über ein Zeitmerkmal

Die Einschränkung der Verteilung anhand eines Zeitmerkmals kann im Einzelfall natürlich gravierend sein. Bei der Partitionierung auf Anwendungsebene greift diese Restriktion nicht mehr, denn diese basiert auf manuell angelegten Info-Cubes für die einzelnen Partitionen, deren Zusammenführung über einen Multi-Provider erfolgt. Die Verbesserung der Abfrageperformance entsteht wieder durch die parallele Verarbeitung. Damit nun nicht alle Info-Cubes bei Abfragen zu involvieren sind, muss der OLAP-Prozessor von den verschiedenen Ausprägungen des Partitionierungsmerkmals Kenntnis erlangen. So können nicht relevante Info-Cubes bereits von vornherein ausgeschlossen werden.

Der Mechanismus im Business Information Warehouse, über den einzelnen Info-Cubes feste Werte für Partitionierungen zugewiesen werden können, basiert auf der Definition von strukturspezifischen Eigenschaften, die eine zusätzliche Eigenschaft von Info-Cubes festlegen. Da dieses Prinzip auf der manuellen Definition von Info-Cubes basiert, eignen sich hierfür insbesondere Merkmale geringer Kardinalität, die keinen häufigen Änderungen unterworfen sind. Für die Befüllung der Cubes mit passenden Daten zu den Partitionen ist der Datenladeprozess verantwortlich, denn dies geschieht im Gegensatz zur Datenbank-Partitionierung nicht automatisch.

Zusammen mit den Möglichkeiten des Aufbaus von Multi-Providern bietet das Partitionierungskonzept einen guten Ansatz zur Optimierung der Info-Cube-Struktur in unserem Marketing-Beispiel, dessen Ergebnis in Abbildung 116 dargestellt ist.



**Abb. 116.** Redesign der Info-Cubes im Marketing-Beispiel

Die grafische Darstellung stellt eine Erweiterung des bisher beschriebenen Ansatzes zur Repräsentation logischer BW-Modelle aus Abschnitt 8.5 dar und ist ebenfalls in den zitierten Schablonen für Visio enthalten.<sup>19</sup>

Als Merkmal für die Verteilung ist der Vertriebsweg nahe liegend, denn dieser hat nur wenige Ausprägungen und neu auftretende Vertriebswege führen ohnehin zu veränderten Anforderungen im Marketing.

Das Konzept basiert auf der Definition von strukturspezifischen Eigenschaften, die in der Abbildung grafisch an die Basis-Cubes angehängt werden. Dieses Konstrukt verbindet die Eigenschaft mit dem Merkmal der

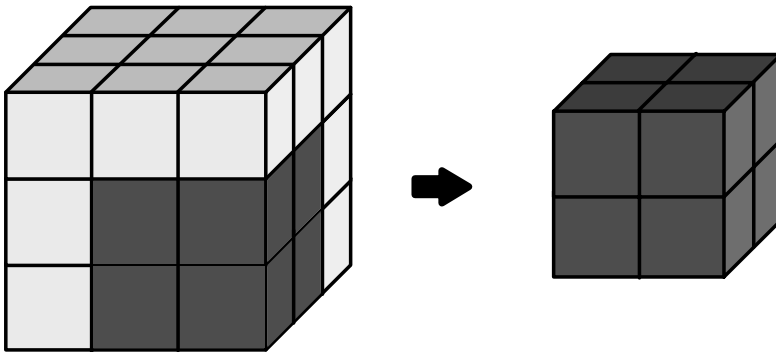
<sup>19</sup> Diese stehen unter [www.hahneonline.de](http://www.hahneonline.de) zum Download bereit.

Partitionierung und dem eingestellten Festwert. Diese sind beispielsweise die Basis-Cubes für die Vertriebswege über den Katalog und den E-Shop im kundenbezogenen Marketing. Ebenfalls erkennbar ist in der Abbildung, dass auch Multi-Provider selbst in der Definition eines Multi-Providers auftreten dürfen. So ergibt sich etwa die aggregierte Marketingsicht als virtuelle Definition über den beiden Multi-Providern für die verschiedenen detaillierten Marketingsichten.

Die Nutzung dieser Form der Partitionierung ist aufgrund der manuellen Pflege der Info-Cubes mit besonderen administrativen Aufwänden verbunden, da insbesondere jeder Info-Cube für sich bzgl. der Performance zu optimieren ist. Dies bezieht sich u. a. auf die Aggregate, die Gegenstand des folgenden Abschnittes sind. Daher ist zwischen den Vorteilen und den Nachteilen sorgsam abzuwägen.

## 9.4 Aggregate

Im Rahmen der Darstellung des allgemeinen Star Schemas ist mit dem Fact Constellation Schema in Abschnitt 7.2.3 bereits auf einen grundlegenden Mechanismus zur Speicherung vorberechneter aggregierter Werte aufmerksam gemacht worden. Einen ähnlichen Ansatz verfolgt die SAP im Business Information Warehouse mit dem Aufbau von Aggregaten, die ebenfalls eine redundante Datenhaltung auf verdichtetem Niveau vorsehen. Für den Benutzer ist dies völlig transparent, da der OLAP-Prozessor über die Eignung und Verwendung von Aggregaten befindet. Wie Abbildung 117 verdeutlicht, stellt ein Aggregat im Business Information Warehouse einen bewusst redundanten Teil eines Info-Cubes dar, der sich aufgrund einer geringeren Detaillierung oder einer Einschränkung ergeben kann.



**Abb. 117.** Aggregat als Teilmenge eines Info-Cubes

Aggregate sind sehr flexibel einsetzbar und können im laufenden Betrieb hinzugefügt oder geändert werden. Die Ausrichtung erfolgt dabei an den Bedürfnissen zur Analyse und Auswertung und stellt damit ein extrem mächtiges Werkzeug zur spezifischen Performance-Optimierung dar. Die Gewinne im Abfragezeitverhalten gehen allerdings mit einem Nachteil im Rahmen des Datenladeprozesses einher, denn der Aufbau von Aggregaten ist recht zeitaufwendig. Dieser Zielkonflikt ist in Abbildung 118 visualisiert.



**Abb. 118.** Zielkonflikt beim Aufbau von Aggregaten

Die Optimierungsmöglichkeit über Aggregate steht lediglich für Basis-Cubes zur Verfügung. Durch die Festlegung einer Teilmenge aller im Basis-Cube verwendeten Info-Objekte wird die Granularität des Aggregats festgelegt. Aggregate können auf Merkmalen, Navigationsattributen und Hierarchieknoten basieren. Eine weitere Verringerung des Datenvolumens ist mit der Festlegung eines Festwertes für ein Info-Objekt erreichbar.

Eine Besonderheit stellen zeitabhängige Navigationsattribute für Aggregate dar, da hierauf keine allgemein gültigen Verdichtungen definierbar sind, sondern diese erst durch Angabe eines Stichtages eindeutig werden. Daher sind solche Aggregate stichtagsbezogen. Für jeden zu berücksichtigenden Stichtag ist dabei ein eigenes Aggregat zu berücksichtigen. Sinnvoll ist dabei aber in jedem Fall, zumindest für den jeweils aktuellen Tag nach Systemdatum ein Aggregat vorzuhalten, da dies im Allgemeinen der meistgenutzte Standardfall für Abfragen sein wird.

Die durch Änderungen in den Stammdaten implizierten notwendigen Anpassungen an Aggregaten, etwa durch eine Änderung des Wertes eines

Navigationsattributes, das in einem Aggregat aufgenommen ist, werden nicht automatisch nachgezogen. Der hierfür notwendige anzustoßende Prozess heißt Change Run und ist oftmals sehr zeitaufwendig, so dass hier eine detaillierte Zeitplanung des Ladeprozesses mit den Abhängigkeiten des Aggregataufbaus zu erfolgen hat.

Die Speicherung von Aggregaten erfolgt im Allgemeinen ebenfalls in Form eines Tabellengerüsts wie bei Basis-Cubes mit den zwei Faktentabellen und den umschließenden Dimensionstabellen. Eine alternative Speichermöglichkeit bietet sich beim Einsatz des Microsoft SQL Servers als zugrunde liegende Datenbank an, da die in den Analysis Services verfügbaren Cubes in Form von MOLAP-Aggregaten für das Business Information Warehouse nutzbar sind. Die Möglichkeiten dieser Speichertechnik sind jedoch gegenüber der herkömmlichen Variante stark eingeschränkt. Innerhalb eines Info-Cubes ist nur eine der beiden Speicherformen möglich.

## 9.5 Weiterführende Literatur

Mit dem Thema SAP Business Information Warehouse befassen sich zahlreiche auch internationale Publikationen, die aber zum Großteil über die Stufe einer Ansammlung von Screenshots mit teilweise sehr detaillierten Erläuterungen zur Administrator Workbench nicht hinauskommen.

Hiervon setzt sich positiv das Buch (Mehrwald 2003) mit einer sehr detaillierten Darstellung der Konzepte und deren Implementierung ab. Insbesondere zu den Aspekten des physischen Modells bietet sich dieses Buch als weiterführende Lektüre an. In (Chamoni et al. 2004) stehen die Konzepte und deren Zusammenspiel im Vordergrund.

Wie auch schon in Kapitel 8 kann an dieser Stelle insbesondere auf den SAP Marktplatz als aktuelle Informationsquelle für sehr detaillierte Informationen zu den verschiedensten Aspekten hingewiesen werden. Dieser steht zugelassenen Benutzern unter [service.sap.com](http://service.sap.com) zur Verfügung.

- Chamoni P, Gluchowski P, Hahne M (2004) Business Information Warehouse. Springer-Verlag, Heidelberg
- Mehrwald C (2003) SAP Business Information Warehouse 3 – Architektur, Konzeption, Implementierung. dpunkt-Verlag, Heidelberg



## 9.6 Zusammenfassung

Auf der physischen Modellebene sind insbesondere die Konzepte der Multi-Provider, die Komprimierung und Partitionierung von Faktentabellen und der Aufbau von Aggregaten als die Hauptdeterminanten des gesamten Zeitverhaltens zu nennen.

Ein Multi-Provider stellt eine virtuelle Schicht ohne persistente Datenablage zur Verfügung. In ihm sind verschiedene andere Info-Provider in einem neuen Kontext zusammengeführt. Die Festlegung der Struktur verläuft analog zur Definition eines Basis-Cubes, so dass insbesondere die Dimensionen und deren Merkmale festzulegen und zu berücksichtigende Navigationsattribute anzugeben sind. Multi-Provider bieten sich für die Kombination unterschiedlicher betriebswirtschaftlicher Bereiche und für die Zusammenführung von technisch bedingt verschiedenen Typen von Info-Providern an. Darüber hinaus sind Multi-Provider eine wesentliche Säule der Performance-Optimierung für das in Abschnitt 10.3 diskutierte 2-Layer-Konzept der Cube-Modellierung.

Durch die automatische Berücksichtigung der Requests des Ladeprozesses in der Datenablage ergibt sich in der Faktentabelle ein größeres Datenvolumen als notwendig. Durch die Komprimierung können die auf den Request bezogenen Informationen teilweise und sogar vollständig aggregiert werden. Dies führt zu kleineren Faktentabellen und somit zu einer verbesserten Performance.

Bei der Partitionierung erfolgt die Aufteilung einer Tabelle in mehrere physisch getrennte Tabellen anhand der Ausprägungen einer Spalte mit dem Ziel der Performance-Steigerung durch Parallelisierung. Im Business Information Warehouse gibt es hierzu die zwei Varianten der datenbankseitigen Partitionierung sowie der Partitionierung auf Anwendungsebene.

Während die datenbankseitige Partitionierung auf die Verteilung anhand eines Zeitmerkmals eingeschränkt ist, greift diese Restriktion bei der Partitionierung auf Anwendungsebene nicht mehr, denn diese basiert auf manuell angelegten Info-Cubes für die einzelnen Partitionen, deren Zusammenführung über einen Multi-Provider erfolgt. Über strukturspezifische Eigenschaften der Basis-Cubes werden die Partitionen festgelegt, so dass diese vom OLAP-Prozessor optimal angesprochen werden können. Aufgrund des gesteigerten manuellen Administrationsaufwandes ist dieses Konzept insbesondere für Merkmale geringer Kardinalität und Volatilität geeignet.

Aggregate sehen eine bewusst redundante Datenhaltung eines Info-Cubes auf verdichtetem Niveau vor, die für den Benutzer völlig transparent ist, da der OLAP-Prozessor über die Eignung und Verwendung von

Aggregaten befindet. Die Ausrichtung der Aggregate erfolgt dabei an den Bedürfnissen zur Analyse und Auswertung und diese stellen damit ein extrem mächtiges Werkzeug zur spezifischen Performance-Optimierung dar. Die Gewinne im Abfragezeitverhalten gehen allerdings mit einem Nachteil im Rahmen des Datenladeprozesses einher, denn der Aufbau von Aggregaten ist recht zeitaufwendig. Im Rahmen des oftmals sehr zeitintensiven Change Runs erfolgt die Anpassung der Aggregate an veränderte Stammdaten, da etwa die Modifikationen durch Änderungen von Navigationsattributswerten Aggregate invalidieren. Hier ist im Rahmen des Zielkonfliktes zwischen schnellen Abfragen und schnellen Ladeprozessen abzuwägen. Somit besteht auch die Gefahr zu vieler Aggregate und die tatsächliche Nützlichkeit eines Aggregats ist anhand der Statistik-Informationen zu überprüfen.

## 10 Gestaltungsempfehlungen

Die wichtigsten Objekte im Business Information Warehouse zum Aufbau von analytischen Informationssystemen stellen die Info-Cubes dar, in denen die Daten mehrdimensional gespeichert vorliegen. Diese bilden die Grundlage zum Aufbau von Systemen auf Basis der Konzepte des Data Warehousing und OLAP.

Ein wesentlicher Schritt beim Aufbau von solchen Systemen ist in der Modellierung zu sehen, denn durch diese wird die erfolgreiche Nutzung und Leistungsfähigkeit bestimmt. Einerseits bestimmt das Modell die überhaupt möglichen Analysen und Auswertungen und andererseits ist dadurch auch beeinflusst, wie schnell das System diese Anfragen beantwortet, denn das implementierte Datenmodell ist ein wesentlicher Einflussfaktor für die allgemeine Performance.

Ein elementarer Aspekt beim Aufbau mehrdimensionaler Datenmodelle ist in der Abbildung hierarchischer Strukturen zu sehen. Diese definieren Navigationsmöglichkeiten und bestimmen mögliche verdichtete Sichtweisen. Hierbei sind auch Aspekte der Berücksichtigung von Veränderung in Strukturen von zentraler Bedeutung. Die Empfehlungen zur Gestaltung von Dimensionen und deren hierarchischer Strukturen werden in Abschnitt 10.1 zusammengetragen.

Hierarchien definieren die mögliche Aggregation für verschiedene Kennzahlen, die in einem Info-Cube abgebildet sind. Die verschiedenen Formen zur Modellierung von Kennzahlen im Business Information Warehouse werden in Abschnitt 10.2 abgegrenzt.

Der Modellierung von Info-Cubes kommt eine zentrale Bedeutung zu, denn diese stellen den wichtigsten Informationsspeicher im BW dar. Insbesondere Aspekte der Wartbarkeit und der Abfragegeschwindigkeit stehen bei der Diskussion verschiedener Ausgestaltungsformen in Abschnitt 10.3 im Vordergrund.

## 10.1 Modellierung von Dimensionsstrukturen

Zur Datenmodellierung für analytische Informationssysteme ist die bisherige Einteilung in die semantische, logische und physische Ebene hilfreich. Diese Einteilung dient auch als Strukturierungsgrundlage für die Darstellung von Empfehlungen für die Modellierung von hierarchischen Strukturen und für den Aufbau von Dimensionen.

Auf der in Abschnitt 10.1.1 diskutierten semantischen Ebene stehen Aspekte der Ergonomie und Benutzbarkeit sowie die Anforderungen auf fachkonzeptioneller Ebene im Vordergrund. Die logische Modellebene reflektiert die Anforderungen in einem konkreten Modell und wird in Abschnitt 10.1.2 dargestellt. Hauptaspekte sind dabei die grundsätzlichen Möglichkeiten der Abbildung eines Modells und von Dimensionsstrukturen. Aspekte der Abfragegeschwindigkeit und der allgemeinen Performance stehen bei der Diskussion der physischen Modellebene in Abschnitt 10.1.3 im Vordergrund.

### 10.1.1 Aspekte der semantischen Modellierung

Auf der fachkonzeptionellen Ebene stehen Aspekte der Benutzerfreundlichkeit und der Usability im Vordergrund. Damit ein mehrdimensionales Modell für den Fachanwender intuitiv bedienbar wird, sind einige Grundregeln zu beachten. Diese sind recht allgemein gehalten und im Einzelfall kann eine andere Vorgehensweise durchaus sinnvoll sein.

Für den Anwender spiegelt jede Dimension eine Gruppierung sachlogisch zusammengehöriger Umweltobjekte, die den gleichen Geschäftsaspekt betreffen, wider. Damit das Gesamtmodell nicht unübersichtlich erscheint, hat sich in der Praxis eine Anzahl von vier bis acht Dimensionen für ein Modell bewährt. Eine Größenordnung von zehn Dimensionen kann sicherlich als ein noch vertretbarer Maximalwert angesehen werden.

Bei den Dimensionen sind die Ausprägungen der elementbestimmten und der ebenenbestimmten Formen zu unterscheiden. Erstere bestehen aus wenigen aufzählbaren Elementen. Demgegenüber sind bei den Mengendimensionen viele Dimensionselemente zu berücksichtigen. Insbesondere für diese Form einer Dimension spielen die Konsolidierungspfade in Form von Dimensionshierarchien eine wesentliche Rolle. Die verschiedenen Varianten von vorkommenden Strukturen haben die Identifikation von einzelnen Ebenen in der Hierarchie gemeinsam. Für die Anzahl der Hierarchiestufen insgesamt in einer Hierarchie ist eine Obergrenze von sieben anzuraten.

Eine der wichtigsten Operationen zur Analyse in mehrdimensionalen Datenstrukturen stellt die Navigation entlang dieser Konsolidierungshierarchien dar. Bei einem drill-down wird ein Element einer festgelegten Ebene aufgeklappt und die untergeordneten Elemente der nächst tieferen Ebene bilden den Fokus der weiteren Analyse. Die Anzahl der dabei erscheinenden Elemente sollte aus Gründen der Übersichtlichkeit und der intuitiven Navigation auf maximal zwanzig beschränkt sein. Diese definiert somit die maximale Anzahl der Elemente je Konsolidierungselement.

Die Elemente einer Konsolidierungshierarchie stehen Ebenenweise in einer 1:n-Beziehung zueinander. Objekte die in einer 1:1-Beziehung zueinander stehen sind tendenziell ungeeignet für eine Dimension. Es handelt sich dabei meistens um Attribute, deren Modellierung dann auch nicht als eigene Dimension erfolgen sollte. Ebenfalls ungeeignet sind im allgemeinen Objekte, die in einer m:n-Beziehung zueinander stehen und besser als zwei eigenständige Dimensionen abgebildet werden.

### 10.1.2 Empfehlungen auf logischer Modellebene

Viele Faktoren beeinflussen die Art der Modellierung von Dimensionsstrukturen im Business Information Warehouse auf der logischen Modellebene. Diese ist ja noch losgelöst von der physischen Implementierung zu sehen und die wesentlichen Freiheitsgrade der Modellierung bestehen in der Frage der Abbildung von Dimensionsstrukturen in Form von Merkmalen, Attributen oder Externen Hierarchien sowie in der Frage der Dimensionsauswahl bzw. Zusammenfassung von Merkmalen zu Dimensionen.

Um hier Gestaltungsempfehlungen geben zu können sind verschiedene Aspekte zu berücksichtigen. Einen wesentlichen Einfluss haben dabei die Anforderungen der Benutzer. Detailliert zu betrachten sind die Aspekte der Historisierung, womit die Berücksichtigung von Zeitabhängigkeit und der Umgang mit Strukturveränderungen sowie mögliche Datenreorganisationen umschrieben sind. Die Frage der Benutzerführung durch vorgegebene Navigationspfade ist ebenso zu betrachten wie die Möglichkeiten der Modellierung der verschiedenen Typen von Dimensionsstrukturen wie beispielsweise parallele Hierarchien oder unbalancierte Strukturen. Auch die Frage, für welche Info-Cubes diese Modellierung gültig ist, im Business Information Warehouse innerhalb eines Systems also die Frage, ob die Stammdaten oder die Bewegungsdaten die hierarchische Information tragen, ist in Betracht zu ziehen.

Die Analyse und Berichtsanforderungen definieren unter dem Gesichtspunkt der Zeitabhängigkeit, welche Zeitszenarien zu berücksichtigen sind. Das Datenmodell muss demzufolge genau diese Szenarien ermöglichen.

Die hier wesentlichen zu differenzierenden Varianten sind die Bewegungsdatensicht und die Stammdatensicht. Aus dem semantischen Modell ist die Frage der Historisierung im Beispiel einer Darstellung mit der Methode T-ADAPT bereits im Kern ableitbar. Handelt es sich bei der Hierarchie um eine transaktionsorientierte historisierte Variante, kommt im Business Information Warehouse lediglich die Implementierung über Merkmale in einer Dimension in Frage und jede Ebene der Hierarchie korrespondiert dann mit einem Merkmal in der Dimension. Die beiden Varianten der Modellierung auf Basis von Stammdaten in Form von Attributen oder mit Externen Hierarchien sind prinzipiell nicht in der Lage, diese Sicht zu ermöglichen. Diese stellen die jeweils aktuelle Sicht auf hierarchische Strukturen zur Verfügung. Das Berichtsszenario der aktuellen Sichtweise ist damit standardmäßig abgebildet. Die Berücksichtigung von definierten historischen Varianten ist in beiden Formen über Zeitabhängigkeit modellierbar, deren Auswahl anhand des Schlüsseldatums der Abfrage erfolgt. Jede Abfrage bestimmt damit, welcher Zeitpunkt der Gültigkeit aller zeitabhängigen Objekte heranzuziehen ist. Bei Externen Hierarchien gibt es darüber hinaus die Möglichkeit, einzelne Versionen direkt zu definieren und in der Abfrage auszuwählen.

Neben der Frage der möglichen Berichtsanforderungen ist auch der Aspekt der Datenreorganisation im Falle der Strukturveränderung zu betrachten. Bei den Formen der Modellierung in den Stammdaten ist eine Veränderung leicht möglich, da lediglich eine Externe Hierarchie oder ein Attribut eines Merkmals zu verändern ist. Dies ist sowohl manuell wie auch über die Ladeprozesse des Business Information Warehouse sehr leicht zu implementieren. Strukturelle Veränderungen der Merkmale einer Dimension selbst gehen immer einher mit einer aufwendigen Reorganisation des Info-Cubes, da hier die Bewegungsdaten neu geladen werden müssen.

Die Unterscheidung der Sichtweise auf Basis von Bewegungs- und Stammdaten impliziert einen weiteren Unterschied in den verschiedenen Modellierungsformen von hierarchischen Strukturen. Im Business Information Warehouse sind die Stammdaten systemweit gültig, sie beziehen sich demzufolge gleichermaßen auf alle Info-Provider. Externe Hierarchien und Navigationsattribute hängen im Datenmodell des BW an den Info-Objekten. Erstere stehen damit in allen Info-Providern als Hierarchien eines Info-Objektes zur Auswertung zur Verfügung. Bei den Navigationsattributen gibt es noch die Möglichkeit, diese je Info-Provider als sicht- und damit nutzbar zu aktivieren. Die konkreten Ausprägungen sind aber für alle Info-Provider identisch.

Die Frage der Benutzerführung ist vor allem bei der Navigation entlang von Konsolidierungshierarchien relevant. Nur Externe Hierarchien haben

eine fixe vorgegebene Navigationsstruktur im Sinne einer auf- und zuklappbaren Hierarchie. Für Merkmale und Attribute ist dies nur in der Query über die hierarchische Ansicht von Zeilen oder Spalten möglich, womit aber die Flexibilität in der Kombination von hierarchischen und flachen Elementen innerhalb des Queryaufbaus teilweise verloren geht. Hier hängt viel davon ab, ob eine vorgegebene Navigationsstruktur hierarchisch dargestellt werden soll oder eine tabellarische Sichtweise bevorzugt wird. Bei der Implementierung über Merkmale und Attribute ist auch die Selektion einzelner Ebenen der Hierarchie in der Darstellung möglich, dies geht bei Externen Hierarchien nicht.

Auf der semantischen Modellebene erfolgte die Darstellung verschiedener möglicher Typen von Dimensionsstrukturen. Die einfachen klassischen Strukturen wie balancierte Baum- und Waldstrukturen sind im Business Information Warehouse problemlos mit allen drei Formen der Modellierung über Merkmale, Attribute und Externe Hierarchien abbildbar. Unbalancierte Strukturen sind optimal in Form Externer Hierarchien implementierbar. Bei Navigationsattributen und Merkmalen impliziert eine unbalancierte Struktur, dass für die Elemente, denen das Konsolidierungselement einer Ebene fehlt und die somit direkt einer höheren Ebene zugeordnet sind, das entsprechende Element der ausgelassenen Ebene keinen Wert hat. In Abfragen auf dieser Ebene erscheinen diese Elemente dann unter den nicht zugeordneten Elementen. Werden diese in der Darstellung nicht mit berücksichtigt, gibt es Abweichungen in der Ergebniszeile, die unter Umständen für Verwirrung beim Anwender sorgen können. Die Varianten der Modellierung auf Basis von Merkmalen und Attributen sind somit für die Darstellung unbalancierter Strukturen eher ungeeignet.

Ein weiterer Dimensionstyp basiert auf m:n-Beziehungen in der Struktur zwischen einzelnen Ebenen. Dies ist sehr gut über Externe Hierarchien abbildbar, da hierbei auch eine konsistente Verdichtung gewährleistet ist. Dies allerdings unter der Voraussetzung, dass die Verdichtung nicht auf einer anteiligen Verrechnung basiert. Auch über Merkmale in einer Dimension ist diese Struktur im Business Information Warehouse theoretisch implementierbar, da dies gerade eine der Stärken des erweiterten Star Schemas ist. Eine konsistente Verdichtung ist dabei allerdings nur dann einfach zu implementieren, wenn es sich um eine anteilige Verrechnung handelt, bei der der Wert eines Elementes komplett auf die übergeordneten Ebenen anteilig verteilt wird und Zeitabhängigkeit nicht zu berücksichtigen ist. Die Verteilungsschlüssel sind dabei über die Bewegungsdaten festgelegt und können im Nachhinein nicht geändert werden. Über Attribute ist die Modellierung von m:n-Beziehungen grundsätzlich nicht möglich. Die Frage, ob es sich bei dieser Struktur eher um zwei Dimensionen als

um eine Hierarchie handelt, ist bereits auf der semantischen Ebene zu entscheiden.

### 10.1.3 Aspekte des physischen Modells

Die Empfehlungen zur Dimensionierung mit Fokus einer performanten physischen Implementierung stehen oftmals in einem Zielkonflikt zu einer geeigneten Modellierung auf semantischer Ebene. Dies stellt aber kein unlösbares Problem dar, vielmehr gibt es im Business Information Warehouse auch hierfür adäquate Ansätze, um Aspekte der physischen Ebene mit den Anforderungen des semantischen Modells konfliktfrei zu vereinbaren. Hierzu wird in Abschnitt 10.3 ein Konzept vorgeschlagen.

Für eine effiziente Implementierung eines mehrdimensionalen Modells im Business Information Warehouse sind zwei Aspekte ausschlaggebend: Das Abfragezeitverhalten und die Performance beim zumeist periodischen Datenaktualisierungsvorgang. Die Antwortzeiten für Abfragen und Analysen sind aufgrund der direkten Betroffenheit der Endbenutzer ein besonders kritisches Kriterium für die Akzeptanz beim User sowie seine erfolgreiche Nutzung des Systems. Ein zeitnahe Aufbau aktueller Datenbestände wird vielfach vorausgesetzt. Dabei wird aber oftmals vergessen, dass die hierfür zur Verfügung stehende Zeit limitiert ist und daher ein möglichst schneller Aktualisierungslauf notwendig ist.

Die Ausgestaltung von Dimensionshierarchien hat im Wesentlichen Einfluss auf das Abfragezeitverhalten. Aspekte des ETL-Prozesses spielen allerdings auch für die Veränderungen in Konsolidierungshierarchien eine Rolle.

Im Rahmen der Diskussion der Modellierung von Dimensionshierarchien im Business Information Warehouse sind mehrere Facetten zu betrachten. Zunächst spielt die Anzahl von Dimensionen eines Info-Cubes und deren Zusammensetzung auf Basis von Merkmalen eine Rolle. Des Weiteren sind die verschiedenen Formen der Abbildung von hierarchischen Strukturen zu analysieren.

Zur Beurteilung der Vor- und Nachteile bezogen auf das Abfragezeitverhalten von Modellen ist es zunächst erforderlich zu verstehen, wie der BW-Server eine Berichts- oder Navigationsanforderung verarbeitet. Innerhalb der gesamten Architektur ist der OLAP-Prozessor für die Übersetzung dieser Anforderungen in Abfragen an die zugrunde liegende Datenbank zuständig. Da die physische Ebene der Datenspeicherung eine relationale Datenbank ist, bestehen diese Abfragen aus SQL-Statements (*Structured Query Language, Standardabfragesprache bei relationalen Datenbanken*). Damit können verschiedene Modellierungsvarianten anhand der entste-



henden verschiedenen SQL-Abfragen beurteilt werden. Für die Beurteilung der Performance dieser Statements gibt es zwei signifikante Kriterien. Zum einen steigt die benötigte Abfragezeit überproportional mit der Anzahl der über einen Join verknüpften Tabellen. Der zweite ausschlaggebende Faktor ist die Größe der beteiligten Tabellen. Tendenziell sinkt die Performance mit der Anzahl der Tupel in einer Tabelle. Die verschiedenen Formen der Abbildung von Dimensionsstrukturen im Business Information Warehouse führen auch zu unterschiedlichen Abfragewegen, deren differenzierte Analyse im Rahmen der Diskussion verschiedener Zeitszenarien in Abschnitt 8.3 erfolgte.

Aus dem komplexen Aufbau des erweiterten Star Schemas ist direkt ableitbar, wie viele Tabellen bei Abfragen mit zu berücksichtigen sind. Die zentrale Faktentabelle ist dabei im Allgemeinen verknüpft mit anderen Tabellen beteiligt. Die Komplexität liegt bei diesem Modell jedoch in der Ausgestaltung der Dimensionen. Zunächst erfolgt über die Dimensionstabelle eine Umsetzung der künstlichen Schlüssel von der Dimension zu den beteiligten Merkmalen. Weitere beteiligte Tabellen sind je nach der Art der abgefragten Objekte verschiedene SID-Tabellen die ihrerseits weiter verbunden sind mit den eigentlichen Stammdatentabellen. Ein wesentlicher Gewinn im Abfragezeitverhalten ist bei dieser Konstruktion dadurch erzielbar, dass die Dimensionstabelle im Falle eines einzelnen Merkmals in der Dimension keine wirkliche Funktion mehr hat und daher obsolet ist. Diesem Umstand trägt das System dadurch Rechnung, dass derartige Dimensionen als Line-Item deklariert werden können. Dann entfällt im physischen Modell diese Umsetzungstabelle. Da insbesondere Merkmale mit vielen Ausprägungen performancekritisch sind, ist es ein guter Ansatz, möglichst viele der größten Merkmale in einer eigenen Line-Item Dimension zu modellieren. Dabei ist es aus Sicht des physischen Modells kein Nachteil, die maximale Anzahl von Dimensionen im Business Information Warehouse komplett auszuschöpfen. Dass dies für den Anwender transparent gestaltbar ist ergibt sich in Abschnitt 10.3.

Sofern die Variante der Dimensionsmodellierung nicht bereits aufgrund der Anforderungen fix vorgegeben ist, hat hier auch die Performance einen Einfluss auf die Entscheidung, welche Form in einem spezifischen Anwendungsfall besser geeignet ist. Die Bewegungsdatensicht ist tendenziell eine sehr performante Implementierung hierarchischer Strukturen, aber die Entscheidung für diese Sichtweise ergibt sich zumeist schon direkt aus der Anforderung einer transaktionsorientierten Sichtweise im Kontext der Zeitabhängigkeit. Von den beiden Varianten basierend auf einer Stammdatensicht sind die Navigationsattribute im Allgemeinen performanter als die Externen Hierarchien, deren Einsatz im konkreten Einzelfall bezogen auf die Performanz kritisch zu hinterfragen ist, da mit diesen deutlich mehr

Lesezugriffe bei Abfragen einhergehen. Zur Verbesserung des Abfragezeitverhaltens im Falle der Attribute und der Externen Hierarchien ist u. a. der Aufbau von geeigneten Aggregaten empfehlenswert und oftmals zwingend notwendig.

Die Aspekte der Datenaktualisierung spielen für die Modellierung von Dimensionshierarchien eine untergeordnete Rolle. Zu berücksichtigen ist in jedem Fall, dass eventuell notwendige Aggregate aufgrund einer Modellierung über Attribute oder Externe Hierarchien den Aktualisierungsprozess verlangsamen und dieser dadurch u. U. eine zeitkritische Größe erreicht. Reorganisationsläufe der Daten sind nach Möglichkeit grundsätzlich zu vermeiden.

Einen sehr großen Einfluss hat der Prozess der Datenaktualisierung aber im Rahmen der Gestaltung von Info-Cubes und wird daher in Abschnitt 10.3 wieder aufgegriffen.

## **10.2 Abbildung von Kennzahlen**

Die Modellierung von Kennzahlen und der funktionalen Abhängigkeiten zwischen ihnen, die in Form von Kennzahlensystemen beschrieben sind, erfolgt im Wesentlichen auf der semantischen Ebene. Dort wird festgelegt, welche Kennzahlen als elementare Basisgröße auftreten und welche sich aus einer Berechnung von anderen Größen ergeben.

Die Abbildung der Basiskennzahlen erfolgt im Business Information Warehouse in jedem Fall als eigene Kennzahl, die in die Info-Cubes aufgenommen wird. Für besondere Additivitätseigenschaften steht mit der Ausnahmeaggregation ein Mechanismus zur Verfügung, diese ebenfalls zu implementieren.

Für die Abbildung abgeleiteter Größen etwa eines umfassenden Kennzahlensystems gibt es im Business Information Warehouse alternative Vorgehensweisen. Ein sehr mächtiges Werkzeug stellen die berechneten und eingeschränkten Kennzahlen im Rahmen der Query-Definition bereit. Diese sind aber für die Berechnung limitiert auf den innerhalb der Query selektierten Datenbereich. Da die Abarbeitung im OLAP-Prozessor zur Laufzeit stattfindet, sind hier umfangreiche Kalkulationen mit tendenziell schlechterer Performance verbunden. Darüber hinaus erfolgt nach jedem Navigationsschritt eine erneute Berechnung.

Bei Berechnungsvorschriften, die sich bzgl. der Summation als Standard-Verdichtung verträglich verhalten, können alternativ auch schon im Rahmen der Datenladeprozesse implementiert werden. Hierzu sind diese Größen als Basis-Kennzahl mit aufzunehmen, deren Berechnung beim Be-

füllen der Info-Cubes erfolgt. Die damit einhergehende steigende Größe der betroffenen Faktentabellen ist dabei als Nachteil abzuwägen gegen den Vorteil der vorberechneten Kennzahlenwerte. Eine Veränderung der Berechnungslogik ist bei diesem Konzept allerdings mit einer umfassenden Datenreorganisation verbunden.

Ein alternativer Ansatz der Modellierung von Basis-Kennzahlen ergibt sich durch die Modellierung einer Kennzahlendimension und die Aufnahme der Werte in einer künstlichen Kennzahl im Info-Cube. Dieser Ansatz ist nur dann möglich, wenn alle Kennzahlen des betrachteten Systems die gleichen Eigenschaften bzgl. des Datentyps, der Einheit und des Aggregationsverhaltens haben. Dies ist im Allgemeinen recht selten der Fall. Diese Form ist verbunden mit einer längeren und zugleich schmaleren Faktentabelle, bietet aber gegenüber der herkömmlichen Modellierungsvariante den Vorteil, dass evtl. auftretende Nulleinträge für nicht vorhandene Kennzahlenwerte wegfallen. Hier bieten sich aber andere Optimierungsmöglichkeiten auf Basis von Partitionierungen und Multi-Providern als grundsätzliches Cube-Konzept eher an, so dass auf die Modellierung von Kennzahlendimensionen im Business Information Warehouse im Allgemeinen verzichtet werden kann.

## **10.3 Gestaltung von Info-Cubes**

Der Aufbau eines soliden Info-Cube Konzeptes ist eine elementare Voraussetzung für die erfolgreiche Nutzung eines BW-Systems. Im Spannungsfeld der fachlichen Anforderung einerseits und der performanten Datenhaltung andererseits kommt es auf ein stimmiges Gesamtkonzept an.

Die Darstellung des hier vorgeschlagenen Grundprinzips einer zweischichtigen Struktur erfolgt in Abschnitt 10.3.1. Die erste dargestellte Ebene in diesem Konzept hat die Anforderungen des Fachanwenders im Fokus und wird in Abschnitt 10.3.2 behandelt. Die im Abschnitt 10.3.3 dargelegte zweite Ebene des diskutierten Ansatzes stellt die physische Optimierung und Fragen der Performance in den Mittelpunkt.

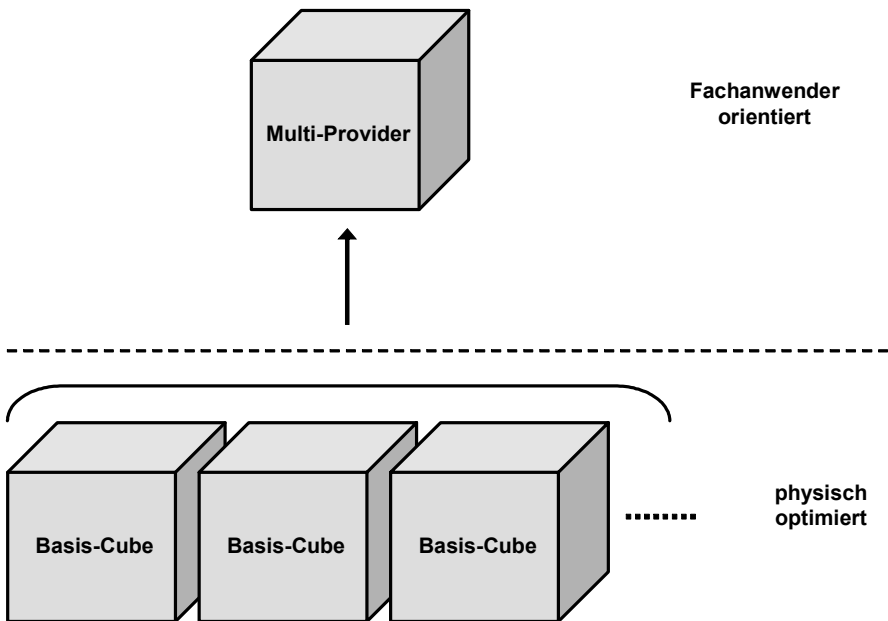
### **10.3.1 Grundprinzip der Cube-Modellierung**

Für den Anwender stellen die Info-Provider einen in sich geschlossenen thematischen Bereich dar, auf denen Abfragen und Analysen aufsetzen. Dem mehrdimensionalen Verständnis der Fachanwender kommen dabei die Info-Cubes am nächsten und diese umspannen eine betriebswirtschaftliche Anwendungsdomäne. Demzufolge erwartet der Benutzer eine Daten-

struktur als Grundlage für seine Analysen und Auswertungen die seinem fachlichen Verständnis entspricht.

Zu diesem fachlich getriebenen Anspruch steht die Frage der performanten Implementierung in Form einer persistenten Datenablage im BW oftmals im Zielkonflikt. Aspekte der physischen Implementierung sind stark technisch getrieben. Daher macht es Sinn, diese zwei Ebenen strikt zu trennen. Das hierzu vorgeschlagene Referenzmodell basiert im Kern auf der Möglichkeit des Aufbaus von Multi-Providern und der Differenzierung der Ebenen der fachlichen Anforderung und der physischen Ablage.

Wie bereits in Abschnitt 9.1 ausgeführt wurde, bietet das Business Information Warehouse mit dem Konzept des Aufbaus von Multi-Providern ein mächtiges Gestaltungselement. Multi-Provider bieten die Möglichkeit, Daten verschiedener Info-Provider in einen neuen gemeinsamen Kontext zu stellen. Den Info-Cubes als zentraler Datenablage im Business Information Warehouse kommt auch in diesem Konzept eine tragende Rolle zu. Die vormals in älteren Versionen des BW als Multi-Cubes bezeichneten Zusammenschlüsse verschiedener Info-Cubes ermöglichen im Kern bereits den Aufbau entsprechender Konzepte (siehe Abbildung 119).



**Abb. 119.** 2-Layer-Konzept der Cube-Modellierung

Die Fragen der Modellierung und der Gestaltungsempfehlungen sind demzufolge für die von den fachlichen Anforderungen und der physisch

optimierten Datenablage getriebenen Ebenen getrennt zu betrachten. Wie die Prinzipdarstellung in Abbildung 119 verdeutlicht, wird im Rahmen der Referenzstruktur die Ebene der Info-Cubes für den Fachanwender durch Multi-Provider implementiert. Zum einen bietet dieses die Möglichkeit, einzelne fachliche Sichten zu implementieren, es können aber auch verschiedene Anwendungsbereiche integriert werden.

Dass im Rahmen dieses Konzeptes die fachlichen Anforderungen adäquat Berücksichtigung finden können basiert auf der Möglichkeit und zugleich auch Notwendigkeit, dass die Frage der Modellierung von Dimensionen und der Gruppierung von Merkmalen zu Dimensionen für jeden einzelnen Multi-Provider erneut zu beantworten ist. Der Aufbau der Basis-Cubes kann sich demzufolge komplett an den Anforderungen der physisch optimierten Datenablage orientieren und hat demnach Aspekte der Performance und der Wartbarkeit im Vordergrund.

### 10.3.2 Modellierung auf der Fachanwender orientierten Ebene

Für den Benutzer ist es unerheblich, ob seine Analysen und Abfragen auf Basis eines Multi-Providers oder eines Basis-Cubes erfolgen. Seine Ansprüche an den thematisch umrissenen Bereich hat er im Rahmen des semantischen Modells definiert und demzufolge ist es die primäre Aufgabe der Info-Provider, auf denen der Benutzer seine Auswertungen durchführt, diese Anforderungen abzudecken. Die Gestaltungsrichtlinien der Modellierung auf der semantischen Ebene sind somit hier als Grundlage zu betrachten.

Das Fachmodell definiert eine mehrdimensionale Struktur als Grundlage für Analysen und Berichte, die betriebswirtschaftliche Kennzahlen nach Dimensionen aufgliedert. Jede der Dimensionen ihrerseits beinhaltet nun eine oder mehrere verschiedene Hierarchien, in denen die Informationen über Konsolidierungspfade hinterlegt sind. Die Frage der Modellierung von Kennzahlen stand in Abschnitt 10.2 im Vordergrund. Für die Modellierung der Dimensionen sind die Empfehlungen des semantischen Modells übertragbar.

Jede Dimension des semantischen Modells korrespondiert auch im Multi-Provider im Business Information Warehouse mit einer Dimension. Die Merkmale dieser Dimension sind gerade die Ebenen der Hierarchien, die sich aus der Bewegungsdatensicht ergeben. Die Frage der Zeitabhängigkeit und der zu berücksichtigenden Szenarien der Berichtsanforderungen im Kontext struktureller Änderungen in Konsolidierungshierarchien ist somit auch für die Dimensionierung der Cubes im BW essentiell.

Hierarchien, die eine nicht-transaktionsorientierte Sichtweise implementieren, sind im Business Information Warehouse in Form von Attributen und als Externe Hierarchien in den Stammdaten abgebildet. Dies ist detailliert in Abschnitt 8.2 dargestellt. Die Ableitung eines semantischen Modells auf Basis der Methode ADAPT bzw. T-ADAPT stand in Abschnitt 8.5 im Vordergrund und die dort aufgeführten Grundsätze sind hier ebenso anzuwenden.

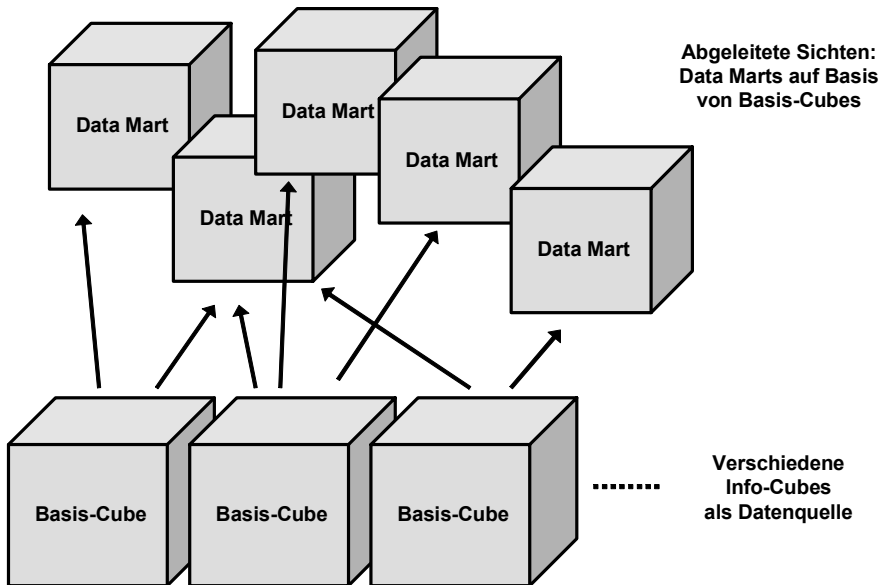
Die Ebene des in Form von Multi-Providern implementierten, am Fachanwender orientierten und auf einen betriebswirtschaftlichen Anwendungsbereich eingegrenzten Modells ist somit streng am semantischen Modell ausgerichtet. Dabei ist es im Business Information Warehouse kein Nachteil, wenn viele Multi-Provider aufgebaut werden. Es besteht also ohne Performance-Nachteile die Möglichkeit, viele verschiedene Sichtweisen in Form dedizierter Multi-Cubes aufzubauen. Maßgeblich hierfür sind ausschließlich die Anforderungen des Anwenders.

Für den Aufbau dieser fachlich fokussierten Cubes gelten somit wieder die folgenden Empfehlungen:

- Im Business Information Warehouse sind in jedem Cube bereits drei Dimensionen vordefiniert, wovon jedoch die Paket- und Einheiten-Dimension technischen Charakter haben und für die Sicht des Fachanwenders nicht relevant sind. Die Zeitdimension ist obligatorisch und von den verbleibenden 13 Dimensionen ist eine Zahl von fünf bis sieben Dimensionen als optimal anzusehen.
- Merkmale, die in keiner 1:n-Beziehung stehen sind ungeeignet für die Aufnahme in einer Dimension, 1:1-Beziehungen implizieren den Charakter eines Attributes und m:n-Beziehungen zwischen Merkmalen deuten auf verschiedene Dimensionen zur Modellierung hin.
- Für alle Formen hierarchischer Strukturen im Business Information Warehouse, somit also sowohl für die Abbildung in den Stammdaten über Navigationsattribute und über Externe Hierarchien als auch für die hierarchischen Beziehungen zwischen Merkmalen in einer Dimension, sollten maximal sieben Ebenen und optimalerweise 15-20 Elemente je Konsolidierungselement modelliert werden.

Neben der Möglichkeit, die physische Implementierung eines Info-Cubes von seiner Darstellung für den Fachanwender zu trennen, bietet das Konzept der Multi-Provider einen weiteren wesentlichen Nutzen. Durch die Kombination verschiedener Info-Cubes, oder ganz allgemein verschiedener Info-Provider, ist die Abbildung aggregierter übergeordneter Sichten, die verschiedene fachliche Sachverhalte miteinander in Beziehung setzen, möglich.

Auch auf der Ebene des semantischen Modells auf Basis der Methode T-ADAPT standen in diesem Zusammenhang bereits Modelle bzw. abgeleitete Teilsichten als weitere Modellierungsbausteine zur Verfügung. Diese entsprechen im Allgemeinen im Business Information Warehouse ebenfalls den Multi-Providern. Für komplexere Berechnungen abgeleiteter Sichten ist dies Mittel allerdings nicht ausreichend und es ist über ein Konzept mit Data Marts nachzudenken, demzufolge auf Basis bestehender Info-Cubes weitere Basis-Cubes mit Daten per Fortschreibung gefüllt werden. Dies ist in Abbildung 120 in der Übersicht dargestellt. Für diese neuen Objekte der persistenten Datenablage sind dann wieder die Regeln für die Modellierung getrennt anzuwenden.



**Abb. 120.** Data Mart-Konzept

Die beim Aufbau von Data Marts entstehende Datenredundanz ist gewollt und der Berechnung dieser abgeleiteten Sicht auf Ebene der Abfragen an das Business Information Warehouse aus Gründen der Performance oftmals vorzuziehen, sofern die zu berücksichtigende Transformationslogik überhaupt im Rahmen einer Query implementierbar ist.

Data-Marts können, sofern auf den zugrunde liegenden Info-Cubes keine Abfragen stattfinden, auch direkt aus ODS-Objekten als Datenquelle befüllt werden, da für diese die Fortschreibung in die Info-Cubes der Data-Marts besonders effizient stattfinden kann.

### 10.3.3 Cube-Modellierung auf der physischen Ebene

Mit den Prinzipien der Partitionierung, des Aufbaus von Aggregaten sowie der Komprimierung von Faktentabellen stehen bereits mächtige Möglichkeiten zum Aufbau eines performanten gesamten Cube-Konzeptes zur Verfügung. Die einzelnen persistenten Datenspeicher stellen dabei insbesondere die Basis-Cubes dar, für deren Modellierung in diesem Abschnitt einige Hinweise zusammengestellt werden. Eine detaillierte Betrachtung der Aspekte der Partitionierung, der Aggregatbildung und der Komprimierung kann an dieser Stelle ausbleiben.

Ein wesentlicher Optimierungsschritt ist die möglichst optimale Verteilung von Merkmalen auf Dimensionen. Der Mechanismus der Line-Item Dimension bietet sich insbesondere für Merkmale mit einer großen Kardinalität an, da hier der Vorteil besonders zum Tragen kommt. Daher sollte sich die Verteilung von Merkmalen auf einzelne Dimensionen nicht an der sachlogischen Gruppierung ausrichten, sondern an den physischen Anforderungen. Zunächst sind möglichst viele der großen Merkmale in eigenständigen Dimensionen abzubilden. Merkmale mit sehr geringer Kardinalität (beispielsweise Kennzeichen-Merkmale wie Fakturaart, Wertart etc.) können sehr gut in einer Dimension kombiniert werden. Das hierbei entstehende kartesische Produkt ist in seiner Gesamtheit immer noch sehr klein und performant im Business Information Warehouse in einer Dimension kombiniert abbildbar.



# Abkürzungsverzeichnis

ABAP	Advanced Business Application Programming
ADAPT	Application Design for Processing Technologies
API	Application Programming Interface
ASAP	AcceleratedSAP
BAPI	Business Application Programming Interface
BEx	Business Explorer Analyzer
BW	Business Information Warehouse
ETL	Extract, Transform and Load
HTML	Hypertext Markup Language
MDX	Multidimensional Expressions
MOLAP	Multidimensionales OLAP
ODBO	OLE DB for OLAP
ODS	Operational Data Store
OLAP	On-Line Analytical Processing
OLE	Object Linking and Embedding
OLTP	On-Line Transaction Processing
PSA	Persistent Staging Area
RFC	Remote Function Call
ROLAP	Relationales OLAP
SEM	Strategic Enterprise Management
SID	Surrogat-ID, Stammdaten-ID
SQL	Structured Query Language
T-ADAPT	Temporal Application Design for Processing Technologies
URL	Uniform Resource Locator
WWW	World Wide Web
XML	Extensible Markup Language

# Abbildungsverzeichnis

<b>Abb. 1.</b> Klassifizierung von Management Support Systemen.....	7
<b>Abb. 2.</b> Data Warehouse Referenzarchitektur .....	9
<b>Abb. 3.</b> Ebenen der Modellierung .....	17
<b>Abb. 4.</b> ANSI 3-Schichten-Konzept.....	17
<b>Abb. 5.</b> Slicing: Herausschneiden von Ebenen aus Würfeln.....	20
<b>Abb. 6.</b> Drehen von Würfeln .....	21
<b>Abb. 7.</b> Navigieren in Dimensionsstrukturen .....	22
<b>Abb. 8.</b> Strukturbestandteile in Dimensionen.....	23
<b>Abb. 9.</b> Datenart-Dimension.....	24
<b>Abb. 10.</b> Balancierte Baumstruktur als Dimensionshierarchie.....	24
<b>Abb. 11.</b> Waldstruktur als Dimensionshierarchie.....	25
<b>Abb. 12.</b> Unbalancierte Dimensionshierarchie.....	26
<b>Abb. 13.</b> Beispiel einer parallelen Hierarchie.....	27
<b>Abb. 14.</b> Beispiel einer Heterarchie (Ausschnitt).....	28
<b>Abb. 15.</b> Ebenenbezogene Gruppenbildung .....	29
<b>Abb. 16.</b> Typisierte Darstellung von Attributen.....	30
<b>Abb. 17.</b> Ausschnitt aus dem Du Pont-Kennzahlensystem .....	31
<b>Abb. 18.</b> Beziehung zwischen Dimensionen und Kennzahlensystem .....	32
<b>Abb. 19.</b> Modellierung von dedizierten Würfeln je Kennzahl .....	33
<b>Abb. 20.</b> Mehrere Kennzahlensysteme in einem Datenmodell .....	33
<b>Abb. 21.</b> Strukturelle Veränderungen in Konsolidierungshierarchien .....	35
<b>Abb. 22.</b> Architektur des Business Information Warehouse 3.1 .....	42
<b>Abb. 23.</b> Stufen im Transformationsprozess .....	46
<b>Abb. 24.</b> Transformation mit Übertragungsregeln .....	46
<b>Abb. 25.</b> Verschiedene Typen von Info-Cubes .....	49
<b>Abb. 26.</b> Datenfluss und Integrationsarchitektur im SAP BW .....	50
<b>Abb. 27.</b> Konzeptionelle Schichten-Architektur im Data Warehousing ...	51
<b>Abb. 28.</b> Ebenen der Query-Ausführung.....	54
<b>Abb. 29.</b> Strukturdefinition im Query Designer.....	56
<b>Abb. 30.</b> Darstellung des Query-Ergebnisses im BEx Analyzer .....	58
<b>Abb. 31.</b> Komponenten der Web Applications.....	59
<b>Abb. 32.</b> Web-Standardbericht auf Basis einer Query-Definition.....	60
<b>Abb. 33.</b> Gestaltung eines Vertriebsberichtes im WAD.....	61
<b>Abb. 34.</b> Organigramm der Modellunternehmung .....	66

<b>Abb. 35.</b> ERM des Vertriebsbereiches .....	70
<b>Abb. 36.</b> ERM der Materialwirtschaft .....	71
<b>Abb. 37.</b> ERM des Produktionsbereiches .....	72
<b>Abb. 38.</b> ER-Modellierung mehrdimensionaler Datenstrukturen .....	76
<b>Abb. 39.</b> Erweiterte ER-Darstellung von Dimensionshierarchien .....	77
<b>Abb. 40.</b> Darstellungsobjekte im ME/R-Modell .....	78
<b>Abb. 41.</b> Beispiel eines ME/R-Modells .....	78
<b>Abb. 42.</b> Beispiel im Dimensional Fact Modeling .....	79
<b>Abb. 43.</b> Aggregation im Dimensional Fact Modeling .....	80
<b>Abb. 44.</b> Acht Dimensionen zur Differenzierung von Kennzahlen .....	81
<b>Abb. 45.</b> Beispiel im Multidimensional Data Model ( <i>MD</i> ) .....	83
<b>Abb. 46.</b> Exemplarische Produktdimension in ADAPT Notation .....	86
<b>Abb. 47.</b> Elementbestimmte Dimension .....	88
<b>Abb. 48.</b> Darstellung eines Cubes in ADAPT .....	89
<b>Abb. 49.</b> Modellierung einer abgeleiteten Teilsicht eines Cubes .....	90
<b>Abb. 50.</b> Vollständige Historisierung über Gültigkeitszeit .....	91
<b>Abb. 51.</b> Transaktionsorientierte Sicht auf Hierarchien .....	92
<b>Abb. 52.</b> Historienverfolgung mit dedizierten Versionen .....	92
<b>Abb. 53.</b> Weitere Formen der Historisierung .....	93
<b>Abb. 54.</b> Controlling Würfel in ADAPT-Notation .....	94
<b>Abb. 55.</b> Modellierung von Scopes in der Zeitdimension .....	95
<b>Abb. 56.</b> Unbalancierte Produktdimension in ADAPT .....	96
<b>Abb. 57.</b> ADAPT-Cube für die aggregierte Marketingsicht .....	97
<b>Abb. 58.</b> Partner-Marketing-Cube in der Gesamtdarstellung .....	99
<b>Abb. 59.</b> Gesamtmodell für das Kunden-Marketing .....	100
<b>Abb. 60.</b> Einfaches Star Schema .....	106
<b>Abb. 61.</b> Attribute in der Zeitdimension .....	111
<b>Abb. 62.</b> Herkömmliche Produktdimension .....	112
<b>Abb. 63.</b> Normalisierte Produktdimension .....	113
<b>Abb. 64.</b> Partitionierte Dimensionstabelle .....	114
<b>Abb. 65.</b> Kennzahlen in der Faktentabelle .....	114
<b>Abb. 66.</b> Ausschnitt aus einem Fact Constellation Schema .....	119
<b>Abb. 67.</b> Snow Flake Schema für den Bereich Kunden-Marketing .....	120
<b>Abb. 68.</b> Modell der Galaxie des Beispiels .....	121
<b>Abb. 69.</b> Überlagerungstabelle für das Partner-Marketing .....	123
<b>Abb. 70.</b> Modell mit demographischer Minidimension .....	124
<b>Abb. 71.</b> Rekursive Beziehung einer Kostenstellenhierarchie .....	125
<b>Abb. 72.</b> Erweitertes Star Schema im SAP BW .....	132
<b>Abb. 73.</b> Ablage von Texten im erweiterten Star Schema .....	133
<b>Abb. 74.</b> Verschiedene Stammdatentabellen im BW .....	133
<b>Abb. 75.</b> Stammdatentabellen bei geklammerten Info-Objekten .....	134

---

<b>Abb. 76.</b> Hierarchietabellen.....	135
<b>Abb. 77.</b> Line-Item Dimension.....	136
<b>Abb. 78.</b> Hierarchische Beziehungen über Bewegungsdaten .....	137
<b>Abb. 79.</b> Navigationsattribute als Grundlage für Dimensionsstrukturen .....	139
<b>Abb. 80.</b> Dimensionsstrukturen auf Basis Externer Hierarchien.....	140
<b>Abb. 81.</b> Beispiel zur Analyse der Berichtsszenarien .....	142
<b>Abb. 82.</b> Berichtsszenario „aktuelle Struktur“ .....	143
<b>Abb. 83.</b> Aktuelle Struktur über Navigationsattribute.....	144
<b>Abb. 84.</b> Aktuelle Struktur über Externe Hierarchie .....	144
<b>Abb. 85.</b> Berichtsszenario „historische Struktur“ .....	145
<b>Abb. 86.</b> Historische Struktur über zeitabhängige Navigationsattribute .....	146
<b>Abb. 87.</b> Historische Struktur über zeitabhängige Externe Hierarchie .....	147
<b>Abb. 88.</b> Berichtsszenario „historische Wahrheit“ .....	149
<b>Abb. 89.</b> Historische Wahrheit über Merkmale.....	149
<b>Abb. 90.</b> Berichtsszenario „vergleichbare Elemente“ .....	150
<b>Abb. 91.</b> Realisierung vergleichbarer Resultate .....	151
<b>Abb. 92.</b> Dimension mit einem Merkmal .....	154
<b>Abb. 93.</b> Line-Item Dimension.....	155
<b>Abb. 94.</b> Anzeigeattribute eines Merkmals .....	155
<b>Abb. 95.</b> Navigationsattribute eines Merkmals .....	156
<b>Abb. 96.</b> Rollen über Referenzen auf Info-Objekte.....	156
<b>Abb. 97.</b> Klammerung von Info-Objekten.....	156
<b>Abb. 98.</b> Hierarchie über Merkmale.....	157
<b>Abb. 99.</b> Hierarchien über Navigationsattribute.....	157
<b>Abb. 100.</b> Externe Hierarchien.....	158
<b>Abb. 101.</b> Darstellung eines Basis-Cubes .....	159
<b>Abb. 102.</b> Repräsentation eines Multi-Providers.....	159
<b>Abb. 103.</b> Info-Cube für den Controllingbereich .....	161
<b>Abb. 104.</b> Zeitdimension der aggregierten Marketingsicht.....	164
<b>Abb. 105.</b> Dimension Produkt der aggregierten Marketingsicht.....	165
<b>Abb. 106.</b> Modell für die aggregierte Marketingsicht .....	166
<b>Abb. 107.</b> Filialdimension im Partner-Marketing .....	167
<b>Abb. 108.</b> Dimension Werbung im Partner-Marketing .....	168
<b>Abb. 109.</b> Verdichtet dargestellter Basis-Cube des Partner-Marketings .....	169
<b>Abb. 110.</b> Kundendimension im BW .....	170
<b>Abb. 111.</b> Verdichtet dargestellter Basis-Cube des Kunden-Marketings .....	170
<b>Abb. 112.</b> Multi-Provider für die aggregierte Marketingsicht.....	177
<b>Abb. 113.</b> Datenfluss zur Komprimierung einer Faktentabelle .....	179
<b>Abb. 114.</b> Komprimierung einzelner Requests .....	179
<b>Abb. 115.</b> Datenbank-Partitionierung über ein Zeitmerkmal .....	181
<b>Abb. 116.</b> Redesign der Info-Cubes im Marketing-Beispiel .....	182
<b>Abb. 117.</b> Aggregat als Teilmenge eines Info-Cubes.....	183

<b>Abb. 118.</b> Zielkonflikt beim Aufbau von Aggregaten.....	184
<b>Abb. 119.</b> 2-Layer-Konzept der Cube-Modellierung .....	198
<b>Abb. 120.</b> Data Mart-Konzept .....	201

# Tabellenverzeichnis

<b>Tabelle 1.</b> Kennzahlenwerte des Rechenbeispiels.....	36
<b>Tabelle 2.</b> Berichtsergebnis nach historischer Struktur.....	36
<b>Tabelle 3.</b> Berichtsergebnis nach aktueller Struktur .....	37
<b>Tabelle 4.</b> Berichtsergebnis mit historischer Wahrheit .....	37
<b>Tabelle 5.</b> Berechnungen nur für vergleichbare Elemente .....	37
<b>Tabelle 6.</b> ADAPT Objekte zur Dimensionsmodellierung .....	84
<b>Tabelle 7.</b> Verbindungsobjekte für die Dimensionsmodellierung.....	85
<b>Tabelle 8.</b> Beziehungstypen in ADAPT .....	88
<b>Tabelle 9.</b> Objekte zur Historisierung von Hierarchien .....	91
<b>Tabelle 10.</b> Dimensionstabelle Vertriebsweg.....	107
<b>Tabelle 11.</b> Dimensionstabelle Zeit.....	108
<b>Tabelle 12.</b> Ausschnitt der Dimensionstabelle Produkt .....	109
<b>Tabelle 13.</b> Zeitdimension mit Kalender- und Geschäftsjahressicht.....	110
<b>Tabelle 14.</b> Dimensionstabelle Zeit mit Aggregattupeln.....	117
<b>Tabelle 15.</b> Dimensionstabelle Kostenstelle .....	126
<b>Tabelle 16.</b> Heterarchie in der Dimensionstabelle Kostenstelle.....	126
<b>Tabelle 17.</b> Dimensionstabelle mit kompletten Versionen.....	127
<b>Tabelle 18.</b> Dimensionstabelle mit zusätzlichem Aktuell-Feld.....	128
<b>Tabelle 19.</b> Zeitmerkmale des Business Content .....	162

# Literaturverzeichnis

- Brodie ML (1984) On the developement of data models. In: Brodie ML, Mylopoulos J, Schmidt J (Hrsg) On conceptual modeling. Springer-Verlag, New York, S 19–47
- Bulos D (1996) A New Dimension. Database Programming & Design 9:33–38
- Bulos D (1998) OLAP Database Design – A New Dimension. In: Chamoni P, Gluchowski P (Hrsg) Analytische Informationssysteme: Data Warehouse, On-Line Analytical Processing, Data Mining. Springer, Berlin, S 251–261
- Cabibbo L, Torlone R (1997a) Querying Multidimensional Databases. In: Sixth International Workshop on Database Programming Languages (DBPL-97), S 319–335
- Cabibbo L, Torlone R (1997b) A Systematic Approach to Multidimensional Databases. In: Schreiber F (Hrsg) Quinto Convegno Nazionale su Sistemi Evoluti per Basi di Dati (SEBD-97)
- Chamoni P, Gluchowski P, Hahne M (2004) Business Information Warehouse. Springer-Verlag, Heidelberg
- Chamoni P, Gluchowski P (1999a) Analytische Informationssysteme – Einordnung und Überblick. In: Chamoni P, Gluchowski P (Hrsg) Analytische Informationssysteme: Data Warehouse, On-Line Analytical Processing, Data Mining, 2. Aufl. Springer, Berlin, S 3–25
- Chamoni P, Gluchowski P (1999b) Entwicklungslinien und Architekturkonzepte des OLAP. In: Chamoni P, Gluchowski P (Hrsg) Analytische Informationssysteme: Data Warehouse, On-Line Analytical Processing, Data Mining, 2. Aufl. Springer, Berlin, S 261–280
- Chamoni P, Gluchowski P, Hahne M (2004) Business Information Warehouse. Springer-Verlag, Heidelberg
- Chamoni P, Stock S (1998) Temporale Daten in Management Support Systemen. Wirtschaftsinformatik 6:513–519
- Codd EF, Codd SB, Sally CT (1993) Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT-Mandate. White Paper
- Egger N (2004) Praxishandbuch SAP BW 3.1. Galileo Press, Bonn
- Färber G, Kirchner J (2002) mySAP Technology – Einführung in die neue Technologie-Plattform der SAP. Galileo Press, Bonn
- Gabriel R, Chamoni P, Gluchowski P (2000) Data Warehouse und OLAP – Analyseorientierte Informationssysteme für das Management. Zeitschrift für betriebswirtschaftliche Forschung 52:74–93

- Gabriel R, Gluchowski P (1998) Grafische Notation für die semantische Modellierung multidimensionaler Datenstrukturen in Management Support Systemen. *Wirtschaftsinformatik* 6:493–502
- Gluchowski P (1997) Data Warehouse-Datenmodellierung – Weg von der starren Normalform. *Datenbank-Fokus* 11:62–66
- Gluchowski P, Gabriel R, Chamoni P (1997) Management-Support-Systeme: computergestützte Informationssysteme für Führungskräfte und Entscheidungsträger. Springer, Berlin
- Golfarelli M, Maio D, Rizzi S (1998a) The Dimensional Fact Model: a Conceptual Model for Data Warehouses. In: *International Journal of Cooperative Information Systems* 2-3:215–247
- Golfarelli M, Maio D, Rizzi S (1998b) Conceptual design of data warehouses from E/R schemes. In: El-Rewini H (Hrsg) *Proceedings of the 31<sup>st</sup> Hawaii International Conference on System Sciences (HICCS-31)*, S 334–343
- Groffmann HD (1992) Kennzahlendatenmodell (1992) als Grundlage aktiver Führungsinformationssysteme. In: Rau KH, Stickel E (Hrsg) *Daten- und Funktionsmodellierung – Erfahrungen – Konzepte – Perspektiven*. Gabler, Wiesbaden
- Günzel H, Bauer A (2001) *Data-Warehouse-Systeme – Architektur, Entwicklung, Anwendung*. dpunkt, Heidelberg
- Hahne M (1999) Logische Modellierung für das Data Warehouse – Bestandteile und Varianten des Star Schemas. In: Chamoni P, Gluchowski P (Hrsg) *Analytische Informationssysteme – Data Warehouse, On-Line Analytical Processing, Data Mining*. Springer-Verlag, Berlin, S 145–170
- Hahne M (2002) Logische Modellierung mehrdimensionaler Datenbanksysteme. Deutscher Universitäts-Verlag, Wiesbaden
- Hahne M (2002a) Datenmodellierung für SAP Business Information Warehouse: Aufbau mehrdimensionaler Datenstrukturen im BW3.0 (Teil 1). *itFokus* 11:12–18
- Hahne M (2002b) Datenmodellierung für SAP Business Information Warehouse: Hierarchische Dimensionsstrukturen im BW3.0 (Teil 2). *itFokus* 12:56–59
- Hahne M (2002c) Transformation mehrdimensionaler Datenmodelle. In: von Maur E, Winter R (Hrsg) *Vom Data Warehouse zum Corporate Knowledge Center*. Physica-Verlag, Heidelberg, S 399–420
- Hahne M (2003a) Logische Datenmodellierung zur Abbildung mehrdimensionaler Datenstrukturen im SAP Business Information Warehouse. In: Weikum G, Schöning H, Rahm E (Hrsg) *BTW 2003 Datenbanksysteme für Business, Technologie und Web, Tagungsband der 10. BTW-Konferenz, 26.–28. Februar 2003*, Leipzig, Köllen Druck + Verlag, Bonn, S 630–647
- Hahne M (2003b) Time aspects in SAP Business Information Warehouse. In: Jardim-Gonçalves R, Cha J, Steiger-Garção A (Hrsg) *Concurrent Engineering*. A.A. Balkema Publishers, Lisse, S 69–74
- Haupt J (2003) *Enterprise Data Warehousing with SAP BW – An Overview*. SAP
- Heinemann R, Rau C (2003) *SAP Web Application Server*. Galileo Press, Bonn



- Holthuis J (1997) Multidimensionale Datenstrukturen. In: Mucksch H, Behme W (Hrsg) Das Data Warehouse Konzept – Architektur – Datenmodelle – Anwendungen. Gabler, Wiesbaden
- Inmon WH (1996) Building the Data Warehouse, 2nd edn. John Wiley & Sons, New York
- Kimball R (1996) The Data Warehouse Toolkit. John Wiley & Sons, New York
- Kütting K (1983) Kennzahlensysteme in der betrieblichen Praxis. WiSt
- Lockemann PC, Radermacher K (1990) Konzepte, Methoden und Modelle zur Datenmodellierung. HMD 152:3–16
- Luft AL (1990) Datenmodelle. In: Mertens P (Hrsg) Lexikon der Wirtschaftsinformatik, 2. Aufl. Berlin, S 132–134
- Mehrwald C (2003) SAP Business Information Warehouse 3 – Architektur, Konzeption, Implementierung. dpunkt-Verlag, Heidelberg
- Meyer C (1994) Betriebswirtschaftliche Kennzahlen und Kennzahlensysteme, 2. Aufl. Poeschel, Stuttgart
- Michel R (1999) Komprimiertes Kennzahlen-Know-How: Analysemethoden, Frühwarnsysteme, PC-Anwendungen, Checklisten. Gabler, Wiesbaden
- Mucksch H, Behme W (2000) Das Data Warehouse-Konzept – Architektur - Datenmodelle - Anwendungen, 4. Aufl. Gabler, Wiesbaden
- Nußdorfer R (1998a) Star Schema, das E/R-Modell steht Kopf – Teil 1: Modellieren von Faktentabellen. Datenbank-Fokus 10:22-28
- Nußdorfer R (1998b) Star Schema – Teil 1: Modellierung von Dimensionstabellen. Datenbank-Fokus 11;16-23
- Pendse N, Creeth R (1995) The OLAP-Report; Succeeding with On-Line Analytical Processing. Business Intelligence 1
- SAP (2000a) Modeling with BW. ASAP for BW Accelerator
- SAP (2000b) Hierarchies in SAP BW. ASAP for BW Accelerator
- Sapia C, Blaschka M, Höfling G, Dinter B (1998) Extending the E/R Model for the Multidimensional Paradigm. In: Kambayashi Y, Lee DL, Lim E, Mohania M, Masunaga Y (Hrsg) Advances in Database Technologies, ER '98 Workshops on Data Warehousing and Data Mining, Mobile Data Access and Collaborative Work Support and Spatio-Temporal Data Management. Proceeding Nr. 1552 der Reihe "Lecture Notes in Computer Science (LNCS)", Springer, Berlin, S 105–116
- Schelp J (2000) Modellierung mehrdimensionaler Datenstrukturen. Deutscher Universitäts-Verlag, Wiesbaden
- Stock S (2001) Modellierung zeitbezogener Daten im Data Warehouse. Deutscher Universitäts-Verlag, Wiesbaden
- Stock S, Beekmann F (2003) Temporal aspects of data warehousing. In: Jardim-Gonçalves R, Cha J, Steiger-Garção (Hrsg) Concurrent Engineering. A.A. Balkema Publishers, Lisse, S 53–59
- Totok A, Jaworski R (1998) Modellierung von multidimensionalen Datenstrukturen mit ADAPT. Arbeitsbericht 98-11, Technische Universität Braunschweig
- Vetschera W (1995) Informationssysteme der Unternehmensführung. Springer, Berlin

- Vossen G (1999) Datenmodelle, Datenbanksprachen und Datenbank-Management-Systeme, 3. Aufl. Addison-Wesley, Bonn
- Wolf F (2003) SAP Web Application Server – Internet-Anwendungen entwickeln mit ABAP, HTML und JavaScript. dpunkt-Verlag, Heidelberg

# Stichwortverzeichnis

2-Layer-Konzept .....	176, 198
ADAPT .....	83
Attribute .....	85
Beziehungstypen .....	88
Dimension .....	85
Hierarchy .....	85
Level .....	85
lockere Beziehung .....	85
many-many-Beziehung .....	85
Member .....	85
Model .....	85
rekursive Beziehung .....	85
Scope .....	85
strenge Beziehung .....	85
additiv .....	34
Additivität .....	34
Aggregat .....	183
Aggregation .....	77
anteilige Verrechnung .....	27
Anwendungs-Partitionierung .....	180
Anzeigeattribut .....	133, 155
Attribut .....	43
Ausnahme .....	<i>Siehe</i> Exception
Ausnahmeaggregation .....	152
Basis-Cube .....	48, 158
Baumstruktur	
balancierte .....	24
unbalancierte .....	26
bebuchbare Knoten .....	141
Bedingung .....	<i>Siehe</i> Condition
Bestandsgröße .....	152
Bestandsveränderung .....	152
Bewegungsdatensicht .....	148, 157
BEx	
Analyzer .....	42, 57
Query Designer .....	42
Web Application .....	58

Web Application Designer .....	42, 58
BEx Web Application .....	61
Beziehung .....	75
Blattelement .....	28
Bookmark .....	60
Business Content .....	162
BW	
Architektur .....	42
cannibalization .....	122
caption .....	154
Change Run .....	185
characteristic .....	154
Condition .....	53, 57
Data Mart .....	9, 201
Data Warehouse .....	7
Real-Time .....	51
Schichten-Architektur .....	51
unternehmensweites .....	51
Data-Source .....	45
DATEFROM .....	146
Datenbank	
Design .....	15
Schema .....	17
Datenbank-Partitionierung .....	180
Datenfluss .....	50
Datenmodell .....	15
logisches .....	16
physisches .....	17
semantisches .....	16
Datenstruktur .....	18
mehrdimensionale - .....	19
Datentyp .....	18
Datenunabhängigkeit	
logische .....	18
physische .....	18
Datenwürfel .....	19
DATETO .....	147

- 
- decision support system ..... 5
  - dice..... 20
  - Dimension ..... 19, 154
    - ebenenbestimmte..... 23
    - elementbestimmte ..... 23
  - Dimensional Fact Modeling..... 79
    - Dimension..... 79
    - Fakt ..... 79
    - Hierarchie..... 79
  - Dimensionselement
    - abgeleitetes..... 23
    - verdichtetes ..... 23
  - Dimensionsgraph ..... 28
  - Dimensionsstruktur ..... 23
  - Dimensionstyp
    - ebenenbestimmt ..... 88
    - elementbestimmt ..... 88
  - Display Attribute..... 155
  - drill down..... 57
  - drill-down..... 21
  - DSS ..... 5
  - Ebene ..... 111
  - EIS ..... 5
  - Entität..... 75
  - Entity..... 75
  - Entity Relationship-Modell.... 16, 75
  - Entscheidungsunterstützungssystem
    - ..... 5
  - ERM..... 16, 75
  - ER-Modell
    - Mehrdimensionales ..... 77
  - erweitertes Star Schema ..... 131
    - Einheitendimension..... 132
    - Paketdimension ..... 132
    - Zeitdimension ..... 132
  - ETL-Prozess..... 8
  - EUS..... 5
  - Exception ..... 53, 57
  - exception reporting ..... 6
  - executive information system ..... 5
  - Externe Hierarchie*..... 134, 139, 158
  - Extraktstruktur ..... 45
  - Fact Constellation Schema..... 118
  - Factless Fact Table..... 122
  - Fakt ..... 132
  - Faktentabelle
    - E-Tabelle..... 178
    - faktenlose ..... *Siehe* Factless Fact Table
    - F-Tabelle..... 178
    - Komprimierung..... 179
  - FASMI ..... 11
  - FIS..... 5
  - Fortschreibungsregel..... 45
  - Führungsinformationssystem ..... 5
  - Galaxie ..... 121
  - generation..... 28
  - Generation..... 111
  - Granularität ..... 23
  - Gültigkeitszeit ..... 134
  - Heterarchie..... 27
  - Hierarchie..... 23
    - parallele..... 26
  - Hierarchieattribut ..... 28
  - Hierarchieversion..... 147
  - Historisierung..... 127
  - Inclusion Table ..... 140
  - Info-Cube ..... 48, 158
  - Info-Objekt..... 43, 132
    - Klammerung ..... 44
    - Referenz ..... 44
  - Info-Provider
    - Basis-Cube ..... 48
    - Info-Cube ..... 48
    - Multi-Provider..... 48
    - Remote-Cube ..... 48
  - Informationssystem
    - analyseorientiert..... 7
  - Info-Source ..... 45
  - Integrationsarchitektur ..... 50
  - Internet Graphic Server..... 59
  - Internet Transaction Server ..... 58
  - Kardinalität ..... 76
  - Kennzahl ..... 31, 132, 152
    - berechnete ..... 54
    - eingeschränkte..... 54
    - nicht additiv ..... 152
  - Kennzahlendatenmodell..... 81
  - Kennzahlendimension..... 153
  - Kennzahlensystem ..... 31, 152
  - Klammerung ..... 134, 156
  - Knoten

- 
- Höhe..... 28
  - innerer ..... 28
  - Tiefe ..... 28
  - Kommunikationsstruktur ..... 45
  - Komprimierung..... 179
  - Konsolidierungspfad ..... 23
  - level..... 28, 125
  - lift..... 122
  - Line-Item Dimension 135, 138, 155, 177
  - M/ER
    - Fakt-Beziehungstyp ..... 78
    - Roll-Up ..... 78
  - M/ER..... 77
    - Dimensionsebene ..... 78
  - management information system ... 5
  - Management Support System ..... 5
  - Managementinformationssystem ... 5
  - MD*..... 82
  - Merkmal..... 132, 137, 154
  - Merkmalsknoten ..... 140
  - Minidimension ..... 124
  - MIS ..... 5
  - Modell..... 15
  - Modellierung
    - ANSI-Architektur..... 17
    - Ebenen..... 17
  - MSS ..... 5
  - Multi-Cube..... 175
  - multidimensional..... 10
  - Multidimensional Data Model ..... 82
    - f-graph..... 82
    - f-node ..... 82
  - Multi-Provider..... 48, 159, 175
  - Navigation Attribute ..... 155
  - Navigationsattribut.... 133, 138, 155, 157
    - zeitabhängig ..... 139
  - nicht-additiv ..... 34, 80
  - Normalform..... 112
  - Normalisierung ..... 113
  - ODS-Objekt ..... 47
  - OLAP ..... 10
    - Datenbank ..... 9
    - Regeln ..... 10
  - OLAP-Prozessor ..... 53
  - OLTP..... 10
  - On-Line Analytical Processing ..... 9
  - On-Line Transaction Processing .. 10
  - Operational Data Store..... 47
  - Paket..... *Siehe* Request
  - Partitionierung..... 114, 180
  - Partitionierungsmerkmal ..... 181
  - Persistent Staging Area ..... 44
  - Personal Information Management 7
  - PIM ..... 7
  - PSA ..... 44
  - Query..... 49, 53
    - Condition..... 53
    - Exception ..... 53
    - Struktur ..... 53
    - Variable..... 55
  - Query Key Date ..... 134, 141, 146
  - ranging ..... 20
  - Referenz ..... 156
  - rekursive Beziehung..... 125
  - Relationship ..... 75
  - Remote-Cube ..... 48, 159
  - Request..... 178
  - roll up ..... 57
  - Rollen ..... 156
  - roll-up..... 21
  - rotation ..... 20
  - Schablonen..... 154
  - Schema
    - externes ..... 17
    - internes ..... 17
    - konzeptionelles..... 17
  - Schichten-Architektur
    - Data Mart layer ..... 52
    - Data Warehouse layer ..... 52
    - Operational Data Store..... 52
    - Staging layer ..... 52
  - Schlüsseldatum..... 141, 146
  - semi-additiv..... 34, 80
  - SID-Schlüssel..... 147
  - SID-Tabellen..... 134
  - slice ..... 20
  - Slowly Changing Dimensions. *Siehe* Historisierung
  - Snow Flake Schema ..... 119
  - sprachabhängig..... 155

Staging-Szenario .....	51	Übertragungsregel .....	45
Stammdatenattribut .....	133	Variable .....	55
stammdatenbezogen .....	148	Verhältniszahl .....	152
stammdatentragendes Merkmal ..	133	Versionierung	
Star Schema .....	76, 105	selektive .....	90
Attribute .....	111	transaktionsorientiert.....	92
balancierte Struktur .....	107	vollständige .....	90
erweitertes .. <i>Siehe</i> erweitertes Star		WAD .....	61
Schema		Waldstruktur	
flache Struktur .....	107	balancierte .....	25
Heterarchie .....	110	unbalancierte .....	26
Kennzahlen .....	106	Web Application Designer.....	61
Kennzahlensystem .....	114	Web Application Server.....	42, 58
parallele Hierarchie .....	109	Web Item .....	61
unbalancierte Struktur .....	108	Web Template .....	58, 61
Struktur .....	53	Würfel .....	19
Strukturanomalie .....	26	Wurzelement .....	25, 28
strukturspezifischen Eigenschaft	182	zeitabhängige Gesamthierarchie	147
Strukturveränderung .....	35	zeitabhängige Hierarchiestruktur	147
Texte .....	154	zeitabhängige Stammdaten .....	134
Textknoten .....	140, 145	zeitabhängiges Navigationsattribut	
Texttabelle .....	132	.....	146
time shifting .....	122	Zeitabhängigkeit .....	138
transaktionsorientiert.....	148	Zeitbezug .....	34
Transferstruktur .....	45	Zeitmerkmal .....	162
TS <i>Siehe</i> Internet Transaction Server			
Überlagerungstabelle .....	123		



#### ZUM AUTOR

Studium der Mathematik an der Ruhruniversität Bochum, Schwerpunkte Kristallographie und Wirtschaftsinformatik. Promotion über mehrdimensionale Datenbanken an der Ruhr-Universität Bochum, Fakultät für Wirtschaftswissenschaft, Lehrstuhl für Wirtschaftsinformatik.

Dr. Hahne leitet den Geschäftsbereich »BI-Consulting Produkte der SAP AG« bei der cundus AG, einem auf Business Intelligence-Lösungen spezialisierten Beratungsunternehmen ([www.cundus.de](http://www.cundus.de)).