

SAP Business Information Warehouse

Mehrdimensionale Datenmodellierung

Michael Hahne

SAP Business Information Warehouse

Mehrdimensionale Datenmodellierung

Mit 120 Abbildungen und 19 Tabellen

 Springer

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

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

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

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

Springer ist ein Unternehmen von Springer Science+Business Media

springer.de

© Springer-Verlag Berlin Heidelberg 2005
Printed in Germany

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

Umschlaggestaltung: Erich Kirchner, Heidelberg

SPIN 11008743

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

Vorwort

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

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

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

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

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

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

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

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

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

Michael Hahne

Inhaltsverzeichnis

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

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

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

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

1 Einleitung

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

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

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

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

² Die Gruppierung und Begriffsbildung unterliegt häufiger Änderungen von Seiten der SAP AG.

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

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

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

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

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

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

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

2 Analyseorientierte Informationssysteme

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

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

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

2.1 Management Support Systeme

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

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

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

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

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

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

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

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

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

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

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

Abb. 1. Klassifizierung von Management Support Systemen

2.2 Data Warehouse

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

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

- Themenorientierung

- Vereinheitlichung
- Zeitorientierung
- Beständigkeit

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

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

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

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

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

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

erstmaligen Befüllens sowie der regelmäßigen periodischen Aktualisierungen zu unterscheiden.

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

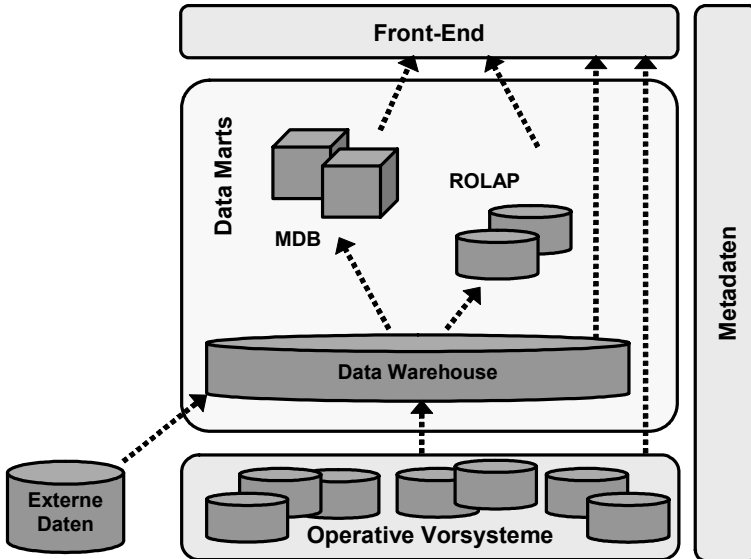


Abb. 2. Data Warehouse Referenzarchitektur

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

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

2.3 OLAP und mehrdimensionale Datenbanken

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

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

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

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

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

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

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

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

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

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

2.4 Weiterführende Literatur

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

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

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

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

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

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

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

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

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

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

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

2.5 Zusammenfassung

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

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

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

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

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

3 Datenmodellierung und Mehrdimensionalität

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

3.1 Datenmodelle und Datenmodellierung

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

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

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

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

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

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

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

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