

# Logische Datenmodellierung zur Abbildung mehrdimensionaler Datenstrukturen im SAP Business Information Warehouse

Michael Hahne

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

**Abstract:** Innovative technische Konzepte des Data Warehousing, OLAP und Data Mining begleiten eine zunehmende strategische Ausrichtung der Informationsverarbeitung mit einer stärkeren Fokussierung auf Aspekte der Analyse und Entscheidungsunterstützung. In diesem Segment positioniert die SAP AG ihre Lösung mySAP Business Intelligence, die auf dem aktuellen Release des Business Information Warehouse (BW) als technologischer Kernkomponente basiert. Das BW ermöglicht den Aufbau von Systemen mit einer konsistenten und einheitlichen Datenbasis für die vielfältigen betrieblichen Anwendungen von Fach- und Führungskräften. Die Leistungsfähigkeit und erfolgreiche Nutzung solcher Systeme werden maßgeblich durch die Möglichkeiten der Modellierung und die Mächtigkeit des zu Grunde liegenden Datenmodells bestimmt. Hierarchischen Dimensionsstrukturen kommt dabei eine zentrale Rolle zu, da die hierdurch festgelegten Konsolidierungspfade die Grundlage für die Navigation in analytischen Datenbeständen sowie zahlreiche OLAP-Operationen bilden. Die Abbildung dieser Strukturbestandteile mehrdimensionaler Datenmodelle im BW ist durch Variantenreichtum und innovative Ansätze gekennzeichnet und wird in diesem Artikel eingehend betrachtet.

## 1 Einleitung

Mit dem *Business Information Warehouse (BW)* 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 Hauptkomponenten 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. Das Strategic Enterprise Management, neuerdings Bestandteil der Lösung mySAP Financials beinhaltet Funktionen für Unternehmensplanung, Performance Management und Konsolidierung. Die Datenhaltungskomponente des SEM ist ebenfalls das BW.<sup>1</sup>

Für viele Lösungen aus dem Hause SAP ist das Business Information Warehouse die wesentliche technologische Basiskomponente zur Datenspeicherung, auf deren Basis 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 zu Grunde 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, wie z. B. 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.

In Abschnitt 2 erfolgt die Darstellung der Architektur und des spezifischen Datenmodells des Business Information Warehouse. Auf die Möglichkeiten der Abbildung hierarchischer Strukturen mit den Mitteln dieses Modells wird anschließend in Abschnitt 3 eingegangen. Die Zusammenfassung der gewonnenen Arbeitsergebnisse erfolgt in Abschnitt 4.

## **2 Business Information Warehouse**

Die Architektur des Business Information Warehouse ist angelehnt an die allgemeinen 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). 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 im BW erfolgt. Im Mittelpunkt der Architektur steht der

---

<sup>1</sup> Die Gruppierung und Begriffsbildung ändert sich bei der SAP AG häufiger.

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 der Persistent Staging Area (PSA), der Operational Data Store (ODS) und die Info-Cubes. Während der 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 Business 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 Abb. 1.

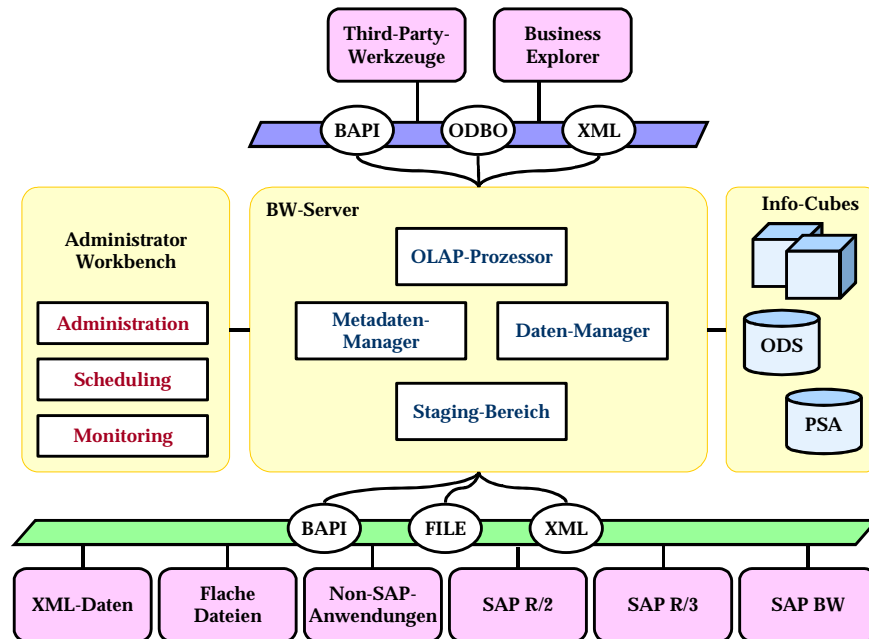


Abb. 1: Architektur des Business Information Warehouse 3.0

Integraler Bestandteil des Business Information Warehouse sind die Werkzeuge Business Explorer Analyzer als Excel-basiertes OLAP-Werkzeug und Crystal Reports für formatierte druckoptimierte Berichte.<sup>2</sup> 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 [Ha99]. Zunächst erfolgt daher in Abschnitt 2.1 eine kurze Darstellung mehrdimensionaler Datenstrukturen und deren Abbildung in Star Schema-Modellen. Das spezifische Datenmodell, auf dem das Business Information Warehouse basiert, wird anschließend in Abschnitt 2.2 behandelt. Die Diskussion der besonderen Eigenschaften und Leistungspotentiale dieses Modells erfolgt in den Abschnitten 2.3 und 2.4.

## 2.1 Mehrdimensionale Datenstrukturen im Star Schema

Fundamentales Paradigma eines Data Warehouse ist die Mehrdimensionalität, die abgelegten Informationen basieren auf mehrdimensionalen Datenstrukturen. Diesem Ansatz folgend werden Datenwerte in einem mehrdimensionalen Datenwürfel aufgespannt betrachtet, deren Zellinhalte die abgelegten Daten darstellen. Die Würfelkanten, im Sprachgebrauch mehrdimensionaler Datenstrukturen Dimensionen genannt, gliedern diese Werte auf und bilden eine wesentliche Grundlage für Analysemöglichkeiten. Betriebswirtschaftliche Kennzahlen wie beispielsweise Umsatz werden entlang dieser Dimensionen als Aufgliederungsrichtung, so z. B. über Kunden, Artikel und Regionen, betrachtet.

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 Abb. 2 veranschaulicht.

Typische Dimensionsstrukturen sind balancierte Baumstrukturen wie in Abb. 2 dargestellt, die dadurch gekennzeichnet sind, dass die Basiselemente alle über gleich viele Stufen bis zu einem obersten Konsolidierungselement verbunden sind und dass es genau ein oberstes (Wurzel-)Element gibt.

---

<sup>2</sup> Crystal Reports ist komplett in die Werkzeuge des Business Information Warehouse integriert, insbesondere sind die speziellen Bedürfnisse des Listen-Reportings im Query Designer des BW integriert. Lizenzrechtlich ist das Produkt aber gesondert zu betrachten, ein unternehmensweiter Einsatz ist durch die my SAP-Lizenz nicht mit abgedeckt.

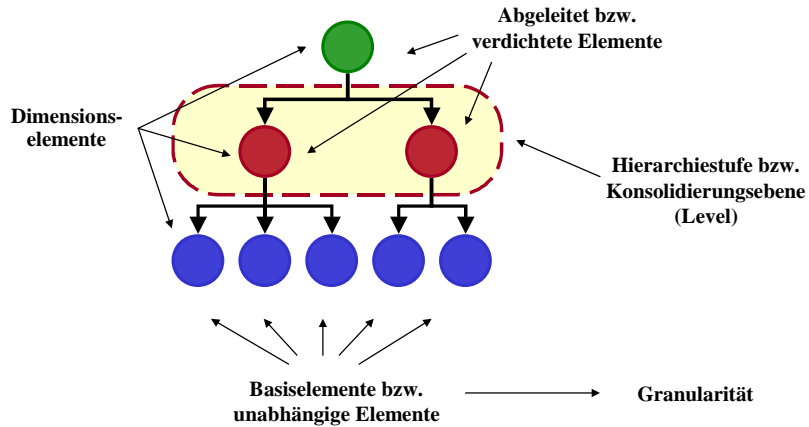


Abb. 2: Strukturbestandteile in Dimensionen

Die balancierte Dimensionshierarchie ist eine idealtypische Strukturform, neben der es weitere Formen gibt. 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.

Strukturen, bei denen die Bedingung einer 1:n-Beziehung der Ebenen in einer Dimension fallengelassen wird, 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 [Ha02]. 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, sind die Anteilsverhältnisse bei der Konsolidierung zu berücksichtigen. Diese Formen von Hierarchien mit m:n-Beziehungen zwischen einzelnen Dimensionsebenen stehen unter dem Begriff der anteiligen Hierarchie in der Diskussion.

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. 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 (siehe Abb. 3).

Zur Abbildung mehrdimensionaler Datenstrukturen in relationalen Datenbanken sind verschiedene Ansätze eingeführt worden, die im wesentlichen unter dem Begriff *Star Schema* zusammengefasst und diskutiert werden [Ha99].

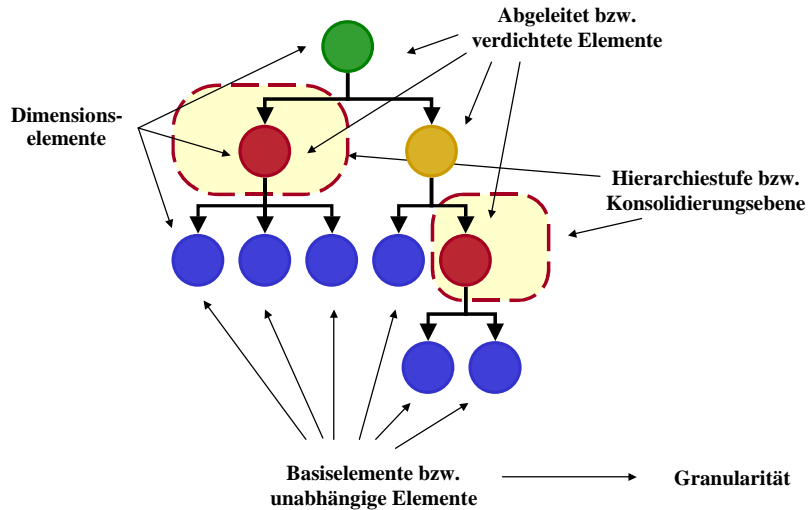


Abb. 3: unbalancierte Dimensionsstruktur

Die Faktentabelle mit den Kennzahlenwerten bildet den Mittelpunkt eines Star Schemas, um den herum die Dimensionstabellen als Satellitentabellen angeordnet sind [Nu98a]. Der identifizierende Schlüssel in der Faktentabelle ist aus unterschiedlichen Komponenten zusammengesetzt, die auf die Dimensionstabellen verweisen. Während in der Faktentabelle die Bewegungsdaten enthalten sind, beinhalten die Dimensionstabellen die Stammdaten, welche die Bewegungsdaten beschreiben. Sie definieren Suchkriterien und legen entlang der Hierarchien die Verdichtungsstufen fest [Nu98b]. Dimensionstabellen im Star Schema sind bei dieser Struktur im allgemeinen denormalisiert [Gl97]. Weitergehende Varianten des Star Schemas lassen diese einschränkende Bedingung fallen und es entstehen Modelle, die als Snow Flake-Schema bezeichnet werden.

Insbesondere aus Anforderungen an ein annähernd gleich bleibendes Antwortzeitverhalten heraus entstanden Modellvarianten, die diesen Aspekt besser berücksichtigen. Zentraler Aspekt ist dabei die Vorberechnung von verdichteten Werten, die beispielsweise in eigenen Aggregattabellen vorgehalten werden [Ha99].

Der Ansatz des Star Schemas zur Abbildung mehrdimensionaler Datenstrukturen in relationalen Datenbanken ist im wesentlichen auch die Grundlage des Datenmodells, das im Business Information Warehouse implementiert ist. Dieses wird im nächsten Abschnitt dargestellt.

## 2.2 Das erweiterte Star Schema der SAP

Die Datenhaltung des Business Information Warehouse basiert ebenfalls auf der relationalen Datenbanktechnik, die Speicherung mehrdimensional strukturierter Informationen folgt dem Ansatz eines Star Schemas, wobei zur Verbesserung in einigen Schwachpunk-

ten dieser Modellansatz entsprechend modifiziert wurde.<sup>3</sup> Wesentliche Verbesserungspotentiale ergeben sich dabei in der Behandlung von m:n-Beziehungen, von unblancierten Hierarchien sowie dem Umgang mit strukturellen Änderungen in Dimensionen und der Behandlung von Dateneingaben auf unterschiedlichen Verdichtungsstufen, wie sie beispielsweise bei der Ablage von Ist-Daten auf Tagesebene und von Plan-Daten auf Jahresebene auftreten.

Das zentrale Objekt zur Speicherung mehrdimensionaler Daten im BW ist der Info-Cube. Das Grundkonzept sieht dabei vor, dass die Informationen aus den Dimensionen für mehrere Info-Cubes gemeinsam genutzt werden, die Stammdaten somit übergreifend gültig sind [Sa00a].

Auch im erweiterten Star Schema bildet die Faktentabelle den Mittelpunkt des Modells. Die Fakten der Faktentabelle heißen im BW auch Kennzahlen oder Key Figures (z.B. Umsatz oder Absatz). Die Faktentabelle ist von Dimensionen umgeben, die aus der Dimensionstabelle und den Stammdatentabellen bestehen, exemplarisch ist dies in Abb. 4 für ein Modell mit den drei Dimensionen Kunde, Produkt und Zeit dargestellt.

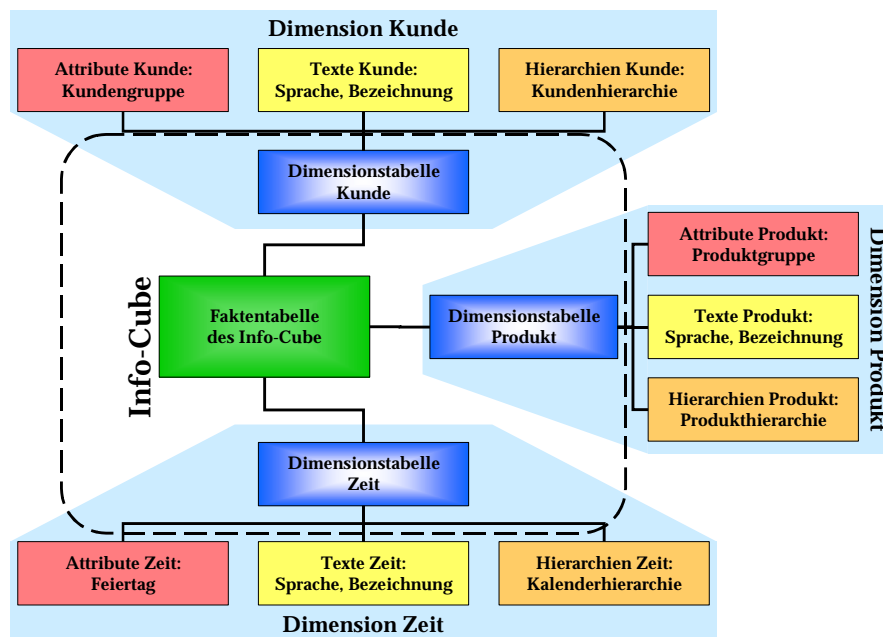


Abb. 4: Erweitertes Star Schema im SAP BW

<sup>3</sup> Das erweiterte Star Schema im BW ist gekennzeichnet durch eine große Anzahl von Tabellen, über die die Informationen verteilt abgelegt sind, und ist gegenüber klassischen Ansätzen aufgrund dieser Verteilung deutlich komplexer.

Der Begriff der Dimensionstabelle ist im BW mit einer besonderen Bedeutung versehen. Im erweiterten Star Schema bildet die Dimensionstabelle den Verknüpfungspunkt von der Faktentabelle zu den abgekoppelten Stammdatentabellen, in denen die eigentlichen Dimensionswerte und deren beschreibenden Attribute abgelegt sind.<sup>4</sup>

Die Basisinformationsbausteine im Business Information Warehouse sind die Info-Objekte. Sie werden global im BW-System definiert und sind über ihren technischen Namen eindeutig identifizierbar. 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. Merkmale (Characteristics) und Kennzahlen (Key Figures) basieren beide auf dem Grundbaustein eines Info-Objektes, durch das damit Stamm- und Bewegungsdaten in strukturierter Form abgebildet werden. Auch die weiteren Objekte des BW basieren auf den Info-Objekten als elementare Komponenten, so sind auch die Felder in ODS-Objekten oder die Definitionen von Info-Sources durch sie beschrieben.

### 2.3 Stammdatenkonzept im BW

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, im Beispiel der Artikelname. Die Ablage dieser beschreibenden Texte von Merkmalen erfolgt im BW in einer separaten Texttabelle, wobei die Texte in verschiedenen im System definierten Sprachen erfasst sein können. Beispielsweise sind zu einem Info-Objekt, das einen Länderschlüssel darstellt, mehrsprachige Landesbezeichnungen abbildbar.

Die Berücksichtigung von Dimensionsattributen ist im erweiterten Star Schema über Stammdaten-Attribute 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 diesen in Berichten darstellbar. Anzeigeattribute als ergänzende Information zu einem Merkmal können nur zusammen mit diesem in Berichten verwendet werden.

Hierarchien eines Merkmals, z. B. eine Länderhierarchie auf Basis des Info-Objektes Länderschlüssel, können in separaten Hierarchietabellen gespeichert werden. Diese Hierarchien werden Externe Hierarchien genannt (z. B. Standard-Kostenstellenhierarchie aus dem R/3-CO für das Merkmal Kostenstelle) und stellen vordefinierte Konsolidierungspfade dar. Die Benutzung des Begriffes Hierarchie ist im BW potenziell missver-

---

<sup>4</sup> Die Ablage der Informationen, die in Dimensionstabellen eines allgemeinen Star Schemas abgelegt sind, erfolgt im BW in den sog. Stammdaten-Tabellen, die eigentliche Dimensionstabelle im BW bildet nur die Schnittstelle von der Faktentabelle zu den verschiedenen Tabellen, in denen die eigentlichen Dimensionsinformationen abgelegt sind. Demzufolge korrespondiert der allgemeine Begriff der Dimensionstabelle im erweiterten Star Schema mit der eigentlichen Dimensionstabelle sowie den Tabellen für die Stammdaten-Attribute, die mehrsprachigen Texte und die Externen Hierarchien.



ständig, denn im normalen Verständnis ist eine Hierarchie eine Sequenz von rekursiven Beziehungen zwischen Merkmalen. Demzufolge gibt es Hierarchien sowohl in der Dimensionstabelle über Merkmale, in der Stammdatentabelle über Attribute als auch in der eigentlichen Hierarchietabelle [Sa00a].

Die Stammdatentabellen sind nicht direkt an die Info-Cubes gebunden, die Verbindung erfolgt über SID-Tabellen, die mit künstlichen Schlüsseln arbeiten (Surrogate Ids). Über die SID-Tabellen, die Zeiger- oder Übersetzungstabellen, werden die Stammdatentabellen mit den Info-Cubes verbunden (siehe Abb. 5). Hierdurch können auch Merkmale auf Basis einer m:n-Beziehung wie z. B. Material und Farbe in einer Dimension abgebildet werden.

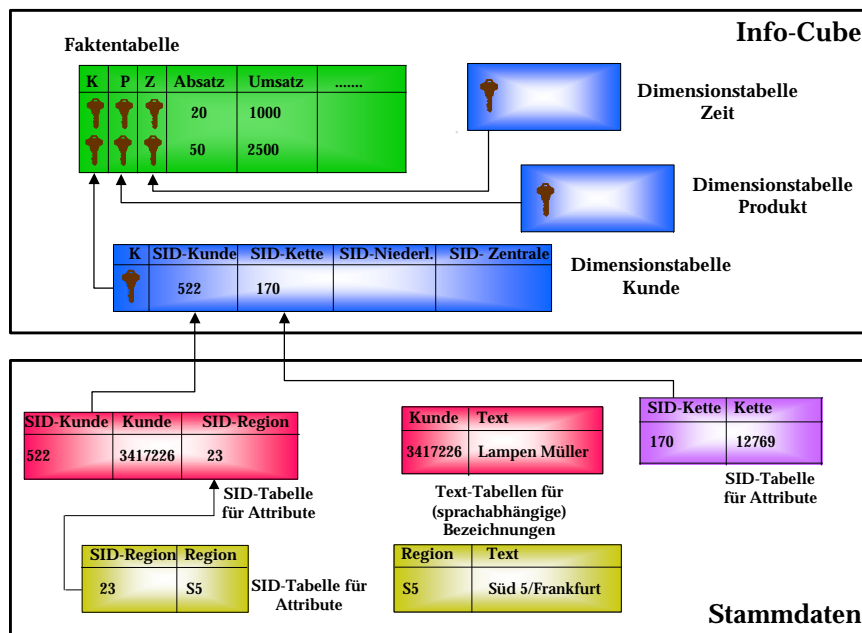


Abb. 5: Trennung von Stammdaten und Bewegungsdaten im BW

Alle Stammdaten im Business Information Warehouse sind als zeitabhängig deklarierbar. Dadurch erhalten Attribute, Texte oder Hierarchien die zusätzliche Information, von wann bis wann sie genau gelten.<sup>5</sup> Diese Zuordnung erfolgt vom System eindeutig während der Ladevorgänge zum Befüllen des Data Warehouse. Die Verknüpfung von Info-Cubes mit zeitabhängigen Stammdaten erfolgt über gesonderte SID-Tabellen, die ebenfalls der Verbindung von Bewegungs- und Stammdaten dienen. Neben diesen gibt es noch die reinen Stammdaten-Tabellen, in denen nicht die SIDs, sondern die konkreten Attributwerte gespeichert werden (siehe Abb. 6).

<sup>5</sup> 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.

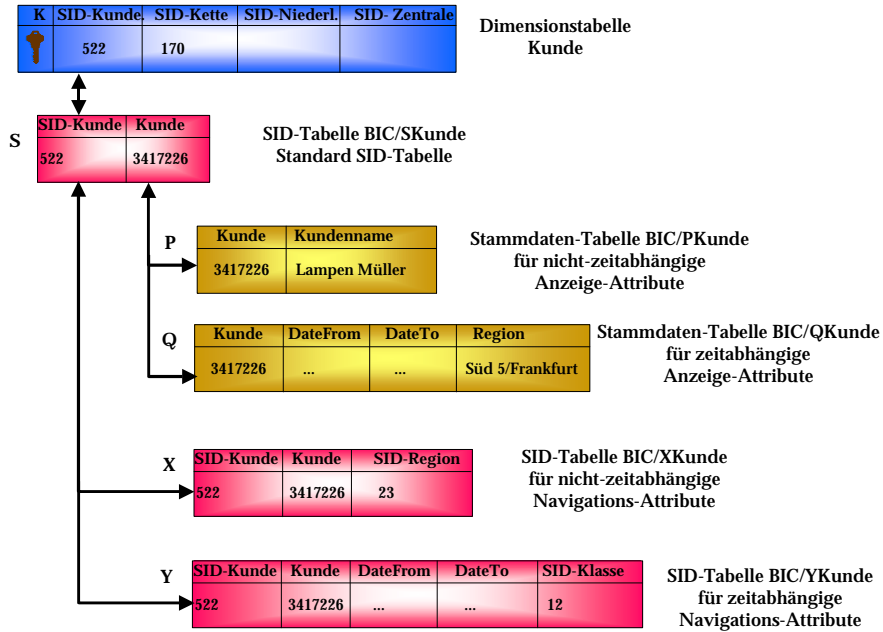


Abb. 6: Zusammenhang der verschiedenen Stammdaten-Tabellen

Der Zugriff auf das Modell bei konkreten Abfragen erfolgt immer von den Dimensionstabellen ausgehend, je nach konkreter Fragestellung dienen unterschiedliche Stammdaten-Tabellen zur näheren Eingrenzung und Bestimmung der qualifizierten Tupel aus der eigentlichen Dimensionstabelle, über die der Join mit der Faktentabelle durchgeführt wird.

## 2.4 Spezielle Eigenschaften des erweiterten Star Schemas

Eine wesentliche Metadaten-Information zu einem Info-Objekt als elementarstem Grundobjekt im Business Information Warehouse ist dessen Datentyp. Bei numerischen Datentypen gibt es neben den klassischen Formen die Möglichkeit, Mengen und Beträge zu definieren. Die Werte werden dann zusammen mit einer Einheit im System gespeichert, Geldbeträge haben dann zum Beispiel die Währung mit als Einheit.

Im Modell werden die Einheiten durch eine automatisch generierte Dimension abgebildet. Hierdurch ist das BW in der Lage, neben den Umrechnungen bei konstanten Verhältnissen, wie beispielsweise der Umrechnung von Mengenangaben mit Einheiten wie etwa Kilogramm und Tonnen, auch mit variablen Wechselkursen umgehen zu können. Die verschiedensten Umrechnungsvarianten sind durch selektierbare Kursumrechnungsarten darstellbar. Wird diese Funktionalität nicht benötigt, ist durch klassische numerische Datentypen der Overhead der Einheiten-Verwaltung vermeidbar und resultiert in schlankeren Modellen.

Die Einheiten-Umrechnung und Mehrwährungsfähigkeit sind wesentliche Leistungspotentiale des erweiterten Star Schemas, wie es im SAP BW zum Einsatz kommt.

### **3 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. Dieser Aspekt ist schon bei der Modellierung mit zu berücksichtigen.

#### **3.1 Alternative Lösungsmöglichkeiten bei Strukturbrüchen**

Der Zeitbezug ist für die Daten in mehrdimensionalen Modellen elementar, was sich auch in der besonderen Bedeutung der Zeitdimension ausdrückt, denn durch diese erfolgt inhärent die Historisierung der Fakten bzw. Bewegungsdaten. Die im Modellierungsprozess immer wieder auftretende Frage nach dem Umgang mit strukturellen Veränderungen in Dimensionen ist für den Aufbau von Data Warehouse- und OLAP-Systemen entscheidend, da hierdurch die Möglichkeiten der Analysen festgelegt sind.

Eine häufig anzutreffende Berichtsanforderung ist die Analyse aller Daten nach der jeweiligen aktuellen Struktur, nach einer historischen Struktur 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“). Zur Abbildung dieser Berichtsanforderungen stehen die folgenden vier Möglichkeiten zur Verfügung:

1. Anpassung des historischen Datenmaterials an neue Strukturen
2. Separate Speicherung des historischen Datenbestandes zusätzlich zum Komplettbestand mit neuen Strukturen
3. Aufbau paralleler Hierarchien mit alten bzw. neuen Strukturen
4. Temporale Datenbanken und Gültigkeitsstempel

Die Anpassung des alten Datenmaterials an die neuen Strukturen hat den Vorteil, dass der Datenbestand nicht aufgebläht wird und die Datenstrukturen überschaubar bleiben. Da die alten Strukturen aber verloren sind, können nicht mehr alle Berichtsanforderungen abgedeckt werden, nur Berichte entsprechend der aktuellsten gültigen Struktur sind abrufbar. Dies ist die einfachste Form des Umgangs mit Strukturbrüchen in Dimensionen.

Die Berücksichtigung von alten und neuen Strukturen wird möglich, wenn eine getrennte Speicherung der Daten nach den jeweiligen Strukturen erfolgt. Bei dieser Vorgehensweise können alte Auswertungen abgerufen werden, dies impliziert aber ein größeres Datenvolumen und aufwendigere Verfahren der Aktualisierung. Durch den Aufbau paralleler Hierarchien sind nicht nur die alten Daten nach alten Strukturen abrufbar, vielmehr können alle Zahlen mit beliebigen Strukturen angezeigt werden. Die dadurch entstehenden Dimensionsstrukturen sind aber eher unübersichtlich und die Darstellung

neuer Zahlen nach alten Strukturen kann für Endbenutzer verwirrend sein. Die Historisierung über Zeitstempel wie in temporalen Datenbanken ermöglicht ein umfassendes Berichtswesen mit beliebigen historischen Varianten von Strukturen. Nachteilig sind aber die tendenziell schlechtere Performance und teilweise noch unausgereifte Konzepte.

### 3.2 Hierarchische Struktur zwischen Merkmalen innerhalb einer Dimension

Im Business Information Warehouse sind die Daten konsequent in die zwei Bereiche der Bewegungsdaten und der Stammdaten getrennt, die über jeweils unterschiedlichen Ladeprozesse in das BW transportiert werden. Die Bewegungsdaten beziehen sich immer auf einen 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 (siehe Abb. 7). Ein typisches Beispiel hierfür im BW ist eine Zeithierarchie mit den Merkmalen Tag, Monat und Jahr oder eine Hierarchie über die Merkmale Kunde, Region und Land. Die Merkmale in einer Dimension qualifizieren eine Wertinformation in einem Bewegungsdatensatz.

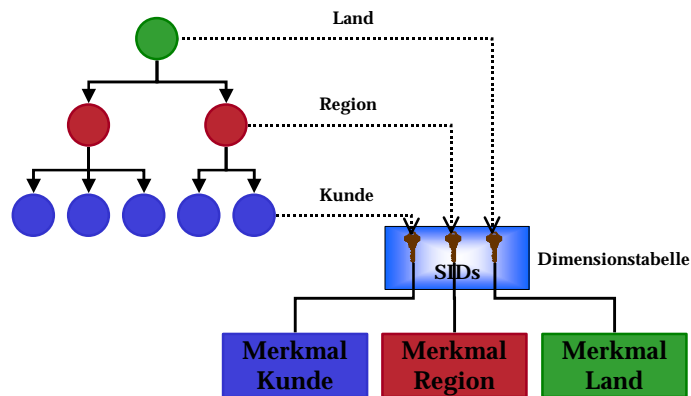


Abb. 7: Dimensionsstrukturen über Bewegungsdaten

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 Kunde, Region und Land 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, 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.

### **3.3 Modellierung von Hierarchien über Navigationsattribute**

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 detailliert 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 mit einem Attribut zu dem Merkmal der Basiselemente (siehe Abb. 8). 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 Hierarchie-Information in den Stammdaten hat zur Konsequenz, dass diese allen Info-Cubes gleichermaßen zur Verfügung steht, 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. Wie auch bei der Modellierung über Merkmale in der Dimension direkt bieten die Navigationsattribute keine vordefinierten Konsolidierungspfade, bei der Navigation können daher Ebenen übersprungen werden, der Anwender muss aber von den Attributen, die die Hierarchie formen, Kenntnis haben.

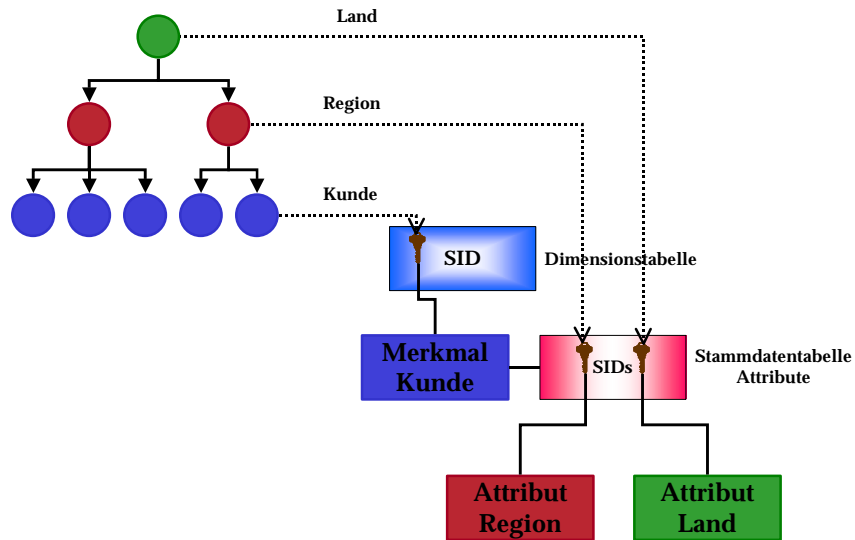


Abb. 8: Navigationsattribute als Grundlage für hierarchische Beziehungen

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 nicht mehr möglich ist.

### 3.4 Externe Hierarchien in den Stammdaten

Die bzgl. der Dimensionsstruktur flexibelste Art der Modellierung von hierarchischen Strukturen in Dimensionen sind im BW die Externen Hierarchien, da sie auf einer Darstellung in Form von rekursiven Beziehungen basiert [Sa00b]. 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.

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 zugrundelie-

genden 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 abgelegt werden. In dieser Inclusion Table genannten Tabelle werden die Beziehungen paarweise über Schlüssel abgelegt. Die Merkmalsknoten werden über den SID-Schlüssel des zugrundeliegenden Merkmals identifiziert, Textknoten bekommen einen identifizierenden Schlüssel mit negativem Vorzeichen, deren Texte in einer eigenen Tabelle gespeichert sind (siehe Abb. 9). Die Identifikation von Konsolidierungsebenen in dieser Form von Hierarchie erfolgt im BW nur auf Basis der Ebene, d. h. der Anzahl der Kanten ausgehend von der Wurzel bis zum Dimensionselement. Die Abbildung 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.

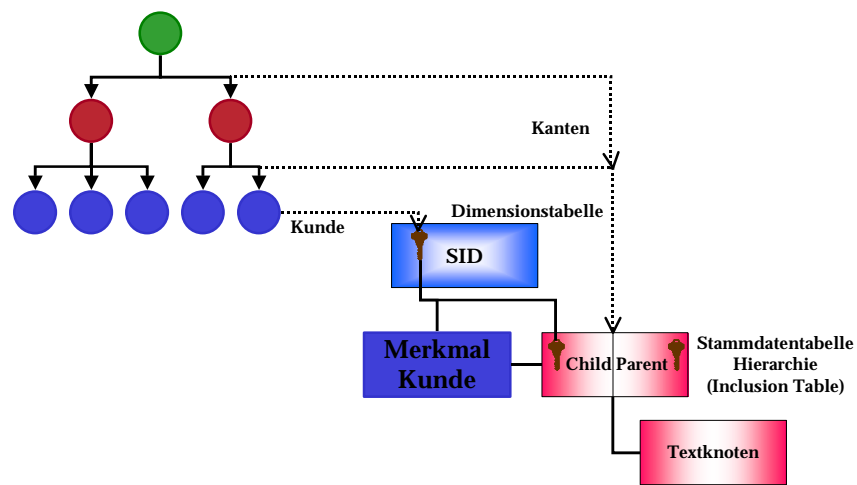


Abb. 9: Explizite Modellierung von Dimensionsstrukturen in Externen Hierarchien

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 festmachen. Dieses Datum ist dann in einer Abfrage für alle zeitabhängigen Strukturbestandteile gültig.

### 3.5 Entscheidungshilfen zur Dimensionsmodellierung

Die dargestellten Varianten im Kontext der Modellierung hierarchischer Dimensionsstrukturen verdeutlichen die Komplexität des BW-Datenmodells. Als Hilfe zur Wahl der Modellierungsvariante im Rahmen des Modellierungsprozesses erfolgt nun eine Gegenüberstellung der Alternativen anhand einzelner Kriterien.

Als eine erste elementare Entscheidungshilfe kann die Form der zu berücksichtigenden *Historisierung* angeführt werden. Nur in der Abbildung über Merkmale in den Bewegungsdaten kann die historische Sicht zum Buchungszeitpunkt implementiert werden. Die aktuell gültigen Strukturen sind bei Navigationsattributen und externen Hierarchien der Standard, wobei in beiden Fällen durch den Übergang zu zeitabhängigen Strukturen eine Historisierung ermöglicht wird. Bei externen Hierarchien ist es darüber hinaus auch möglich, einen zeitlichen Bezug über verschiedene Versionen oder Namen abzubilden.

Das Stammdatenkonzept des BW führt zu einem zweiten Anhaltspunkt, an dem sich die Wahl der Modellierungsform orientieren kann. Der *Wirkungsbereich* von Merkmalen in Bewegungsdaten ist gerade der Info-Cube, die damit abgebildete Struktur ist so zunächst nicht in anderen Info-Cubes verfügbar. Demgegenüber sind externe Hierarchien wie auch Navigationsattribute Bestandteil der Stammdaten und damit für alle Info-Cubes eines Systems gültig.

Unabhängig vom Gesichtspunkt der Transformation eines semantischen Datenmodells in das BW-Modell gibt es immer Aspekte des physischen Datenmodells, die einen signifikanten Einfluss auf die Abbildung der Datenstrukturen haben. Im SAP BW spielt die *Performance* eine herausragende Bedeutung, da diese einen kritischen Faktor darstellt. Dieser legt den Verzicht auf zeitabhängige Strukturen ebenso nahe, wie den sparsamen Einsatz von Navigationsattributen. Auf externe Hierarchien kann oftmals ohne Aggregatgar nicht performant zugegriffen werden.

Das Konzept der mehrdimensionalen Datenmodellierung impliziert eine Vorstellung von den *Navigationspfaden* innerhalb der Dimensionshierarchien. Nur in externen Hierarchien sind diese direkt berücksichtigt, in der Darstellung von hierarchischen Strukturen über Merkmale oder Navigationsattribute ist dieser Zusammenhang nicht modelliert, sondern muss sich aus der Kenntnis des Modells ergeben.

Die möglichen abbildbaren *Dimensionsstrukturen* bilden ein weiteres Entscheidungskriterium, denn nur externe Hierarchien bieten die volle Flexibilität. Da Merkmale und Navigationsattribute jeweils mit einer festen Ebene in der Dimensionsstruktur korrespondieren, sind damit im allgemeinen nur balancierte Strukturen darstellbar.

Dieses Kriterium erfährt noch eine Verschärfung durch die Anforderung, auch *m:n-Beziehungen* in Dimensionen zu unterstützen. Externe Hierarchien erlauben Blätter mit mehreren übergeordneten Elementen, in Bewegungsdaten ist dies nur in der Form, wie es gebucht wurde, möglich. Über Navigationsattribute ist dies nicht implementierbar.

Diese angeführten Kriterien beziehen sich auf die statischen Aspekte, aber die Anforderungen im Rahmen des ETL-Prozesses zur Befüllung des Data Warehouse resultieren in einem weiteren Anhaltspunkt zur Entscheidungshilfe. Die performante *Reorganisation* von Datenbeständen bei Strukturveränderungen ist insbesondere bei großen Installationen, in denen nur ein kleines Ladefenster für die ETL-Prozesse verbleibt, sehr wichtig. Bei Bewegungsdaten-basierten Strukturen ist dies nur mit einem kompletten Neuaufbau möglich. Externe Hierarchien lassen sich hingegen sehr schnell flexibel verändern.



## 4 Zusammenfassung

Analyseorientierte Informationssysteme, deren Fokus in der zeitnahen Versorgung betrieblicher Entscheidungsträger mit relevanten Informationen zu Analyse Zwecken liegt, zielen auf die Unterstützung der dispositiven und strategischen Prozesse in Unternehmen ab. Genau diesen Bereich adressiert das Business Information Warehouse als zentrale Data Warehouse-Lösung und bildet damit die Grundlage für vielfältige analytische Anwendungen. Wichtige Eigenschaften des Systems der SAP sind dabei die Möglichkeiten der Währungsumrechnung und das Stammdatenkonzept. Von anderen Lösungen hebt es insbesondere der mitgelieferte Business Content ab, denn dieser bietet gerade für R/3-Quellsysteme vorgefertigte betriebswirtschaftliche Schablonen zum schnellen Aufbau von analyseorientierten Applikationen auf Basis einer Datenhaltung im Business Information Warehouse.

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. Die Möglichkeiten zur Abbildung vielfältiger Strukturen für Konsolidierungspfade sind ein Aspekt, der die Analysemöglichkeiten stark beeinflusst. Hierbei spielen auch Aspekte der Zeitabhängigkeit und der Umgang mit Änderungen in Dimensionen eine Rolle.

Der Modellierung hierarchischer Dimensionsstrukturen kommt im Kontext analyseorientierter Informationssysteme eine besondere Bedeutung zu, daher unterstützt das BW vielfältige Dimensionsstrukturen.

Bei der Abbildung von Hierarchien im Business Information Warehouse gibt es die drei grundsätzlichen Möglichkeiten der Darstellung über Merkmale in einer Dimension, über Navigationsattribute des Basismerkmals sowie über Externe Hierarchien. Dies ist ein Freiheitsgrad der Datenmodellierung. Für die Entscheidung, welche der Varianten in einer konkreten Anwendung vorzuziehen ist, sind verschiedene Kriterien heranzuziehen. Erstes Unterscheidungsmerkmal ist der Wirkungsbereich, denn der Ansatz des unternehmensweiten Data Warehouse mit einem zentralen konsistenten Modell propagiert die Modellierung in Stammdaten, die für alle Info-Cubes gültig sind. Ein sehr bedeutendes Kriterium ist die Performance, wobei vom Grundprinzip her in Dimensionen direkt modellierte hierarchische Beziehungen tendenziell performanter sind als Navigationsattribute und Externe Hierarchien. Die Kriterien der flexiblen Reorganisation sowie der Vielfalt an unterstützten Dimensionstypen sprechen in vielen Fällen für die Modellierung über Externe Hierarchien. Die drei Varianten schließen sich aber nicht gegenseitig aus, auch die gleichzeitige Abbildung auf verschiedenen Wegen ist im Modell möglich. Ein weiteres wesentliches Kriterium berücksichtigt die Anforderungen an die Unterstützung für Dimensionsstrukturänderungen und deren unterschiedlichen Anforderungen an ein Berichtswesen. Die Historisierungen und Nachverfolgung von Strukturänderungen werden unter dem Stichwort Zeitabhängigkeit diskutiert. Hier bietet das Business Information Warehouse Möglichkeiten, über die zeitabhängige Ausgestaltung von Stammdaten verschiedene Berichtsszenarien abzubilden. Die hinzukommende Flexibilität in den Berichtsanforderungen wird allerdings mit einer erhöhten Komplexität und schlechterer Performance erkauft.

## Literaturverzeichnis

- [Gl97] Gluchowski, P.: Data Warehouse-Datenmodellierung – Weg von der starren Normalform. In: Datenbank-Fokus, Nr 11, 1997; S. 62-66.
- [Ha99] Hahne, M.: 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, 1999; S. 145-170.
- [Ha02] Hahne, M.: Logische Modellierung mehrdimensionaler Datenbanksysteme. Deutscher Universitäts-Verlag, Wiesbaden, 2002.
- [Nu98a] Nußdorfer, R.: Star Schema, das E/R-Modell steht Kopf – Teil 1: Modellieren von Faktentabellen. In: Datenbank-Fokus, Nr 10, 1998; S. 22-28.
- [Nu98b] Nußdorfer, R.: Star Schema – Teil 1: Modellierung von Dimensionstabellen. In: Datenbank-Fokus, Nr 11, 1998; S. 16-23.
- [Sa00a] SAP: Multidimensional Modeling with BW. ASAP for BW Accelerator, 2000.
- [Sa00b] SAP: Hierarchies in SAP BW. ASAP for BW Accelerator, 2000.