



**Data Warehousing mit dem
SAP Business Information Warehouse**
- Universitäres Berichtswesen -

Theorieskript

SAP und mySAP.com sind Marken der SAP Aktiengesellschaft Systeme, Anwendungen, Produkte in der Datenverarbeitung, Neurottstraße 16, D-69190 Walldorf. Der Herausgeber bedankt sich für die freundliche Genehmigung der SAP Aktiengesellschaft, das Warenzeichen im Rahmen des vorliegenden Titels verwenden zu dürfen. Die SAP AG ist jedoch nicht Herausgeberin des vorliegenden Titels oder sonst dafür presserechtlich verantwortlich.

R/3, Business Information Warehouse, mySAP Business Intelligence und weitere im Text erwähnte SAP-Produkte sowie die entsprechenden Logos sind Marken oder eingetragene Marken der SAP AG.

Alle anderen Produkte sind Marken oder eingetragene Marken der jeweiligen Firmen.

Vorwort

Unsere Lehremission

Wir lehren das Fach Wirtschaftsinformatik an der Technischen Universität München, damit unsere Studentinnen und Studenten eine fachlich fundierte Ausbildung im Umfeld der Entwicklung, Implementierung und Nutzung von Informations- und Kommunikationstechniken in Unternehmen und Verwaltung erhalten.

Hierbei legen wir großen Wert auf eine theoriefundierte, methodisch ausgerichtete und praxisorientierte Ausbildung unserer Studierenden. Darunter verstehen wir die Entwicklung flexibler und teamfähiger Persönlichkeiten, die in der Lage sind, ihr Fachwissen auf reale, soziotechnische Problemstellungen anzuwenden und dabei eine wissenschaftliche Herangehensweise an praktische Problemstellungen als Methode beherrschen.

Wir stellen uns vor, dass unsere Absolventinnen und Absolventen

- in der Lage sind, sowohl die Aufgaben als **System- und Anwendungsspezialist**, **Projektmanager** oder **Informationsmanager** im Unternehmen als auch als **Unternehmensberater** mit dem Schwerpunkt des Informationstechnikeinsatzes auszufüllen,
- Grundlagen beherrschen, um **qualifizierte Wissenschaftler** im Bereich der Wirtschaftsinformatik zu werden und
- **Führungsaufgaben**, auch in der Unternehmensführung, anstreben.

Wir achten darauf, dass bei der Vermittlung des Fachwissens methodisches und theoretisch fundiertes Wissen dominieren. In verschiedenen Lehrformen vermitteln wir ein Grundverständnis in der Beherrschung komplexer Anwendungssysteme. Zusätzlich wird das Lehrangebot durch Veranstaltungen abgerundet, die aktuelle Fragen wie beispielsweise Ubiquitous Computing, Service Engineering, Computer Supported Cooperative Work (CSCW) und Informatikrecht behandeln.

Zusätzlich veranstalten wir im Rahmen einer Public Private Partnership Arbeitskreise und Übungen unter Verwendung von SAP®-Software.

Vor allem in den Arbeitskreisen und Übungen wird die Aktivität des Einzelnen in der Gruppe gefordert und gezielt gefördert.

Wir wollen die zukünftigen Absolventinnen und Absolventen so gut und umfassend ausbilden, dass sie in allen Unternehmen die strategische Führungsaufgabe Informationsmanagement entscheidend mit gestalten und übernehmen können.

Kontakt

Bei Fragen und Anregungen zum vorliegenden Dokument wenden Sie sich bitte an:

Matthias Mohr

Technische Universität München

Lehrstuhl für Wirtschaftsinformatik (I17) – Prof. Dr. H. Krcmar

Boltzmannstr. 3

85748 Garching b. München

matthias.mohr@in.tum.de

Inhaltsverzeichnis

1	Das BW als Data Warehouse der SAP AG.....	1
2	Verbesserung der betrieblichen Informationslogistik durch Data Warehouses	1
2.1	Was ist ein Data Warehouse ?	1
2.2	Business Intelligence als Rahmenwerk für Data Warehousing	3
2.3	Nutzenpotenziale beim Data Warehousing	4
2.4	Eine kurze Vorstellung des Business Information Warehouse	4
2.4.1	Grobaufbau des SAP BW	5
2.4.2	Werkzeuge und Instrumente des SAP BW	5
3	Der Lebenszyklus eines Data Warehouse.....	6
3.1	Das Business Dimensional Lifecycle Modell als Beschreibungsrahmen für Data Warehouses.....	6
3.2	Projektplanung und Projektmanagement für eine DW-Einführung	8
3.3	Anforderungsdefinition.....	10
3.4	Design der technischen Architektur und Produktauswahl.....	11
3.4.1	Architekturvarianten beim Aufbau eines Data Warehouse	11
3.4.2	Auswahl eines Data Warehouse Produkts	12
3.4.3	Technische Anforderungen des SAP BW.....	13
3.5	Modellierung multidimensionaler Datenstrukturen.....	14
3.5.1	Ebenen der Datenmodellierung.....	14
3.5.2	Strukturen multidimensionaler Daten	15
3.5.3	Semantische Datenmodellierung für Data Warehouses.....	19
3.5.4	Logische Datenmodellierung zur Erstellung von Datenbankschemata	24
3.5.5	Umsetzung von Datenbankschemata im Rahmen der physischen Modellierung	32
3.5.6	Spezialitäten und Problemfelder der Modellierung im BW	33
3.5.7	Vorkonfigurierte Informationsmodelle in Form des Business Content.....	39
3.5.8	Daten über Daten	40
3.6	Design und Implementierung der Datenbereitstellung.....	41
3.6.1	Persistente und nicht persistente Stagingszenarien.....	41
3.6.2	Elemente der Datenbereitstellung.....	43
3.6.3	Laden von Stammdaten.....	44
3.6.4	Vom Quellsystem zum Datenziel	45
3.7	Spezifikation und Entwicklung von Endbenutzeranwendungen	47
3.7.1	Unterschiedliche Anwender benötigen unterschiedliche Analysewerkzeuge	47
3.7.2	Möglichkeiten des Datenzugriffs im BW	48
3.7.3	Datenauswertungen mit Queries im BEx Analyzer	49
3.7.4	Einfache Informationsdistribution durch Webreporting.....	52
3.7.5	Abdeckung der Benutzeranforderungen im BW-Reporting	53
3.8	Aufgaben im Rahmen der Verteilung, Wartung und Ersetzung eines Data Warehouse	54
4	Ist das SAP BW ein klassisches Data Warehouse ?.....	56
5	Literaturverzeichnis	57
I	Anhang	61
	Screenshots.....	61
	Metadatenmodell des SAP BW.....	62
	Internetquellen.....	63

Abbildungsverzeichnis

Abbildung 1: Abgrenzungen des DW-Begriffs (Quelle: in Anlehnung an Schinzer/Bange/Mertens 1999, 16).....	3
Abbildung 2: Architektur des SAP BW (Quelle: eigene Darstellung)	5
Abbildung 3: Erweitertes Business Dimensional Lifecycle Modell (Quelle: in Anlehnung an Kimball/Ross 2002, 332)	7
Abbildung 4: Klassifizierung von Geschäftsprozessen (Quelle: Kimball/Ross 2002, 347) ...	10
Abbildung 5: Dreidimensionaler InfoCube (Quelle: eigene Darstellung)	16
Abbildung 6: Klassische und SAP-spezifische Strukturelemente eines MDM (Quelle: Eigene Darstellung).....	18
Abbildung 7: Notationselemente des MERM (Quelle: in Anlehnung an Totok (2000), 192) 20	
Abbildung 8: Beispiel eines MERM (Quelle: Eigene Darstellung).....	20
Abbildung 9: Beispiel-ERM mit zwei Geschäftsprozessen (Quelle: Eigene Darstellung).....	22
Abbildung 10: Bestimmung der Granularität (Quelle: Eigene Darstellung)	23
Abbildung 11: Starschema (Quelle: Eigene Darstellung).....	25
Abbildung 12: Verwendung von Surrogat-IDs im erweiterten Starschema (Quelle: Eigene Darstellung).....	28
Abbildung 13: Galaxyschema (Quelle: Eigene Darstellung).....	29
Abbildung 14: Snowflakeschema (Quelle: Eigene Darstellung)	29
Abbildung 15: Zeitdimension mit Merkmal <i>Semester</i> (Quelle: Eigene Darstellung).....	34
Abbildung 16: Faktenlose Faktentabelle am Beispiel von universitären Wahlen (Quelle: Eigene Darstellung).....	36
Abbildung 17: Referenzmerkmal (Quelle: Eigene Darstellung).....	37
Abbildung 18: Schwach besetzte Matrix (Quelle: Eigene Darstellung)	38
Abbildung 19: MultiCubes (Quelle: Eigene Darstellung)	39
Abbildung 20: Stagingszenarien (Quelle: Eigene Darstellung, in Anlehnung an SAP 2000a)	42
Abbildung 21: Klassifikation der Daten im BW (Quelle: Eigene Darstellung).....	45
Abbildung 22: Phasen des Datenbereitstellungsprozesses aus Flatfiles (Quelle: Eigene Darstellung).....	46
Abbildung 23: Anwendergruppen und ihre Anforderungen (Quelle: SAP 2001f, 2-12).....	48
Abbildung 24: Metadatenmodell Querydefinition und Query (Quelle: Eigene Darstellung). 49	
Abbildung 25: Formelvariable mit Ersetzungspfad (Quelle: Eigene Darstellung).....	51
Abbildung 26: Metadatenmodell des SAP BW (Quelle: eigene Darstellung).....	62

Tabellenverzeichnis

Tabelle 1: Werkzeuge des SAP BW und ihre Zielgruppen (Quelle: eigene Darstellung)	6
Tabelle 2: Planungsphase (Quelle: eigene Darstellung)	10
Tabelle 3: Führende Anbieter von Data Warehouse Produkten (Quelle: Mertens/Bange/Schinzer 2000a, 35)	12
Tabelle 4: Entwurfsebenen der multidimensionalen Modellierung (Quelle: in Anlehnung an Böhnlein/Ulbrich-vom Ende o.J., 4)	15
Tabelle 5: OLTP und OLAP im Vergleich (Quelle: eigene Darstellung).....	15
Tabelle 6: Schritt für Schritt vom ERM zum MDM (Quelle: Eigene Darstellung).....	23
Tabelle 7: Umsetzung logischer Datenbankschemata im SAP BW (Quelle: Eigene Darstellung).....	30
Tabelle 8: Abdeckung der Benutzeranforderungen im BW-Reporting (Quelle: Eigene Darstellung).....	54

Verzeichnis der Screenshots

Screenshot 1: ABAP-Routine zur Buchstabenkonvertierung (© SAP AG)	61
Screenshot 2: Request-Administration im InfoCube (© SAP AG)	61

Abkürzungsverzeichnis

ABAP	Advanced Business Application Programming
ADAPT	Application Design for Analytical Processing Technologies
APO	Advanced Planner and Optimizer
ASP	Application Service Providing / Provider
AWB	Administrator Workbench
BAPI	Business Application Programming Interface
BARC	Business Application Research Center
BC	Business Content
BDS	Business Document Service
BEx	Business Explorer
BEx Analyzer	Business Explorer Analyzer
BEx Browser	Business Explorer Browser
BI	Business Intelligence
BIW	Business Information Warehouse
BW	Business Information Warehouse
CIO	Chief Information Officer
CSA	Client Server Architektur
CSS	Cascading Style Sheets
CSV	Comma Separated Values
DBMS	Datenbank-Management-System
DSS	Decision Support System
DW	Data Warehouse / Data Warehousing
EIS	Executive Information System
ERM	Entity-relationship-Modell
ETL	Extraktion, Transformation, Laden
EUS	Entscheidungsunterstützungssystem (siehe DSS)
FIS	Führungsinformationssystem (siehe EIS)
GIS	Geographisches Informationssystem / Geoinformationssystem
ITS	Internet Transaction Server
MDM	Multidimensionales Datenmodell / Multidimensionale Datenmodellierung
ME/R Model	Multidimensionales Entity relationship Model
MERM	Multidimensionales Entity relationship Model

MIS	Management Information System
MOLAP	Multidimensionales Online analytical processing
ODBO	OLE-DB-for-OLAP
ODS	Operational Data Store
OLAP	Online analytical processing
OLTP	Online transaction processing
PDA	Personal Digital Assistant
PSA	Persistent Staging Area
ROLAP	Relationales Online analytical processing
SAP	Systeme, Anwendungen und Produkte in der Datenverarbeitung (Firmenname)
SAPGui	SAP Graphical User Interface
SDWM	Semantisches Data Warehouse Modell
SID	Surrogat-ID
VBA	Visual Basic for Applications
WTS	Windows Terminalserver

1 Das BW als Data Warehouse der SAP AG

Die folgenden Kapitel geben einen Überblick über die Grundideen von Data Warehouses (DW) und das Data Warehouse Produkt der SAP AG *Business Information Warehouse*. Diese beiden inhaltlichen Stränge werden parallel geführt, d.h. passend zur Erläuterung des theoretischen Data Warehouse Backgrounds werden Idee, Aufbau und Funktionsweise des BW vorgestellt. Dadurch wird auch die Beurteilung vereinfacht, inwiefern das BW ein klassisches Data Warehouse ist und in welchen Punkten es sich von diesem Konzept unterscheidet. Das Kernstück dieser Betrachtung, der Lebenszyklus eines Data Warehouse, wird auf dem *Lifecycle-Modell* nach RALPH KIMBALL aufgebaut, das zuvor kurz beschrieben wird. Im Anhang ist ein Metadatenmodell aufgeführt, das den Zusammenhang der wichtigsten Strukturelemente des SAP BW verdeutlicht.¹

2 Verbesserung der betrieblichen Informationslogistik durch Data Warehouses

Das DW-Konzept wird aufgrund der vielen verschiedenen Definitionen und Deutungen oft falsch verstanden. Daher wird es in diesem Kapitel formal beschrieben und anschließend in das übergeordnete Konzept *Business Intelligence* eingeordnet.

2.1 Was ist ein Data Warehouse ?

Die aktuelle Situation in vielen Unternehmen ist durch steigende Datenflut bei gleichzeitigem Informationsdefizit gekennzeichnet (Behme/Mucksch 1996, 9). Obwohl die Informationsversorgung v.a. des Managements einen wichtigen Wettbewerbsfaktor darstellt, fehlt häufig die richtige Information in der richtigen Menge am richtigen Ort zur richtigen Zeit. Dies ist der Ansatzpunkt einer betrieblichen Informationslogistik, die mit Hilfe von Data Warehouses verbessert werden soll (Behme/Mucksch 1996, 20).

Ein Data Warehouse ist kein Produkt, sondern ein Konzept, das sich der Datenproblematik von managementunterstützenden Systemen annimmt. W. H. INMON, oft als Vater des Data Warehousing bezeichnet, hat folgende Definition geprägt, die hier als Arbeitsdefinition Verwendung finden wird: „A data warehouse is a subject-oriented, integrated, nonvolatile², time-variant collection of data in support of management's decision“ (Behme 1996, 31; Inmon 2001a). Hierbei bedeuten die vier Hauptmerkmale:

1. subject-oriented: Die Themenausrichtung an **Sachverhalten des Unternehmens**, z.B. Kunden- oder Produktkriterien, wird im BW durch das konsequente Einordnen aller Daten in Fachbereiche und durch die Bezugnahme auf Geschäftsprozesse realisiert (Seemann/Schmalzridt/Lehmann 2001, 18). Im Gegensatz dazu sind operative Daten

¹ SAP (2001a) liefert eine übersichtliche Auflistung der Bedeutung aller BW-spezifischen Ausdrücke.

² Volatilität: Flüchtigkeit (Langenscheidt 1995). Hier: Grad, mit dem sich Daten im Laufe der Nutzung verändern.

immer auf einzelne betriebliche Funktionen bezogen (Schinzer/Bange/Mertens 1999, 14; Bange/ Schinzer o.J., 1).

2. integrated: Mit dem DW-Konzept wird eine unternehmensweite **Integration von Daten** in einem einheitlich gestalteten System angestrebt (Mucksch/Behme 2000, 11). Vereinheitlichung und Integration externer und interner Daten bedeutet weniger die physische Zentralisierung der Daten in einem einzigen Datenpool, sondern deren logische Verbindung. Integration bedeutet konsistente Datenhaltung im Sinne einer Struktur- und Formatvereinheitlichung durch Maßnahmen wie Vergabe eindeutiger Bezeichnungen, Anpassung der Datenformate und Herstellung einer semantischen Integrität (Mucksch/Behme 2000, 11ff.). Ebenso tragen Elemente wie *einheitliche Merkmale* und *standardisierte Kennzahlen* zu einer Datenintegration bei.
3. nonvolatile: Bei einem DW handelt es sich um eine **dauerhafte Sammlung** von Informationen, auf die im Gegensatz zu OLTP-Systemen (online transaction processing³) nur in Form von Lese- und Einfügeoperationen zugegriffen werden darf, um die Nicht-Volatilität der Daten sicherzustellen.⁴ Dieser Forderung kann jedoch nur bedingt zugestimmt werden, da Korrekturen von aus Quellsystemen geladenen Daten auf jeden Fall möglich sein müssen (Behme 1996, 31). Das BW bietet hierfür eine Eingangsablage in Form der Persistent Staging Area (PSA)⁵, in der manuelle Korrekturen zur Validierung und Fehlerbehebung nach dem Extraktionsvorgang durchgeführt werden können (SAP 2000a, 1; SAP 2000b).
4. time-variant: Während bei operativen Systemen eine zeitpunktgenaue Betrachtung der Daten im Mittelpunkt steht, liegt das Interesse bei Auswertungen im DW eher in einer Zeitraumbetrachtung, z.B. einer Trendanalyse (Behme 1996, 31). Der **Zeitraumbezug** ist daher impliziter oder expliziter Bestandteil der Daten in einem DW. Ein Ansatz zur Herstellung dieses Zeitraumbezugs im BW ist die obligatorische Verwendung einer Zeitdimension in jedem Informationsspeicher.

Diese eng eingegrenzte DW-Definition wurde von vielen Autoren ergänzt. So sind einige der Meinung, dass ein DW (im engeren Sinne) neben der eigentlichen Datensammlung und ihrer Verwaltung um Aspekte wie Anbindung, Extraktion und Transformation von Fremddaten ergänzt werden sollte. Andererseits erfolgt eine Ausdehnung in Richtung Analyse und Präsentation mit Hilfe entsprechender Werkzeuge (Schinzer/Bange/Mertens 1999, 15). Folgende Abbildung veranschaulicht die beiden DW-Definitionen:

³ Vgl. Kapitel 3.5.2.1

⁴ Während des normalen DW-Betriebs werden daher keine sonst üblichen Locking-Mechanismen benötigt (Mucksch/Behme 2000, 13).

⁵ Vgl. Kapitel 3.6.2.3

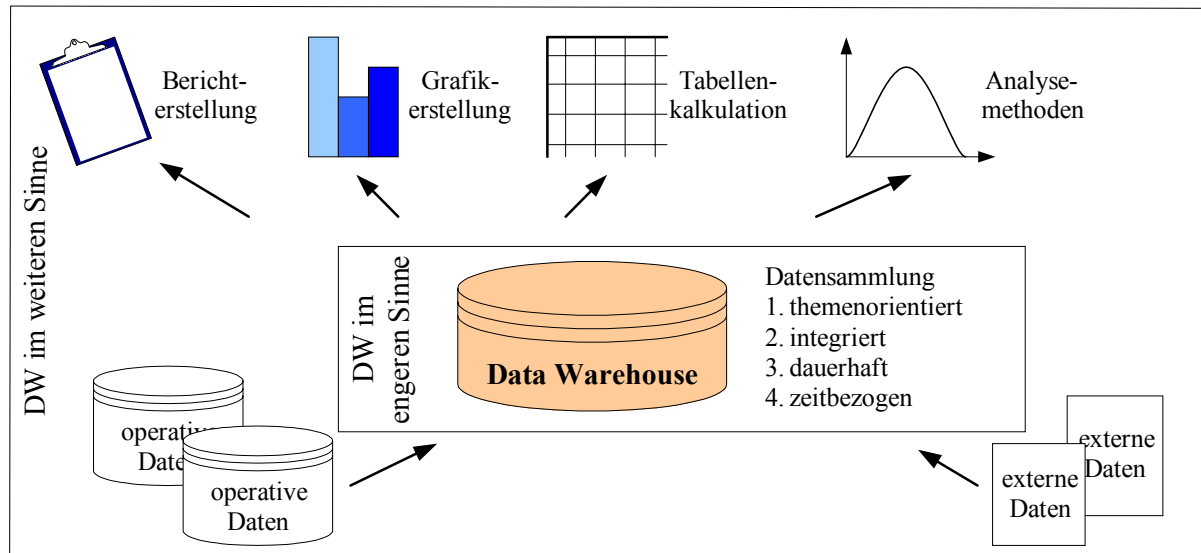


Abbildung 1: Abgrenzungen des DW-Begriffs (Quelle: in Anlehnung an Schinzer/Bange/Mertens 1999, 16)

Wie in späteren Kapiteln beim Vorstellen der Funktionalität des SAP BW deutlich werden wird, entspricht dieses Produkt der Definition eines DW im weiteren Sinne. Die Datensammlung auf dem BW-Server wird ergänzt durch umfassende Funktionalitäten zum Extrahieren operativer und anderer externer Daten und im Funktionsumfang des BW ist durch den in Microsoft Excel integrierten BEx Analyzer ein Reportingtool samt Visualisierungsmöglichkeiten enthalten, in dem durch das individuelle Gestalten der Abfragen mittels Formeln eine Vielzahl von Analysemethoden anwendbar sind. Man spricht auch von einem dreistufigen DW-Konzept, bestehend aus Datenbereitstellung, Datenhaltung, Informationsanalyse (Seemann/Schmalzridt/Lehmann 2001, 25).

2.2 Business Intelligence als Rahmenwerk für Data Warehousing

Die Archivierung der integrierten Daten in einem Data Warehouse alleine bringt noch keine Wettbewerbsvorteile, sondern erst deren kreative und intelligente Verwendung (Behme 1996, 30). Diese Art der Nutzung von unternehmensweit verfügbarem Wissen wird heute trotz einiger Kontroversen⁶ als *Business Intelligence (BI)* bezeichnet und bildet somit als nochmalige Ausdehnung der weiten DW-Definition sozusagen das Frontend eines Data Warehouse. Vorläufer dieser Idee waren Konzepte wie Management Information Systeme (MIS), Decision Support Systeme (DSS/EUS) und Führungsinformationssysteme (FIS/EIS) (Schinzer/Bange/Mertens 1999, 5ff.). Der Begriff Business Intelligence wurde 1989 von der Gartner Group geprägt und folgendermaßen definiert: "Business Intelligence is the process of transforming data into information and, through discovery, into knowledge." (Behme 1996, 37). Die SAP AG bietet mit ihrem Produktpaket mySAP Business Intelligence, das Teil der umfassenden Strategie mySAP.com ist, eine entsprechende Lösung für das Informationsmanagement und die Entscheidungsunterstützung an. Data Warehousing ist das Kernstück dieses Pakets, das aus vielen Bausteinen von der Etablierung eines Berichtswesens für die Unternehmensführung bis hin zur Unterstützung von strategischen Aufgaben besteht: Data

⁶ Der Begriff BI ist immer wieder in Diskussionen aufgekommen und oft wird behauptet, BI als Dachbegriff habe versagt, was jedoch nicht der Fall ist (Sexl/Bange 2002).

Warehousing, Reporting und Analyse, Information Deployment über den mySAP Workplace, Planung und Simulation, Balanced Scorecard, Web Content Management und analytische Applikationen wie z.B. Customer Relationship Analytics (Seemann/Schmalzridt/Lehmann 2001, 16f.). Im Bereich Data Mining bietet SAP keine eigene Lösung an, sondern bindet mit einer zertifizierten Schnittstelle den IBM Intelligent Miner an mySAP BI an (Seemann/Schmalzridt/Lehmann 2001, 20).

2.3 Nutzenpotenziale beim Data Warehousing

Der Aufbau und Betrieb einer Datensammlung ist kein Selbstzweck, sondern muss durch entsprechenden Nutzen gerechtfertigt werden. Es lassen sich technische und betriebswirtschaftliche Nutzenpotenziale unterscheiden (Reiser/Holthuis 1996, 121-128).

Aus **technischer** Sicht liegt der Hauptnutzen von DWs in einer integrierten Datenbasis für managementunterstützende Systeme, wodurch auch die horizontale Datenintegration verbessert wird, was den Entscheidungsträgern die Aufgabe abnimmt, Konsistenz und Qualität der Daten selbst prüfen zu müssen. Zudem werden die operativen Anwendungen entlastet, da die Daten in einem DW im Normalfall redundant gehalten werden. Durch die multidimensionalen Datenstrukturen sind schnelle Abfragen und Reports möglich, die durch OLAP-Techniken schnell und einfach aufbereitet, sortiert und gruppiert werden können. Die Trennung von Datenhaltung und -präsentation ermöglicht die verschiedensten Zugriffsmöglichkeiten vom Excel-Report am lokalen PC bis hin zur Informationsübermittlung per SMS auf ein mobiles Gerät.

Daneben bieten DWs mehrere **betriebswirtschaftliche** Nutzenpotenziale. So sollte ein DW die Informationsbereitstellung für die Entscheidungsträger aller Ebenen verbessern. Durch die frühzeitige Erkennung von Trends mit Hilfe von DWs kann die Wettbewerbsfähigkeit des Unternehmens erhöht werden. Ein als Frühwarnsystem ausgestaltetes DW kann zügig auf Umweltveränderungen reagieren. Schließlich werden Kundenservice und -zufriedenheit verbessert, indem man die in DWs enthaltenen Kundendaten mit den angebotenen Unternehmensleistungen abgleicht und dadurch kundenspezifische Strategien entwickelt. Die Harmonisierung der Interpretationsunterschiede von betriebswirtschaftlich relevanten Begriffen ist ein weiterer wichtiger Nebeneffekt der Planung eines DW (Frie 2000, 11).

Die geschilderten Erfolgspotenziale sind jedoch nur dann erreichbar, wenn die Einführung eines Data Warehouse sauber geplant und durchgeführt (Reiser/Holthuis 1996, 127) und das Data Warehouse während seines „Lebens“ professionell administriert wird. Aus diesem Grund wird im nächsten Hauptkapitel der typische Lebenszyklus eines DW vorgestellt und seine Phasen daraufhin genauer analysiert.

2.4 Eine kurze Vorstellung des Business Information Warehouse

Bevor das Data Warehouse Konzept und das SAP BW in den einzelnen Phasen des Lebenszyklus genauer betrachtet werden, erfolgt in diesem Kapitel ein kurzer Blick auf die BW-Architektur und die zur Verfügung stehenden Werkzeuge, damit die nachfolgend erläuterten Aktivitäten konkreten Bereichen und Instrumenten zugeordnet werden können.

2.4.1 Grobaufbau des SAP BW

Die Architektur des BW orientiert sich an dem bereits dargestellten dreistufigen Aufbau von Data Warehouses: Datenbereitstellung, Datenhaltung und Informationsanalyse. Folgende Abbildung zeigt den groben Aufbau des BW mit den wichtigsten Bausteinen.

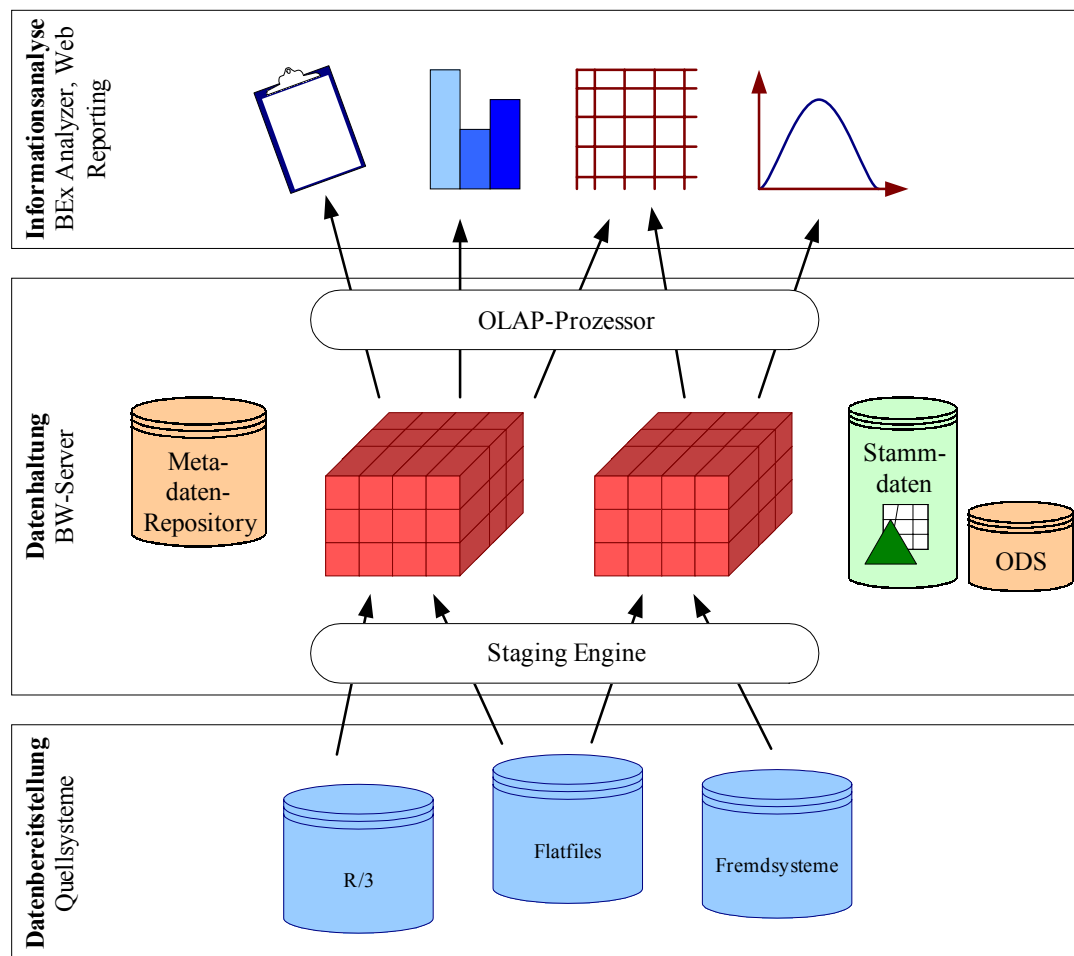


Abbildung 2: Architektur des SAP BW (Quelle: eigene Darstellung)

2.4.2 Werkzeuge und Instrumente des SAP BW

Für die Datenbereitstellung und -haltung steht mit der **Administrator Workbench (AWB)** ein zentrales Verwaltungswerkzeug zur Verfügung, das Aufbau, Pflege und Betrieb des BW-Systems ermöglicht (Seemann/Schmalzridt/Lehmann 2001, 26). Ebenso ermöglicht die AWB die Überwachung des Datenladeprozesses (SAP 2001b, 12). Wie der Name schon aussagt, ist die AWB eher für den Data Warehouse Administrator als für den Anwender gedacht. Die AWB ist eine Transaktion⁷, die in einem SAPGui aufgerufen wird.

Der Business Explorer ist die Komponente des BW, die flexible Reporting- und Analysewerkzeuge zur strategischen Analyse und Entscheidungsunterstützung im Unternehmen zur Verfügung stellt (SAP 2001c, 12). Der **Business Explorer Analyzer (BEx Analyzer)** ist das Reportingwerkzeug auf der Basis von Microsoft Excel⁸ für die Definition von Reports sowie

⁷ Transaktionscode der AWB: RSA1

⁸ Nach Erfahrungen des Autors unterstützen Microsoft Excel 2000 und XP den BEx Analyzer problemlos.

für deren Aufbereitung und Präsentation sowohl in Microsoft Excel als auch in einem Webbrowser (Webreporting). Ergänzend gibt es den Business Explorer Analyzer on the web als webbasierte Lösung des BEx Analyzer (Seemann/Schmalzridt/Lehmann 2001, 26).

Für den rein informationskonsumierenden Endanwender ist der **Business Explorer Browser** entwickelt worden, der die Verwaltung und das Ausführen von Reports unterstützt. Ergänzt wird diese Funktionalität um die Möglichkeit, auf externe, auch unstrukturierte Informationsquellen wie Webseiten, lokale Dateien oder SAP R/3 Transaktionen zu verweisen. Dadurch wird der dem BEx Browser zugrunde liegende Portalgedanke deutlich (Seemann/Schmalzridt/Lehmann 2001, 82).

Werkzeug	Funktionalität	Zielgruppe	Integration
Administrator Workbench	Administration des BW-Systems	Systemadministrator	SAPGui
Business Explorer Analyzer	Definition, Aufbereitung und Präsentation von Reports, Webreporting	Berichtsautoren, Analysten, Konsumenten	Microsoft Excel
Business Explorer Browser	Verwaltung und Ausführen von Reports, Integration weiterer Informationsquellen	Konsumenten	Eigenständige Anwendung

Tabelle 1: Werkzeuge des SAP BW und ihre Zielgruppen (Quelle: eigene Darstellung)

3 Der Lebenszyklus eines Data Warehouse

3.1 Das Business Dimensional Lifecycle Modell als Beschreibungsrahmen für Data Warehouses

Zur Vorstellung von Data Warehouses im allgemeinen und des SAP BW im speziellen wird im Rahmen des vorliegenden Lehrmaterials der DW-Lebenszyklus als Beschreibungs- und Gliederungsrahmen verwendet, d.h. alle in den nachfolgenden Kapiteln angesprochenen Aspekte werden in einen chronologischen Ablauf von der Projektplanung bis zum Wachstum eines DW eingeordnet.

Für dieses Vorgehen wird eine erweiterte Version des *Business Dimensional Lifecycle Modells* nach RALPH KIMBALL verwendet, das ausführlich in Kimball et al. (1998) und in komprimierter Form in Kimball/Ross (2002) dargestellt wird. Seine Bezeichnung leitet sich von dem Gedanken ab, dass ein DW-Projekt die Bedürfnisse des *Geschäftslebens* mit der Hilfe *multidimensional* strukturierter Daten in einem *Implementierungszyklus* mit festem Beginn und festem Ende befriedigen soll (Kimball/Ross 2002, 332). Die folgende Abbildung stellt die Hauptphasen des Modells grafisch dar:

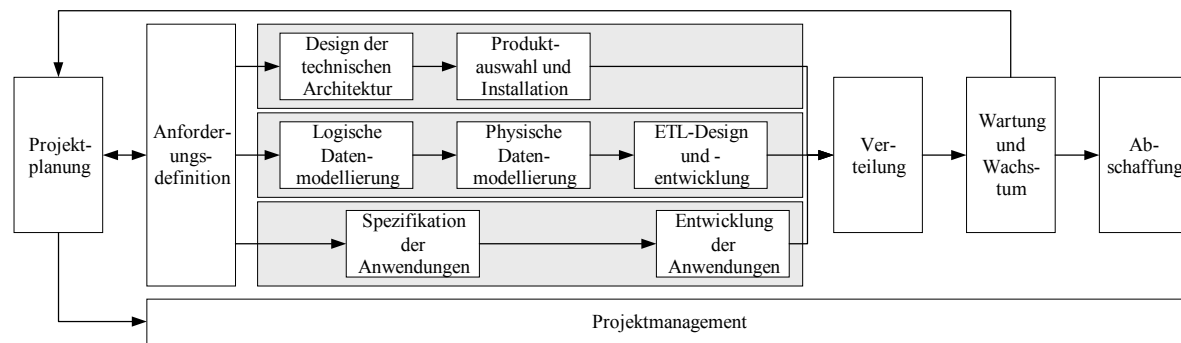


Abbildung 3: Erweitertes Business Dimensional Lifecycle Modell (Quelle: in Anlehnung an Kimball/Ross 2002, 332)

Die erste Phase des Lebenszyklus ist die **Projektplanung**, bei der das Unternehmen auf die Einführung des DW vorbereitet wird, indem Fragen wie Rechtfertigung, Dimensionierung und personelle Ausstattung des Projektes erörtert werden. Neben der Projektplanung stellt fortwährendes und begleitendes **Projektmanagement** einen unverzichtbaren Bestandteil dar, um die planmäßige Durchführung des Projektes sicherzustellen und eventuelle Abweichungen vom Zeitplan frühzeitig erkennen zu können (Kimball/Ross 2002, 334-340).

Essentiell für den Projekterfolg ist das Verstehen und Berücksichtigen der Wünsche der DW-Benutzer. Daher wird während der **Anforderungsdefinition** die Zielgruppe definiert, deren Anforderungen mittels geeigneter Methoden ermittelt und aufbereitet. Wichtig ist die durch den Doppelpfeil zwischen Projektplanung und Anforderungsdefinition angedeutete enge Abhängigkeit dieser zwei Phasen. Nach der Anforderungsdefinition teilt sich der Lebenszyklus in drei Bereiche. Der obere Bereich behandelt Fragen der Data Warehouse Technik. Der mittlere befasst sich mit allen Fragestellungen rund um logische und physische Datenmodellierung, wobei hier zusätzlich die semantische Modellierungsebene betrachtet wird. Der untere Bereich spiegelt schließlich die Definition und Entwicklung von Endbenutzeranwendungen wider. Diese drei Bereiche laufen teilweise parallel ab (Kimball/Ross 2002, 340-347).

Als erster Schritt werden im **Technik-Bereich** die Anforderungen an die technische Architektur des DW definiert, indem Fragen wie Verteilung der Daten, Performance, Datenvolumen u.ä. angesprochen werden. Darauf aufbauend wird im zweiten Schritt ein geeignetes Produkt ausgewählt, das diesen Anforderungen genügt. Hierbei können Methoden und Instrumente wie Produktbewertungsmatrizen, Marktanalysen und Prototypen zum Einsatz kommen (Kimball/Ross 2002, 347-353).

Der **Daten-Bereich** des Modells befasst sich mit der Frage, wie die zu speichernden Daten strukturiert werden sollen und wie sie in das Data Warehouse gelangen. Hierfür wird zunächst eine semantische und logische Datenmodellierung durchgeführt, bei der das Konzept der multidimensionalen Modellierung zum Einsatz kommt, denn im Gegensatz zu OLTP-Datenbanken werden die Daten in Data Warehouses meistens in mehrdimensionaler Form gespeichert, um einen schnellen und einfachen Zugriff zu gewährleisten. Im Rahmen der physischen Modellierung werden die logischen Datenmodelle am System umgesetzt. Aufgabe des ETL-Designs (Extraktion, Transformation, Laden) ist es dafür zu sorgen, dass sowohl Stamm- als auch Bewegungsdaten problemlos aus den Quellsystemen in das Data Warehouse gelangen (Kimball/Ross 2002, 353-362).

Hauptaufgabe im Bereich „**Anwendungen**“ ist das Spezifizieren und Entwickeln von auf das DW aufsetzenden Endbenutzeranwendungen, welche die zu Beginn des Projekts definierten Anforderungen der Benutzer erfüllen. Neben Ergonomie und Bedienbarkeit tauchen in dieser Phase Begriffe wie Web Access und Informationsportale auf (Kimball/Ross 2002, 362-364).

Zusammengeführt werden die drei Bereiche Technik, Daten und Anwendungen vor der **Verteilung** und Bekanntmachung des DW im Unternehmen. Das Data Warehouse wird den Benutzern nicht „nackt“ serviert, sondern durch Schulungsmaßnahmen und Supportangebote begleitet. Dies ist deshalb von entscheidender Bedeutung für das Gelingen des gesamten Projektes, weil sich in dieser Phase entscheidet, ob das Produkt akzeptiert wird oder nicht (Kimball/Ross 2002, 364f.).

Als letzten Schritt betrachtet das Modell die Problematik von **Wartung und Wachstum** eines DW, wodurch ein Folgeprojekt ausgelöst werden kann und somit der Zyklus von neuem beginnt. Hier geht es vor allem um die Gestaltung des laufenden Supports für die Benutzer des Data Warehouse. (Kimball/Ross 2002, 365f.).

Das Business Dimensional Lifecycle Modell von KIMBALL ermöglicht eine konsequente und konsistente Darstellung der Aufgaben, die bei der Einführung und dem Betrieb eines Data Warehouse relevant sind. Lediglich die Frage der **Abschaffung** eines Data Warehouse wird vom Modell nicht beantwortet, es liefert z.B. keine Hilfestellungen zur Entscheidung, zu welchem Zeitpunkt ein DW-System durch ein neues abgelöst werden sollte (Krcmar 2000, 112). Da es sich hierbei um eine interessante Fragestellung handelt, wird das Modell um diesen Aspekt erweitert.

3.2 Projektplanung und Projektmanagement für eine DW-Einführung

Die Einführung eines Data Warehouse Projekts beginnt optimalerweise mit einer umfassenden Projektplanung und wird durch ein professionelles Projektmanagement begleitet (Kimball et al. 1998, 41ff.).

Erste Hauptaufgabe der Projektplanungsphase ist es, das Unternehmen für das DW-Projekt **vorzubereiten**, wofür KIMBALL folgende Erfolgsfaktoren nennt: Zunächst muss ein *Sponsor* im Unternehmen gefunden werden, der persönlich vom Wert des Projektes überzeugt ist und es daher vorantreiben und unterstützen wird. Parallel zum Sponsoring als persönlichem Treiber sollte ein weiterer Treiber existieren, der das Projekt fachlich rechtfertigt. Kann das DW kritische Probleme im Unternehmen lösen, ist dieser Treiber schnell gefunden. Einen weiteren wichtigen Erfolgsfaktor in dieser Phase stellt die Durchführbarkeit des Projekts dar, d.h. technische, personelle und datenbezogene Probleme müssen lösbar sein. Ein gutes Verhältnis zwischen Fach- und IT-Abteilung trägt das seine zum Gelingen des Projekts bei, die IT-Abteilung sollte die Arbeit der Fachabteilungen verstehen und respektieren. Letztendlich leistet das Verständnis der analytischen Arbeitsweise innerhalb des Unternehmens einen erheblichen Beitrag zur optimalen Vorbereitung auf das DW-Projekt. Es muss herausgefunden werden, ob Entscheidungen eher auf Basis von quantitativen Informationen getroffen werden oder vorwiegend intuitiv.

Die nächste Aufgabe der Projektplanung besteht darin, den **Umfang** des Projekts festzulegen. Es bietet sich an, den Fokus zu Beginn auf die Abbildung eines einzigen Geschäftsprozesses zu legen und mit fortschreitenden Kenntnissen und Erfahrungen langsam den Horizont zu erweitern. Oft wird der Umfang exogen vorgegeben, z.B. durch die Anzahl der zu berücksichtigenden Abteilungen oder Tochterunternehmen, durch die Art und Beschaffenheit der zu speichernden Daten, durch einen bestimmten Zeitrahmen oder durch ein Projektbudget.

Der Aufbau eines Data Warehouse ist kein Selbstzweck, sondern muss durch eine Kosten-Nutzenanalyse **gerechtfertigt** werden.⁹ Während die durch Hardware, Software (Inmon 2001b) und die langwierigen und aufwändigen Datenaufbereitungs- und -vereinheitlichungsarbeiten entstehenden immensen Kosten (Manning 2002) relativ einfach zu bestimmen sind, wird der Nutzen eines Data Warehouse oft durch gestiegene Erlöse oder Gewinne und eine verbesserte Entscheidungsfindung repräsentiert, was jedoch schwierig nachzuweisen und zu quantifizieren ist. Aufgrund der Wertschöpfungsferne kann der Zusatznutzen eines DW häufig nicht in Zahlungsgrößen ausgedrückt werden.

Data Warehousing stellt hohe Anforderungen an die Projektorganisation eines Unternehmens, denn die notwendige Einbeziehung vieler Organisationseinheiten und somit auch unterschiedlicher betriebswirtschaftlicher Disziplinen erhöht die Komplexität des Projekts maßgeblich. Diese wird zusätzlich durch das Bestreben verstärkt, eine hohe Integration der Daten zu erreichen (Schwarz 1999, 16f.). Das **DW-Projektteam** wird aus Mitarbeitern sowohl der Fachabteilungen als auch der IT-Abteilung¹⁰ zusammengestellt, wobei es durchaus vorkommen kann, dass eine Person gleichzeitig mehrere Rollen wahrnimmt. Die Rollenverteilung¹¹ hängt von Projektgröße und Verfügbarkeit und Fähigkeiten der einzelnen Beteiligten ab.

Um im Rahmen eines effektiven Projektmanagements alle anfallenden Aufgaben inhaltlich und zeitlich besser koordinieren zu können, ist ein **Projektplan** ein wertvolles Hilfsmittel. Änderungen in den Anforderungen und im geplanten Ablauf müssen genauestens überwacht werden, bevor sie lawinenartig anwachsen und das gesamte Projekt gefährden.

Folgende Abbildung fasst die wichtigsten Aktivitäten im Rahmen der Projektplanung und des Projektmanagements zusammen:

Aktivitäten der Planungsphase	Beschreibung
Vorbereiten des Unternehmens	Sponsoren, Fachliche Rechtfertigung, Durchführbarkeit, Verhältnis zwischen Fachabteilung und IT-Abteilung, „Analytische Kultur“
Umfang des Projekts festlegen	Mit einem Geschäftsprozess beginnen, dann ausdehnen
Projekt rechtfertigen	Kosten-Nutzen-Analyse
Projektteam zusammenstellen	Fachabteilungen, IT-Abteilung, externe Beteiligte
Projektplan erstellen	Planung, Steuerung, Kontrolle, Erkennen von Abweichungen

⁹ Vgl. Kapitel 2.3

¹⁰ Selbstverständlich gehören auch Mitarbeiter externer Partner zum Projektteam.

¹¹ Die SAP AG stellt im Dokument ETBTD006 (Q78) eine Art Stellenbeschreibung für verschiedene Rollen innerhalb eines BW-Projektteams zur Verfügung.

Tabelle 2: Planungsphase (Quelle: eigene Darstellung)

3.3 Anforderungsdefinition

Die Anforderungen der Benutzer des DW zu erkennen und beim Systemaufbau umzusetzen, ist grundlegend für den Projekterfolg (Kimball/Ross 2002, 340-347).

Nach entsprechenden Vorarbeiten müssen die Benutzer dazu bewogen werden, über ihre Arbeit und vor allem über ihre Arbeitsweise zu reden, wobei es nicht um technische Details gehen sollte. Typischerweise wird nach den Verantwortlichkeiten und dem organisatorischen Aufbau gefragt. Im weiteren Verlauf des Gesprächs können Fragen nach wichtigen **Kennzahlen** und der Art und Weise der **Erfolgsmessung** gestellt werden. Die Antworten fließen später direkt in ein multidimensionales Modell ein.

Im Zentrum der Befragung stehen stets die wichtigsten **Geschäftsprozesse** des Unternehmens, denn diese dienen in späteren Phasen als Ausgangspunkt zur Modellierung der zu speichernden Daten. Sind die zentralen Geschäftsprozesse ermittelt, können sie anhand des folgenden Schemas klassifiziert werden:

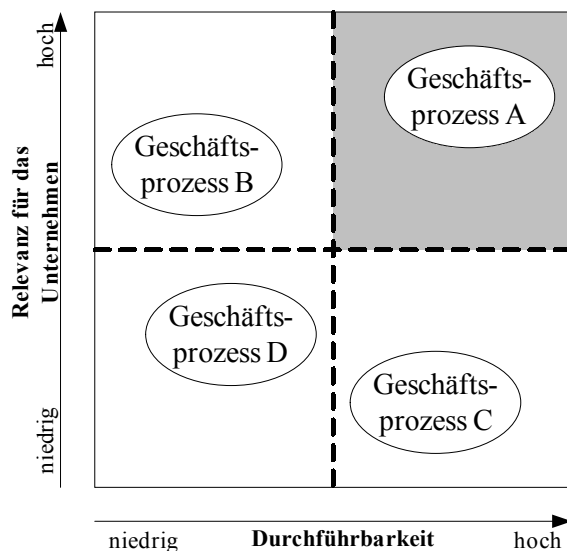


Abbildung 4: Klassifizierung von Geschäftsprozessen (Quelle: Kimball/Ross 2002, 347)

Geschäftsprozesse vom Typ D sollten bei einer Abbildung in einem DW weitestgehend umgangen werden, denn weder stiften sie dem Unternehmen besonders hohen Nutzen noch ist ihre Umsetzung einfach. Typ B und Typ C stellen Zwischenfälle dar.¹² Geschäftsprozesse vom Typ A hingegen verdienen eine eingehende Betrachtung, da sie einerseits relativ einfach umsetzbar sind und andererseits eine hohe Bedeutung für das Unternehmen haben. In Kapitel 3.5.3.2 wird die Umsetzung von Geschäftsprozessen in ein Datenmodell genauer analysiert.

¹² Prozesse vom Typ B könnten z.B. einer genauen Kostenanalyse unterzogen werden, da sie wichtig für das Unternehmen sind. Prozesse vom Typ C könnten daraufhin untersucht werden, ob sie nicht doch einen gewissen Nutzen stiften und daher ins DW miteinbezogen werden sollten.

3.4 Design der technischen Architektur und Produktauswahl

Ein wichtiges Ergebnis der Anforderungsdefinition sind die aus den fachlichen Wünschen ableitbaren Vorgaben für den technischen Aufbau des DW. So wirken sich Wünsche der Fachabteilungen wie z.B. „ständige, weltweite Verfügbarkeit aller Reports“ entscheidend auf die technische Architektur aus.

3.4.1 Architekturvarianten beim Aufbau eines Data Warehouse

Wie bereits erwähnt, entspricht die Architektur des SAP BW dem klassischen dreistufigen Konzept von Data Warehouses im weiteren Sinne, bestehend aus Datenbereitstellung, Datenthaltung und Informationsanalyse.¹³

Als Quellsysteme zur Datenbereitstellung sind mySAP.com-Komponenten wie R/3-Systeme, andere BW-Systeme, SAP CRM Systeme und externe Systeme wie Nicht-SAP-Systeme und Dateien, z.B. Flatfiles, möglich (SAP 2001d, 2-11). Dieser Aspekt wird eingehender in Kapitel 3.6 besprochen.

Aufbauend auf den operativen Vorsystemen befindet sich auf der nächsten Ebene der DW-Architektur das eigentliche Data Warehouse, hier: der BW-Server. Neben der Verwaltung von Metadaten ist seine Hauptaufgabe, die aus den Quellsystemen extrahierten Anwendungsdaten zu speichern. Der Aufbau des Datenservers entspricht in den meisten Fällen einem der folgenden Architekturtypen (Schinzer/Bange/Mertens 1999, 19-23):

- Das Konzept des **virtuellen DW** besteht im direkten Zugriff der Endanwenderapplikationen auf die operativen Datensysteme, was eigentlich dem Grundgedanken widerspricht, die operativen Systeme zu entlasten. Ein DW im engeren Sinne, als gesonderte Datensammlung, ist nicht vorhanden, daher die Bezeichnung *virtuell*. Der Vorteil liegt in der schnellen und kostengünstigen Realisierung, da die operativen Daten nicht transformiert und keine neue technische Infrastruktur angeschafft werden muss. Mit den sog. RemoteCubes können im SAP BW Datenspeicher erzeugt werden, die selbst keine Daten speichern, sondern direkt auf die Quellsysteme zugreifen und daher das Konzept des virtuellen DW repräsentieren.
- Bei einem **zentralen DW** handelt es sich um eine physische Datenbasis, die parallel zu den operativen Datenbeständen existiert. Ist von einem DW die Rede, wird damit automatisch ein zentrales DW assoziiert. Mit BasisCubes werden im SAP BW Datenspeicher erzeugt, die selbst Daten enthalten und damit die Bausteine eines zentralen DW darstellen.
- Ein **Data Mart** ist ein subjekt- oder abteilungsspezifisches Data Warehouse. Hier werden im Unterschied zum zentralen DW nicht alle Daten eines Unternehmens gespeichert, sondern dezentral auf mehreren kleinen DW-Servern, den Data Marts, verteilt. Dem zusätzlichen Verwaltungsaufwand z.B. durch die notwendige Synchronisation der Daten steht der Vorteil gegenüber, dass die Fachabteilungen flexibel auf ihre oft sehr speziellen Daten zugreifen können. Ein komplettes SAP BW System kann als Data Mart aufgefasst und über das *Data Mart Interface* mit anderen BW-Systemen konnektiert werden, die dann zusammen als ein einziges DW aufgefasst werden.

¹³ Vgl. Kapitel 2.1

Das SAP BW lässt sich als flexibles DW-System bezeichnen, da es sowohl Elemente von virtuellen, zentralen und dezentralen (Data Marts) Data Warehouses vereint.

3.4.2 Auswahl eines Data Warehouse Produkts

Die technischen Anforderungen und die daraus generierte Architektur dienen in gewisser Weise als „Einkaufsliste“ für die Suche nach einem geeigneten Data Warehouse Produkt, das den formulierten Anforderungen gerecht wird. Dieser Suchprozess entspricht in den Grundzügen der Beschaffung jedes beliebigen IT-Produkts, welche in einigen Unternehmen nach fest vorgegebenen Regeln abläuft. So muss der betriebliche Beschaffungsprozess bekannt sein; eine Produktbewertungsmatrix kann als Hilfsmittel zum Vergleich der verschiedenen Produkte dienen, die über ein systematisches Absuchen des DW-Marktes gefunden werden. Selbstverständlich spielen bisherige Erfahrungen mit bestimmten Produkten oder Anbietern eine große Rolle, wie auch persönliche Beziehungen zu Lieferanten. Übertrifft ein bestimmtes Produkt die Konkurrenten bei weitem, ist die Auswahl einfach. In Zweifelsfällen kann es sinnvoll sein, einen Prototyp zu testen, wobei aufgrund des hohen Aufwands nicht mehr als zwei Produkte verglichen werden sollten (Kimball/Ross 2002, 351ff.).

Der DW-Markt ist durch eine Vielzahl von Anbietern gekennzeichnet, die umfassende Produktsuiten für den Aufbau einer DW-Lösung anbieten. Das Spektrum der Produkte umfasst einen großen Funktionsumfang, wodurch eine fundierte Auswahlentscheidung erschwert wird (Mertens/Bange/Schinzer 2000a, 34). Das Business Application Research Center (BARC) hat 12 führende Anbieter von DW-Produkten identifiziert:

Anbieter	Produkt	URL
Cognos	BI Platform	www.cognos.com
Hyperion Software	Hyperion Essbase	www.hypion.com
IBM	Visual Warehouse	www.ibm.com
Information Builders	SmartMart	www.informationbuilders.com
Microsoft	OLAP Services	www.microsoft.com
MicroStrategy	MicroStrategy	www.microstrategy.com
NCR	TerraData	www.ncr.com
Oracle	Warehouse	www.oracle.de
Pilot	DSS Suite	www.pilotsw.com
SAP	BW	www.sap.de
SAS Institute	SAS System	www.sas.com
Seagate Software	Holos	www.seagatesoftware.com

Tabelle 3: Führende Anbieter von Data Warehouse Produkten (Quelle: Mertens/Bange/Schinzer 2000a, 35)

Je nach Einsatzbedingungen und Aufgabenschwerpunkt beim einführenden Unternehmen eignen sich einige Werkzeuge besser, andere dagegen weniger gut für eine bestimmte Problemstellung. So zeichnen sich z.B. die Lösungen von IBM, Information Builders, Oracle und SAS durch eine gute Unterstützung der Datenintegration und umfangreiche Transformations- und Metadatenverwaltungsmöglichkeiten aus.

Das **SAP BW** eignet sich der BARC-Untersuchung zufolge besonders gut für die Auswertung und Analyse der in einem Data Warehouse enthaltenen Daten. Die unterschiedlichen vorgefertigten Applikationen und die Werkzeuge zur Erstellung individueller Anwendungen

(BEx Analyzer und Web Reporting) sind weitere Stärken dieses Produkts (Mertens/Bange/Schinker 2000a, 38). Auch beim Einführungsprozess ist das BW den Konkurrenten voraus: Das Konzept der SAP ist im Vergleich zu den klassischen DW-Einführungen mit dem Wandel von der Individual- zur Standardsoftware gleichzusetzen (Hecht/Bange/Schinker 2000, 42). Der Hauptvorteil jedoch besteht in den vorkonfigurierten Informationsmodellen, dem Business Content¹⁴, der bisher in der Data Warehousing Welt einmalig ist. Dieser Vorteil kommt besonders bei der Überführung von R/3-Daten in das BW zum Tragen, denn dieser Prozess wird durch die vorkonfigurierten Extraktoren und den Business Content erheblich vereinfacht¹⁵ (Michel 1999, 105). Nach Auskunft der Unternehmensberatung KPMG wird der Einsatz des BW als Data Warehouse dann empfohlen, wenn der Anteil der R/3-Quellsysteme bei mindestens 50% liegt. Als eigenständiges Produkt ist das BW zwar nicht an den Einsatz von R/3 gebunden, es profitiert davon jedoch erheblich.

3.4.3 Technische Anforderungen des SAP BW

In dieser Veranstaltung wird der BW-Server von einem HCC betrieben, daher werden lediglich die clientseitigen technischen Anforderungen für ein SAP BW System Release 2.1C¹⁶ kurz beschrieben.

Das **BW Frontend** ist eine BW-spezifische Frontendkomponente, welche zuzüglich zum SAPGui lokal auf dem Frontend-PC oder dem Windows Terminalserver (WTS) installiert wird (SAP 2002a). Beim Ausführen des Setup-Programms besteht die Möglichkeit, neben dem „normalen“ R/3-Frontend, APO und anderen Komponenten, die BW-Komponenten zur Installation zu selektieren. Nach der Installation stehen der SAPGui für die Ausführung der AWB, der BEx Analyzer und der BEx Browser zur Verfügung. Die BW Frontend Komponenten sind Bestandteil des SAP Frontends. Das SAP Frontend für Windows wird seit Juli 1999 in sog. *Compilations* ausgeliefert. Eine neue Compilation wird immer dann zusammengestellt, wenn sich eine darin enthaltene Komponente geändert hat, z.B. das BW oder APO AddOn (SAP 2002b).

Das BW Frontend 2.1C wird auf den **Betriebssystemen** Windows 95, Windows 98, Windows 2000, Windows NT und Windows Terminalserver in Verbindung mit SAPGui 4.6D, 6.10 oder höher unterstützt. Windows XP wird nur in Verbindung mit SAPGui 6.10 oder höher für 2.1C oder höher unterstützt (SAP 2002a).

Für die **Hardwareausstattung** des Frontend-PC empfiehlt die SAP AG einen Pentium PC mit 300 MHz oder einen anderen Prozessor mit entsprechender Leistung. Der Hauptspeicher sollte mindestens 128 MB umfassen. Schließlich wird eine Grafikkarte mit 32.768 Farben und Unterstützung einer Auflösung von 1.024 · 768 Pixel empfohlen. Für die Installation auf der lokalen Festplatte sind 50 MB Speicherplatz notwendig (SAP 2002a).

Für den Einsatz des **Webreporting**¹⁷ ist die lokale Installation eines Internet Transaction Server (ITS)¹⁸ notwendig. Dieser Server wird lokal, d.h. bei der Institution, die das BW nutzt,

¹⁴ Vgl. Kapitel 3.5.7

¹⁵ Ein modernes R/3-System besteht aus ca. 20.000 Tabellen (Seemann/Schmalzridt/Lehmann 2001, 327).

¹⁶ Zum Zeitpunkt der Erstellung dieses Lehrmaterials wurde von den HCCs ein BW-System 2.1C angeboten.

¹⁷ Vgl. Kapitel 3.7.4

installiert und nicht bei einem HCC. Eine Auflistung der Hardwarevoraussetzungen und eine genaue Schritt-für-Schritt-Anleitung zur Installation des ITS liefert SAP (2001e). Der Betrachter von webbasierten Reports muss clientseitig über einen Browser verfügen, der dem HTML-Standard 3.2 genügt und zumindest die elementaren Funktionen von Cascading Style Sheets (CSS) 1.0 unterstützt. Hierbei handelt es sich in der Regel um einen Browser mit einer Version höher als 3.0. Für Browser mit JavaScript-Implementierung werden automatisch grafische Verbesserungen zugeschaltet und es werden kompaktere HTML-Seiten erzeugt (SAP 2002c).

3.5 Modellierung multidimensionaler Datenstrukturen

3.5.1 Ebenen der Datenmodellierung

Da bei Data Warehouse Entwicklungen primär eine datenorientierte Sicht eingenommen wird, sind Datenmodelle das wichtigste Entwicklungswerkzeug. Datenmodelle dienen zum einen als Grundlage für Diskussionen mit den Entscheidungsträgern, zum anderen nutzt sie der Datenbankdesigner als Grundlage zur Überführung in ein Datenbankmodell (Holthuis 2000, 159). Außerdem sind Datenmodelle ein wichtiges Instrument der Standardisierung von Begrifflichkeiten innerhalb eines großen Unternehmens (Maier 1998, 134).

Beim Datenbankentwurf von OLTP-Systemen hat sich die Unterscheidung in die Entwurfsebenen des konzeptuellen (semantischen), logischen und physischen Entwurfs mit den korrespondierenden Entwurfsergebnissen konzeptuelles, logisches und physisches Schema durchgesetzt. Diese Trennung kann auch auf den OLAP-Bereich übertragen werden (Böhnlein/Ulbrich-vom Ende o.J., 4). Daher wird in diesem Kapitel auf die Konzeption der Datenbasis im Rahmen der semantischen und logischen Modellierung eingegangen und daraufhin die konkrete Umsetzung dieses logischen Schemas in Form einer physischen Modellierung am BW-System dargestellt. Somit wird die Problematik der Datenmodellierung auf den Informationssystem-Beschreibungsebenen *Fachkonzept*, *DV-Konzept* und *Implementierung* analysiert (Scheer 1998, 14ff.). Folgende Tabelle gibt einen Überblick über die bekannten Entwurfsmethoden, auf die nachfolgend teilweise genauer eingegangen wird.

Entwurfsebene	Entwurfsmethoden
Konzeptueller (semantischer) Entwurf	<ul style="list-style-type: none"> ▪ Semantisches Data Warehouse Modell ▪ Multidimensionales ERM ▪ Dimensional Fact Modeling ▪ Application Design for Analytical Processing Technologies
Logischer Entwurf	<ul style="list-style-type: none"> ▪ Starschema ▪ Erweitertes SAP-Starschema ▪ Fact/Constellation Schema ▪ Galaxy Schema ▪ Snowflake Schema ▪ Partial Snowflake Schema
Physischer Entwurf	<ul style="list-style-type: none"> ▪ Speicherungsstrukturen

¹⁸ Der ITS kann kostenlos von der SAP AG bezogen werden. Somit fallen für die Verwendung des Web-reporting keine zusätzlichen Lizenzkosten an.

	<ul style="list-style-type: none"> ▪ Zugriffsmechanismen ▪ Datenbanktuning ▪ usw.
--	--

Tabelle 4: Entwurfsebenen der multidimensionalen Modellierung (Quelle: in Anlehnung an Böhnlein/Ulbrich-vom Ende o.J., 4)

3.5.2 Strukturen multidimensionaler Daten

3.5.2.1 Operationelle und analytische Informationssysteme

In der Regel wird ein Data Warehouse getrennt von den operativen Anwendungssystemen betrieben. Dies ermöglicht eine effizientere Datenverwaltung und gibt dem Benutzer einen flexibleren und schnelleren Zugriff auf die benötigten internen und externen Unternehmensdaten. Operationelle Datenverarbeitung (online transaction processing, OLTP) bezieht sich auf Systeme, die das Tagesgeschäft von Unternehmen erledigen, indem Transaktionen akkurat und effizient verarbeitet werden (Poe/Reeves 2000, 40). Das beste Beispiel für Transaktionssysteme sind die verschiedenen Komponenten von SAP R/3, z.B. Auftragsbearbeitung, Produktionsplanung und Finanzbuchhaltung. Im Gegensatz dazu zielt die in Data Warehouses angewendete analytische Datenverarbeitung (online analytical processing, OLAP) auf die Unterstützung von Entscheidungen in den Bereichen Strategie und Unternehmenssteuerung (Poe/Reeves 2000, 40). Folgende Tabelle zeigt die wichtigsten Unterschiede:

	OLTP	OLAP
Ziel	Effizienz durch Automation	Wettbewerbsvorteile durch Wissensgenerierung
Inhalt der Daten	Anwendungsbezogen, funktionsbezogen	Themenbezogen
Art der Daten	Transaktionsdaten	Aggregierte Daten
Alter der Daten	Aktuell, zeitnah: 30-60 Tage	Historisch (oft 8-10 Jahre alt), aktuell, zukünftig
Datenvolumen	Klein	Sehr umfangreich
Hauptfunktionalität	Häufige Änderungen	Zeitabhängige Auswertungen
Datenintegration	Wenig mit anderen Anwendungen integriert	Integrierte Daten aus einer Menge von Anwendungen
State of the Art beim Datenbanksystem	Relationale Datenbanken	Relationale und multidimensionale Datenbanken
Datenmodell	Normalisiert (häufig 3. Normalform)	Denormalisiertes Datenmodell
Semantische Modellierungsmethode	Entity Relationship Modell	Multidimensionales ERM
Erlaubte Operationen auf den Datenbestand	Einfügen, Aktualisieren, Löschen, Lesen	Lesen

Tabelle 5: OLTP und OLAP im Vergleich (Quelle: eigene Darstellung)

RALPH KIMBALL formuliert die Notwendigkeit von OLAP-Datenbanken folgendermaßen: „We have created transaction-oriented databases that cannot be queried.“ (Kimball et al. 1998, 142).

3.5.2.2 Multidimensionalität von Daten

Das Hauptcharakteristikum von Daten, die in Data Warehouses gespeichert werden, ist deren Multidimensionalität. Dies bedeutet, dass die Daten nicht wie in OLTP-Systemen tabellenartig (maximal zweidimensional) dargestellt, sondern von beliebig vielen analyserelevanten Kriterien bestimmt werden, um die Daten so genau wie möglich zu beschreiben (Jahnke/Groffmann/Kruppa 1996, 321). Zur Veranschaulichung wird oft ein dreidimensionaler Datenwürfel¹⁹ abgebildet, jedoch nur deshalb, weil in einer Grafik nicht mehr als drei Dimensionen darstellbar sind. Tatsächlich unterliegt ein Datenwürfel aus logischer Sicht keinerlei derartigen Beschränkungen (Holthuis 2000, 151). Physische Grenzen setzt lediglich das verwendete DW-System. So erlaubt das SAP BW die Verwendung von maximal 16 Dimensionen, von denen jedoch 3 fest vorgegeben sind und daher vom Benutzer nicht vergeben werden können. Im Vergleich mit anderen DW-Produkten klingen 13 Dimensionen recht spärlich, jedoch nur aufgrund der Unterschiedlichkeit in den verwendeten Begriffen. Im BW kann jede Dimension bis zu 248 Merkmale enthalten, wobei von vielen anderen Anbietern Merkmale als Dimensionen bezeichnet werden. Diese Merkmale wiederum können eine Vielzahl an Navigationsattributen enthalten. Somit stehen pro Datenwürfel mindestens $248 \cdot 13 = 3.224$ Merkmale zur Verfügung (SAP 2000c, 34).

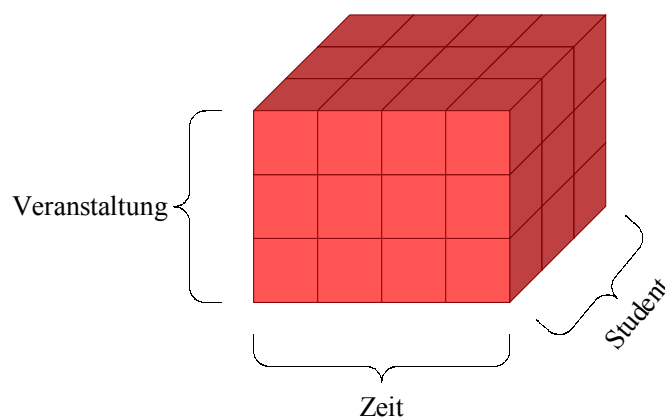


Abbildung 5: Dreidimensionaler InfoCube (Quelle: eigene Darstellung)

Die multidimensionale Modellierung weist eine Menge von Vorteilen auf, die auch letztendlich zu ihrer Entstehung geführt haben. So ist deren Struktur um einiges **leichter nachvollziehbar** und beim Auswerten **einfacher handhabbar** als die oft sehr komplexe OLTP-Struktur mit ihrer Vielzahl an Entitätstypen und Beziehungstypen. Somit ist eine intuitive Datenbearbeitung durch den Anwender möglich. Da die in einer multidimensionalen Matrix gespeicherten Daten einen hohen Grad an inhärenter Organisation aufweisen, wird zudem die

¹⁹ Oft begegnet man den Bezeichnungen Würfel, Datenwürfel, Cube, Hypercube, die alle dieselbe Bedeutung haben. Die SAP AG verwendet wiederum eine eigene Bezeichnung: InfoCube.

Performanz von Analysen deutlich gesteigert, was sich in deutlich kürzeren Antwortzeiten der Systeme bei Abfragen bemerkbar macht²⁰ (Holthuis 2000, 152).

3.5.2.3 InfoObjects und InfoCubes als Bausteine multidimensionaler Datenmodelle im BW

Das multidimensionale Datenmodell (MDM) unterscheidet grundsätzlich zwischen den Strukturkomponenten *Kennzahlen* und *Merkmale*²¹. Im BW werden diese beiden Auswertungsobjekte als **InfoObjects** bezeichnet. InfoObjects sind die Basis-Informationsträger des BW. Durch sie werden die Informationen in strukturierter Form abgebildet, die zum Aufbau von Datenzielen wie den **InfoCubes** benötigt werden (SAP 2000b). Jedes InfoObject ist einem InfoObject-Katalog und einer InfoArea zugeordnet, welche zur Gliederung aller Metaobjekte im Business Information Warehouse dienen.

Inhalt der Zellen der multidimensionalen Datenwürfel und damit zentrales Element jedes MDM sind die **Kennzahlen**²², die messbare Kenngrößen aus der Realität darstellen. Ihre Semantik wird jedoch erst durch die Art der verwendeten Dimensionen bestimmt. Viele Kennzahlen sind numerischen Charakters, wobei zwischen Fluss- und Bestandsgrößen unterschieden wird (Seemann/Schmalzridt/Lehmann 2001, 118). Theoretisch sind auch Textwerte als Kennzahlen möglich, dies wird jedoch vom SAP BW nicht unterstützt (SAP 2000a, 2). Häufig anzutreffende Beispiele für Kennzahlen aus dem betriebswirtschaftlichen Umfeld sind Umsatz, Kosten und Leistungen. Für ein universitäres Berichtswesen sind Kennzahlen wie Noten, Leistungspunkte, Anzahl Studenten und finanzielle Landesmittel denkbar. Die am einfachsten handhabbaren Kennzahlen sind numerisch und additiv (Kimball et al. 1998, 144).

Aus Sicht des Anwenders, der eine multidimensionale Datenbasis auswertet, spiegeln sich die direkten Komponenten dieser Abfrage in den **Dimensionen** wieder. Jede für eine bestimmte Kennzahl relevante Einflussgröße wird als Dimension dargestellt und stellt sozusagen das „Eintrittstor“ zur Auswertung der Kennzahlen dar. Unter einer Dimension versteht E.F. CODD die höchste Ebene eines Datenkonsolidierungspfades. Art und Anzahl der Dimensionen bestimmen die Komplexität der Datenstruktur, weshalb der Planung und Gestaltung von Dimensionen eine besondere Bedeutung zukommt. In technischer Hinsicht können Dimensionen als Index für den Zugriff auf die Werte (Kennzahlen) in einer Matrix gesehen werden (Holthuis 2000, 164). Im BW-Kontext jedoch sind die Dimensionen selbst „nur“ eine inhaltliche Zusammenfassung der sie konstituierenden **Merkmale**. Somit erlangen die Merkmale im BW-Bereich größtenteils die Bedeutung und Funktion der Dimensionen bei anderen DW-Produkten. Es muss beachtet werden, dass sich der Ansatz der SAP AG in diesem Punkt in den Begrifflichkeiten, nicht inhaltlich, von dem der meisten anderen DW-Anbieter unterscheidet. Während nicht hierarchische Dimensionen eine einfache interne Struktur ohne vertikale Beziehungen widerspiegeln, bestehen bei hierarchischen Dimensionen vertikale Beziehungen zwischen Dimensionspositionen, wodurch eine Hierarchie mit unterschiedlichen Verdichtungsstufen erkennbar wird (Holthuis 2000, 166f.). Im BW können Merkmalshier-

²⁰ Interessant ist in diesem Zusammenhang, dass die Entwicklung multidimensionaler Datenstrukturen nicht von einer Person planmäßig vorangetrieben wurde, sondern dass sie ein zwangsweises Resultat des Wunsches nach Verständlichkeit und hoher Performanz eines Datenmodells ist (Kimball et al. 1998, 143).

²¹ Man beachte die oft analoge Verwendung der Bezeichnungen *Dimension* und *Merkmal* im BW-Umfeld.

²² Kennzahlen werden oft als *Facts*, *Measured Facts* oder *Fakten* bezeichnet.

archien aus einem Quellsystem geladen werden²³ oder auch direkt im Business Information Warehouse angelegt werden, wobei auch mehrere Hierarchien pro Merkmal möglich sind (SAP 2000b). In nahezu allen betriebswirtschaftlichen Anwendungsbereichen anzutreffende Merkmale sind Zeit, Wertetyp (Istzahlen, Sollzahlen, Planzahlen) und Maßeinheiten (Währung, Stück). Ebenso häufig werden Merkmale wie Organisationseinheit, Region, Kunde, Artikel/Produkt benötigt (Holthuis 2000, 164ff.). Speziell im universitären Umfeld sind Veranstaltung, Lehrstuhl und Student sinnvolle Auswertungsmerkmale.

Die zentralen Datenspeicher für alle Auswertungen und Berichte sind die Datenwürfel, die im SAP-Jargon **InfoCubes** genannt werden und ebenso wie die InfoObjects in InfoAreas abgelegt werden. Sie sind multidimensional aus den Datentypen Kennzahlen und Merkmale aufgebaut. Wie bereits erwähnt, beträgt die maximale Anzahl an Merkmalen über 3.224, pro Datenwürfel können 233 Kennzahlen angelegt werden. Drei Dimensionen sind bereits durch *Zeit*, *Einheit* und *InfoPackage* vorgelegt. Die Zeitdimension ist wegen ihrer überragenden Bedeutung in einer Datenbank mit großen Mengen an historischen Daten unerlässlich, die Dimension *Einheit* kennzeichnet die im InfoCube gespeicherten Kennzahlen dahingehend, ob es sich um DM, € oder Stückzahlen handelt. In der Dimension *InfoPackage* wird beim ETL-Prozess eine eindeutige Identifikationsnummer für die geladenen Datenpakete hinterlegt. Auf der physischen Entwurfsebene betrachtet, sind im BW der InfoCube und die zugehörigen Dimensionstabellen relational über die Dimensionsschlüssel miteinander verbunden. Dimensionsschlüssel werden vom System je Merkmalskombination in der Dimensionstabelle bereitgestellt (SAP 2001d, 4-18).²⁴

Folgende Abbildung verdeutlicht den Zusammenhang der Strukturelemente und den Unterschied zwischen SAP Terminologie und klassischer Bezeichnungsweise.

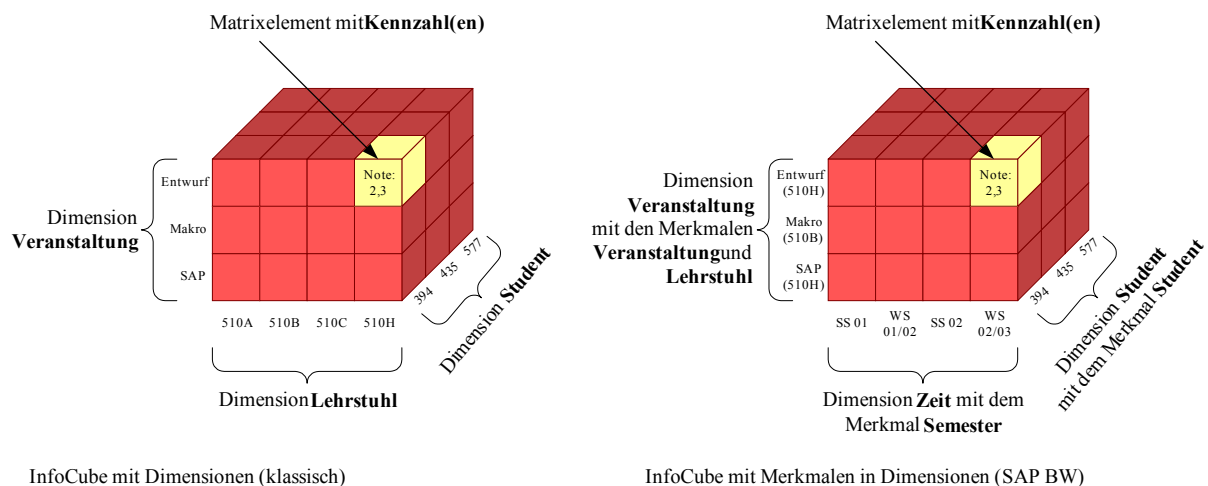


Abbildung 6: Klassische und SAP-spezifische Strukturelemente eines MDM (Quelle: Eigene Darstellung)

3.5.2.4 Möglichkeiten der Analyse multidimensionaler Daten

Für detaillierte Fragestellungen des Anwenders stehen im multidimensionalen Datenmodell verschiedenartige Operationen zur Manipulation des Datenwürfels zur Verfügung. Hierbei

²³ Ein typisches Beispiel für eine Merkmalshierarchie ist die Einordnung des Merkmals Kostenstelle in die Kostenstellenstandardhierarchie.

²⁴ Vgl. Kapitel 3.5.4.1

handelt es sich überwiegend um einen Wechsel von Dimensionen und Verdichtungsstufen, d.h. um eine Navigation im Datenraum (Holthuis 2000, 154). Diese Analysemöglichkeiten werden im BEx Analyzer z.B. über das Kontextmenü im Ergebnisbereich angeboten, an den OLAP-Prozessor weitergegeben und von diesem interpretiert und auf den Datenbestand angewendet. Die Technik des **Drill Down** erlaubt es, das Zustandekommen hochverdichteter Kennzahlen nachvollziehen zu können, indem sich der Anwender in einer Dimensionshierarchie zu Elementen mit niedrigerem Verdichtungsniveau bewegt. Beispielsweise kann der Anwender einen Notendurchschnitt aus einer Veranstaltung näher analysieren, indem er die Einzelnoten anfordert. **Roll Up** kennzeichnet die Bewegung in die entgegengesetzte Richtung, d.h. es erfolgt ein Wechsel zu einer höheren Verdichtungsstufe. **Slicing** beschreibt die Auswahl einer Scheibe aus dem Datenwürfel, z.B. Lehrstuhl 510H. Durch diese auch als Rotation bezeichnete Auswahl unterschiedlicher Sichten wird der Würfel aus einem anderen Blickwinkel betrachtet, ohne dass eine Neuordnung oder Sortierung der Daten erforderlich ist. **(Data) Dicing** schließlich beschreibt die Möglichkeit, entlang einer bestimmten Dimension eine gewünschte Position auszuwählen und dadurch die Daten des Würfels auf eine Teilmenge einzugrenzen (Holthuis 2000, 154ff.).

3.5.3 Semantische Datenmodellierung für Data Warehouses

3.5.3.1 Möglichkeiten der Notation semantischer multidimensionaler Datenmodelle

Ein semantisches Modell dient zur Begriffsklärung, zur Informationsbedarfsanalyse bei den Fachabteilungen, zur Dokumentation und zur Datendefinition. Daher sollte ein großes Augenmerk auf die Notation gelegt werden, mit deren Hilfe das Modell beschrieben wird. Eine grafische Notation für multidimensionale Informationssysteme muss in der Lage sein, dessen Basiskonstrukte (Bausteine) adäquat abzubilden (Totok 2000, 190). Während mit der ER-Modellierung eine allgemein anerkannte grafische Notationsform für operative, transaktionsorientierte Datenbanken gegeben ist, hat sich bislang für multidimensionale Systeme noch keine allgemein akzeptierte Abbildungstechnik etablieren können (Gabriel/ Gluchowski 1998, 502). Somit besteht hier noch **Forschungsbedarf**. Lediglich auf der logischen und physischen Ebene steht eine Vielzahl an Modellierungsansätzen zur Verfügung (Böhnlein/ Ulbrich-vom Ende o.J., 1).

Eine relativ einfache Möglichkeit der grafischen Notation multidimensionaler Datenmodelle stellt das **multidimensionale Entity-Relationship-Model** (ME/R Model, MERM) dar²⁵. Die ERM-Notation bedient sich hauptsächlich der Elemente *Entitätstyp* als ein „Etwas“ aus der realen Welt, *Beziehungstyp* als Art der Beziehung zwischen Entitäten und *Attribut* als Eigenschaften der Entitäten. Diese ERM-Notation wird im MERM um drei neue Elemente ergänzt: Faktenrelation, Dimensionsebene²⁶ und hierarchische Beziehung. Bei Faktenrelation und hierarchischer Beziehung handelt es sich im Prinzip um spezialisierte Beziehungstypen, denn im Rahmen eines ERM kann ein InfoCube (Faktenrelation) als Beziehungstyp verstanden werden, der die Relationsmenge unterschiedlicher Dimensionen repräsentiert (Gabriel/

²⁵ Das MERM ist eine Modellierungsnotation, die von der Forschungsgruppe Wissensbasen des bayrischen Forschungszentrums für wissensbasierte Systeme (Forwiss) im Rahmen eines Projekts namens *System 42* entwickelt wurde (Totok 2000, 191). Näheres unter <http://www.forwiss.tu-muenchen.de/~system42>.

²⁶ Die Bezeichnungen Dimensionsebene und –feld werden oft uneinheitlich verwendet.

Gluchowski 1998, 497f.). Dimensionsebenen sind eine besondere Ausprägung von Entitätstypen. Für die MERM-Notation wurde das Prinzip der Minimalität angewendet, d.h. Notationselemente werden so sparsam wie möglich benutzt. Daher gibt es auch keine Elemente für bestimmte Dimensionstypen oder -elemente. Ebenso werden normale und hierarchische Beziehungen nicht durch das traditionelle Rautensymbol des ERM dargestellt, sondern werden vom Verbindungselement impliziert (Totok 2000, 192).

Somit stehen folgende Konstruktionselemente zur Verfügung:

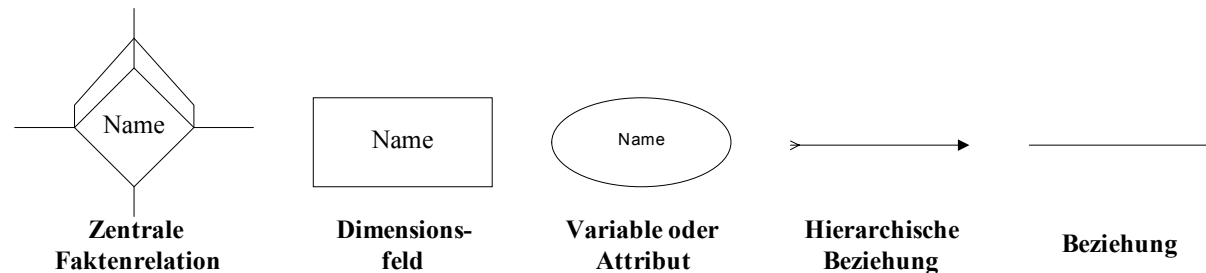


Abbildung 7: Notationselemente des MERM (Quelle: in Anlehnung an Totok (2000), 192)

Folgende Abbildung zeigt ein einfaches Beispiel der grafischen Notation eines multidimensionalen Datenmodells zur Klausurverwaltung:

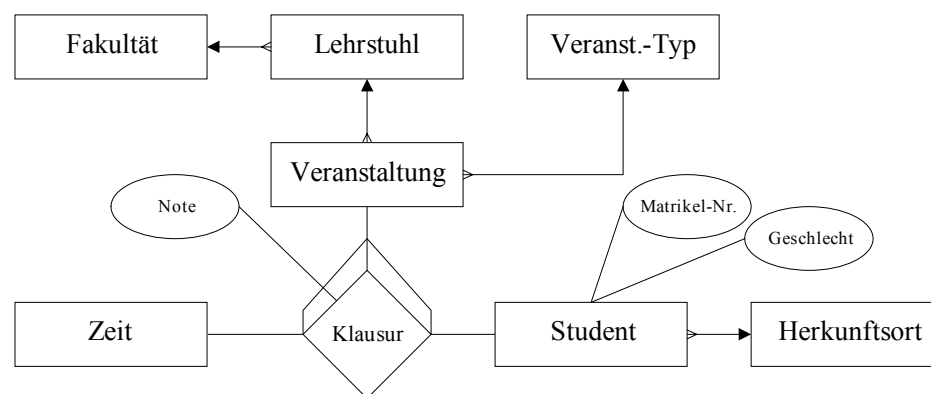


Abbildung 8: Beispiel eines MERM (Quelle: Eigene Darstellung)

Das Prinzip der Minimalität hat sowohl Vor- als auch Nachteile. Auf der einen Seite bleiben die Grafiken sehr übersichtlich und die Modellkonstrukte sind, v.a. im Lehbetrieb, leicht vermittelbar. Auf der anderen Seite ist die Semantik von bestimmten Notationselementen nicht immer eindeutig und wird nur im konkreten Zusammenhang klar. Nicht unterschieden wird z.B. zwischen Attributen der Faktentabelle, hier Note, und Attributen von Dimensionselementen (Totok 2000, 193).

Als weitere Notationsvariante multidimensionaler Informationssysteme lässt sich das **Dimensional Fact Modeling** nennen, das den grafischen Modellen im Prinzip die gleiche Aussagekraft verleiht wie das MERM, jedoch erweitert um die explizite Betrachtung der Additivität von Kennzahlen (Gabriel/Gluchowski 1998, 498f.). Da auf diese Problematik in der hier vorgestellten Lehrveranstaltung jedoch nicht auf Modellierungsebene eingegangen wird, unterbleibt eine weitergehende Betrachtung dieses Notationsansatzes.

Die von BULOS vorgeschlagene Methode namens **ADAPT** (Application Design for Analytical Processing Technologies) umfasst ein speziell auf die Belange analytischer

Anwendungen ausgerichtetes Modellierungsinstrumentarium. Dabei bietet ADAPT eine breit gefächerte Palette unterschiedlicher Beschreibungselemente²⁷, mit denen sich die einzelnen Bestandteile multidimensionaler Datenmodelle darstellen lassen (Gabriel/Gluchowski 1998, 499). Jedoch gibt die große Menge der zur Verfügung stehenden Abbildungsobjekte Anlass zur Kritik, denn ein intuitiver Zugang für den Anwender bleibt verborgen und der Aufwand, sich in die Spezifikation einzuarbeiten, ist enorm.

M. BÖHNLEIN und A. ULBRICH-VOM ENDE schlagen mit dem **Semantischen Data Warehouse Modell** (SDWM) eine weitere Möglichkeit zur konzeptuellen Modellierung multidimensionaler Datenstrukturen vor. Ihr Hauptaugenmerk liegt auf der sauberen Unterscheidung der verschiedenen Entwurfsebenen und dem Einsatz von Modellierungssichten zur Komplexitätsbewältigung (Böhnlein/Ulbrich-vom Ende o.J., 1ff.).

3.5.3.2 Vom ERM zum MDM

Obwohl sich die transaktionsorientierte Modellierung mit Hilfe eines ERM äußerlich sehr stark von der analyseorientierten Modellierung eines MDM unterscheidet, stellt die Transformation eines bestehenden Entity Relationship Modells eine elegante Möglichkeit dar, ein multidimensionales Datenmodell für die spätere Umsetzung im BW zu erstellen. Der Grund dafür liegt in der engen Beziehung zwischen ERM und MDM: **Ein einziges ERM lässt sich in ein oder mehrere MDMs abbilden** (Kimball et al. 1998, 146). Der Grund für die oft überwältigende Komplexität der ERMs ist darin zu sehen, dass sie in einer einzigen Darstellung oft mehrere Geschäftsprozesse gleichzeitig abbilden. So werden nicht selten Auftrags- eingang, Warenausgang, Rechnungsstellung und Zahlungseingang in einem ERM dargestellt, obwohl diese Prozesse nie zur gleichen Zeit ablaufen können. Genau hier setzt die Transformation in ein MDM an. Der erste Schritt dieser Transformation besteht darin, das ERM in seine einzelnen **Geschäftsprozesse aufzuspalten**²⁸ und jeden in einem separaten MDM abzubilden (Kimball et al. 1998, 146). Folgende Abbildung zeigt ein ERM aus einem Universitätsverwaltungssystem mit den zwei identifizierten Geschäftsprozessen *Klausur- geschehen* und *Lehrstuhlbetrieb*.

²⁷ Unter <http://www.symcorp.com> findet man die Website der Firma *Symetry Corporation*, die den kostenlosen Download einer Visio-Schablone für die Datenmodellierung mit ADAPT anbietet.

²⁸ Die Prozesse müssen nicht unbedingt überschneidungsfrei sein.

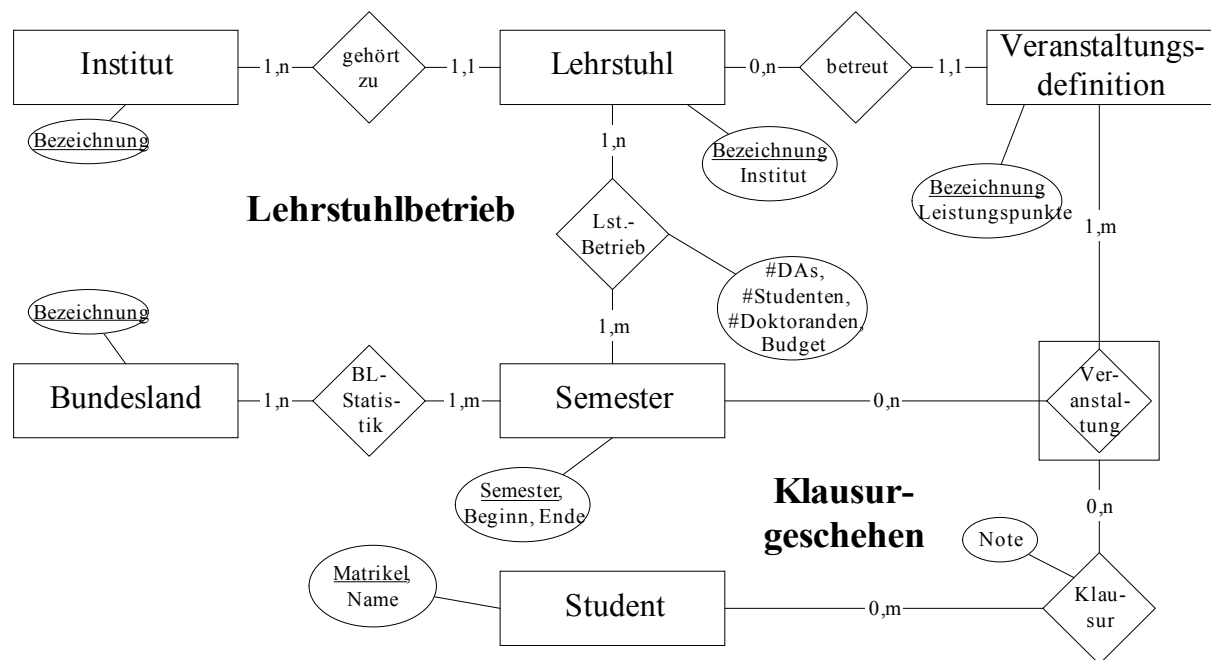


Abbildung 9: Beispiel-ERM mit zwei Geschäftsprozessen (Quelle: Eigene Darstellung)

Nach der Identifizierung der Hauptgeschäftsprozesse besteht der zweite Schritt darin, diejenigen **n-m-Beziehungen**, die numerische und, wenn möglich, additive Kennzahlen als Attribute enthalten, als **Faktenrelation** zu kennzeichnen. Von besonderem Interesse sind hier die n-m-Beziehungen zwischen starken Entitäten, da diese Entitäten das gesamte Modell umspannen, alle anderen Entitäten von ihnen abhängig sind und die Beziehungen oft durch Kennzahlen beschrieben sind, welche dann sehr einfach in die Faktentabellen integriert werden können. Man spricht auch von Überschneidungsentitäten, denn die meisten Geschäftsprozesse resultieren aus der Zusammenführung der starken Entitäten (SAP 2001d, 3-11). Überschneidungsentitäten stellen häufig Transaktionen, z.B. geschriebene Klausuren oder Prüfungsanmeldungen, oder (Kontroll-)Dokumente wie Leistungsscheine dar.

Im dritten Schritt werden die **Dimensionen** betrachtet: Alle verbleibenden Entitäten werden zunächst zu Gruppen zusammengefasst, die inhaltlich zusammengehören und jeweils mindestens eine starke Entität enthalten. Die so entstehenden Gruppen werden als Dimensionen bezeichnet. Im Beispiel aus Abbildung 9 würde die Entität *Student* eine eigene Dimension *Student* bilden, ebenso würde die Entität *Semester* z.B. zu einer *Zeitdimension*. Aus den beiden Entitäten *Lehrstuhl* und *Institut* könnte die Dimension *Lehrstuhl* generiert werden. Es ist nahe liegend, die Bezeichnung der starken Entität als Name für die komplette Dimension zu übernehmen, wie aus der im eben angeführten Beispiel verwendeten Nomenklatur ersichtlich wird. Der Grund hierfür ist darin zu sehen, dass die starke Entität der gesamten Dimension eine ganz bestimmte Semantik verleiht, ihr sozusagen ihren „Stempel aufdrückt“. So ist eine Dimension *Kunde*, die sich aus den Entitäten *Kunde* (starke Entität), *Stadt*, *Region* und *Land* zusammensetzt, in ihrem Wesen eben eine Kundendimension, jedoch mit zusätzlichen Informationen zur geographischen Herkunft.

Die Anzahl der aus den eben dargestellten Schritten resultierenden multidimensionalen Datenmodelle liegt bei großen Data Warehouses oft zwischen 10 und 25, wobei jedes Datenmodell wiederum 5 bis 15 Dimensionen enthält. Wenn die Modelle sauber erstellt wurden,

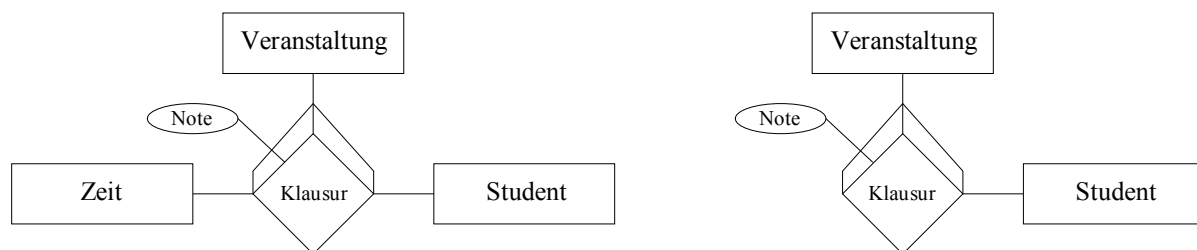
teilen sie sich möglichst viele der Dimensionen²⁹ (Kimball et al. 1998, 147). Tabelle 6 fasst die notwendigen Schritte bei der Transformation eines ERM in ein MDM zusammen.

Schritt	Bezeichnung	Beschreibung
1	Geschäftsprozesse identifizieren	Aufspaltung eines ERM in einen oder mehrere Geschäftsprozesse
2	Faktenrelation erzeugen	n-m-Beziehungen zwischen starken Entitäten ergeben die Faktenrelation, die numerischen Attribute sind Kandidaten für Fakten
3	Dimensionen bilden	Inhaltliche Zusammenfassung der verbleibenden Entitäten zu Gruppen, die von starken Entitäten dominiert werden

Tabelle 6: Schritt für Schritt vom ERM zum MDM (Quelle: Eigene Darstellung)

3.5.3.3 Bestimmung der Datengranularität als Modellierungsaufgabe

Mit dem Begriff **Granularität** wird das Detail einer Datenbank im Data Warehousing beschrieben. Daten mit hoher Granularität sind sehr detaillierte Daten, es sind zahlreiche Merkmale vorhanden, mit denen die Daten beschrieben werden. So ist die Granularitätsstufe „nach Veranstaltung“ weniger detailliert und damit weniger granular als „nach Veranstaltung und Zeit“. Die Granularität ist das fundamentale Kriterium dafür, wie weit ein Drill Down in den Daten möglich ist (SAP 2001d, 3-24). Gleichzeitig wirkt sich die Granularität jedoch auf die Größe der Datenbank aus, denn im Fall einer Modellierung von Kennzahlen auf Veranstaltungs- und Zeitebene ist die Datenmenge größer als auf Veranstaltungsebene. Entsprechend werden auch Abfrage-Performance³⁰ und notwendige Ladezeit der Daten von dieser Modellierungsfrage beeinflusst (SAP 2000c, 41). Folgendes Beispiel veranschaulicht die Problematik.



Fall 1: hohe Granularität

Fall 2: niedrige Granularität

Abbildung 10: Bestimmung der Granularität (Quelle: Eigene Darstellung)

Im ersten Fall werden die Klausurnoten pro Student, Veranstaltung und Semester (Zeit) gespeichert, d.h. jede Einzelnote wird in der Faktentabelle abgelegt und steht für Auswertungen zur Verfügung. Im zweiten Fall dagegen werden alle Noten eines Studenten, die er in einer bestimmten Veranstaltung erzielt hat, aggregiert: Sobald neue Klausurergebnisse in die Faktentabelle geschrieben werden, die kein Zeitmerkmal enthält, wird die Semesterangabe der neuen Klausuren sozusagen ignoriert und mit den schon vorhandenen Klausurergebnissen

²⁹ Vgl. Kapitel 3.5.4.4

³⁰ Die Festlegung der Granularität hat die größte Auswirkung auf Plattenplatz und Performance (SAP 2001d, 3-25).

des betreffenden Studenten in der jeweiligen Veranstaltung aggregiert, d.h. die Reduzierung der Granularität bedeutet Informationsverlust zugunsten von Speicherplatzgewinn. Die Aggregation würde hier am sinnvollsten in einer Mittelwertbildung bestehen. Es können keine Auswertungen angefertigt werden, die die Einzelnoten enthalten, da diese aufgrund des Modellierungsansatzes überhaupt nicht in der Faktentabelle untergebracht werden könnten. Die Vor- und Nachteile hoher und niedriger Granularität müssen gegenübergestellt und daraufhin eine Entscheidung getroffen werden. Auf die Frage der „richtigen“ Granularität gibt es keine allgemeingültige Antwort. Letztendlich entscheiden der Informationsbedarf des Anwenders und die Möglichkeit, an detaillierte Daten zu gelangen, darüber, wie detailliert die Daten modelliert werden.

Als Alternative zur Modellierung eines relativ detaillierten InfoCube bietet das BW die Verwendung des **Operational Data Store (ODS)** an. Ein ODS-Objekt dient der Ablage von konsolidierten und bereinigten Bewegungsdaten auf Belegebene. Im Gegensatz zur multidimensionalen Datenablage bei InfoCubes werden die Daten in ODS-Objekten in transparenten, flachen Datenbanktabellen³¹ abgelegt. Fakten- und Dimensionstabellen werden nicht angelegt (SAP 2000b). In dieser Lehrveranstaltung ist jedoch gerade die multidimensionale Datenmodellierung im DW-Bereich von Interesse, daher wird auf das ODS-Objekt nicht weiter eingegangen. Eine weitere Variante besteht in der Verknüpfung einer größeren mit einer detaillierten Abfrage durch Verwendung der Bericht-Berichts-Schnittstelle innerhalb des Analysetools BEx Analyzer.

3.5.4 Logische Datenmodellierung zur Erstellung von Datenbankschemata

Während das im vorigen Kapitel beschriebene konzeptuelle Datenbankmodell (MDM) mit der Zielsetzung entwickelt wurde, Verständlichkeit für den Anwender zu erreichen, ist es Aufgabe des nun beschriebenen Modellierungsschrittes, „Verständlichkeit“ für das zugrunde liegende Data Warehouse System zu erzielen. Hierfür wird ein logisches Datenbankschema entwickelt.

3.5.4.1 Das klassische Starschema als einfaches und effizientes Datenbankschema

Das Starschema ist die wohl bekannteste Variante, ein logisches Datenmodell für relationale DW-Systeme zu erstellen. Da die InfoCubes im BW eine auf dem klassischen Starschema basierende Variante des Starschemas realisieren (SAP 2001d, 3-21), wird es nachfolgend kurz vorgestellt.

Im Starschema wird jede Dimension durch eine Tabelle repräsentiert, deren Spalten aus den Primärschlüsseln und weiteren Attributen der zugrunde liegenden Entitäten gebildet werden. In diesen **Dimensionstabellen** werden sozusagen alle Felder „untergebracht“, die inhaltlich etwas mit der Dimension zu tun haben. So kann eine Dimension *Lehrstuhl* beispielsweise die Felder Lehrstuhl, Fachgebiet, Institut, Fakultät usw. enthalten. Letztendlich handelt es sich

³¹ Bei ODS-Objekten ist es möglich, Datenfelder zu überschreiben. Dies ist insbesondere im Kontext von belegnahen Strukturen wichtig, denn werden Belege im Quellsystem geändert, umfassen diese Änderungen nicht nur numerische Felder wie Auftragsmenge, sondern auch nichtnumerische Felder wie Warenempfänger, Status und Lieferdatum. Damit diese Änderungen auch im BW in den ODS-Objekten abgebildet werden können, müssen die entsprechenden Felder in den ODS-Objekten auch überschrieben werden und auf den aktuellen Wert gesetzt werden (SAP 2000b).

hierbei um eine Denormalisierung von Teilen des Datenmodells. Die so entstehenden Dimensionen werden mit einem (künstlichen) Primärschlüssel, der Dimensions-ID³², versehen und mit dessen Hilfe an die Faktentabellen gebunden. Die Dimensionstabellen umgeben somit die im Zentrum stehende **Faktentabelle**, die die Kennzahlen enthält. Verknüpfungen bestehen nur mit der Faktentabelle, die Dimensionstabellen sind untereinander nicht verknüpft. Hieraus entsteht eine sternförmige Anordnung der Tabellen mit der Faktentabelle als Zentrum des Sterns und den Dimensionstabellen als Endpunkte der Zacken, woraus sich die Bezeichnung Starschema ableitet. Die Faktentabellen sollten möglichst schmal gestaltet werden, d.h. wenige Spalten enthalten, umfassen dabei aber nicht selten mehr als 10 Millionen Datensätze. Die mit mehreren Spalten eher breit ausgelegten Dimensionstabellen sind zwar von der Anzahl der enthaltenen Datensätze her vergleichsweise kleiner als die Faktentabellen, im Falle einer Artikeldimension eines Warenhauses sind aber auch hier bis zu 100.000 Datensätze denkbar (Behme/Holthuis/Mucksch 2000, 226).

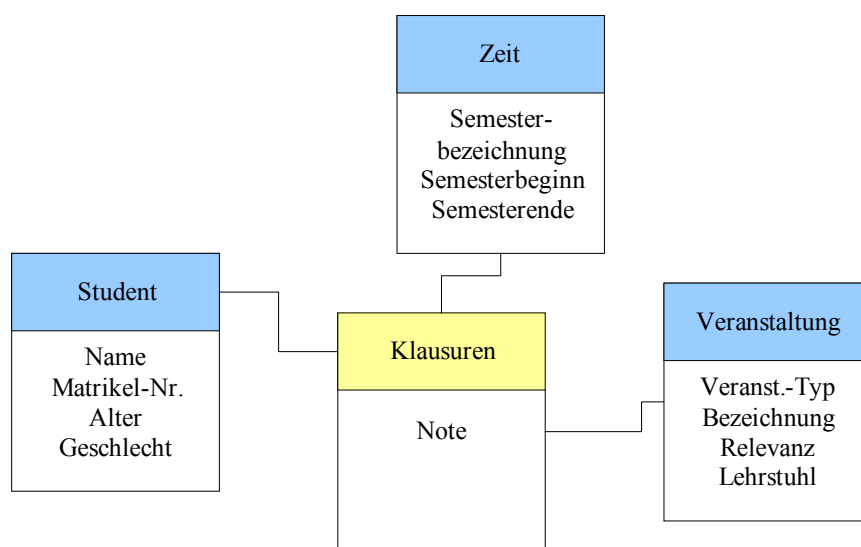


Abbildung 11: Starschema (Quelle: Eigene Darstellung)

Die Strukturierung von Data Warehouse Informationen nach dem Starschema soll zu einer hohen Abfrageleistung führen und die Möglichkeit schaffen, das Modell einfach an veränderte Unternehmensanforderungen anzupassen (SAP 2001d, 3-21). Ein hoch effektives Verhalten bei Abfragen wird dem Starmodell auch von BILL INMON zugesprochen, er sieht jedoch ein Trade-Off zwischen Abfrageleistung und Flexibilität. Seiner Meinung nach ist das Starmodell sehr unflexibel hinsichtlich veränderter Benutzerwünsche. Er empfiehlt es daher für das Design kleinerer Data Marts, nicht jedoch bei der Planung kompletter Unternehmens-DWs (Inmon 1999).

3.5.4.2 Das erweiterte Starschema der SAP AG

Mit der Entwicklung des *erweiterten Starschemas* will die SAP AG den vielfältigen Problemen begegnen, die die Verwendung des klassischen Starmodells mit sich bringt. So wird die heutzutage immer wichtiger werdende **Mehrsprachigkeit** für Attribute nicht unterstützt.

³² Es handelt sich hierbei um einen ganzzahligen numerischen Wert, der beim Einfügen eines neuen Dimensions-Datensatzes automatisch hochgezählt wird.

Die **alphanumerischen Fremdschlüssel** erschweren den Datenzugriff auf die Kennzahlen. Zudem wird **Zeitabhängigkeit** von Stammdaten nicht explizit unterstützt: Ändern sich Attribute im Laufe der Zeit, besteht keine Möglichkeit, sowohl die alten als auch die neuen Werte für das Attribut zu pflegen. Außerdem muss jedes Starschema die Daten duplizieren, die für alle möglichen Berichte der Benutzer benötigt werden. Mehrfach verwendete Stammdaten wie z.B. die Kundendimension werden auch mehrfach angelegt. Schließlich müssen die **Hierarchiebeziehungen** der Daten als Attribute einer Dimensionstabelle modelliert werden³³. Daneben sind einige Hierarchietypen nicht möglich. Somit können oft keine strukturierten Drill Downs durchgeführt werden (SAP 2001d, 4-25).

Das erweiterte Starschema baut auf dem klassischen Starschema auf. So gleichen die Faktentabellen in ihrem Aufbau im Prinzip denen des klassischen Starschemas. Die Dimensionstabellen und ihre Bausteine hingegen werden genauer analysiert: Die in den Dimensionen enthaltenen Merkmale bestehen neben dem eigentlichen Datenfeld aus den folgenden drei Segmenten (SAP 2001d, 4-26):

- **Texte:** Textliche Beschreibungen (Kurz-, Mittel- und Langtext) eines Merkmals können in den Texttabellen zeit- und sprachabhängig definiert werden.
- **Hierarchien:** Merkmalen können beliebig viele externe Hierarchien für den strukturierten Datenzugriff mittels Drill Down zugewiesen werden.
- **Stammdaten:** Die von einem Merkmal abhängigen Eigenschaftsfelder (**Attribute**) können in einer separaten Stammdatentabelle abgelegt werden. Die Attribute selbst stellen auch InfoObjects dar. Attribute müssen nicht unbedingt Merkmals-InfoObjects sein, es ist ohne weiteres möglich, eine Kennzahl als Attribut eines Merkmals zu deklarieren. Dies ist z.B. dann sinnvoll, wenn zu einem Produkt-Merkmal der Preis des Produkts gespeichert werden soll. Geklammerte Attribute stellen einen Spezialfall von Attributen dar: In manchen Fällen ist ein Merkmal durch das Datenelement selbst noch nicht eindeutig gekennzeichnet, daher ist ein zweites Attribut notwendig, um einen eindeutigen (mehrteiligen) Primärschlüssel zu generieren. Eine Kostenstelle beispielsweise ist nur im Zusammenhang mit dem Kostenrechnungskreis eindeutig adressiert. Daher wird dem Merkmal *Kostenstelle* ein geklammertes Attribut *Kostenrechnungskreis* beigelegt.

Attribute, Texte und Hierarchien sind nicht obligatorisch und können bei der Definition des betreffenden Merkmals aktiviert oder deaktiviert werden. Entscheidend für das erweiterte Starschema ist, dass diese drei Datensegmente getrennt von den InfoCubes angelegt werden können. Eine Faktentabelle bildet zusammen mit den sie umgebenen Dimensionstabellen den **lösungsabhängigen** Bereich, da hier die Kennzahlen und Merkmalskombinationen (als Dim-ID) abgelegt sind, die sich auf einen ganz bestimmten Informationsbereich beziehen. Der **lösungsunabhängige** Bereich wird durch die Merkmale samt ihren Stammdatentabellen gebildet, da diese unabhängig von einer Faktentabelle für das gesamte Data Warehouse zur Verfügung stehen. Die strikte Trennung von lösungsabhängigem und -unabhängigem Bereich macht sich auch in den unterschiedlichen Definitionsbereichen innerhalb des Modellierungsbereiches in der AWB bemerkbar.

³³ Hierarchien werden in relationalen Datenbanksystemen (DBMS) i.a. bisher nicht explizit unterstützt. 1:n-Beziehungen können über Primär-/Fremdschlüssel-Konstruktionen abgebildet werden. (Q97, S. 1)

Bei der Benennung der Segmente geht die SAP AG nicht unbedingt konsistent vor: „Im BW können die Merkmale Stammdaten tragen. Bei Stammdaten handelt es sich um Attribute, Texte oder Hierarchien.“ (SAP 2000b). Oft werden jedoch schon Attribute alleine als Stammdaten bezeichnet. Mit folgender Definition ist man auf der sicheren Seite: *Merkmale können Stammdaten enthalten, die Attribute, Texte und Hierarchien sein können.*

Die technische Verknüpfung sowohl zwischen InfoCube und Merkmal als auch zwischen Merkmal und zugehörigen Attributs-, Text- und Hierarchietabellen wird durch **Surrogat-IDs**³⁴ (SID) hergestellt. Jeder Ausprägung eines Merkmals wird eine eindeutige 4-Byte-Ganzzahl zugewiesen, welche fortan die Rolle des (künstlichen) Primärschlüssels³⁵ übernimmt. Der natürliche Primärschlüssel, das Datenelement selbst, wird um eine wesentlich performantere SID ergänzt. Diese SID wird zum einen als Fremdschlüssel in den Dimensionstabellen eingetragen und stellt somit die Verbindung zwischen Dimensionstabelle und Merkmalstabelle her. Zum anderen wird auch die Beziehung zwischen einem Merkmal und seinen Attributs-, Text- und Hierarchietabellen mittels SID realisiert. SID-Tabellen sind sozusagen Zeigertabellen, die den lösungsabhängigen InfoCube-Bereich vom lösungsunabhängigen Stammdatenbereich des erweiterten Starschemas abtrennen (SAP 2000c, 20).

Die SID-Technik löst zudem noch ein Problem der Zusammenarbeit zwischen OLTP- und OLAP-Systemen: Wird für einen eindeutigen Schlüssel eines Attributes der Primärschlüssel der entsprechenden Tabelle aus dem OLTP-Bereich benutzt, spricht man von einem Produktivschlüssel. Dieses Vorgehen scheitert jedoch bereits daran, dass die Daten eines Merkmals oft aus mehreren operationalen Systemen mit unterschiedlichen Primärschlüsseln geladen werden. Dieser und weitere Gründe sprechen dafür, anstelle des übernommenen Produktivschlüssels eine künstlich generierte SID zu verwenden. Künstliche Schlüssel lassen sich auf minimalen Speicherplatzbedarf hin optimieren, sie können durch einen allgemeinen Aufbau im gesamten Warehouse eingesetzt werden und sie sind grundsätzlich eindeutig. Ein operationales System hingegen kann einen Primärschlüssel immer wieder neu vergeben.³⁶

Folgende Grafik zeigt zur Veranschaulichung dieses komplexen Phänomens ein stark vereinfachtes BW-Starschema mit SID-Feldern ohne Berücksichtigung zeit- oder sprachabhängiger Stammdaten.

³⁴ Surrogat: Ersatz, Ersatzmittel, Behelf (Langenscheidt 1995).

³⁵ Da die SID eine automatisch inkrementierte Ganzzahl ist, handelt es sich um einen nicht sprechenden Schlüssel. Es lässt sich lediglich eine Rangfolge erkennen, wann ungefähr der betreffende Datensatz angelegt wurde.

³⁶ Wird z.B. ein Student mit der Matrikelnummer 273511 exmatrikuliert und die entsprechenden Daten werden aus der Personalverwaltung gelöscht, könnte dem nächsten neu immatrikulierten Student eventuell diese Matrikelnummer erneut zugewiesen werden. Ist in der Personalverwaltung die Matrikelnummer der Primärschlüssel, kann dieser im Data Warehouse nicht verwendet werden, sondern muss dort durch eine Studenten-SID ersetzt werden (Inmon 2002).

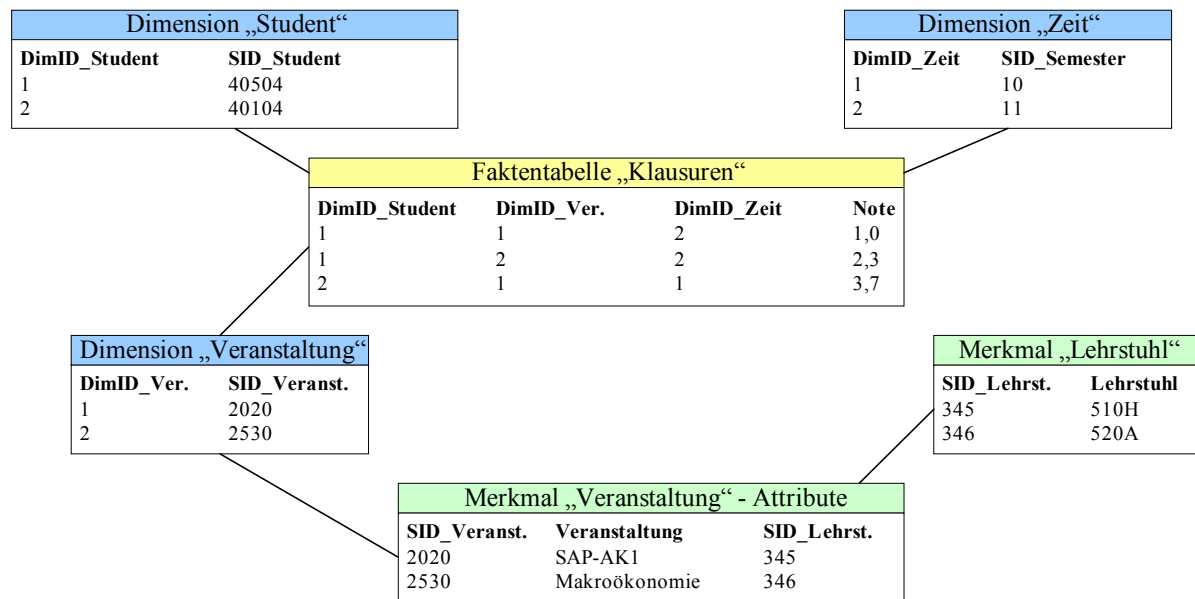


Abbildung 12: Verwendung von Surrogat-IDs im erweiterten Starschema (Quelle: Eigene Darstellung)

Zur übersichtlichen Darstellung der durch die SID-Technik und die Struktur des erweiterten Starschemas im BW-System erzeugten Tabellen eignet sich die Transaktion **Listschema**. Sie liefert eine geschachtelte Auflistung aller InfoCube-Tabellen und unterstützenden Tabellen, wobei durch die integrierte Data Browser Funktion der Tabelleninhalt direkt betrachtet werden kann (SAP 2001d, 4-37).

3.5.4.3 Weitere Modellierungsschemata für multidimensionale Strukturen

Wenn mehrere Fakten durch genau dieselben Dimensionen beschrieben werden können, reicht für die Modellierung ein einziges Starschema aus. Die Geschäftssituation ist in der Regel jedoch komplexer, da sehr viele Fakten mit sehr unterschiedlichen Dimensionen existieren. Daher werden alle Fakten, die die gleiche Dimensionierung haben, in einer Faktentabelle zusammengefasst und Fakten mit unterschiedlichen Dimensionen in getrennten Faktentabellen platziert. Das so entstehende Schema nennt man Multi-Faktentabellen-Schema oder **Galaxie** (Behme/Holthuis/Mucksch, 226f.). In Fällen, in denen eine Dimension mit mehr als einer Faktentabelle verknüpft ist, wird diese Dimension in allen MDM-Schemata aufgeführt, jedoch physikalisch nur einmal erzeugt. Man spricht dann von *einheitlichen Dimensionen*³⁷. Im folgenden Beispiel stellen die beiden Faktentabellen *Klausuren* und *Lehrstuhlaktivitäten* die Zentren der Galaxie dar, *Zeit* ist eine einheitliche Dimension, die von beiden Faktentabellen benutzt wird.

³⁷ Oft wird auch die Bezeichnung *shared dimension* oder *conformed dimension* verwendet.

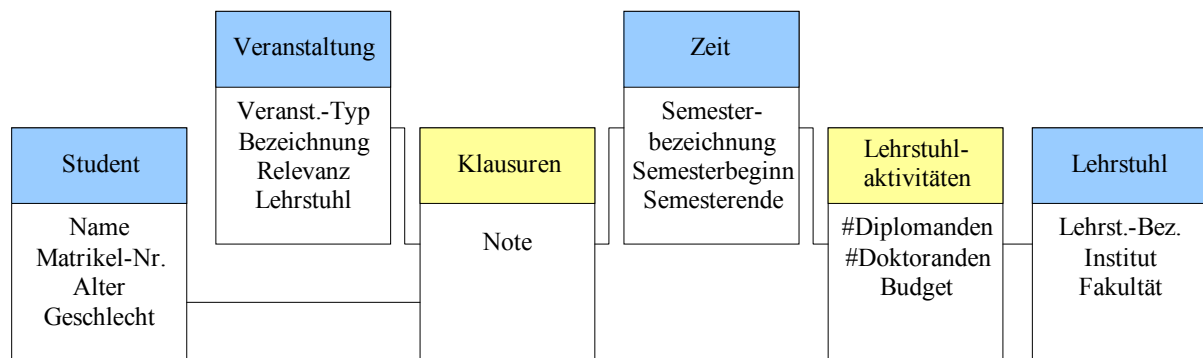


Abbildung 13: Galaxy Schema (Quelle: Eigene Darstellung)

Die Abbildung eines Galaxy Schemas im BW ist problemlos realisierbar, indem ein einmal definiertes Merkmal bei der Definition beliebig vieler InfoCubes mit aufgenommen wird.

Im Starschema, dem erweiterten Starschema der SAP AG und dem Galaxy Schema sind die Dimensionstabellen denormalisiert³⁸. Um die großen Datenbestände in den Dimensionstabellen zu verringern, kann man durch Normalisierung sehr großer Dimensionstabellen ein Starschema in ein **Snowflakeschema** überführen. Der Name leitet sich aus der zusätzlichen strukturellen Komplexität ab (Behme/Holthuis/Mucksch, 2027), wie in Abbildung 14 deutlich wird. Wenn dies auch wie ein neues Konzept klingt, handelt es sich in Wirklichkeit um eine einfache Variation des klassischen Starschemas (Poe/Reeves 2000, 147). Der Vorteil dieser Struktur liegt in kürzeren Zugriffszeiten und gewissen Speicherplatzeinsparungen. Jedoch steht diesen Vorteilen der Nachteil der höheren Komplexität gegenüber, die es v.a. den Endanwendern erschwert, durch die Snowflake-Struktur zu navigieren (Behme/Holthuis/Mucksch, 2029). MARKUS LUSTI ist sogar der Meinung, dass gerade die Normalisierung einer Datenbank die Abfrageeffizienz senkt (Lusti 2002, 189).

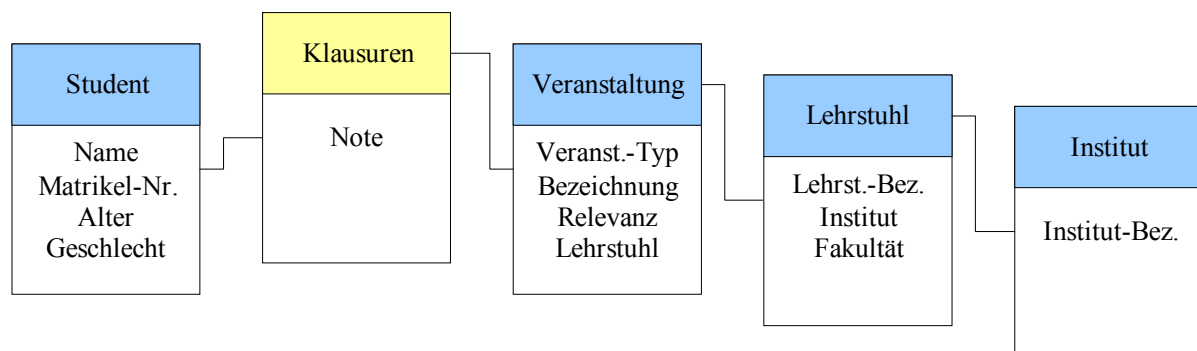


Abbildung 14: Snowflakeschema (Quelle: Eigene Darstellung)

Auch ein Snowflakeschema kann im BW umgesetzt werden: Beim Anlegen eines Merkmals können diesem diverse Attribute zugewiesen werden³⁹. Diese Attribute können inhaltlich als Fremdschlüssel interpretiert werden, welche die Verbindung zu anderen Merkmalstabellen herstellen. So kann z.B. dem Merkmal *Lehrstuhl* das Attribut *Institut* zugewiesen werden, welches als eigenständiges Merkmal existiert und evtl. noch Stammdatentabellen besitzt. Da

³⁸ Im Sinne der 3. Normalform

³⁹ Um einen gewissen Grad an Normalisierung zu erreichen, sollten die Attribute voll funktional vom Datenelement abhängig sein.

ein Merkmalsattribut selbst ein Merkmal darstellt, ist diese Schlüsselbeziehung unmittelbar einzusehen: Attribute sind schon existierende InfoObjects, die dem neuen Merkmal logisch zugeordnet sind (SAP 2000b). Durch die Möglichkeit, Merkmale als Navigationsmerkmale zu kennzeichnen, können selbst solche Merkmale, die sich in der normalisierten Snowflake-Struktur weiter entfernt vom eigentlichen InfoCube befinden, dem Endanwender für die Navigation im Datenbestand zugänglich gemacht werden, ohne dass dieser mit der komplexen Merkmalsstruktur direkt konfrontiert wird.

Zusammenfassend lässt sich sagen, dass das SAP BW nicht nur die Möglichkeit bietet, das klassische und das erweiterte Starschema umzusetzen, sondern auch Elemente für die Implementierung von Galaxy- und Snowflakeschemata bereitstellt.

Datenbankschema	Umsetzung im BW
Klassisches Starschema	<ul style="list-style-type: none"> ▪ Faktentabellen und Dimensionstabellen
Erweitertes Starschema	<ul style="list-style-type: none"> ▪ Entkopplung der Stammdaten von den Faktentabellen ermöglicht Attributstabellen, Mehrsprachigkeit, Zeitabhängigkeit und die Anbindung von Hierarchien. ▪ SID-Technik
Galaxyschema	<ul style="list-style-type: none"> ▪ Beliebig viele Faktentabellen ▪ Einheitliche Merkmale können von beliebig vielen Faktentabellen genutzt werden.
Snowflakeschema	<ul style="list-style-type: none"> ▪ Merkmalsattribute verweisen auf andere Merkmale.

Tabelle 7: Umsetzung logischer Datenbankschemata im SAP BW (Quelle: Eigene Darstellung)

3.5.4.4 Unternehmensweite Datenstandardisierung

Die Aufgabe, ein unternehmensweites Data Warehouse zu implementieren, ist oft entmutigend. Gerade in größeren Unternehmen sind DW-Projekte gigantische Unterfangen, vor allem aufgrund der Fülle an Anforderungen und Themenbereiche, die ein solches DW abdecken sollte. Eine Hilfestellung bietet hier das Konzept der **Data Marts**.⁴⁰ Ein DW wird nicht als ein monolithischer Block konstruiert, sondern in kleinere, vorwiegend nach thematischen oder organisatorischen Gesichtspunkten getrennte Einheiten zerlegt, die Data Marts. Hierbei muss es sich neben der logischen Separation nicht zwingenderweise auch um eine technische Isolation der Datenbestände handeln. Ein Problem bei der Verwendung von Data Marts ist die redundante Datenhaltung, die zu dem ernststen Problem der Dateninkonsistenz führen kann. Pflegen z.B. das Prüfungsamt und die Lehrstühle einer Universität die Daten von Studenten in separaten Merkmalstabellen, ist aufgrund fehlender oder mangelhafter Synchronisationsmechanismen nicht gewährleistet, dass ein Student in beiden Bereichen immer korrekt mit den aktuellen Adress- und Prüfungsleistungsdaten geführt wird.

Eine Lösung dieses Problems bietet das Konzept einheitlicher Merkmale und standardisierter Kennzahlen. **Einheitliche Merkmale** sind Merkmale, die in Verbindung mit allen möglichen Kennzahlen dieselbe Bedeutung haben und daher auch in allen Data Marts in allen Faktentabellen genutzt werden können, wie es auch das Galaxyschema vorsieht. Ebenso wie Merkmale bedürfen auch Kennzahlen einer Vereinheitlichung zu **standardisierten Kennzahlen**.

⁴⁰ Vgl. Kapitel 3.4.1

Der Begriff „Leistungspunkte“ sollte in jeder möglichen Verwendung dieselbe Bedeutung besitzen, ansonsten versagt das Data Warehouse in seiner Aufgabe als Kommunikationsmedium zwischen Menschen.

Es ist eine der Hauptaufgaben des zentralen DW-Designteams, einheitliche Merkmale und standardisierte Kennzahlen zu definieren, bereitzustellen und ihre Verwendung durchzusetzen (Kimball et al. 1998, 156). Ohne ein striktes Festhalten an diesen Vorgaben kann kein DW als Ganzes funktionieren (Kimball et al. 1998, 157). Da die Verwendung allgemein obligatorischer Merkmale und Kennzahlen eine ebenso politische wie technische Entscheidung darstellt, sollte sie von höchster Stelle im Unternehmen unterstützt werden, z.B. durch einen Chief Information Officer (CIO) (Kimball et al. 1998, 158).

3.5.4.5 Optimierung von Datenstrukturen

Die Speicherung von multidimensionalen Daten in einem Starschema oder einer Variante davon bedeutet, dass es aufgrund der vielfältigen Verknüpfungsmöglichkeiten unter den Dimensionen zu einer enorm großen Anzahl von Datensätzen in den Faktentabellen kommen kann, was sich negativ auf die Abfrageeffizienz auswirkt (Behme/Holthuis/Mucksch, 229). Mit den Konzepten der Vorverdichtung und der Partitionierung werden Optimierungsstrategien bereitgestellt, die speziell für relationale Datenbanken entwickelt wurden.

Ziel der Erstellung von **Vorverdichtungen** ist es, auf Basis konsistenter Dimensionen Summenstrukturen zu erzeugen, die vorberechnet werden und im DW bewusst redundant gehalten werden. Durch eine Verringerung der abzufragenden Datenmenge bei Queries wird ein Performancegewinn erzielt. Diese Strategie basiert auf der Annahme, dass die meisten Abfragen lediglich eine verdichtete Untermenge der Daten betreffen und selten auf den größtmöglichen Detaillierungsgrad heruntergehen (Behme/Holthuis/Mucksch, 230f.). Vorverdichtungen werden im BW durch das Konzept der **Aggregate** realisiert, die in der AWB zu den InfoCubes angelegt werden. Im Prinzip sind Aggregate selbst kleine InfoCubes, die jedoch nicht eigenständig administriert werden, sondern fest zum Ursprungs-Cube gehören⁴¹. Im BW können Aggregate nach gleichen Merkmalsausprägungen, nach Knoten einer Hierarchiestufe oder nach Festwerten gebildet werden (SAP 2001d, 7-31). Wenn Daten für eine Abfrage abgerufen werden, bestimmt der OLAP-Prozessor automatisch, ob ein Aggregat zur Verfügung steht, das die Anforderungen der Abfrage erfüllt (SAP 2001d, 7-29). Der Anwender kommt mit den Aggregaten nicht in Berührung, er profitiert aber von der gesteigerten Performance der Abfrage. Aggregate für zeitabhängige Stammdaten sind nicht möglich (SAP 2000d, 4). Zur Optimierung von Aggregaten können Daten aus einer vom BW-System geführten Statistik eingelesen werden und daraus Vorschläge für neue und zu löschende Aggregate abgeleitet werden (SAP 2001d, 7-49).

Ebenso wie die Aggregate dem Anwender verborgen bleiben, ist auch die **Partitionierung** eine Optimierungsstrategie, von der der Anwender nur die Effizienzgewinne bei Abfragen mitbekommt. Hier werden sehr große Datenbanktabellen, im Normalfall die Faktentabellen, in mehrere physische Tabellen aufgeteilt, logisch jedoch weiterhin als eine Einheit betrachtet.

⁴¹ Nach der reinen Relationentheorie ist das Abspeichern aggregierter Daten nicht notwendig und führt nur zu Problemen hinsichtlich der Integrität der Datenbank (Becker/Priemer/Wild 1994, 432). Jedoch wird im BW die Integrität zwischen InfoCube und Aggregaten systemseitig gewährleistet.

Oft werden die Tabellen nach einem Zeitmerkmal partitioniert, so dass z.B. die Daten eines Monats oder eines Jahres in separate Partitionen abgelegt werden (Behme/Holthuis/Mucksch, 231). Auch das BW bietet die Funktionalität zur Partitionierung von Faktentabellen an⁴². Voraussetzung ist allerdings, dass die Partitionierung anhand eines der beiden Partitionsmerkmale Kalendermonat (0CALMONTH) oder Geschäftsjahr/Periode (0FISCPER) vorgenommen wird. Mindestens eines dieser beiden InfoObjects muss also im InfoCube enthalten sein (SAP 2000b).

3.5.5 Umsetzung von Datenbankschemata im Rahmen der physischen Modellierung

3.5.5.1 SAP BW als physisch relationales Data Warehouse System

Es wird zwischen physisch multidimensionalen und physisch relationalen Datenbanksystemen unterschieden, für die unterschiedliche logische Entwurfsschemata entwickelt wurden. Bei der Vorstellung logischer Datenbankschemata im vorigen Kapitel wurden daher bewusst nur die Konzepte vorgestellt, die sich für physisch relationale Datenbanksysteme eignen.

Unter **physisch multidimensionalen** Datenbanksystemen (MOLAP, multidimensionales OLAP) werden Systeme verstanden, welche die auf konzeptueller Ebene dargestellten multidimensionalen Datenstrukturen auch in ihren physischen Datenbank- und Speicherstrukturen umsetzen. Bislang besteht noch kein Standard, welche Grundfunktionalitäten ein multidimensionales Datenbanksystem enthalten muss. Beispiele für am Markt erhältliche Produkte sind *Express* von Oracle, *Holos* von Seagate Software und *Essbase* von Applix. Erste Implementierungen zeigen, dass diese Systeme keine besonders großen Datenbestände verwalten können (Behme/Holthuis/Mucksch, 216f.).

Die grundlegende Prämisse bei der multidimensionalen Modellierung im **relationalen Umfeld** (ROLAP, relationales OLAP) ist die Klassifikation der Daten in Fakt- und Dimensionsdaten. Beide Datengruppen werden, wie in relationalen Datenbank-Management-Systemen (DBMS) üblich, in Tabellen gehalten, die über Schlüssel miteinander verbunden sind. Die Primärschlüssel der Dimensionstabellen sind Fremdschlüssel der Faktentabelle, wobei die Gesamtheit aller Fremdschlüssel den zusammengesetzten Primärschlüssel der Faktentabelle bildet (Behme/Holthuis/Mucksch, 224f.). Das BW ist ein klassisches relationales Data Warehouse System (SAP 2000a, 2).

3.5.5.2 Namenskonventionen beim Implementieren von Datenstrukturen

An dieser Stelle soll kein Nomenklatur-Schema für eine BW-Implementierung vorgestellt werden. Es ist prinzipiell im Ermessen des Designers, wie er die InfoObjects, InfoCubes usw. benennen möchte. Lediglich auf drei Punkte soll hier hingewiesen werden.

⁴² Nur bestimmte Datenbankanbieter unterstützen diese Funktion (z.B. Oracle und Informix). Wenn eine Datenbank verwendet wird, die diese Funktion nicht unterstützt, wird diese Funktion vom BW-System nicht angeboten (SAP 2000b).

1. Es dient der Übersichtlichkeit, Nachvollziehbarkeit und der leichteren Wartbarkeit eines DW, wenn die Benennung der Objekte nach einem logischen Schema erfolgt. Wichtig hierbei ist, dass die Anwendung des Schemas konsequent eingehalten wird.
2. Original-SAP-Objekte werden im BW nach dem Schema **0XXX** bis **9XXX** benannt. Selbstverständlich sind Änderungen an diesen Objekten untersagt und dieser Namensraum ist für kundenspezifische Objekte gesperrt.
3. Speziell im HCC-Umfeld gelten einige besondere Namensregelungen. Alle individuell angelegten Objekte müssen mit einem vom zuständigen HCC vergebenen eindeutigen Präfix⁴³ beginnen. In den Bereichen Datenziele und InfoObjects erhält jede Institution eigene InfoAreas, um Konflikte zu vermeiden⁴⁴ (HCC 2002a).

3.5.6 Spezialitäten und Problemfelder der Modellierung im BW

3.5.6.1 Additive, semiadditive und nicht additive Fakten

Wenn möglich, sollten in Faktentabellen verwendete Kennzahlen numerisch und vollständig additiv sein. Vollständig additiv bedeutet, dass die Standardaggregationsmethode *Summation* über alle Dimensionen hinweg sinnvoll ist. Flussgrößen wie Umsätze oder Kosten sind im Normalfall vollständig **additiv**, sie können z.B. über den Zeitverlauf und über verschiedene Regionen hinweg sinnvoll addiert werden. Bestandsgrößen wie ein Kontostand oder Lagerbestände sind meistens über alle Dimensionen außer der Zeit additiv aggregierbar, man nennt sie daher **semiadditive** Fakten. Andere Kennzahlen wiederum können nach keiner denkbaren Dimension sinnvoll aufsummiert werden. Es macht beispielsweise keinen Sinn, eine gemessene Temperatur im Zeitverlauf oder über Orte hinweg aufzusummieren. Es handelt sich hierbei um **nicht additiv** Fakten (Kimball et al. 1998, 193f.). Diese Sachverhalte können im BW prinzipiell auf zwei Wegen abgebildet werden:

1. Bei der Definition einer Kennzahl in der AWB besteht die Möglichkeit, eine Ausnahmeaggregation bezüglich eines Ausnahmemerkmals festzulegen. So kann hier beispielsweise bei der Kennzahl *Kontostand* hinterlegt werden, dass bezüglich der Zeit eine abweichende Aggregationsmethode angewandt wird, beispielsweise *Letzter Wert*. Wird ein neuer Detaildatensatz in die Faktentabelle geschrieben, der hinsichtlich aller Merkmale mit einem bereits existierenden Datensatz übereinstimmt, wird der alte Kontostand durch den neuen ersetzt.
2. Im Rahmen der Definition einer Query auf Anwendungsebene kann selbst für Kennzahlen, die in der AWB als additiv definiert wurden, eine Ausnahmeaggregation vorgegeben werden. Dies gilt selbstverständlich nur für Aggregationen, welche über die durch die Modellierung vorgegebene Granularitätsstufe hinausgehen.

Ähnlich verhält es sich mit der Kennzahl **Note**. Eine Note (1...6) kann weder sinnvoll im Zeitverlauf addiert werden noch über Studenten oder Veranstaltungen hinweg. Die Aggregationsmethode, die in diesem Fall den meisten Sinn macht, ist die Bildung eines Mittelwertes. Die so entstehenden Mittelwerte pro Student (aktueller Notenschnitt), pro Veranstalt-

⁴³ Beispielsweise beginnen alle von der Universität Hohenheim angelegten Objekte mit dem Buchstaben *S*.

⁴⁴ Die InfoAreas der Universität Hohenheim heißen *BW Training Uni Hohenheim*.

ung (Klausurdurchschnitt) oder pro Semester (Semesterschnitt) haben durchaus eine Aussagekraft.⁴⁵

3.5.6.2 Modellierung einer Zeitdimension

Die Zeitdimension nimmt in jedem DW eine spezielle Position ein, denn jede Faktentabelle spiegelt Tatbestände in Form von Kennzahlen im Zeitverlauf wider. RALPH KIMBALL schlägt eine standardisierte Zeitdimension mit der Granularitätsstufe *Tag* vor, die auch Sachverhalte wie Jahreszeiten, Geschäftsjahre, nationale Feiertage u.ä. berücksichtigt (Kimball et al. 1998, 175f.).

Für ein einfaches universitäres Berichtswesen reicht eventuell die grobe Strukturierung der Zeitdimension durch die einzelnen Semester. So kann ein Merkmal *Semester* angelegt werden, das die textliche Beschreibung des Semesters (WS02/03, SS03 usw.) darstellt und zumindest das Anfangsdatum als Attribut enthält. Dieses Attribut ist notwendig, um in Abfragen ein Feld für die korrekte Sortierung zur Verfügung zu haben, die für Zeitreihenanalysen unerlässlich ist. Dennoch muss im InfoCube ein Standard-Zeitmerkmal eingebunden werden. Folgende Abbildung zeigt die Verwendung des Merkmals *Semester* in einer Zeitdimension:

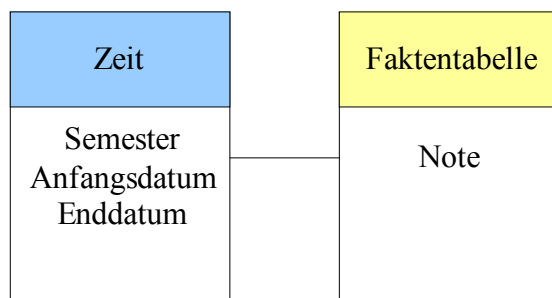


Abbildung 15: Zeitdimension mit Merkmal *Semester* (Quelle: Eigene Darstellung)

3.5.6.3 Erzwingung der Mandantenfähigkeit des BW

Das Business Information Warehouse wurde im Gegensatz zu R/3-Systemen mandantenunabhängig entwickelt. Es liegt ja gerade im Wesen von Data Warehouses, dass möglichst alle Daten integriert gesammelt werden. Diese Unabhängigkeit von Mandanten ist im Datenmodell daran erkennbar, dass im BW in keiner Tabelle das Feld *Mandt* (Mandant) enthalten ist, welches in R/3-Systemen das erste Schlüsselfeld für alle Tabellen darstellt (SAP 2001, 247). Dennoch fließt in das BW an einigen Stellen mandantenabhängige Information ein (z.B. Nummernkreise für Dimension und Stammdaten-IDs). Aus der sich ergebenden Mandantenproblematik folgt, dass das BW nur in einem einzigen Mandanten betrieben werden darf.⁴⁶

Besteht dennoch aus irgendeinem Grund die Anforderung, Daten im BW einem bestimmten Mandanten zuweisen zu müssen, kann dies durch einfache Nachbildung des Merkmals *Mandant* realisiert werden. Es wird ein Merkmals-InfoObject *Mandant* definiert, das allen anderen InfoObjects sinnvollerweise als geklammertes Attribut angehängt wird und somit die

⁴⁵ Vgl. Kapitel 3.7.3.3

⁴⁶ Für HCC-Kunden ist der Mandant 800 gültig.

Unterscheidung mehrerer Mandanten im BW ermöglicht. Der große Nachteil dieser Lösung ist, dass das Mandantenmerkmal bei allen InfoObjects manuell berücksichtigt werden muss, was einen hohen administrativen Aufwand mit sich bringt.

3.5.6.4 Modellierung einer Kennzahl als Faktum oder als Attribut

Im Normalfall ist die Unterscheidung zwischen Fakt und Attribut einfach. Bei einigen Kennzahlen jedoch ist es ein Problem zu entscheiden, ob eine Kennzahl in einer Faktentabelle oder als Kennzahlen-Attribut eines Merkmals im BW abgebildet wird.

Beispielsweise sind die **Preise** eines Produktes auf der einen Seite Beschreibungen der Produkte, ebenso wie auch Hersteller, Produktgruppe usw. Attribute des Produkt-Merkmals sind. Daher müsste der Preis eines Produkts als Attributsmerkmal in die Produkt-Stammdaten mit aufgenommen werden. Auf der anderen Seite stellen Preise auswertbare Kennzahlen dar. So könnten in Abfragen beispielsweise Gesamtpreise oder Rabatte auf Basis dieser Preise interessieren. Diese Rabatte können nicht unter Verwendung des Preis-Attributs aus den Produkt-Stammdaten kalkuliert werden, wenn eine Abfrage gestartet wird, denn dort steht im Normalfall der aktuelle Preis und nicht der Preis, der zum Zeitpunkt der Entstehung des Detaildatensatzes in der Faktentabelle gültig war. Diese Sichtweise spricht für eine Integration der Preis-Kennzahl in die Faktentabelle, was beispielsweise durch das direkte Übernehmen des Stammdatenattributs Preis in die Faktentabelle im Rahmen der Fortschreibungsregeln⁴⁷ beim Datenladeprozess realisiert werden kann (SAP 2000c, 73).

Ein ähnlich gelagertes Problem stellt sich bei der Modellierung der Kennzahl **Leistungspunkte**. Leistungspunkte sind vom Wesen her eindeutig eine Kennzahl, jedoch ist nicht unmittelbar einsichtig, ob sie im Rahmen einer Modellierung eines Klausuren-InfoCube als Fakt in die Faktentabelle eingehen oder als Attribut zum Merkmal *Veranstaltung* angesehen werden sollen. Einerseits sind Leistungspunkte eine Beschreibung des Merkmals *Veranstaltung*, andererseits werden sie für die Berechnung diverser anderer Kennzahlen, z.B. dem gewichteten Notendurchschnitt, in der Faktentabelle benötigt. Da hier kein Zeitproblem wie bei oben beschriebenen Produktpreisen vorliegt⁴⁸, sind prinzipiell beide Varianten möglich, denn auch Kennzahlenattribute aus Merkmalen können in BEx-Queries durch die Verwendung von sog. Formelvariablen⁴⁹ für Berechnungen herangezogen werden.

3.5.6.5 Event Tracking mit faktenlosen Faktentabellen

Einen Sonderfall bei Faktentabellen stellen die so genannten *Factless Fact Tables* (faktenlose Faktentabellen) dar. Es handelt sich hierbei um Faktentabellen, die keine echten betriebswirtschaftlichen Kennzahlen beinhalten (Behme/Holthuis/Mucksch, 224), sondern ihrer ursprünglichen Herkunft als Überschneidungsentität entsprechend „lediglich“ die Verknüpfungen zwischen den beteiligten Dimensionen abbilden.

RALPH KIMBALL führt als typischen Anwendungsfall das *Event Tracking* (Ereignisverfolgung) an, wofür die Teilnahme an universitären Wahlen ein gutes Beispiel ist. In diesem

⁴⁷ Vgl. Kapitel 3.6.2.2

⁴⁸ Ein Zeitproblem liegt nicht vor, wenn man unterstellt, dass die universitären Veranstaltungen im Zeitverlauf stets mit denselben Leistungspunkten bewertet werden.

⁴⁹ Vgl. Kapitel 3.7.3.3

Fall würde eine Faktentabelle modelliert, die eine Reihe von Fremdschlüsseln als Verweis auf die Dimensionen *Student*, *Wahlanlass*, *Zeit*, *Raum* u.ä., jedoch kein Faktum enthielte. Der Grund für die „Faktenlosigkeit“ kann darin liegen, dass aus datenschutzrechtlichen Gründen beispielsweise das Wahlergebnis des betreffenden Studenten nicht im DW abgespeichert werden darf oder das Ergebnis selbst überhaupt nicht interessiert, weil lediglich Wahlverhalten und -beteiligung für diverse Auswertungen von Interesse sind.

Die direkte Abbildung faktenloser Faktentabellen ist im BW jedoch leider nicht möglich, denn jeder InfoCube muss mindestens eine Kennzahl enthalten. Einen Ausweg aus diesem nicht ganz nachvollziehbaren Zustand bietet folgende Vorgehensweise:

1. Eine numerische, ganzzahlige Dummy-Kennzahl *Zähler* wird definiert.
2. Man integriert die Kennzahl *Zähler* in einen bislang faktenlosen InfoCube.
3. Beim Laden der Bewegungsdaten in den InfoCube wird der Kennzahl *Zähler* der konstante Wert „1“ zugewiesen.
4. Bei Auswertungen auf den InfoCube kann der *Zähler* dazu verwendet werden, die Anzahl der Ereignisse, hier der Wahlbesuche, darzustellen.

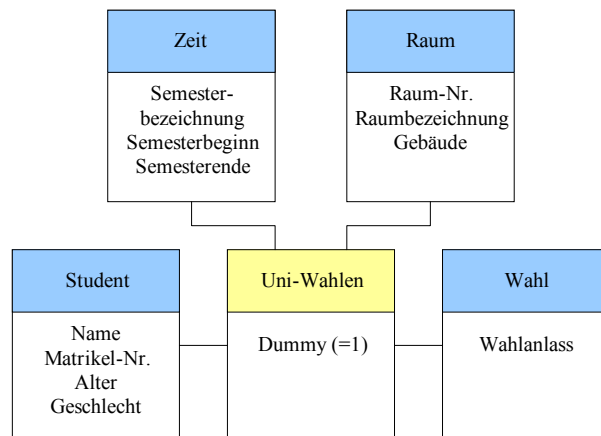


Abbildung 16: Faktenlose Faktentabelle am Beispiel von universitären Wahlen (Quelle: Eigene Darstellung)

Nicht nur deshalb, weil die Umsetzung am BW-System relativ umständlich ist, sollten faktenlose Faktentabellen möglichst vermieden werden. In vielen Fällen lassen sich andere, aussagekräftigere Kennzahlen als ein Dummy finden (Kimball et al. 1998, 249).

3.5.6.6 Das Rollenspiel der Merkmale

Innerhalb eines multidimensionalen Datenmodells kann es Fälle geben, in denen dasselbe Merkmal mehrmals, jedoch in verschiedenen Rollen auftaucht. So ist es denkbar, dass hinsichtlich universitärer Veranstaltungen ein Lehrstuhl zwei unterschiedliche Aufgaben wahrnimmt: Zum einen tritt er als *verantwortlicher* Lehrstuhl auf, der eine bestimmte Veranstaltung inhaltlich und organisatorisch plant und vorbereitet. Tatsächlich ausgeführt wird die Veranstaltung jedoch von anderer Seite, von einem *durchführenden* Lehrstuhl.

Eine mögliche Variante der Implementierung dieses inhaltlich nicht außergewöhnlichen Modells besteht darin, zwei separate Merkmale *Lst_Verantwortlich* und *Lst_Durchführend* anzulegen, die dann beide in ein Datenmodell integriert werden. Jedoch müssten dann auch

zwei separate Stammdatentabellen gepflegt werden, obwohl es sich um dieselben Lehrstühle handelt, die eben im Einzelfall in verschiedenen Rollen auftreten. Dies bedeutet doppelten Administrationsaufwand und höhere Anforderungen an den Speicherplatz.

Das BW ermöglicht für diesen Sachverhalt eine elegantere Problemlösung mittels **Referenzmerkmalen**. Zunächst wird ein Merkmal *Lehrstuhl* samt Attributen, Texten usw. definiert, welches also selbst Stammdaten enthalten kann. Daraufhin werden zwei weitere Merkmale *Lst_Verantwortlich* und *Lst_Durchführend* definiert, die auf *Lehrstuhl* als Referenzmerkmal verweisen. Das Referenzmerkmal liefert die technischen Eigenschaften zu einem anderen Merkmal. Unter technische Eigenschaften fallen Attribute, Stammdaten, Texte, Hierarchien, Datentyp, Länge, Anzahl und Art der geklammerten Merkmale, Kleinbuchstaben und Konvertierungsroutine. Diese Eigenschaften sind nur beim Referenzmerkmal pflegbar. Somit verweisen die beiden Merkmale auf die Felder des Referenzmerkmals, ohne selbst Daten zu enthalten. Neben den technischen Eigenschaften trägt ein Merkmal auch die betriebswirtschaftliche Semantik. Daher sind Eigenschaften wie Beschreibung, Darstellung, Textauswahl, Berechtigungsrelevanz, Verantwortlicher, Konstante und ausschließlich Attribut auch bei Merkmalen pflegbar, die auf einem Referenzmerkmal basieren (SAP 2000b).

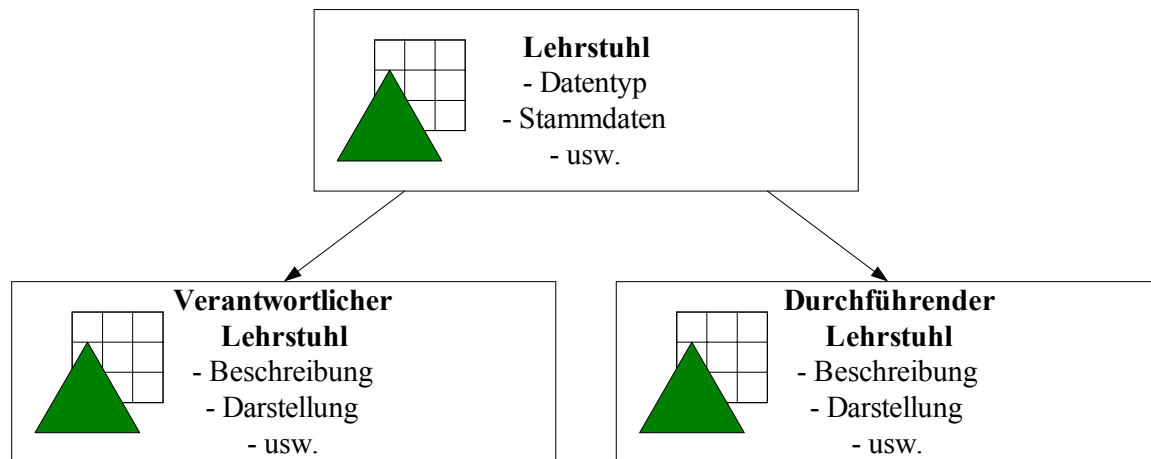


Abbildung 17: Referenzmerkmal (Quelle: Eigene Darstellung)

3.5.6.7 Vermeidung dünn besetzter Matrizen durch MultiCubes

Häufig werden Auswertungen über Daten aus zwei oder mehr Geschäftsprozessen gleichzeitig benötigt, die jedoch in unterschiedlichen InfoCubes gespeichert sind.

Anhand des folgenden Beispiels werden die möglichen Lösungsvarianten für dieses Problem vorgestellt: Um die Hypothese zu testen, ob Lehrstühle mit besseren Klausurergebnissen größeren Zulauf bei den Diplomarbeiten haben, wird eine Auswertung benötigt, die sowohl die Anzahl der geschriebenen Diplomarbeiten eines Lehrstuhls als auch die in diversen Klausuren vergebenen Noten liefert.

Eine relativ umständliche Lösung besteht darin, die Daten **pro InfoCube separat** auszuwerten und dann manuell zusammenzustellen. Dies wird vom Anwender in der Regel verständlicherweise nicht akzeptiert. Daher könnte der Entwickler dazu verleitet werden, die benötigten Daten statt in zwei InfoCubes in einem **einzigen InfoCube** abzulegen. Folgende Abbildung verdeutlicht die Vorgehensweise:

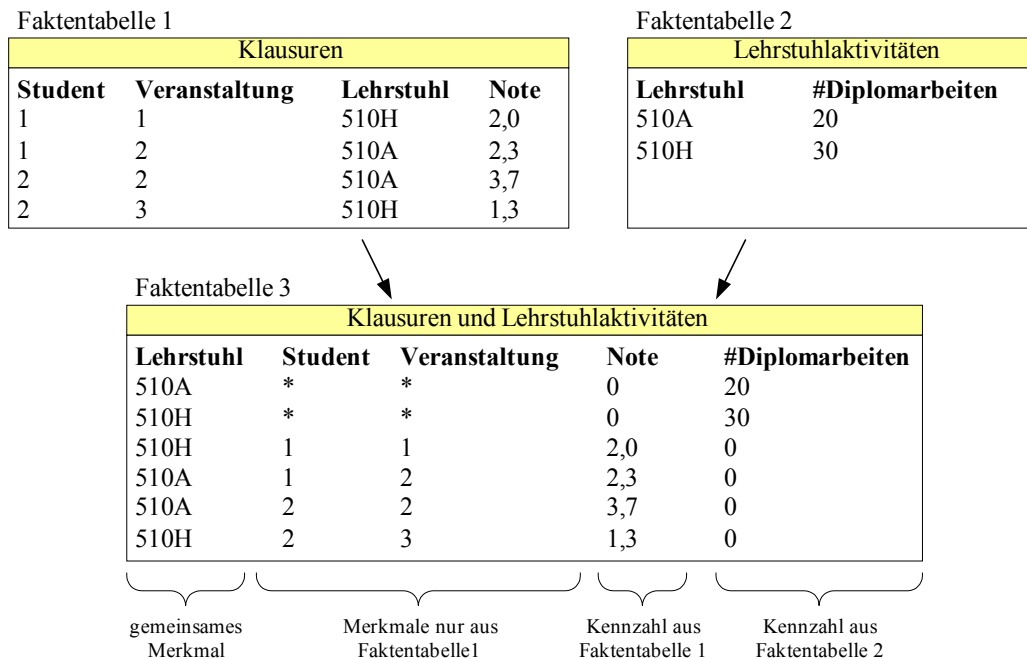


Abbildung 18: Schwach besetzte Matrix (Quelle: Eigene Darstellung)

Die entstehende Faktentabelle sieht wie ein „Schweizer Käse“ aus, denn an Stellen, für die aus den Quelldatensätzen keine Informationen gezogen werden können, werden Initialwerte wie * oder 0 eingetragen. Es handelt sich um eine schwach besetzte Matrix. Durch geeignete Modellierungsüberlegungen könnte eventuell ein eleganteres Schema entwickelt werden.

Das BW bietet mit dem Konstrukt der **MultiCubes** eine speicherplatzsparende, performante und transparente Herangehensweise an die Problematik (SAP 2000c, 68). MultiCubes sind InfoCubes, die selbst keine Daten enthalten, sondern diese aus mehreren anderen InfoCubes zusammenführen und in einen gemeinsamen Kontext stellen. Indem die gemeinsamen Merkmale und die Kennzahlen der zugrunde liegenden InfoCubes im MultiCube zusammengeführt werden, ist ein Reporting über mehrere InfoCubes hinweg möglich. Dieses Konzept bietet erweiterte Analysemöglichkeiten, ohne dass neue, extrem große InfoCubes mit Daten gefüllt werden müssen (SAP 2001d, 9-14 f.).

MultiCubes können mit **Views** verglichen werden, die in relationalen Datenbanksystemen eine dynamische Sicht auf physische Tabellen darstellen.

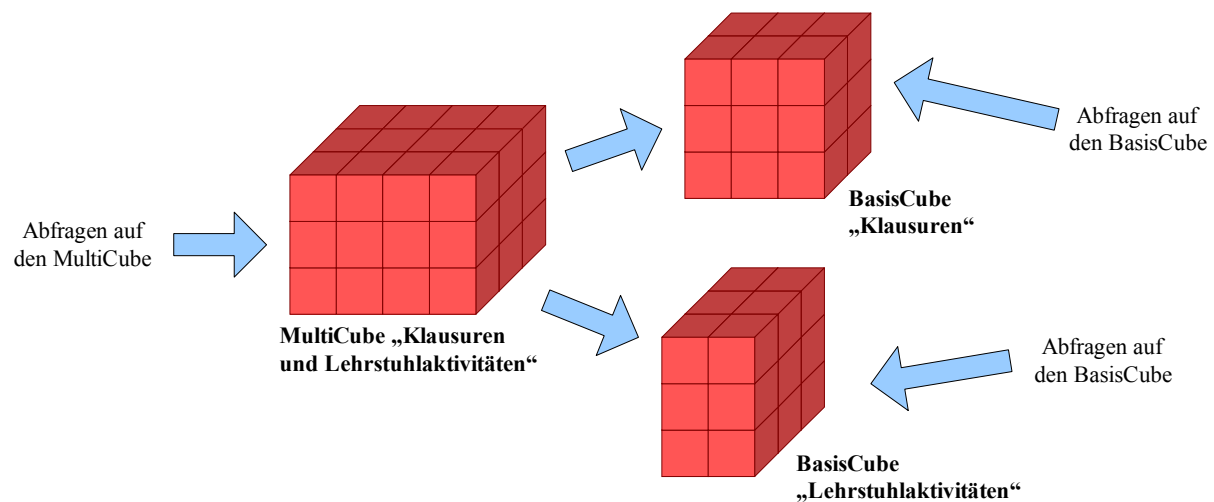


Abbildung 19: MultiCubes (Quelle: Eigene Darstellung)

3.5.7 Vorkonfigurierte Informationsmodelle in Form des Business Content

Wie in den letzten Kapiteln deutlich wurde, ist die Modellierung anforderungsgerechter Datenmodelle eine langwierige und teilweise hoch komplexe Angelegenheit. Der Aufwand ist umso höher, je individueller die Anforderungen sind und je weniger die Entwickler auf bereits existierende Vorlagen zurückgreifen können. An diesem Punkt setzt die Idee des **Business Content** der SAP AG an. Es wurde festgestellt, dass Unternehmen in vielen Fällen immer dieselben Sachverhalte modellieren und abbilden. Mit dem Business Content (BC) werden daher vorkonfigurierte, auf konsistenten Metadaten basierende Informationsmodelle im BW angeboten. Der BC stellt ausgewählten Rollen in einem Unternehmen das Angebot an Informationen zur Verfügung, das diese zur Erfüllung ihrer Aufgaben benötigen. Diese Informationsmodelle umfassen im wesentlichen Rollen, Arbeitsmappen, Queries, InfoSources, InfoCubes, Kennzahlen, Merkmale, Fortschreibungsregeln sowie Extraktoren für diverse Quellsysteme (SAP 2000b). Business Content kann

- ohne Anpassung verwendet werden,
- angepasst⁵⁰, d.h. verfeinert oder vergrößert werden und
- als Vorlage oder Beispiel für selbsterstellten Business Content dienen.

Der Business Content ist das Konzept, mit dem sich das SAP BW am gravierendsten von anderen Data Warehouses unterscheidet, denn das Angebot von 2.500 InfoObjects und 450 vorgefertigten Reports für 60 Rollen⁵¹ ist bislang einmalig in der DW-Produktwelt (Mertens/Bange/Schinzer 2000b, 14). Diese Idee ermöglicht die sehr schnelle Implementierung eines DW, da der aufwändige Prozess der Datenmodellierung und Definition des ETL-Prozesses entfällt. Standarddatenmodelle allein sind allerdings längst kein Alleinstellungsmerkmal von Systemhäusern oder Produktlieferanten mehr. Die inzwischen verfügbare Vielfalt an Publikationen im DW-Umfeld hat dieses Thema immer wieder aufgegriffen und

⁵⁰ Eine ernste Problematik bei Verwendung des Business Content besteht nach Auskunft der KPMG darin, dass, wenn im R/3-Quellsystem kundenspezifische Änderungen durchgeführt wurden (z.B. Änderung der Feldlänge), die Anpassung des BC extrem arbeitsaufwändig ist.

⁵¹ Die Zahlenangaben sind Schätzwerte, denn der Business Content wächst ständig an.

für einen reichen Fundus an publizierten Ergebnissen gesorgt. BILL INMON, Pionier des Data Warehousing, bietet beispielsweise auf seiner Webseite www.billinmon.com frei verfügbar zwölf umfassende Datenmodelle für bestimmte Anwendungsbereiche und Branchen an (Hecht/Bange/Schinzer 2000, 40).

Aus drei Gründen wird in dieser Veranstaltung nicht detaillierter auf den BC eingegangen:

1. Es ist erklärtes Ziel der hier konzipierten Lehrveranstaltung, den Teilnehmern die Ideen und Methoden der multidimensionalen Datenmodellierung nahe zu bringen. Das Arbeiten mit bereits vorgefertigten Datenmodellen kann daher höchstens darin bestehen, die BC-Modelle zu analysieren und Schlüsse für die eigene Modellierung daraus zu ziehen oder die vorgefertigten Queries für einen „Schnellstart“ im BW-System zu nutzen.
2. Im BC, der auf Basis der R/3-Geschäftsmodelle entwickelt wurde, werden hauptsächlich Daten aus R/3-Systemen angebunden. Für Daten aus anderen Systemen müssen genau wie bei anderen Produkten eigene Extraktions- und Ladeprozesse definiert werden (Mertens/Bange/Schinzer 2000b, 14). Da zum Zeitpunkt der Erstellung dieses Theorieskriptes eine Anbindung eines BW-Systems an ein operatives R/3-System von den HCCs nicht unterstützt wurde, bietet sich das detaillierte Analysieren der Business Content Objekte aus Gründen der fehlenden Durchgängigkeit eines Datenladeprozesses nicht an.
3. Ein erklärtes Ziel der Lehrveranstaltung ist es, herstellerunabhängige Data Warehouse Konzepte zu vermitteln.⁵² Da aber gerade der Business Content ein Spezifikum des SAP BW darstellt, sollte dieses Thema bewusst kurz gehalten werden.

Mit dem SAP **Demo Content** werden neben den Metadaten des BC auch Beispieldaten ausgeliefert. Somit können schnell fertige Demonstrationsszenarien aus verschiedenen Unternehmensbereichen gezeigt und der Umgang mit dem BW erlernt werden (SAP 2000b). Demo Content wird für die Bereiche Einkauf, Ergebnisrechnung und Vertrieb bereitgestellt. In der hier konzipierten Lehrveranstaltung wird der DemoContent im Rahmen eines Crashkurses dazu verwendet, den Teilnehmern das rasche Arbeiten mit bereits existierenden Daten zu ermöglichen.

3.5.8 Daten über Daten

Informationen über die Datenstrukturen und ihre Beziehungen sind „Daten über Daten“ und werden als Metadaten bezeichnet. In einer DW-Umgebung sind zwei Arten von Metadaten von Interesse.

- **Technische Metadaten** beinhalten Informationen über das Data Warehouse, die von DW-Administratoren und Designern benötigt werden, um das Data Warehouse zu entwickeln und zu betreiben. Darunter fallen beispielsweise Datenbankfelder, -spalten, -tabellen, Speicherbedarf der Datenbank, Datenmodelle und Mappings (Frie/Strauch 1999, 6). Sie umfassen auch Informationen des operationellen Systems und können z.B. den Originalnamen eines Quellsystems beinhalten (Poe/Reeves 2000, 51).

⁵² Vgl. Kapitel **Fehler! Verweisquelle konnte nicht gefunden werden.**

- **Fachliche Metadaten** enthalten dagegen solche Informationen, die dem Fachanwender eine geschäftliche Sicht auf das Data Warehouse ermöglichen. Darunter fallen z.B. Details über Auswertungen, Fachbegriffe usw. (Frie/Strauch 1999, 6). Fachliche Metadaten weisen Daten aus dem DW dem multidimensionalen Geschäftsmodell und dem Frontend-tool des Endanwenders zu und beinhalten üblicherweise geschäftsinterne Bezeichnungen und Hierarchien (Poe/Reeves 2000, 51).

Ein **Metadatenbanksystem** stellt für den Endbenutzer eine Art Hilfesystem bereit, bestehend aus Informationskatalog und Navigationshilfe. Darüber hinaus unterstützt ein solches System auch die Datenbankadministratoren (Mucksch/Behme 2000, 23). Im BW sind beide Elemente eines so definierten Metadatenbanksystems enthalten: Das *Metadata Repository* verwaltet zentral sämtliche Metadaten des BW-Systems. Mit dem HTML-basierten *Metadata Repository Browser* wurde ein Werkzeug für einen schnellen und komfortablen Zugriff auf die Metadaten geschaffen. Es kann zentral auf Informationen wie etwa die wichtigsten Eigenschaften der Objekte und die Verknüpfungen mit anderen Objekten zugegriffen werden. Der Benutzer kann Metadaten suchen, austauschen, HTML-Seiten exportieren und Grafikdarstellungen der Objekte anzeigen (SAP 2000b). Ferner kann man Dokumentationen zu Metaobjekten⁵³, welche im Business Document Service (BDS) hinterlegt sind, aufrufen und pflegen. Diese Dokumentationen können in Form von Textdateien (z.B. Word, HTML, PDF), Präsentationsfolien (z.B. Powerpoint) und Bildern (z.B. Bitmapgrafiken) auf dem BDS abgelegt werden. Nachteilig ist, dass Dokumentationen nur an Metaobjekte als Ganzes angefügt werden können (z.B. *Lehrstuhl*), nicht jedoch an einzelne Ausprägungen (z.B. Dokumentation zum *Lehrstuhl 510H*) (Seemann/Schmalzridt/Lehmann 2001, 174f.).

3.6 Design und Implementierung der Datenbereitstellung

Nach dem im vorigen Kapitel beschriebenen Prozess der Modellierung der Datenumgebung eines Data Warehouse besteht der nächste Schritt im Lifecycle-Modell nach KIMBALL darin, die notwendigen Vorkehrungen für eine problemlose Datenbereitstellung zu treffen. Dabei wird betrachtet, wo die Daten liegen, in welchen Strukturen sie vorliegen und wie sie in reportingfähige BW-Strukturen gelangen (SAP 2000a, 1).

3.6.1 Persistente und nicht persistente Stagingszenarien

Beim Prozess der Bereitstellung von Bewegungsdaten⁵⁴ für ein DW lassen sich prinzipiell zwei Gruppen von Szenarien unterscheiden: Stagingszenarien⁵⁵ mit persistenter Datenablage und Szenarien ohne persistente Datenablage. Persistent bedeutet, dass die aus dem Quellsystem ins BW-System geladenen Daten über die Dauer einer Transaktion hinaus gespeichert werden. Nichtpersistent bedeutet, dass die Daten immer wieder neu beschafft werden und nur für die Dauer einer Transaktion im BW-System gehalten werden (SAP 2000a, 1).

Ein Stagingszenario mit **nicht persistenter Datenablage** lässt sich im BW durch das Konzept der **RemoteCubes** realisieren. Ein RemoteCube ist ein InfoCube, dessen Bewegungsdaten nicht im BW verwaltet werden, sondern extern, z.B. in einem R/3-System. Es handelt

⁵³ InfoSources, InfoCubes, InfoObjects, Aggregate und Queries können im BW dokumentiert werden.

⁵⁴ Auf das Laden von Stammdaten wird in Kapitel 3.6.3 eingegangen.

⁵⁵ Staging: Datenbereitstellung

sich somit um ein virtuelles DW.⁵⁶ Im BW wird nur die Struktur des RemoteCube definiert. Die Daten werden zum Reporting über ein BAPI (Business Application Programming Interface) aus einem anderen System gelesen (SAP 2000b). Diese Variante des direkten Datenzugriffs auf die Quellsysteme ist dann zu empfehlen, wenn eine zeitnahe Verfügbarkeit von Daten aus dem SAP Quellsystem erforderlich ist, nur sporadisch auf eine kleine Datenmenge zugegriffen werden soll, im Vergleich dazu eine Replikation der Daten ins BW zu aufwändig ist und nur wenige Benutzer gleichzeitig Queries auf dem angeforderten Datenbestand des Quellsystems ausführen (SAP 2000a, 6). Auf RemoteCubes und Stagingszenarien mit nicht persistenter Datenablage wird im Rahmen dieser Veranstaltung nicht weiter eingegangen, da sie einem Grundgedanken des Data Warehousing widersprechen, nämlich der strikten Trennung operativer und analytischer Daten.

Im Rahmen der Stagingszenarien **mit persistenter Datenablage**, d.h. Speicherung der Bewegungsdaten im BW, wird auf folgende Variante des Datenbereitstellungsprozesses explizit eingegangen: Aus einem Quellsystem werden Daten extrahiert und in einer Zwischenablage im BW, der Persistent Staging Area (PSA), unverändert gespeichert. Von der PSA werden die Bewegungsdaten mithilfe der Übertragungsregeln, die auch Datenmodifikationen bewirken können, unter Beachtung gewisser Fortschreibungsregeln direkt in die InfoCubes fortgeschrieben. Dieses Szenario ist unter folgenden Voraussetzungen besonders interessant (SAP 2000a, 4f.):

- Auf die Vorteile des ODS-Objekts (v.a. belegnahe Speicherung) kann verzichtet werden.
- Die Integrität der Daten wird bereits vom Quellsystem garantiert.
- Die Daten werden aus einem einzigen Quellsystem extrahiert.

Wie in der Beschreibung der Fallstudien-Lerneinheit zum Thema „Datenextraktion“⁵⁷ deutlich wird, sind diese Voraussetzungen in der Fallstudie allesamt gegeben, so dass das oben beschriebene Szenario zum Einsatz kommen kann. Daher ist im Folgenden immer dieses Stagingszenario gemeint, wenn von Datenladeprozess die Rede ist.

Einen Überblick über weitere denkbare Szenarien liefert folgende Abbildung.

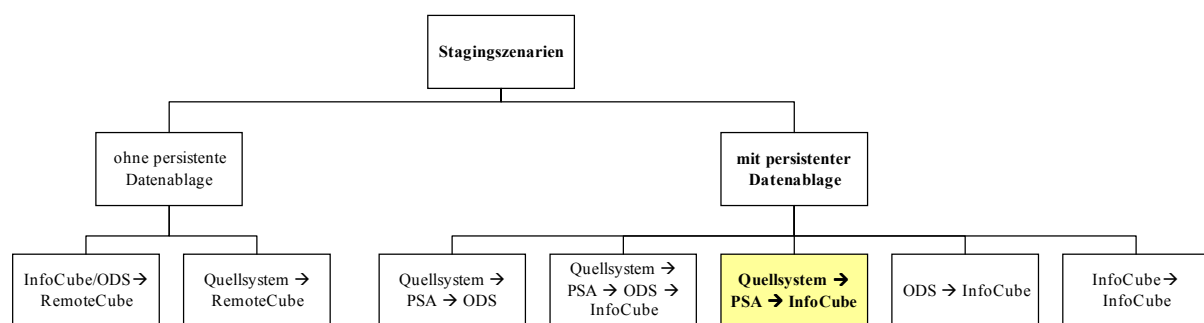


Abbildung 20: Stagingszenarien (Quelle: Eigene Darstellung, in Anlehnung an SAP 2000a)

⁵⁶ Vgl. Kapitel 3.4.1

⁵⁷ Vgl. Kapitel Fehler! Verweisquelle konnte nicht gefunden werden.

3.6.2 Elemente der Datenbereitstellung

3.6.2.1 Quellsysteme als Datenlieferanten

Alle Systeme, die Daten zur Extraktion in das BW liefern, werden als Quellsysteme bezeichnet (SAP 2001d, 5-9). Die Datenbereitstellung des BW ist nicht nur auf Daten aus SAP-Systemen beschränkt (Seemann/Schmalzridt/Lehmann 2001, 140): Im BW können sowohl R/3-Systeme, Textdateien⁵⁸, Fremdsysteme und andere BW-Systeme als Quellsystem definiert werden können (SAP 2000b). Jedoch beschränkt sich das BW auf die Speicherung strukturierter Daten, d.h. auf unstrukturierte Informationsquellen wie Internetquellen⁵⁹, Videodateien usw. kann nicht zugegriffen werden. Für diese Datenwelt bietet SAP das Web Content Management an, wodurch die quantitativen Informationen aus dem BW um Hintergrundinformationen ergänzt werden können (Seemann/Schmalzridt/Lehmann 2001, 18f.).

Voraussetzung für den Datenupload aus den Quellsystemen ist eine passende Metadatendefinition. Bei SAP-Quellsystemen können die Metadaten automatisch in das BW übertragen werden, bei Textdateien müssen sie manuell definiert werden. Der Pflegeaufwand von Metadaten lässt sich durch die Verwendung von Extraktionswerkzeugen von Drittanbietern reduzieren (Seemann/Schmalzridt/Lehmann 2001, 141). Der Prozess der Datenbereitstellung aus Flatfiles entspricht dem der Extraktion aus R/3-Systemen, bis auf die unterschiedliche Behandlung der Metadaten. Im Folgenden wird ausschließlich auf die Datenextraktion aus Flatfiles eingegangen, da in dem vom HCC betriebenen BW-System eine Anbindung an ein R/3-Quellsystem nicht realisierbar ist und daher die Quelldaten in den Fallstudien in Form von Textfiles bereitgestellt werden.

3.6.2.2 InfoSources und DataSources als logisch zusammenhängende Informationsangebote

Im Quellsystem liegen logisch zusammengehörige Daten in Form von **DataSources** vor. DataSources sind also quellsystembezogen. Sie umfassen eine Menge von Feldern, die in einer flachen Struktur (Extraktstruktur) zur Datenübertragung ins BW angeboten werden. In Form einer Auswahl an Feldern der Extraktstruktur, der Transferstruktur, werden die Daten vom Quellsystem in das BW übertragen.

Eine **InfoSource** beschreibt die Menge aller verfügbaren Daten zu einem Geschäftsvorfall oder einer Art von Geschäftsvorfällen. Eine InfoSource ist eine Einheit von logisch zusammengehörigen Informationen, d.h. von InfoObjects, und kann unter Verwendung von Übertragungsregeln Daten aus einer oder mehreren DataSources beziehen. Die Struktur der InfoSource heißt Kommunikationsstruktur. Sie ist im Gegensatz zur Transferstruktur der DataSource unabhängig von den Quellsystemen.

⁵⁸ Als Textdateiformate sind ausschließlich ASCII-Dateien mit fester Feldlänge und CSV-Dateien (comma separated values) mit variabler Feldlänge zulässig.

⁵⁹ Internetquellen sind auf der einen Seite Marktforschungsunternehmen und Wirtschaftsdatenbanken, auf der anderen Seite aufgrund ihrer Position und Marktstellung besonders relevante Unternehmen und Institutionen (Uhr/Kosilek 1999, 461).

Die **Übertragungsregeln** bestimmen, welche Felder der quellsystemabhängigen Transferstruktur in welche Felder der quellsystemunabhängigen Kommunikationsstruktur auf welche Weise übertragen werden. Dazu können detaillierte Transformationsregeln erzeugt werden:

1. Die einfachste Variante der Übertragung besteht darin, die Felder der Transferstruktur ohne Modifikation in die Kommunikationsstruktur zu schreiben.
2. Einem Feld der Kommunikationsstruktur kann ein konstanter Festwert zugeordnet werden, ohne dass die Transferstruktur berücksichtigt wird.
3. Zur gezielten Modifikation von Feldinhalten können lokale ABAP-Übertragungsroutinen angelegt und bestimmten Feldern der Kommunikationsstruktur zugewiesen werden. Hierbei stehen alle Möglichkeiten der ABAP-Programmiersprache zur Verfügung. Diese Übertragungsvariante ist besonders bei folgendem Anwendungsfall sinnvoll einsetzbar: In den Datenelementen der Merkmale-InfoObjects sind nur Großbuchstaben erlaubt, d.h. die aus den Quellsystemen extrahierten Merkmale dürfen keine Kleinbuchstaben enthalten, ansonsten bricht der ETL-Prozess mit einer entsprechenden Fehlermeldung ab. Sollen dennoch Merkmale, die Kleinbuchstaben enthalten, in eine InfoSource übertragen werden, muss innerhalb einer ABAP-Übertragungsroutine sichergestellt werden, dass die Datenelemente der Merkmale in Großbuchstaben konvertiert werden. Der in Screenshot 1 abgebildete Codeausschnitt führt die gewünschte Konvertierung aus.

Fortschreibungsregeln schließlich spezifizieren, wie die Daten (Kennzahlen, Zeitmerkmale, Merkmale) aus der Kommunikationsstruktur einer InfoSource in die endgültigen Datenziele, hier die InfoCubes, fortgeschrieben werden. Sie verbinden also eine InfoSource mit einem InfoCube oder ODS-Objekt. Bei InfoCubes gibt es zwei verschiedene Möglichkeiten, die Fortschreibungsregel für eine Kennzahl zu definieren: *Keine Fortschreibung* oder *Addition, Minimum oder Maximum*.⁶⁰ Zudem können Merkmale in externen Tabellen, z.B. einer Stammdatentabelle, nachgeschlagen werden (SAP 2001d, 5-14).

3.6.2.3 Persistent Staging Area als Zwischenablage für Daten

Die Persistent Staging Area (PSA) stellt innerhalb des BW die Eingangsablage von angeforderten Daten aus verschiedenen Quellsystemen dar. Die angeforderten Daten werden unverändert in Form der Transferstruktur in transparenten, relationalen Datenbanktabellen abgelegt und können somit auch fehlerhaft sein, wenn sie schon im Quellsystem fehlerhaft sind. Die logischen Datenpakete (Requests) können nun auf Qualität, Reihenfolge und Vollständigkeit überprüft und eventuell manuell geändert werden (SAP 2000a, 1).

3.6.3 Laden von Stammdaten

Grundsätzlich werden im SAP BW zwei Arten von Anwendungsdaten unterschieden: **Bewegungsdaten** und **Stammdaten**. Bewegungsdaten entsprechen den Datensätzen der Faktentabelle eines InfoCube, während unter Stammdaten die bei der Analyse des erweiterten Starschemas vorgestellten Segmente Attribute, Texte und Hierarchien verstanden werden. Eine Übersicht über die Datentypen im BW vermittelt Abbildung 21:

⁶⁰ Bei ODS-Datenzielen existiert darüber hinaus noch die Fortschreibungsvariante *Überschreiben* (SAP 2000b).

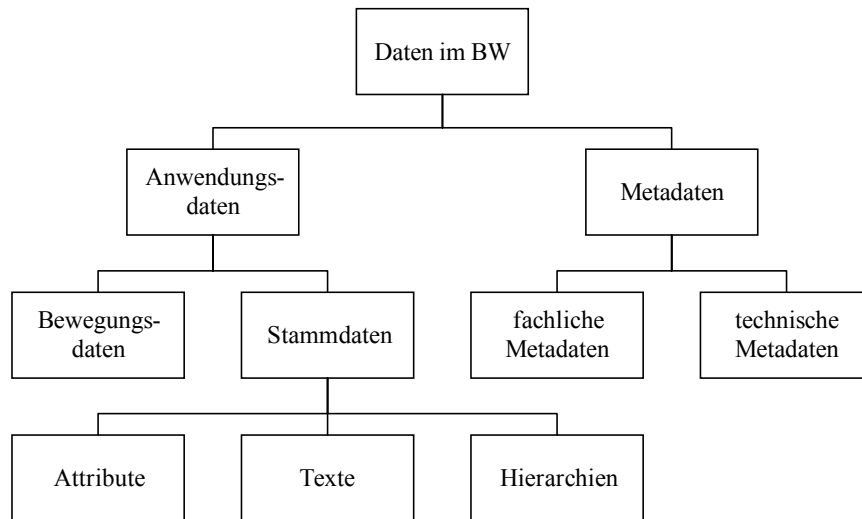


Abbildung 21: Klassifikation der Daten im BW (Quelle: Eigene Darstellung)

Die Behandlung des ETL-Prozesses von Bewegungsdaten wurde im letzten Kapitel erläutert, daher wird hier kurz auf die Unterschiede zum Laden von Stammdaten eingegangen.

Grundlegender Gedanke des erweiterten Starschemas ist es, die Stammdaten eines Merkmals getrennt von den lösungsabhängigen InfoCubes zu speichern.⁶¹ Stammdaten werden daher nicht wie Bewegungsdaten in die InfoCubes fortgeschrieben, sondern in die betreffende Stammdatentabelle des Merkmals. Die Stammdatentabellen eines Merkmals werden also von der Stammdaten-InfoSource des betreffenden Merkmals versorgt. Da Fortschreibungsregeln fest einem Datenziel zugeordnet sind, entfällt dieses Element beim Laden von Stammdaten.

3.6.4 Vom Quellsystem zum Datenziel

Der Datenfluss vom Quellsystem über DataSource, InfoSource, PSA usw. bis in das endgültige Datenziel oder die Merkmals-Stammdaten lässt sich im BW grafisch ansprechend darstellen. Die Datenflussanzeige in der AWB verdeutlicht die Zusammenhänge und Abhängigkeiten zwischen den einzelnen Objekten. Sie kann für Datenziele (InfoCubes und ODS-Objekte) und InfoSources aufgerufen werden (SAP 2000b).⁶²

Folgende Grafik fasst die wichtigsten Etappen des Datenladeprozesses von Bewegungs- und Stammdaten aus Flatfiles noch einmal zusammen:

⁶¹ Vgl. Kapitel 3.5.4.2

⁶² Beim Verändern des Zoomfaktors der Datenflussanzeige ist Vorsicht geboten. Hier kommt es nach Erfahrungen des Autors zu regelmäßigen ABAP-Kurzdumps.

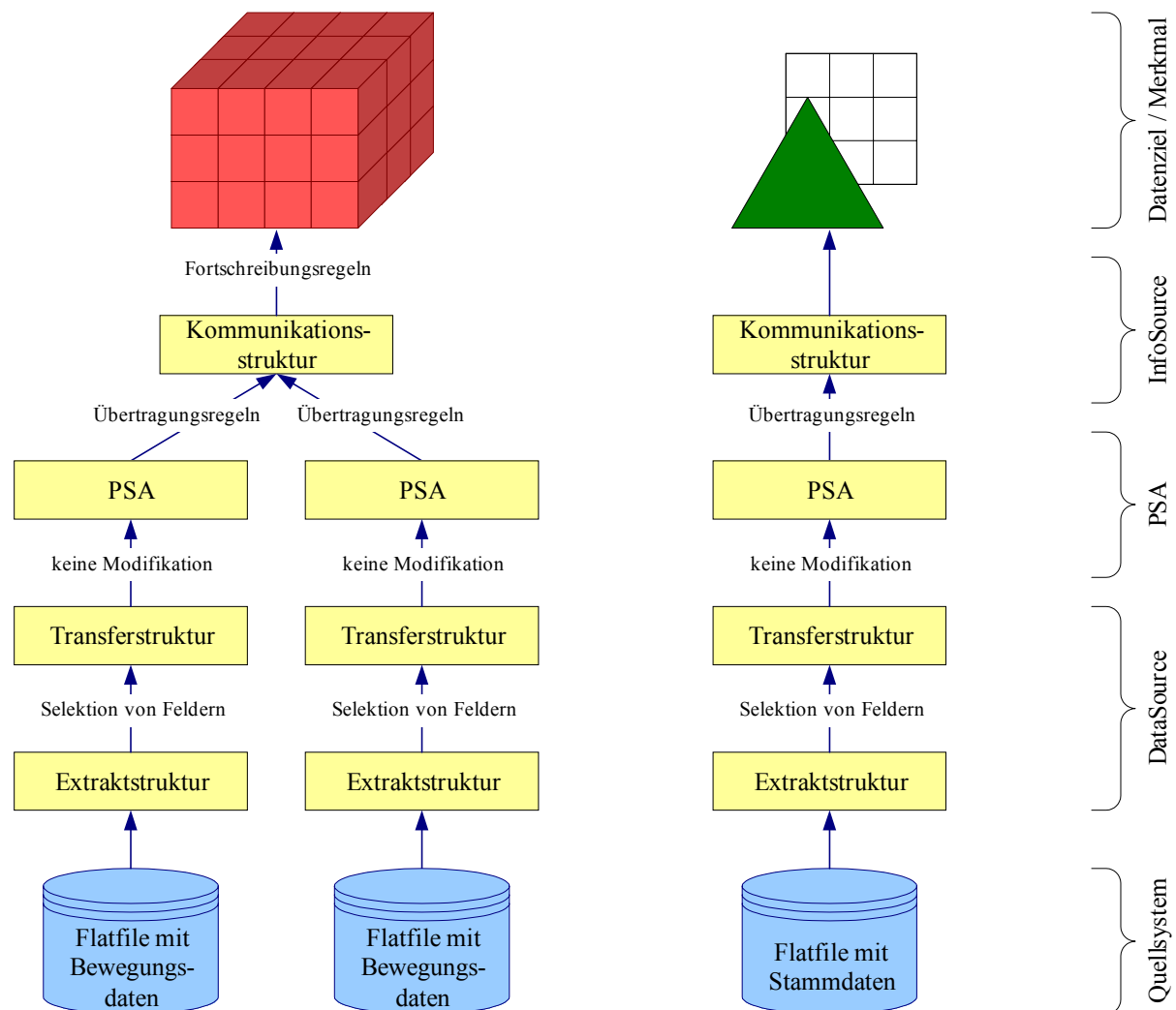


Abbildung 22: Phasen des Datenbereitstellungsprozesses aus Flatfiles (Quelle: Eigene Darstellung)

Prinzipiell ist keine feste Reihenfolge für das Vorgehen beim ETL-Prozess vorgegeben. Dennoch bietet es sich an, zuerst die Stammdaten und daraufhin die Bewegungsdaten zu laden, und zwar aus folgenden Gründen:

1. Den Merkmalsausprägungen wird beim Laden eine SID zugeordnet, auf die sich dann die Dimensionen der Faktentabelle beziehen können. Würden zuerst die Bewegungsdaten geladen, müsste diesen eine **temporäre SID** zugewiesen werden, die später wieder korrigiert würde.
2. Beim Laden der Bewegungsdaten kann **referenzielle Integrität** mit den Stammdaten sichergestellt werden. Dies bedeutet, dass Bewegungsdatensätze zurückgewiesen werden können, die auf Merkmalsausprägungen verweisen, die überhaupt nicht existieren. Dies ist natürlich nur dann möglich, wenn bereits Stammdatensätze vorhanden sind.
3. Manchmal ist es notwendig, innerhalb der Fortschreibungsregeln Stammdatenattribute aus den Merkmalen direkt in die Faktentabelle zu schreiben. Dazu müssen natürlich die Stammdaten zum Zeitpunkt des Bewegungsdatenuploads bereitstehen.

Während mit dem **Scheduler** die Möglichkeit geboten wird, Extraktionsaufträge mit Selektionskriterien (InfoPackages) per Jobplanung⁶³ auf Hintergrundprozesse zu verteilen und so die Datenbeschaffung zu automatisieren, stellt der **Monitor** ein Werkzeug zur Überwachung von Datenanforderungen dar. Die übersichtliche Auflistung aller zu analysierenden Datenanforderungen gibt schnell Aufschluss über Erfolg oder Misserfolg einer Aktion (SAP 2001d, 5-33).

Im Rahmen der **InfoCube-Administration** kann man sich einen Überblick über die bislang in den InfoCube geladenen Requests verschaffen. Jeder Request besitzt eine eindeutige Nummer, die in der systemseitig für alle InfoCubes vorgegebenen Dimension *Package* abgelegt ist. Diesen Sachverhalt verdeutlicht Screenshot 2. Aus der Faktentabelle können keine einzelnen Datensätze, sondern nur komplette Requests gelöscht werden.

3.7 Spezifikation und Entwicklung von Endbenutzeranwendungen

3.7.1 Unterschiedliche Anwender benötigen unterschiedliche Analysewerkzeuge

Neben Formen der Wissensgenerierung wie kennzahlenbasierte Führungsinformationen (EIS) und Erkennung von Informationsmustern (Data Mining) spielt die multidimensionale Analyse der Daten eines DW eine sehr große Rolle. Die dazu eingesetzten Analysewerkzeuge müssen die Anforderungen unterschiedlicher Anwendertypen berücksichtigen, damit sie von diesen akzeptiert und effizient eingesetzt werden können. Üblicherweise unterscheidet man drei Arten von Endanwendern: Konsumenten, Analysten und Autoren (Seemann/Schmalzridt/Lehmann 2001, 36).

- **Konsumenten** sind einfache Anwender des DW, die die analytischen Funktionen des Systems nur in geringem Umfang nutzen, da sie in der Regel bereits vordefinierte Queries über festgelegte Datenmengen ausführen. Diese Anwendergruppe⁶⁴ benötigt daher eine leicht bedienbare, ergonomisch gestaltete Benutzerschnittstelle (SAP 2001f, 2-12).
- Für **Analysten**, die innerhalb von Berichten navigieren und sie aus unterschiedlichen Perspektiven analysieren wollen, muss dagegen ein umfassendes und effektives Instrumentarium an analytischen Möglichkeiten zur Verfügung stehen.
- Die kleine Gruppe der **Autoren** (Queryentwickler) benötigt neben dem vollständigen Zugriff auf vordefinierte Queries und Datenzusammenstellungen die Möglichkeit, neue Queries zu erzeugen, um spezifische Probleme zu lösen (SAP 2001g, 2-8; SAP 2001f, 2-12).

⁶³ Es handelt sich hier um das R/3-eigene Jobsystem. Hier wird nochmals deutlich, dass das BW auf einem R/3-Kern aufgebaut ist.

⁶⁴ Es wird geschätzt, dass bis zu 70 % aller Anwender dieser Gruppe angehören.

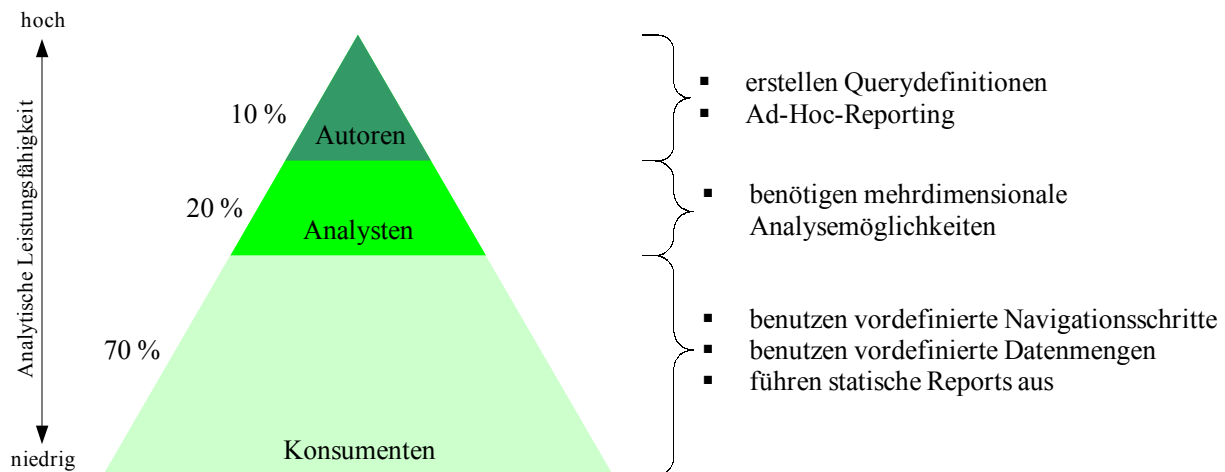


Abbildung 23: Anwendergruppen und ihre Anforderungen (Quelle: SAP 2001f, 2-12)

Das SAP BW trägt diesen unterschiedlichen Bedürfnissen Rechnung, indem diverse Reportingmöglichkeiten und -funktionalitäten bereitgestellt werden. Der Prozess der Generierung eines Berichts kann in die Phasen Konfiguration, Grobgliederung, Kreativphase und Berichtserstellung aufgeteilt werden (Barent et al. 1993).

Auf die diversen Spezialitäten des **BW-Berechtigungskonzepts** soll hier nicht weiter eingegangen werden, denn eine näher Beschreibung würde den Rahmen dieser Veranstaltung sprengen. Zudem wird vom zuständigen HCC ein Benutzerprofil für die Teilnehmer einer BW-Lehrveranstaltung vorgegeben, das die Bezeichnung *BW_TEST_USER* trägt und sämtliche notwendigen Berechtigungen in der AWB und im Reporting besitzt (HCC 2002a).

3.7.2 Möglichkeiten des Datenzugriffs im BW

Die im BW gespeicherten Daten können über eine Vielzahl an Schnittstellen präsentiert und weiterverarbeitet werden (SAP 2001c, 12ff.):

- Der **administrative Datenzugriff** für Datenbankadministratoren besteht im Prinzip aus der Möglichkeit, sich mittels der Transaktion *ListSchema* die relevanten Tabellen (Fakientabelle, Dimensionstabellen, Stammdatentabellen von Merkmalen) anzeigen zu lassen und dann mit Hilfe der integrierten Data Browser Funktionalität den Inhalt der Tabellen zu durchleuchten.
- Das BW liefert den Business Explorer als **Browser** (BEx Browser) zur Auswahl von Berichten und als **Analyzer** (BEx Analyzer) in Form eines Excel-AddIn zur Darstellung und Erzeugung von Berichten und der interaktiven Analyse von Daten. Es handelt sich hierbei um das Standard-Frontend eines BW-Anwendungsentwicklers.
- Im Rahmen des **Webreporting** ist die generische OLAP-Navigation in Webanwendungen sowie die Erstellung von Business Intelligence Cockpits für einfache bis hin zu hochindividuellen Szenarios möglich.
- Um das SAP BW auch als Datengrundlage für Auswertungswerkzeuge von Fremd-anbietern zu öffnen, wird eine **OLE-DB-for-OLAP Schnittstelle** (ODBO) angeboten.
- Mit Hilfe der **BEx Mobile Intelligence** kann der Anwender die benötigten DW-Daten auch unterwegs abrufen. Dazu werden WAP-fähige Mobiltelefone und PDAs (Personal

Digital Assistants) mit Windows CE 3.0 Betriebssystem und Pocket Internet Explorer unterstützt.

Der administrative Datenzugriff wurde bereits im Zusammenhang mit dem erweiterten Starschema⁶⁵ angesprochen. Auf die ODBO-Schnittstelle und die Anbindung mobiler Endgeräte wird in diesem Skript nicht näher eingegangen, da sie Themengebiete darstellen, die aus Zeitgründen nicht in die hier konzipierte Lehrveranstaltung integriert werden können, obwohl es sich durchaus um „schicke“ Themen handelt. Nachfolgend wird daher auf die Definition und Bearbeitung von Queries mittels BEx Analyzer, deren Darstellung im BEx Browser und abschließend kurz auf das Webreporting eingegangen.

3.7.3 Datenauswertungen mit Queries im BEx Analyzer

3.7.3.1 Wie eine Abfrage entsteht

Im BEx Analyzer werden Abfragen (Querydefinitionen) generiert, die sich stets auf genau einen InfoCube beziehen und dessen Daten in geeigneter Weise präsentieren. Der OLAP-Prozessor sorgt für die Weiterleitung der Abfragen an das BW. Dabei berücksichtigt er selbstständig evtl. existierende Aggregate. Dieser Vorgang ist für den Anwender nicht sichtbar, er bestimmt zur Erstellung einer Abfrage nur die gewünschten Informationen, nicht die Datenquellen. Soll eine Querydefinition Daten aus mehr als einem InfoCube darstellen, so sind in der logischen Modellierung MultiCubes⁶⁶ zu verwenden.

Die Querydefinition wird auf dem BW-Server gespeichert. Während die **Querydefinition** in gewisser Weise als Klassendefinition⁶⁷ verstanden werden kann, wird eine **Query** in einer Excel-Arbeitsmappe sozusagen als Instanz dieser Klasse abgelegt und mit der Mappe gespeichert.⁶⁸ Hier kann die Query dann individuell modifiziert, Navigationszustände verändert und exceleigene Funktionalitäten ausgeführt werden. Dieser individuelle Queryzustand kann als View abgespeichert werden. Die Arbeitsmappe selbst kann entweder auf dem lokalen Dateisystem des Client-PC oder zentral im Business Document Service (BDS) abgelegt werden. Folgende Abbildung zeigt den betreffenden Ausschnitt aus dem BW-Metadatenmodell:

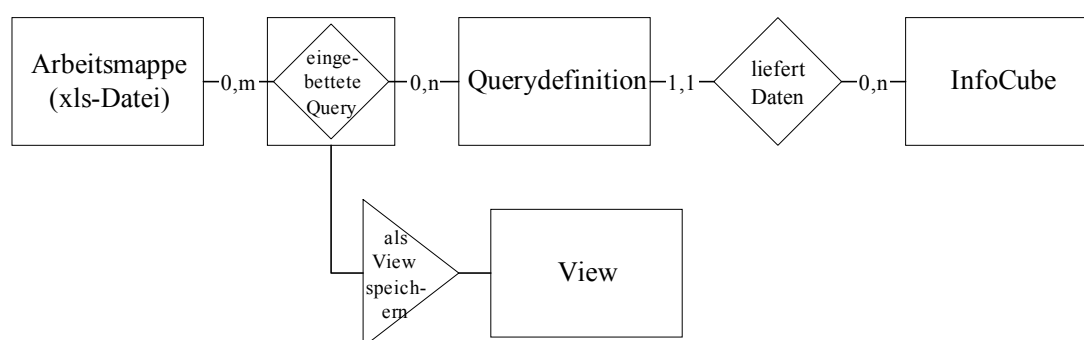


Abbildung 24: Metadatenmodell Querydefinition und Query (Quelle: Eigene Darstellung)

⁶⁵ Vgl. Kapitel 3.5.4.2

⁶⁶ Vgl. Kapitel 3.5.6.7

⁶⁷ Die SAP geht bei der Unterscheidung zwischen Querydefinition und eingebetteter Query nicht immer konsistent vor.

⁶⁸ Eine Arbeitsmappe kann mehrere Queries enthalten.

Nach der Auswahl eines InfoCube wird dem Anwender im *QueryBuilder*, dem in den BEx Analyzer integrierten Werkzeug zur **Definition** einer Abfrage, eine Liste der für die Generierung der Querydefinition zur Verfügung stehenden Komponenten angezeigt: die Kennzahlen der InfoCube-Faktentabelle und die Merkmale der InfoCube-Dimensionen. Es können auch eigene, von den Basiskennzahlen und -merkmalen abgeleitete Komponenten erzeugt werden. Die Komponenten werden per Drag & Drop in die Querydefinition integriert, indem sie in die Zeilen- oder Spaltenbereiche gesetzt werden. Nach dem Abspeichern der Querydefinition auf dem BW-Server kann sie im Excel-Arbeitsblatt ausgeführt und mit BW- und Excel-Funktionalitäten weiterbearbeitet werden. Über die durch das Aktivieren des BW-AddIns in die Excel-Arbeitsumgebung integrierte Symbolleiste kann der Anwender auf alle Funktionen des BEx Analyzers zugreifen.

3.7.3.2 Möglichkeiten der Manipulation von Kennzahlen

Liefern die Basiskennzahlen des zugrunde liegenden InfoCube nicht das gewünschte Ergebnis, können **berechnete Kennzahlen** in die Querydefinition eingebaut werden. Die Definition kann auf der Ebene der Querydefinition sowie auf InfoCube-Ebene⁶⁹ erfolgen. Auf Queryebene ist die berechnete Kennzahl nur für diese Querydefinition gültig, berechnete Kennzahlen auf Cubeebene können in allen auf dem InfoCube aufbauenden Querydefinitionen verwendet werden. Die Festlegung einer Ausnahmeaggregation kann nur für berechnete Kennzahlen erfolgen, die auf Cube-Ebene definiert wurden. Für die Berechnung von Kennzahlen können Grund-, Prozent-, Daten-, mathematische und trigonometrische Funktionen und boolsche Operatoren verwendet werden (SAP 2001f, 5-5). Die wichtigsten Funktionen für berechnete Kennzahlen sind in SAP (2002d) dokumentiert. Durch die geschickte Anwendung berechneter Kennzahlen können komplexe Analysen wie z.B. eine ABC-Analyse implementiert werden, wie in SAP (2001h) ausführlich beschrieben wird.

3.7.3.3 Flexibilisierung von Queries durch die Verwendung von Variablen

Mit Hilfe von Variablen können Queries flexibler gestaltet werden, denn Variablen werden beim Auffrischen einer Query dynamisch mit Werten gefüllt. Je nach Variablentyp und gewählter Verarbeitungsart können sie vor dem Ausführen der Query manuell eingegeben oder automatisch verarbeitet werden.

Einen Spezialfall unter den Variablen stellen die **Formelvariablen** dar, die dann eingesetzt werden, wenn ein Formelbestandteil in der Querydefinition nicht fest vorgegeben werden kann, sondern erst beim Ausführen der Query bekannt ist (SAP 2001f, 5-19). Wird die Formelvariable durch einen **Ersetzungspfad** verarbeitet, so wird sie dynamisch durch Merkmalswerte ersetzt, die durch die Angabe eines Von- und Bis-Wertes oder eines Merkmals bestimmt werden. Diese Variante, Variablen einzusetzen, ist im Zusammenhang mit der Verwendung von Kennzahlen interessant, die als Attribut zu einem Merkmal modelliert wurden. Ist diese Attributs-Kennzahl nämlich Bestandteil der Formel einer berechneten Kennzahl, sind die Werte der Attributs-Kennzahl nur unter Verwendung einer Formelvariablen mit Ersetzungspfad zugänglich. Folgende Abbildung veranschaulicht den Zusammenhang:

⁶⁹ Gemeint ist damit jedoch nicht die Modellierungsebene des InfoCube. Somit existieren drei Ebenen der Definition von Kennzahlen: 1. InfoCube, 2. Cubeebene in einer Querydefinition, 3. Querydefinition

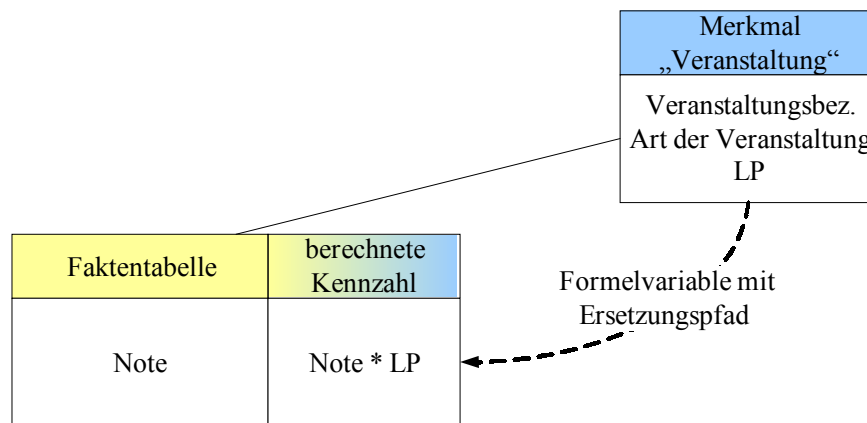


Abbildung 25: Formelvariable mit Ersetzungspfad (Quelle: Eigene Darstellung)

3.7.3.4 Excel-Funktionalitäten zur Verfeinerung von Queries

Der Anwender kann eine Query mit den aus Excel vertrauten Gestaltungsmöglichkeiten weiterbearbeiten. So kann auf Zellbereiche jede denkbare **Formatierung** angewendet werden, wobei beachtet werden muss, dass dies eine manuelle Modifikation des Arbeitsblattes darstellt, die beim nächsten Auffrischen der Query nicht mehr aktuell ist.

Zudem besteht die Möglichkeit, in der Arbeitsmappe Makros zu hinterlegen, die in der officeeigenen prozeduralen Programmiersprache **Visual Basic for Applications** (VBA) verfasst werden. Dies bietet sich dann an, wenn die gewünschte Funktionalität nicht durch Standard-Excel- oder BW-Funktionen abgedeckt wird. Der BEx Analyzer stellt VBA-Befehle für benutzerspezifische Drill Downs, Sprünge zu Views, Sortierungen und das Anbinden von Diagrammen bereit (SAP 2000e, 6ff.). Interessant ist auch die Möglichkeit, jeden OLAP-Befehl per VBA automatisieren zu können, denn OLAP-Befehle sind durch eindeutige Codes gekennzeichnet, die mit der Funktion *SAPBEXfireCommand()* aufgerufen werden können. Mit Hilfe der Logbuch-Funktion kann der Anwender diese Codes ermitteln, denn jeder ausgeführte Befehl wird im Logbuch protokolliert.

3.7.3.5 Visualisierung geographischer Daten

Zahlreiche Merkmale des BW besitzen auch geographische Bedeutung, z.B. Kunde, Verkaufsregion, Bundesland oder Land. Die geographische Information kann in der BEx Map⁷⁰ zusammen mit den betriebswirtschaftlich relevanten Kennzahlen ausgewertet werden, indem die georelevanten Daten grafisch auf einer Landkarte dargestellt werden. Über erweiterte Navigationsmöglichkeiten ("geographischer Drilldown") können regionale Bezüge in unterschiedlichen Granularitätsebenen leichter ausgewertet werden. Da eine Karte die räumliche Nachbarschaft von Orten und Regionen anzeigt, werden geographische Zusammenhänge deutlicher (SAP 2001i, 6). Für die Visualisierung von Daten mittels BEx Map sind folgende Schritte notwendig:

1. Zuerst werden die georelevanten Merkmale (z.B. Bundesland) in der InfoObject-Pflege als Geomerkmale gekennzeichnet.

⁷⁰ Die BEx Map ist das in den BEx integrierte geographische Informationssystem (GIS) des BW.

2. Daraufhin werden die Landkarten in das BW-System geladen, die in Form von **Shapefiles** vorliegen, welche danach dem betreffenden Merkmal zugeordnet werden. Shapefile ist ein gängiger Standard zur Beschreibung von Geodaten, der bei vielen Geoinformationssystemen zum Einsatz kommt. Während detaillierte Shapefiles, die auch demographische Angaben wie Sozialstruktur, Altersstruktur usw. enthalten können, unter Umständen sehr teuer sind, werden im Internet einfache Shapefiles oft kostenlos zum Download angeboten.⁷¹ Für die aktuelle Lehrveranstaltung wird ein einfaches Shapefile mit der Struktur der deutschen Bundesländer bereitgestellt.
3. Schließlich wird eine Query mit Geomerkmalen definiert und in eine Arbeitsmappe eingefügt. Nachdem eine Landkarte angebunden wurde, werden Querydaten mit geographischer Relevanz auf dieser dargestellt. Es ist möglich, auf der Landkarte zu navigieren, um die georelevanten Daten weiter auszuwerten.

3.7.3.6 Rechtzeitige Erkennung von Abweichungen mit Hilfe eines Frühwarnsystems

Potenzielle Probleme in einem Unternehmen sollen möglichst frühzeitig erkannt und behoben werden können. In regelmäßigen Abständen durchgeführte Analysen tragen zur Aufdeckung von Schwachstellen bei. Das im BW implementierte **Exception Reporting** erlaubt das Erkennen, Markieren und Hervorheben außerordentlicher Abweichungen von zuvor festgelegten Kennzahlwerten im Queryergebnis. Außerdem kann bei der Feststellung von Abweichungen eine automatische Folgeverarbeitung angestoßen werden, die aus der Anzeige auf einem Alert-Monitor und der Versendung einer E-Mail bestehen kann. Die Einstellungen für die Hintergrundverarbeitung werden im Reporting Agent hinterlegt (SAP 2001g, 3-7f.). Für BW Release 3.0 ist auch die Anbindung von Workflow-Aktivitäten geplant.

3.7.4 Einfache Informationsdistribution durch Webreporting

Das Webreporting ermöglicht es, Queries, die im BEx Analyzer definiert wurden, über das Intranet und Internet zu publizieren. Die Queries können in beliebige HTML-Seiten eingefügt und präsentiert und die Daten der Web Query durch Navigations- und OLAP-Funktionen ausgewertet werden (SAP 2001j, 6). Voraussetzung für den Einsatz des Webreporting ist ein lokal installierter Internet Transaction Server (ITS), der in Kapitel 3.4.3 beschrieben wurde.

Ein Vorteil der Verteilung von Informationen über das Inter- oder Intranet liegt neben der **ständigen** und **ubiquitären Verfügbarkeit** in der Verwendung eines **Webbrowsers** zur Darstellung der Queries. Ein Webbrowser macht umfangreiche Softwareinstallationen auf lokalen PCs überflüssig und kann einfach und intuitiv bedient werden. Zudem verfügen die meisten PC-Benutzer durch das Arbeiten mit dem Internet über Grundkenntnisse solcher Anwendungen. Die heute am Markt verfügbaren Webbrowser können selbst aufwändige Grafiken darstellen und bieten eine robuste Navigation (SAP 2001k, 2-6).

Es wird prinzipiell zwischen drei Vorgehensweisen unterschieden, mit denen ein web-basierter Zugang zum DW hergestellt werden kann:

⁷¹ Unter <http://www.cdc.gov/epiinfo/EIeurope.htm> (zugegriffen am 01.09.2002) können kostenlose Shapefiles europäischer Staaten heruntergeladen werden.

- Beim **Offline-Ansatz** werden Berichtsdaten in bestimmten zeitlichen Abständen aus dem DW abgefragt und als statische HTML-Dokumente ohne direkte Anbindung an die ursprüngliche Datenquelle auf einem Webserver abgelegt (Mucksch/Behme 2000, 60). Der Reporting Agent des BW bietet die Möglichkeit, Web Templates⁷² im Hintergrund vorzuberechnen und so den Offline-Ansatz zu realisieren.
- Beim Ansatz der **dynamisch generierten** HTML-Seiten werden die Webdokumente erst auf Anfrage des Anwenders hin erzeugt. Dieser Ansatz wird von den meisten Anbietern verfolgt (Mucksch/Behme 2000, 62), so auch von der SAP beim BW: In einem Web Template werden Platzhalter für die eigentlichen Daten untergebracht, die dann zur Laufzeit dynamisch ersetzt werden.
- Bei der Verwendung von **Java-** oder **ActiveX-Applets** erlauben es die vielfältigen Programmiermöglichkeiten, sowohl eine anspruchsvolle Benutzeroberfläche zu generieren, als auch eine direkte Verbindung zum DW-Server herzustellen (Mucksch/Behme 2000, 63). In BW Web Templates besteht die Möglichkeit, per JavaScript Befehle oder Befehlsfolgen auszuführen. Eine weitere Anwendung von JavaScript ist die Erweiterung des Kontextmenüs um kundenspezifische Einträge (SAP 2001c, 312).

Zusammenfassend betrachtet, unterstützt das BW alle drei Ansätze des webbasierten Zugriffs auf DW-Daten und bietet daher eine flexible Webreporting-Lösung. Während dieser Lehrveranstaltung wird die dynamische Variante des Webzugriffs behandelt.

Im Bereich des Internetzugriffs bieten länger am Markt vertretene Anbieter deutlich fortgeschrittenere Lösungen an. Beispiele sind Information Builders, MicroStrategy, Oracle oder SAS (Mertens/Bange/Schinzer 2000b, 17).

3.7.5 Abdeckung der Benutzeranforderungen im BW-Reporting

Dem Portalgedanken folgend, einen einheitlichen Zugang zu allen relevanten Informationen zu schaffen (single point of entry), ermöglicht der **BEx Browser** den Zugriff auf alle Dokumenttypen des BW, die der Rolle des Anwenders oder seinen Favoriten zugeordnet wurden (SAP 2001f, 3-10). Neben Queries können dort auch Arbeitsmappen, Dokumente aus dem BDS, Verknüpfungen auf das Dateisystem, URLs und Transaktionscodes des R/3-Systems hinterlegt werden (SAP 2001f, 3-12). Hervorzuheben ist, dass der BEx Browser die einzige Stelle ist, wo Verzeichnisse zur strukturierten Ablage von Arbeitsmappen angelegt werden können.

Der Ansatz zur Darstellung der Daten in einem Excel-Arbeitsblatt erlaubt es auch unerfahrenen Anwendern, vorgefertigte Analysen durchzuführen und umzugestalten. Die Komplexität der Querydefinitionen ist hinter den Zellen „versteckt“ und verwirrt den Anwender nicht. Gleichzeitig können jedoch auch fortgeschrittene Benutzer innerhalb der gleichen Oberfläche komplexe Abfragen durchführen und modifizieren, so dass das ganze Anforderungsspektrum der Benutzergruppen eines DW abgedeckt wird. Das Webreporting ist eine einfache und unkomplizierte Möglichkeit, ständig und überall Informationen für Konsumenten bereitzustellen, wobei in zukünftigen Releases von SAP BW auch die Definition von Queries im

⁷² Web Templates sind HTML-Seiten.

Webbrowser durchgeführt werden kann. Folgende Tabelle fasst zusammen, wie das BW zur Abdeckung der individuellen Benutzeranforderungen beiträgt:

Benutzergruppe	Abdeckung der Anforderungen durch...
(Informations-) Konsumenten	<ul style="list-style-type: none"> ▪ Nutzung des BEx Browser als einfaches „Portal“ ▪ Benutzung vorgefertigter Queries im BEx Analyzer ▪ Betrachtung von Webreports
Analysten	<ul style="list-style-type: none"> ▪ Nutzung der OLAP-Funktionalitäten des BEx Analyzer ▪ Nutzung der OLAP-Funktionalitäten und Navigationsfunktionen des Webreporting ▪ Nutzung und Navigation in BEx Map Landkarten ▪ Exception Reporting
Autoren	<ul style="list-style-type: none"> ▪ Anlegen neuer Querydefinitionen im QueryBuilder des BEx Analyzer ▪ Zuhilfenahme der Excel-Funktionalitäten ▪ VBA-Programmierung ▪ Erstellung von Webreports, Einbettung in HTML-Seiten, evtl. Erweiterung um JavaScript-Funktionen ▪ Anlegen neuer Querydefinitionen im webbasierten QueryBuilder (erst in zukünftigen Releases)

Tabelle 8: Abdeckung der Benutzeranforderungen im BW-Reporting (Quelle: Eigene Darstellung)

3.8 Aufgaben im Rahmen der Verteilung, Wartung und Ersetzung eines Data Warehouse

Nachdem die drei großen Bereiche Technik, Daten und Anwendungen analysiert und bearbeitet wurden, ist es nun an der Zeit, das DW unter den Anwendern zu **verteilen** und sie mit dem entwickelten Produkt vertraut zu machen. Dazu gehören die Vorbereitung der technischen Umgebung beim Nutzer, eine umfassende Anwenderschulung und die Entwicklung einer Supportstrategie im Rahmen eines Roll-Out, das diese Aktivitäten integriert und professionell durchführt (Kimball et al. 1998, 691ff.).

Genau so wichtig wie die technische Ausstattung des Backend (Hardwareplattform der diversen Server, Datenbankmanagementsystem, Infrastruktur usw.) ist die adäquate **Ausstattung der Arbeitsplätze** der Anwender. Neben der Funktionstüchtigkeit der Hardware muss die Verbindung zum DW-Server gesichert sein, der Zugang zum BW-System muss durch das Anlegen entsprechender Logons und Vergabe von Kennwörtern gewährleistet werden. Es bietet sich an, diese Tätigkeiten genau zu planen und sorgfältig auszuführen, denn sie beziehen sich unmittelbar auf den sensibelsten Faktor in einem DW-Projekt: den Anwender (Kimball et al. 1998, 692ff.).

Das raffinierteste Datenmodell und die ausgefeiltesten Anwendungen bleiben im Verborgenen, wenn der Anwender sie nicht kennt oder sie nicht bedienen kann. Eine erfolgreiche Einführung des DW umfasst fortlaufende, gut konzipierte und implementierte **Schulungsprogramme** für die Benutzergemeinde. Hierbei müssen Themen wie Einführung in das DW-Konzept, Dateninhalte des DW, praktischer Umgang mit Endbenutzeranwendungen und Datenzugriffsmöglichkeiten angesprochen werden. Optimalerweise wird in solchen Schul-

ungen mit Datenmaterial und Berichten gearbeitet, die der Anwender bereits aus seiner täglichen Arbeit kennt (Poe/Reeves 2000, 202).

Wenn die Anwender das DW benutzen, werden auch die Fragen dazu „sprudeln“. Um den Erfolg des Systems zu gewährleisten, muss daher eine **Supportstruktur** entwickelt und implementiert werden. Das Supportteam muss auf Fragen zu den Daten selbst, den Werkzeugen, dem Einloggen, dem Arbeiten mit Berichten und darüber hinausgehenden Themen kompetent Antwort leisten können. Eine einfache Supportmethode besteht in der Etablierung oder Erweiterung eines bestehenden Help Desk. Ein entscheidender Erfolgsfaktor ist, dass das Supportteam möglichst das gesamte inhaltliche Spektrum der auftretenden Fragen abdeckt, von technischen Problemen bis hin zu betriebswirtschaftlichen Fragestellungen (Poe/Reeves 2000, 203).

Eine mögliche Ergänzung des Support besteht darin, eine **Data Warehouse Webseite** zu veröffentlichen, auf der das neue System vorgestellt wird, Standardberichte erläutert werden und eventuell Trainingseinheiten zu bestimmten Themengebieten angeboten werden. Es ist ebenso denkbar, eine Data Warehouse **Community** zu integrieren, um den Anwendern den Kontakt untereinander und evtl. mit dem Entwicklerteam zu ermöglichen (Kimball et al. 1998, 703ff.). Diese Webseite ist zudem eine interessante Möglichkeit, internes **Marketing** für das Data Warehouse zu betreiben. Wird das DW auch nicht frei vermarktet, so muss es dennoch intern „verkauft“ werden, denn schließlich stellt es für das Unternehmen eine erhebliche Investition dar. Dieses Marketing kann durch das Ansprechen wichtiger Gruppen im Unternehmen praktiziert werden. Wenn die technischen Experten und diejenigen mit dem größten Einfluss auf andere Mitarbeiter das Produkt akzeptieren, ist der Erfolg gesichert. Auch eine klar sichtbare Unterstützung durch die Geschäftsleitung sorgt für positive Akzeptanz (Poe/Reeves 2000, 202ff.).

Nach der Verteilung des DW ist es keinesfalls an der Zeit, sich zurückzulehnen, abzuwarten und die Früchte der Arbeit zu genießen: Im Rahmen der **Wartung** eines DW sind vielfältige Aufgaben zu erledigen.

- Der Kontakt zu den Anwendern muss durch ständige Gespräche, fortführenden Support und weiterführende oder wiederholende Schulungsangebote aufrechterhalten werden, was durch die Ausgabe eines Newsletters und die Aktualisierungen der Webseite angekündigt und ergänzt werden kann.
- Die technische Infrastruktur muss ständig gepflegt und evtl. erneuert werden, um Performance zu gewährleisten und Systemausfälle zu vermeiden.
- Das System muss für ein eventuelles Wachstum gerüstet werden, was ein Zeichen von Erfolg und Akzeptanz darstellt.

Auch ein erfolgreiches und ständig genutztes Data Warehouse erreicht einmal das Ende seines Lebenszyklus. Es stellt sich die Frage, wann sich dieser Zustand einstellt und wie die **Abschaffung** eines solchen Systems organisiert wird. Es muss die Entscheidung getroffen werden, zu welchem Zeitpunkt ein System abgeschafft und evtl. durch ein neues ersetzt wird. Auch nach dem produktiven Betrieb kann ein System noch Umstellungskosten und remanente Lizenzkosten verursachen. Da die Nutzungsdauer vom gestifteten Nutzen des Systems abhängt, ist zunächst dieser zu ermitteln (Krcmar 2000, 113). Eine Möglichkeit be-

steht in der Anwendung von Investitionsrechnungen zur Ermittlung der optimalen Laufzeit oder des optimalen Ersatzzeitpunktes eines Investitionsprojekts (Troßmann 1998, 507ff.). Eine weitere Vorgehensweise ist die Bestimmung der Phase des Lebenszyklus, in der sich ein Anwendungssystem befindet. Hinweise darauf geben Art und Umfang von Nutzung, Benutzerakzeptanz, Nutzungsprognose, Geschäftsprozessabdeckung, Konkurrenzprodukten, Adaptierbarkeit, Zukunftsfähigkeit, Wartungskosten usw. Die Quantifizierung dieser Kriterien ist in vielen Fällen so schwierig, dass eine systematische Entscheidung zur Ablösung eines Systems oft nicht getroffen werden kann (Krcmar 2000, 113f.).

4 Ist das SAP BW ein klassisches Data Warehouse ?

Häufig wird dem Walldorfer Softwarehersteller der Vorwurf gemacht, das Business Information Warehouse sei kein „richtiges“ Data Warehouse und es stelle keine Konkurrenz zu etablierten Anbietern wie Oracle oder SAS dar. Ist das BW nun ein klassisches Data Warehouse oder nicht ?

Legt man die in diesem Lehrmaterial vorgestellte **erweiterte DW-Definition**⁷³ für die Beurteilung dieser Frage zugrunde, so ist diese eindeutig mit Ja zu beantworten. Das SAP BW bietet sämtliche Funktionalitäten zur themenbezogenen, integrierten, zeitbezogenen und dauerhaften Sammlung unternehmensrelevanter Daten in logisch multidimensionalen Strukturen, wie in den letzten Kapiteln deutlich wurde. Zudem liefert es Mechanismen zur Bereitstellung von Daten, welche die Extraktion aus diversen Quellsystemen sowie individuelle Transformationsmöglichkeiten umfassen. Ebenso sind Werkzeuge zur Analyse und Darstellung der Informationen im BW enthalten. Der Vorteil oder auch gerade der Nachteil des BW ist diese Konzeption als „eierlegende Wollmilchsau“. SAP hat mit dem BW ein Produkt entwickelt, das versucht, alle Bereiche des Data Warehousing abzudecken und eventuell Gefahr läuft, die Breite auf Kosten der Tiefe überzubewerten.

Das BW ist eine eigenständige Data Warehouse Lösung, die theoretisch nicht vom Kernprodukt R/3 abhängig ist. Faktisch basiert aber der Großteil der BW-Strukturen, speziell des Business Content, auf den R/3-Geschäftsprozessen. Seine volle Leistungsfähigkeit und einen relativ zu anderen DWs geringeren Implementierungsaufwand erzielt das BW daher nur mit R/3-Quellsystemen. Im Bereich der Anbindung anderer Quellsysteme (inklusive SAP R/3) bieten viele andere DW-Werkzeuge deutlich umfangreichere Funktionalitäten (Mertens/Bange/Schinker 2000b, 16). In die „Phalanx der weltweiten Top Ten Anbieter von Business Intelligence Lösungen“ wird SAP dennoch eindringen: das BW wird sich auf Grund der weiten Verbreitung des ERP-Systems R/3 erfolgreich in den vorderen Plätzen einordnen (Schinker/ Bange/Mertens 2000, 10). Somit ist das BW ein Data Warehouse, das seinen Erfolg nicht unbedingt nur der eigenen Leistungsfähigkeit verdankt, sondern auch seiner Nähe zum „großen Bruder“ R/3.

⁷³ Vgl. Kapitel 2.1

5 Literaturverzeichnis

- Bange, C.; Schinzer, H. (o. J.):** Data Warehouse und Business Intelligence – Grundlagen entscheidungsorientierter Informationssysteme. In: <http://www.competence-site.de/bisysteme.nsf>, zugegriffen am 26.07.2002.
- Barent, V.; Scheubrein R.; Krcmar, H.; Habenicht, W. (1993):** Teamorientierte Berichtserstellung für Führungskräfte: Problemanalyse und Systemkonzeption. Hrsg.: Habenicht, W., Universität Hohenheim, Inst. für Betriebswirtschaftslehre, Stuttgart 1993.
- Becker, J.; Priemer, J.; Wild, R. (1994):** Modellierung und Speicherung aggregierter Daten. In: Wirtschaftsinformatik 36 (1994) 5, S. 422-433.
- Behme, W. (1996):** Business Intelligence als Baustein des Geschäftserfolgs. In: Das Data Warehouse-Konzept – Architektur-Datenmodelle-Anwendungen. Hrsg.: Mucksch, H. u. Behme, W., 1. Aufl., Gabler, Wiesbaden 1996, S. 27-46.
- Behme, W.; Holthuis, J.; Mucksch, H. (2000):** Umsetzung multidimensionaler Strukturen. In: Das Data Warehouse-Konzept – Architektur-Datenmodelle-Anwendungen. Hrsg.: Mucksch, H. u. Behme, W., 4., vollst. überarb. und erw. Aufl., Gabler, Wiesbaden 2000, S. 215-242.
- Behme, W.; Mucksch, H. (1996):** Die Notwendigkeit einer unternehmensweiten Informationslogistik zur Verbesserung der Qualität von Entscheidungen. In: Das Data Warehouse-Konzept – Architektur-Datenmodelle-Anwendungen. Hrsg.: Mucksch, H. u. Behme, W., 1. Aufl., Gabler, Wiesbaden 1996, S. 3-26.
- Böhnlein, M.; Ulbrich-vom Ende, A. (o. J.):** Semantisches Data Warehouse-Modell (SDMW) - Ein konzeptuelles Modell für die Erstellung multidimensionaler Datenstrukturen. In: <http://pda15.seda.sowi.uni-bamberg.de/ceus/cver.asp>, zugegriffen am 01.08.2002.
- Bose, I.; Mahapatra, R. K. (2001):** Business Data Mining – a machine learning perspective. In: Information & Management, Vol. 39 Issue 3 (2001), S. 211-226.
- Frie, T. (2000):** Data Warehouse-Architekturkonzept und Migrationspfade bei den Winterthur Versicherungen. In: <http://www.unisg.ch>, zugegriffen am 31.07.2002.
- Frie, T.; Strauch, B. (1999):** Kriterienkatalog für Metadatenmanagement-Werkzeuge. Bericht Nr. BE HSG/CC DWS/03, St. Gallen 1999.
- Gabriel, R.; Gluchowski, P. (1998):** Grafische Notationen für die semantische Modellierung multidimensionaler Datenstrukturen in Management Support Systemen. In: Wirtschaftsinformatik 40 (1998) 6, S. 493-502.
- HCC (2002a):** Richtlinien zum Umgang mit dem Arbeiten mit dem SAP BW. Passau 2002.
- Hecht, H.; Bange, C.; Schinzer, H. (2000):** Warehouse-Konfektion von der Stange ? In: is report, 8/2000 (4. Jg.), S. 40-44.
- Holthuis, J. (2000):** Grundüberlegungen für die Modellierung einer Data Warehouse-Datenbasis. In: Das Data Warehouse-Konzept – Architektur-Datenmodelle-

- Anwendungen. Hrsg.: Mucksch, H. u. Behme, W., 4., vollst. überarb. und erw. Aufl., Gabler, Wiesbaden 2000, S. 149-188.
- Inmon, B. (1999):** Star Joins. In: <http://www.billinmon.com>, zugegriffen am 31.07.2002.
- Inmon, B. (2001a):** Definition of a data warehouse. In: www.billinmon.com, zugegriffen am 31.07.2002
- Inmon, B. (2001b):** Budgeting for the data warehouse. In: www.billinmon.com, zugegriffen am 31.07.2002
- Inmon, B. (2002):** Surrogate Keys. In: <http://www.billinmon.com>, zugegriffen am 31.07.2002.
- Jahnke, B.; Groffmann, D.; Kruppa, S. (1996):** On-Line Analytical Processing (OLAP). In: *Wirtschaftsinformatik* 38 (1996) 3, S. 321-324.
- Kimball, R.; Reeves, L.; Ross, M.; Thornthwaite, W. (1998):** The Data Warehouse Lifecycle Toolkit – Expert Methods for Designing, Developing, and Deploying Data Warehouses. 1. Aufl., Wiley, NewYork 1998.
- Kimball, R.; Ross, M. (2002):** The Data Warehouse Toolkit – The complete Guide to Dimensional Modeling. 2nd edition, Wiley, New York 2002.
- Krcmar, H. (2000):** Informationsmanagement. 2., verb. Aufl., Springer, Berlin u. Heidelberg 2000.
- Küpper, H. (1997):** Controlling. 2., akt. und erg. Aufl., Schäffer-Poeschel, Stuttgart 1997.
- Lusti, M. (2002):** Data Warehousing und Data Mining: Eine Einführung in entscheidungsunterstützende Systeme. 2., überarb. und erw. Aufl., Springer, Berlin u. Heidelberg 2002.
- Maier, R. (1998):** Nutzen und Qualität der Datenmodellierung – Ergebnisse einer empirischen Studie. In: *Wirtschaftsinformatik* 40 (1998) 2, S. 130-140.
- Manning, I. (2002):** Data Warehousing - What is it. In: <http://www.bettermanagement.com>, zugegriffen am 7.5.2002.
- Mertens, H.; Bange, C.; Schinzer, H. (2000a):** Data-Warehouse-Szenarien – Produktbewertung im Überblick. In: *is report*, 6/2000 (4. Jg.), S. 34-39.
- Mertens, H.; Bange, C.; Schinzer, H. (2000b):** Data Warehouse – wie innovativ ist SAP BW ? In: *is report*, 5/2000 (4.Jg.). In: <http://www.competence-site.de>, zugegriffen am 26.07.2002.
- Michel, R. (1999):** What makes BW tick ? In: <http://www.manufacturingsystem.com>, zugegriffen am 01.03.2002.
- Mucksch, H.; Behme, W. (2000):** Das Data Warehouse-Konzept als Basis einer unternehmensweiten Informationslogistik. In: *Das Data Warehouse-Konzept – Architektur-Datenmodelle-Anwendungen*. Hrsg.: Mucksch, H. u. Behme, W., 4., vollst. überarb. und erw. Aufl., Gabler, Wiesbaden 2000, S. 3-82.
- Poe, V.; Reeves, L. (2000):** Aufbau eines Data Warehouse. Prentice Hall, München u.a. 2000.

- Reiser, M.; Holthuis, J. (1996):** Nutzenpotentiale des Data-Warehouse-Konzepts. In: Das Data Warehouse-Konzept – Architektur-Datenmodelle-Anwendungen. Hrsg.: Mucksch, H. u. Behme, W., 1. Aufl., Gabler, Wiesbaden 1996, S. 117-132.
- SAP (2000a):** Staging Szenarien. Business Information Warehouse 2.0B. Version 1.0. In: <http://service.sap.com/bw>, zugegriffen am 20.07.2002.
- SAP (2000b):** SAP-Bibliothek: Business Information Warehouse Release 2.1C, Dezember 2000, SAP Basis Release 4.6D, Juni 2000.
- SAP (2000c):** Multi-Dimensional Modeling with BW - ASAP for BW Accelerator - A background of the techniques used to create SAP BW InfoCubes Document Version 2.0. In: <http://service.sap.com/bw>, zugegriffen am 01.08.2002.
- SAP (2000d):** Performance Tuning for Queries with Aggregates - ASAP for BW Accelerator. In: <http://service.sap.com>, zugegriffen am 01.08.2002.
- SAP (2000e):** BW and Excel Features – How to Guide. In: <http://service.sap.com>, zugegriffen am 01.07.2002.
- SAP (2001a):** BW-Glossar. In: <http://service.sap.com>, zugegriffen am 01.08.2002.
- SAP (2001b):** Administrator Workbench. Release 30A. In: <http://service.sap.com/bw>, zugegriffen am 01.08.2002.
- SAP (2001c):** Business Explorer. Release 30A. In: <http://service.sap.com/bw>, zugegriffen am 01.08.2002.
- SAP (2001d):** BW210 – Business Information Warehouse (BW) - Warehouse Management (SAP-Schulungsordner). Release 2.0B/2.1C, Collection 13.
- SAP (2001e):** SAP@Web Installation Guide. Release 6.10. In: <http://service.sap.com>, zugegriffen am 10.08.2002.
- SAP (2001f):** BW205 – Business Information Warehouse – Analyse (SAP-Schulungsordner). Release 2.0B.
- SAP (2001g):** BW207 – Business Information Warehouse – Reporting Management (SAP-Schulungsordner). Release 2.0B.
- SAP (2001h):** How to... Realize an ABC Analysis – ASAP How to Paper. In: <http://service.sap.com>, zugegriffen am 01.09.2002.
- SAP (2001i):** BEx Map (incl. Geocodierung) – Release 20B. In: <http://service.sap.com>, zugegriffen am 01.08.2002.
- SAP (2001j):** Web Reporting - Update der Online-Dokumentation 2.1C. Release 2.0B. In: <http://service.sap.com>, zugegriffen am 01.08.2002.
- SAP (2001k):** BW209 – BW – Web Reporting (SAP-Schulungsordner). Release 2.0B.
- SAP (2001l):** Aufgaben für die Systemverwaltung. Release 3.0A. In: <http://service.sap.com>, zugegriffen am 01.08.2002.
- SAP (2002a):** Hinweis Nr. 321973. In: <http://service.sap.comhttps://websmp202.sap-ag.de/notes>, zugegriffen am 01.08.2002.

- SAP (2002b):** Hinweis Nr. 166130. In: <http://service.sap.comhttps://websmp202.sap-ag.de/notes>, zugegriffen am 10.08.2002.
- SAP (2002c):** Hinweis Nr. 338532. In: <http://service.sap.comhttps://websmp202.sap-ag.de/notes>, zugegriffen am 01.08.2002.
- SAP (2002d):** Formeloperatoren in BW2.0B. Version 1.0. In: <http://service.sap.com>, zugegriffen am 01.09.2002.
- Scheer, A.-W. (1998):** Wirtschaftsinformatik – Referenzmodelle für industrielle Geschäftsprozesse. Studienausgabe 2. Aufl., Springer, Berlin u.a. 1998.
- Schinzer, H.; Bange, C.; Mertens, H. (1999):** Data warehouse und Data mining: marktführende Produkte im Vergleich. 2. Auflage, Vahlen, München 1999.
- Schinzer, H.; Bange, C.; Mertens, H. (2000):** Wachstum, Trends und gute Produkte - Neue BARC-Studie zum OLAP- und Business Intelligence-Markt. In: is report, 1/2000 (4. Jg.), S. 10-17.
- Schwarz, S. (1999):** Organisationskonzepte im Data Warehousing - Bericht Nr. BE HSG/CC DWS/02, St. Gallen 1999.
- Seemann, A.; Schmalzridt, B.; Lehmann, P. (2001):** SAP® Business Information Warehouse. 1. Aufl., Galileo Press, Bonn 2001.
- Sexl, S.; Bange, C. (2002):** Hat Business Intelligence als Dachbegriff versagt ? In: is report, 4/2002 (6. Jg.), S. 30-32.
- Totok, A. (2000):** Grafische Notationen für die semantische multidimensionale Modellierung. In: Das Data Warehouse-Konzept – Architektur-Datenmodelle-Anwendungen. Hrsg.: Mucksch, H. u. Behme, W., 4., vollst. überarb. und erw. Aufl., Gabler, Wiesbaden 2000, S. 189-214.
- Troßmann (1998):** Investition. Lucius & Lucius, Stuttgart 1998.
- Uhr, W.; Kosilek, E. (1999):** Internet-Quellen zur Integration wirtschaftsrelevanter unternehmensexterner Daten in Management Support Systems. In: Wirtschaftsinformatik 41 (1999) 5, S. 461-466.
- Watson, H. J. (2001):** Current Practices in Data Warehousing. In: Information Systems Management, Vol. 18 Issue 1 (2001), S. 47-56.

I Anhang

Screenshots

Screenshot 1: ABAP-Routine zur Buchstabenkonvertierung (© SAP AG)

Screenshot 2: Request-Administration im InfoCube (© SAP AG)

Metadatenmodell des SAP BW

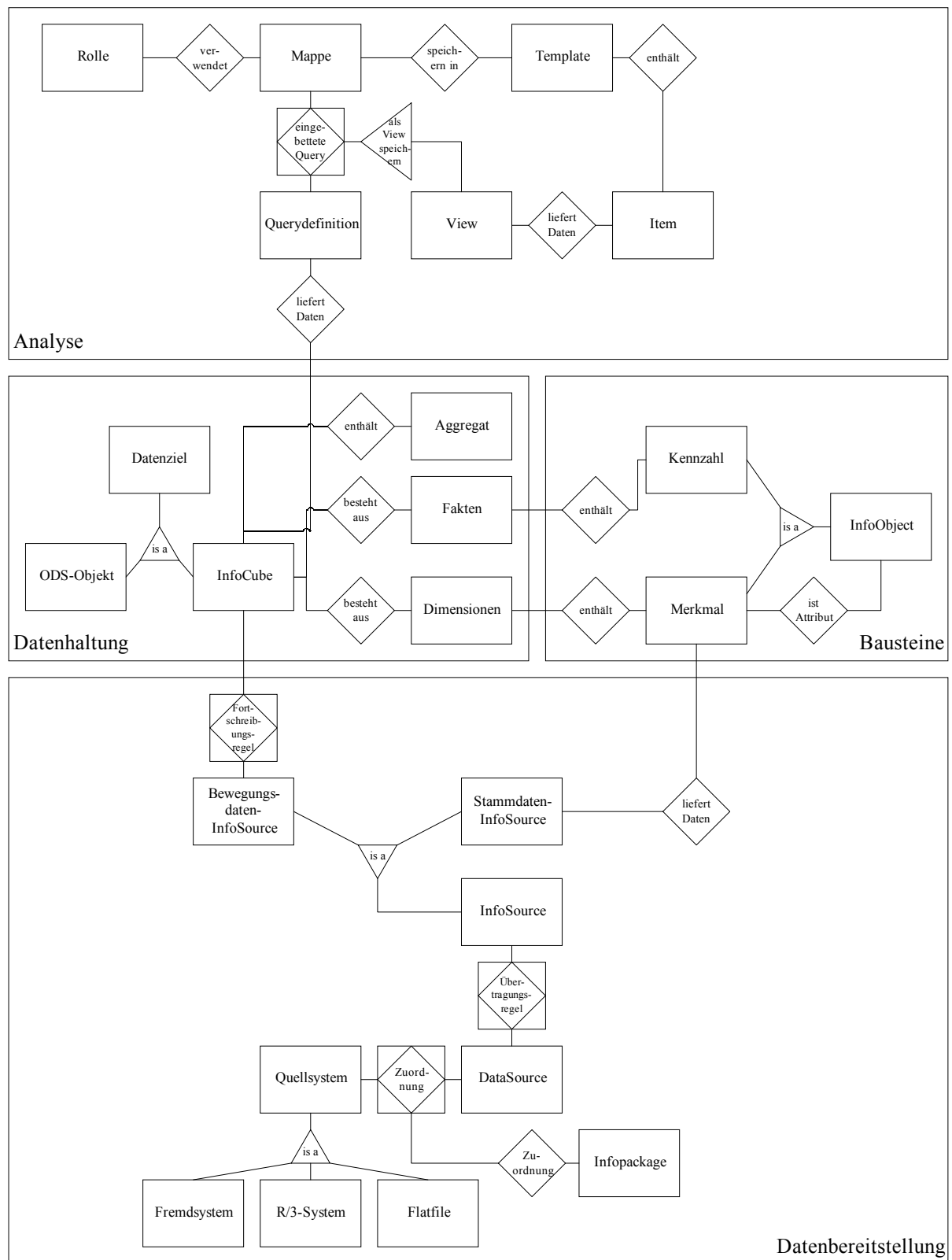


Abbildung 26: Metadatenmodell des SAP BW (Quelle: eigene Darstellung)

Internetquellen

Inhalt	URL
Große Linksammlung zu OLAP und weiteren DW-Themen	http://olap.winf.ruhr-uni-bochum.de/
SAP-Hilfeportal zum BW	http://help.sap.com → SAP Portals → Business Information Warehouse
SAP BW Dokumentationen der SAP AG	http://service.sap.com/bw
Data Warehouse Infocenter	http://www.dwinfocenter.org/
Bill Inmons Webseite <ul style="list-style-type: none"> ▪ Unmengen von Artikeln rund um Data Warehousing ▪ Datenmodelle zum Downloaden 	http://www.billinmon.com
Website der Firma <i>Symmetry Corporation</i> , die den kostenlosen Download einer Visio-Schablone für die Datenmodellierung mit ADAPT anbietet.	http://www.symcorp.com
Website des Projekts <i>System 42</i> , ein Grundlagenprojekt im Bereich Data Warehousing	http://www.forwiss.tu-muenchen.de/~system42/
tse technologieberatung und systementwicklung – SAP-Arbeitspapiere	http://www.tse-hamburg.de/Papers/SAP/Navigation/SAP.html
Jacta Spatial Business Intelligence - Typische Auswertungen und Analysen (Geoauswertungstools)	http://www.jacta.de/solution.htm
Hochschulinformationssystem CEUS	http://pda15.seda.sowi.uni-bamberg.de/CEUS/
Projekt <i>Mistral</i> an der TU München	http://mistral.in.tum.de/
Kostenloser Download von Shapefiles für europäische Staaten	http://www.cdc.gov/epiinfo/Eleurope.htm