

TDWI 05 Deutschland

Data Modelling – Corporate Performance Management auf Basis von Data Warehouses

PD Dr. Peter Gluchowski

Dr. Michael Hahne



15. Juni 2005

Agenda

- 1) Abgrenzung und Einordnung von CPM und BI
- 2) Architekturen und Komponenten analyseorientierter Informationssysteme
- 3) Bestandteile multidimensionaler Datenstrukturen
- 4) Methoden der semantischen Modellierung mehrdimensionaler Datenstrukturen
- 5) Umsetzung mit multidimensionalen Datenbanksystemen
- 6) Varianten und Bestandteile des Star Schemas
- 7) Erweitertes Star Schema des SAP Business Information Warehouse
- 8) Mehrdimensionales Datenmodell des Microsoft SQL Server
- 9) Zusammenfassende Gegenüberstellung relationaler und multidimensionaler Modellierung

Kontakt:

PD Dr. Peter Gluchowski
Brauckmanns Knapp 8c
58313 Herdecke
Germany

Email: gluchows@uni-duesseldorf.de
URL: <http://www.uni-duesseldorf.de/HHU/fakultaeten/wiwi/lehrstuehle/fachgebStatOekon/Personalia/Gluchowski>

Dr. Michael Hahne
Freiherr-vom-Stein-Str. 13a
55559 Bretzenheim
Germany

Email: michael@hahneonline.de
URL: www.hahneonline.de

Agenda

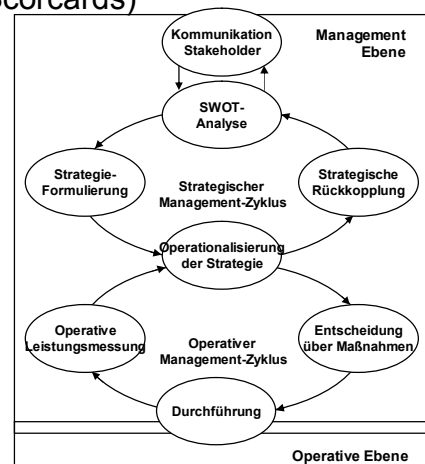
- 1) **Abgrenzung und Einordnung von CPM und BI**
- 2) Architekturen und Komponenten analyseorientierter Informationssysteme
- 3) Bestandteile multidimensionaler Datenstrukturen
- 4) Methoden der semantischen Modellierung mehrdimensionaler Datenstrukturen
- 5) Umsetzung mit multidimensionalen Datenbanksystemen
- 6) Varianten und Bestandteile des Star Schemas
- 7) Erweitertes Star Schema des SAP Business Information Warehouse
- 8) Mehrdimensionales Datenmodell des Microsoft SQL Server
- 9) Zusammenfassende Gegenüberstellung relationaler und multidimensionaler Modellierung

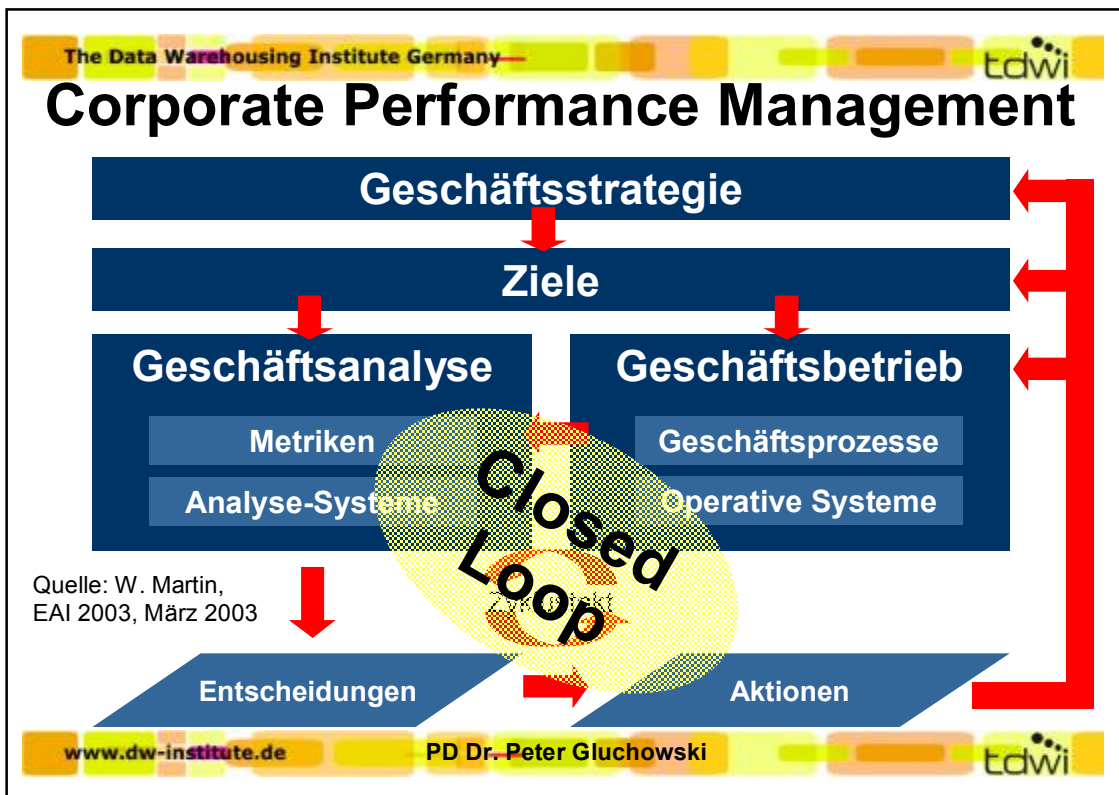
Corporate Performance Management

Corporate Performance Management umfasst

- **Methodologien** (z. B. Balanced Scorecards)
- **Prozesse** (z. B. Planungs- und Budgetierungsprozesse)
- **Metriken** (Performance-Kennzahlen, z. B. ROI) und
- **Systemlösungen** (z. B. Business Intelligence)

zur zielgerichteten Umsetzung von **Unternehmensstrategien**.





The Data Warehousing Institute Germany — tdwi

Business Intelligence

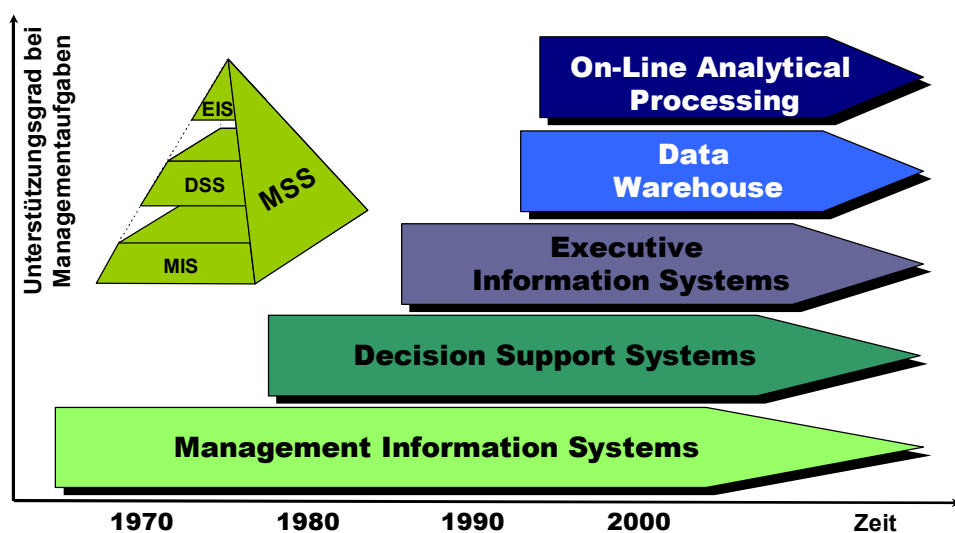
- Gesamtheit aller **Werkzeuge** und **Anwendungen** mit entscheidungsunterstützendem Charakter, die zur besseren **Einsicht in das eigene Geschäft** und damit zum besseren **Verständnis in die Mechanismen relevanter Wirkungsketten**
- **Begriffliche Klammer**, die eine Vielzahl unterschiedlicher Ansätze zur **Analyse geschäftsrelevanter Daten** zu bündeln versucht (kein neues Konzept / Produkt)

www.dw-institute.de PD Dr. Peter Gluchowski tdwi

Agenda

- 1) Abgrenzung und Einordnung von CPM und BI
- 2) **Architekturen und Komponenten analyseorientierter Informationssysteme**
- 3) Bestandteile multidimensionaler Datenstrukturen
- 4) Methoden der semantischen Modellierung mehrdimensionaler Datenstrukturen
- 5) Umsetzung mit multidimensionalen Datenbanksystemen
- 6) Varianten und Bestandteile des Star Schemas
- 7) Erweitertes Star Schema des SAP Business Information Warehouse
- 8) Mehrdimensionales Datenmodell des Microsoft SQL Server
- 9) Zusammenfassende Gegenüberstellung relationaler und multidimensionaler Modellierung

Historie der BI-Systeme



Data Warehouse-Konzept

Data Warehouse (DW) =

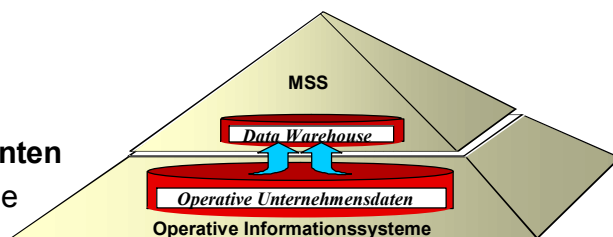
"A subject oriented, integrated, nonvolatile, time variant collection of data organized to support management needs"

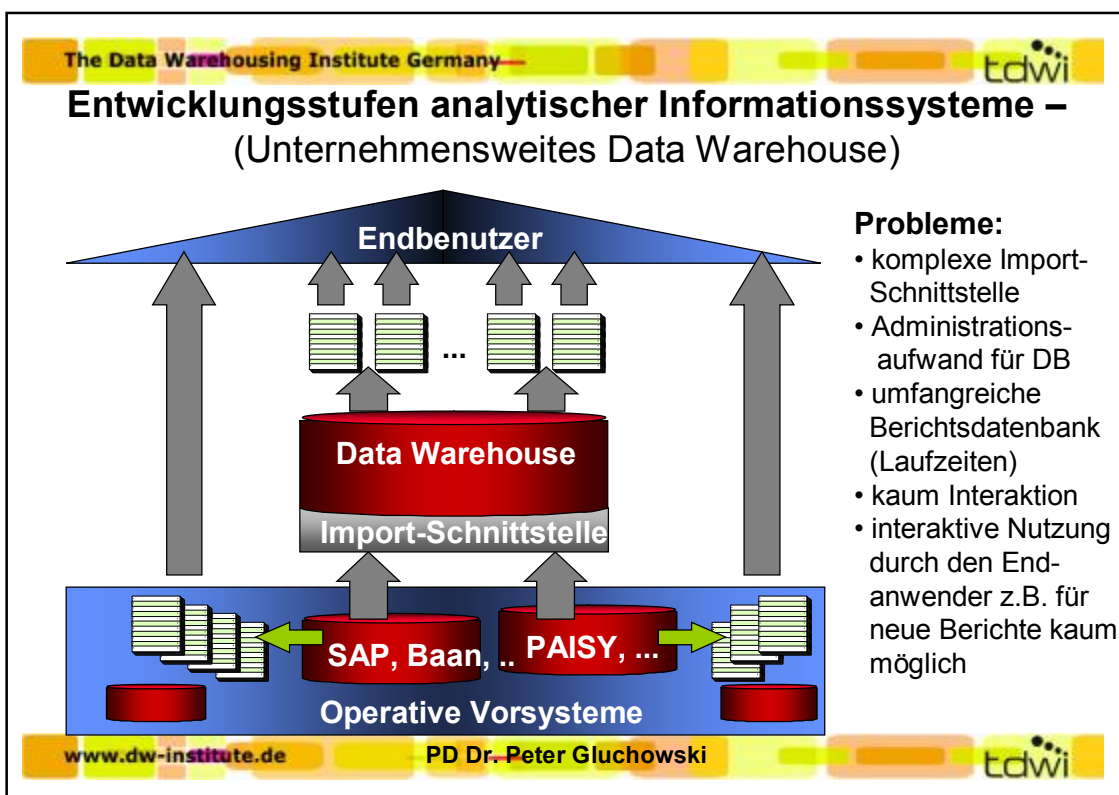
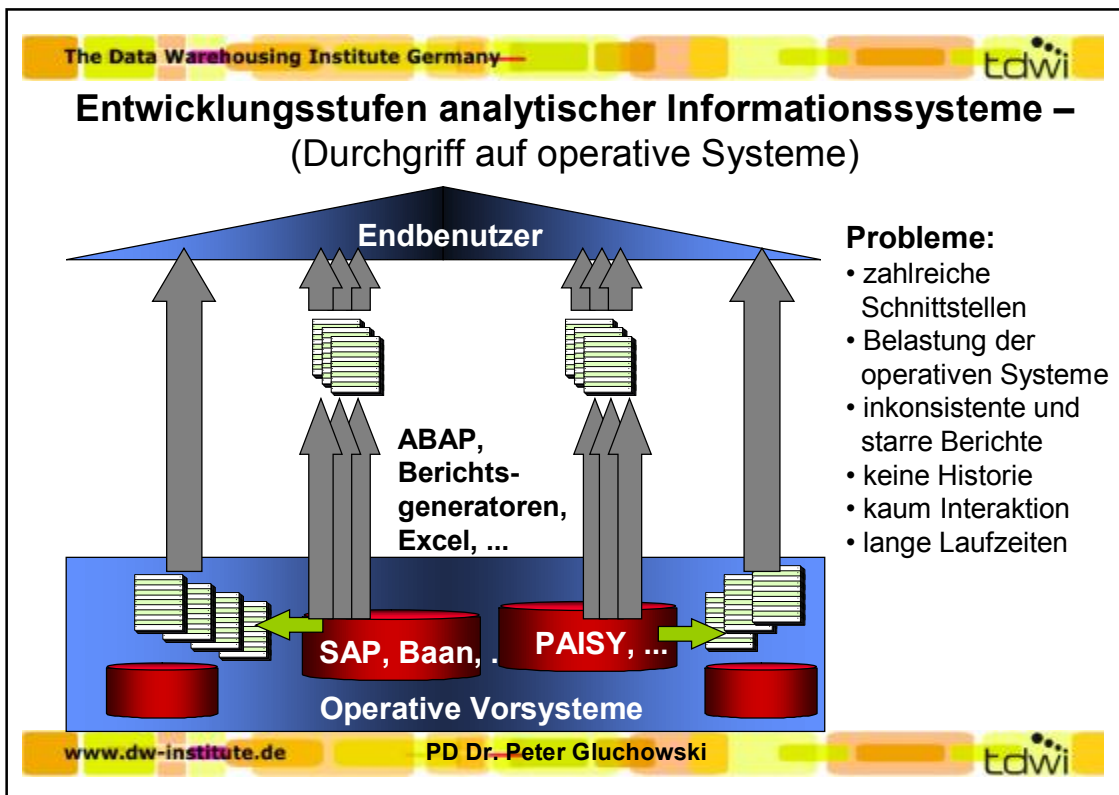
W. H. Inmon, Building the Data Warehouse, New York u. a. 1993, S. 29.

- | | | |
|--------------------|---|--|
| ▪ subject oriented | ➡ | Themenorientierung |
| ▪ integrated | ➡ | Vereinheitlichung |
| ▪ nonvolatile | ➡ | Dauerhaftigkeit, Stabilität |
| ▪ time variant | ➡ | Zeitorientierung der Informationen |
| ▪ management needs | ➡ | Analyse und Entscheidungsunterstützung |

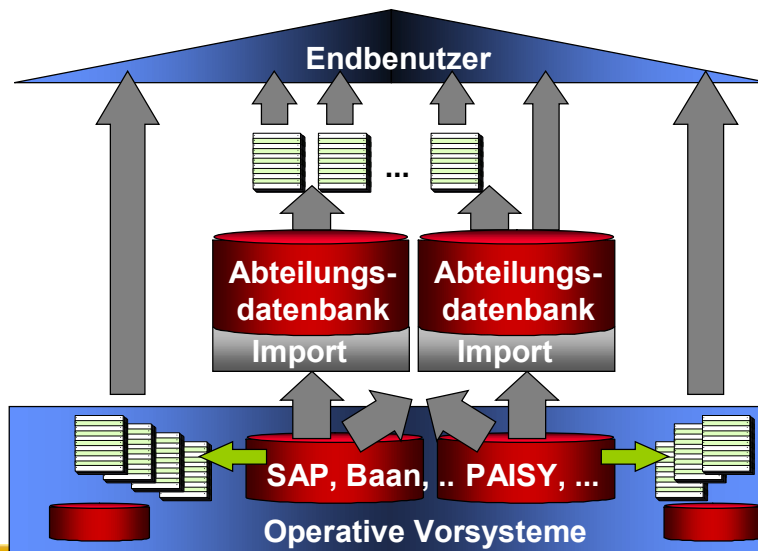
Data Warehouse

- umfasst die **Serverkomponenten** einer Systemlösung, die für die **unternehmensweite Datenversorgung** der Front-End-Systeme zur **Informationsbereitstellung und Entscheidungsunterstützung** betrieblicher Fach- und Führungskräfte zuständig sind,
- ist **physikalisch** von den operativen Vorsystemen **getrennt** und
- baut lediglich zum Zweck der **periodischen Datenaktualisierung** bzw. -ergänzung Verbindungen zu den operativen DV-Systemen auf.





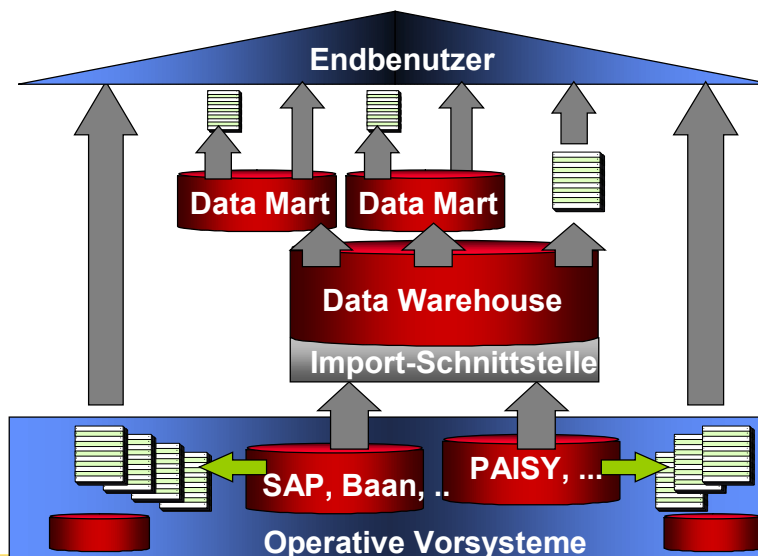
Entwicklungsstufen analytischer Informationssysteme – (Data Marts ohne Data Warehouse)



Probleme:

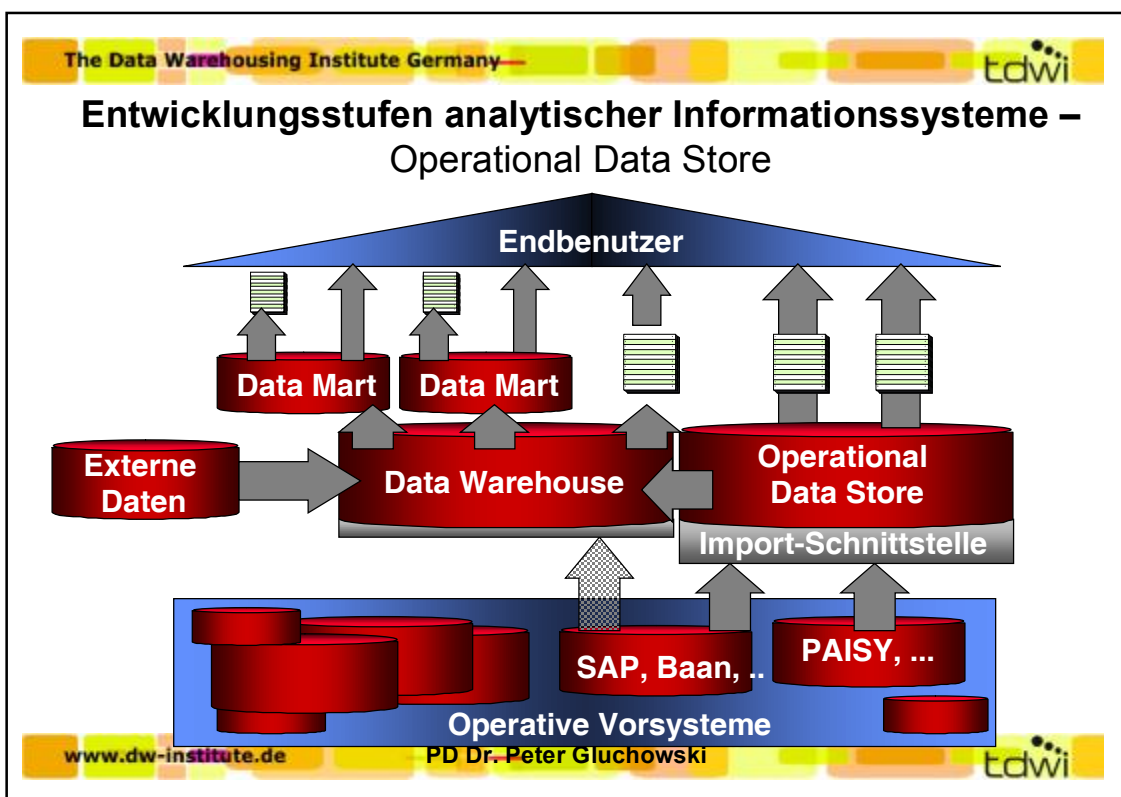
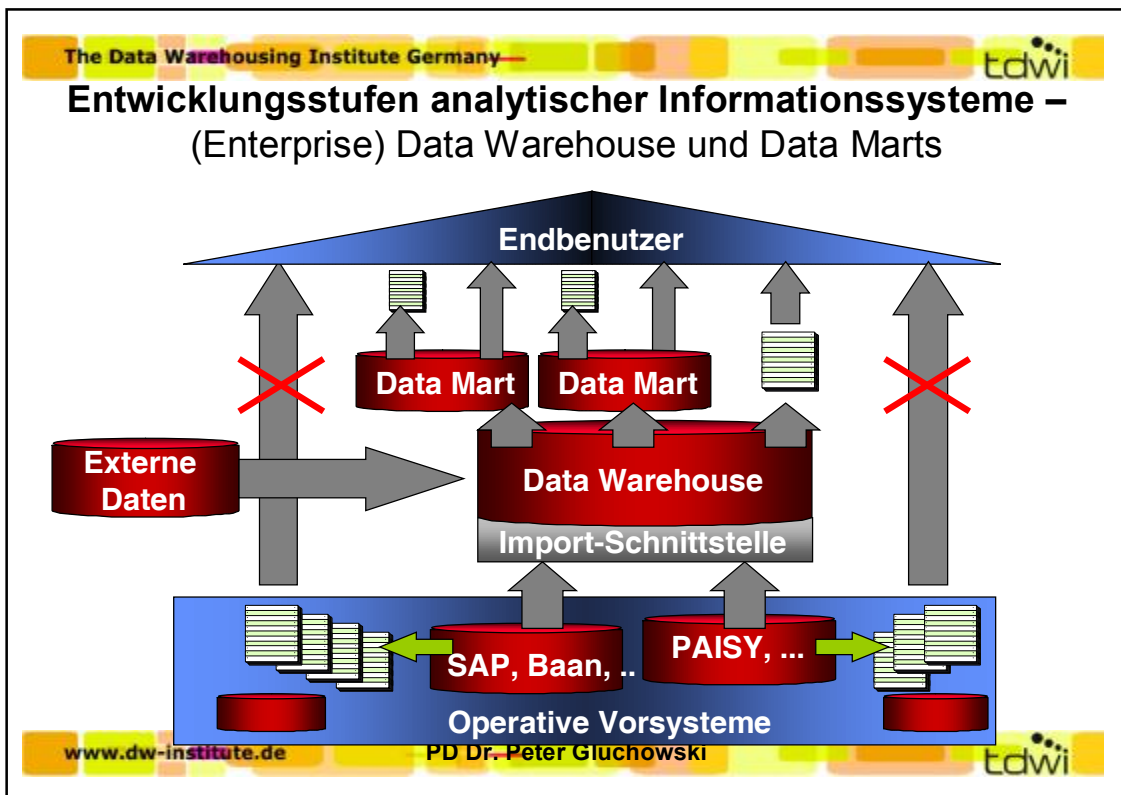
- Anzahl der Schnittstellen
- viele DBs
- Konsistenz der Daten

Entwicklungsstufen analytischer Informationssysteme – (Data Marts mit Data Warehouse)

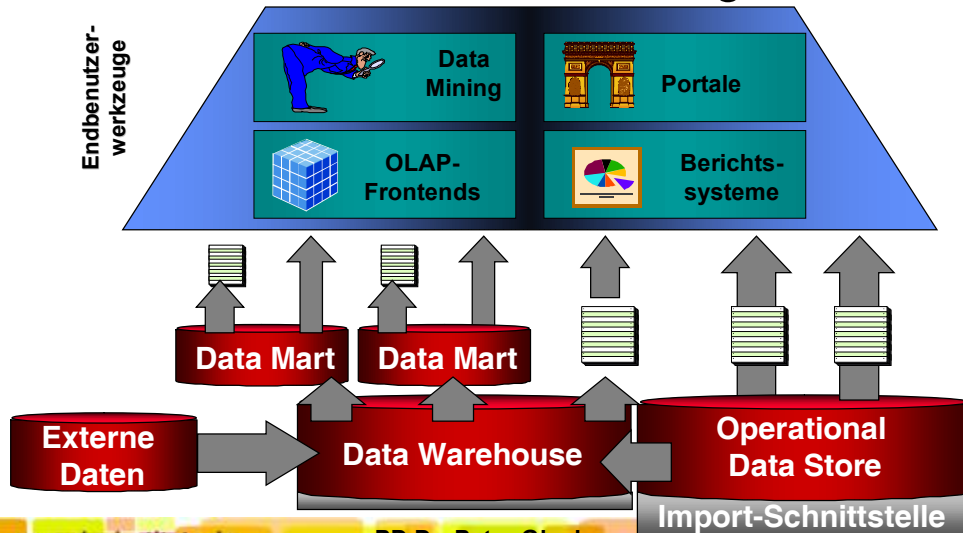


Probleme:

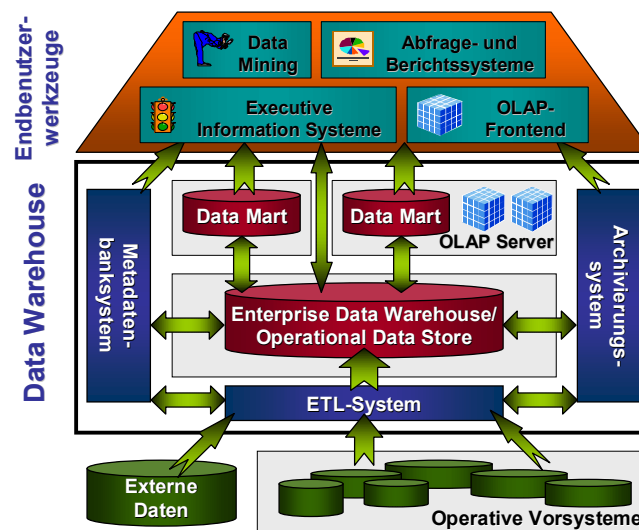
- komplexe Lösung
- lange Projektlaufzeit bis zu ersten Ergebnissen
- viele Einzelkomponenten



Analytische Informationssysteme – Oberflächentechnologien



Aktuelles DW-Architektur-Konzept



On-Line Analytical Processing (OLAP)

- Software-Technologie, die Managern **schnelle, interaktive und vielfältige Zugriffe** auf relevante und konsistente Informationen ermöglicht

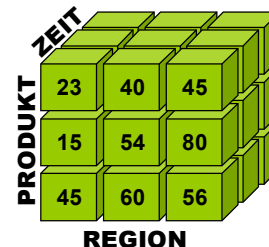


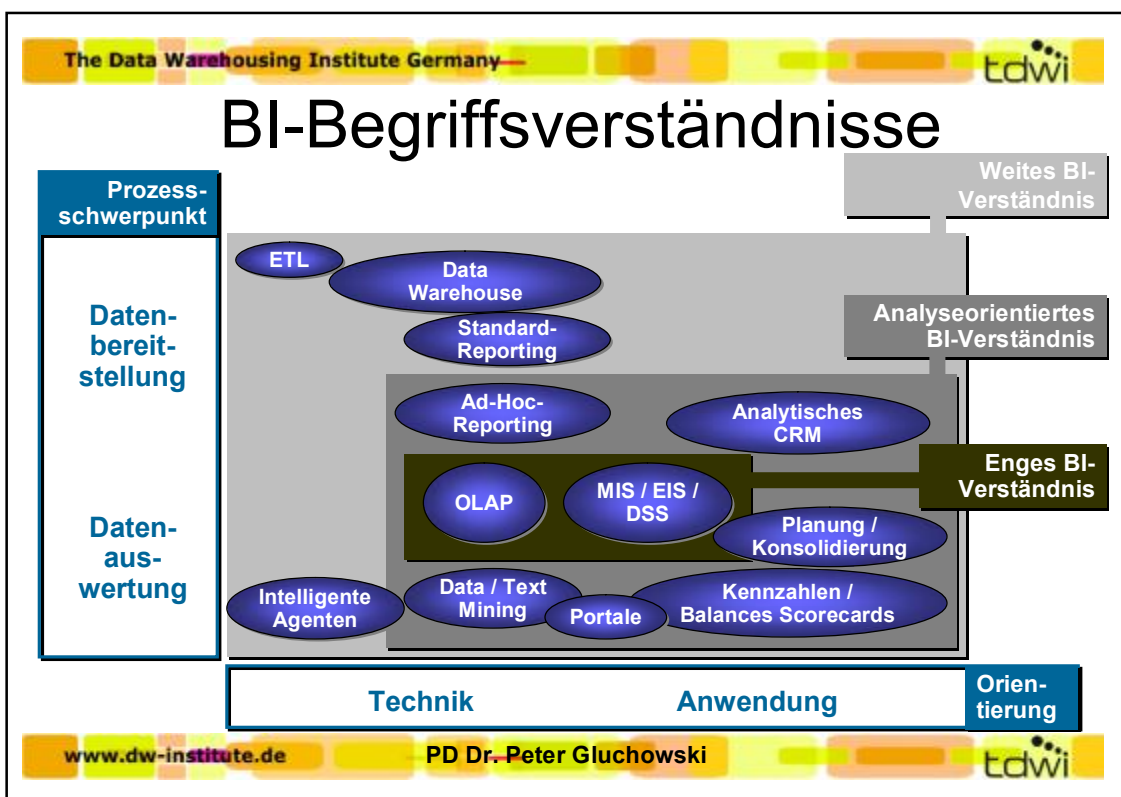
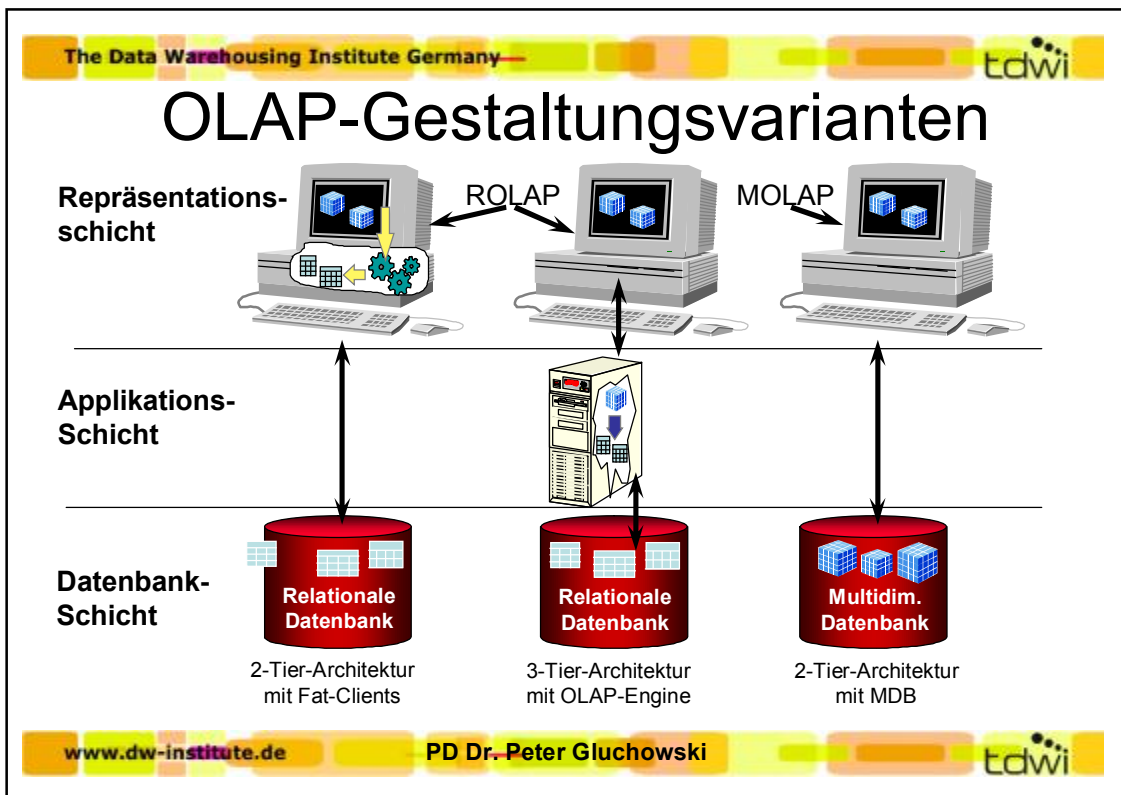
dynamische, multidimensionale
Analysen auf konsolidierten
Unternehmensdatenbeständen

- zwölf Evaluationsregeln nach Codd/Codd/Salley

On-Line Analytical Processing (OLAP)

- **Mehr- / Multidimensionalität:**
 - Anordnung betriebswirtschaftlicher Variablen (z.B. Kennzahlen) entlang mehrerer, unterschiedlicher Dimensionen (z. B. Region, Artikel, Produkte, Zeit, Kunden)

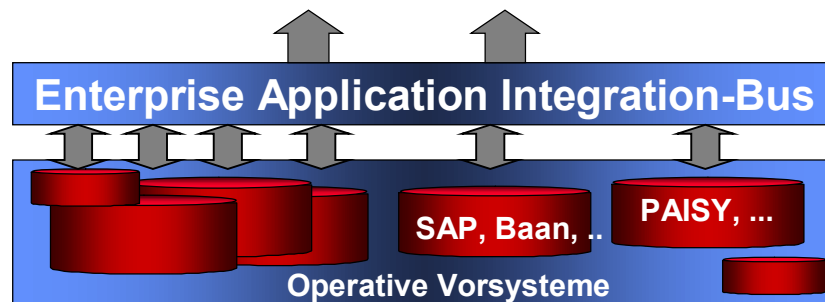




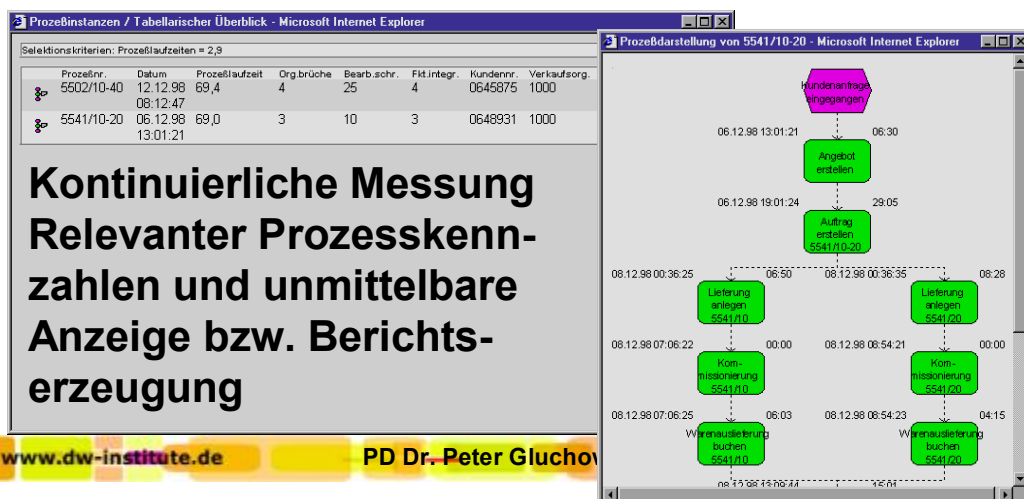
Aktuelle Entwicklungen – Real Time Enterprise

Anforderungen an „Echtzeit-Unternehmen“

- Integrierte Echtzeit-Information über Abteilungs-/Bereichsgrenzen hinweg
- Integration über Unternehmensgrenzen im Rahmen kollaborativer Geschäftsmodelle
- Ausrichtung auf Geschäftsereignisse



Aktuelle Entwicklungen – Process Performance Management bzw. Business Activity Monitoring



Agenda

- 1) Abgrenzung und Einordnung von CPM und BI
- 2) Architekturen und Komponenten analyseorientierter Informationssysteme
- 3) **Bestandteile multidimensionaler Datenstrukturen**
- 4) Methoden der semantischen Modellierung mehrdimensionaler Datenstrukturen
- 5) Umsetzung mit multidimensionalen Datenbanksystemen
- 6) Varianten und Bestandteile des Star Schemas
- 7) Erweitertes Star Schema des SAP Business Information Warehouse
- 8) Mehrdimensionales Datenmodell des Microsoft SQL Server
- 9) Zusammenfassende Gegenüberstellung relationaler und multidimensionaler Modellierung

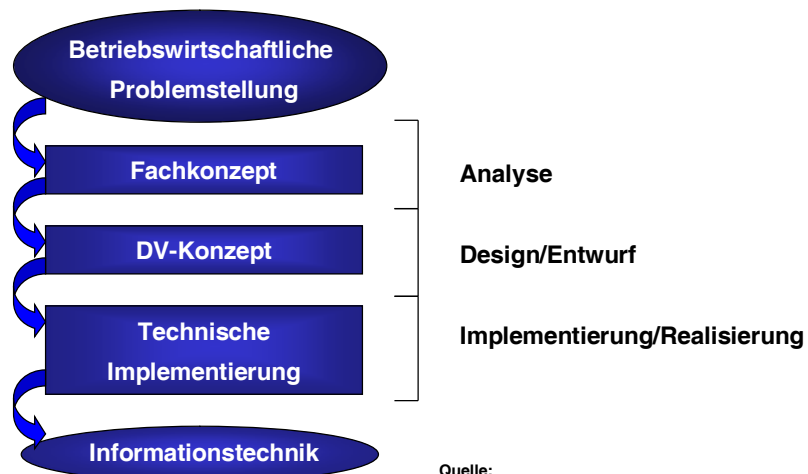
Agenda

- Phasen der Datenmodellierung
- Grundbestandteile mehrdimensionaler Datenstrukturen
- Hierarchische Dimensionsstrukturen
- Gestaltungsempfehlungen für die Modellierung

Agenda

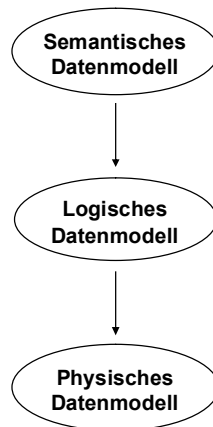
- Phasen der Datenmodellierung
- Grundbestandteile mehrdimensionaler Datenstrukturen
- Hierarchische Dimensionsstrukturen
- Gestaltungsempfehlungen für die Modellierung

Beschreibungsebenen betrieblicher Informationssysteme



Quelle:
Scheer, A.-W.: Wirtschaftsinformatik - Referenzmodelle für
industrielle Geschäftsprozesse, 3. Aufl., Berlin u.a. 1990.

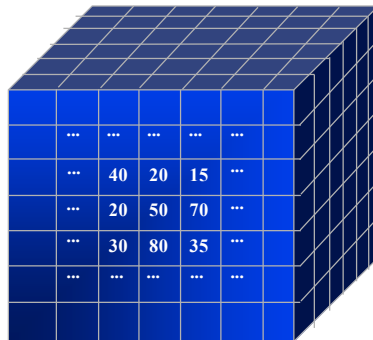
Ebenen der Datenmodellierung



Agenda

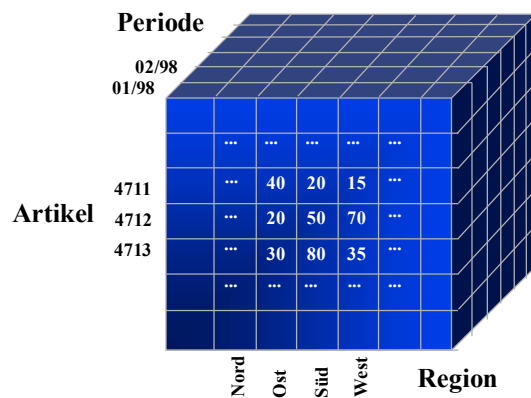
- Phasen der Datenmodellierung
- **Grundbestandteile mehrdimensionaler Datenstrukturen**
- Hierarchische Dimensionsstrukturen
- Gestaltungsempfehlungen für die Modellierung

Zahlenwürfel



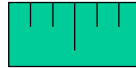
Zahlenwürfel mit betriebswirtschaftlichem Bezug

Absatzmenge

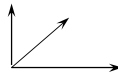


Strukturkomponenten

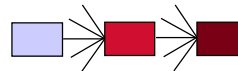
1. Betriebswirtschaftliche Variablen
(Meßgrößen, Kennzahlen, Fakten)



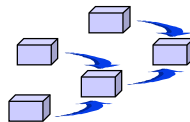
2. Dimensionen



3. Dimensionshierarchien



4. Regeln



Betriebswirtschaftliche Variablen (Kennzahlen)

Kennzahlendefinition

- Zahl, die betriebliche und außerbetriebliche Informationen in aussagekräftiger, komprimierter Form wiedergibt

Kennzahlentypen

- Statistische Form

Absolute Kennzahlen
(z.B. Beschäftigte, Umsatz, Bilanzsumme)

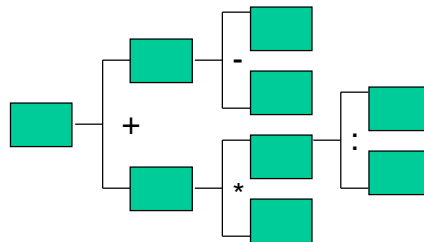
Relative Kennzahlen

- Gliederungszahlen
(z.B. Anlagevermögen/Gesamtvermögen)
- Beziehungszahlen
(z.B. Gewinn/Eigenkapital)
- Indexpzahlen
(z.B. Lohnkosten heute * 100 / Lohnkosten Vorjahr)

- Gegenstand (Mengen / Werte, Zeitpunkt / Zeitraum)
- Objektbereich (Gesamtwirtschaftlich, Konzern, Teilbetrieb)

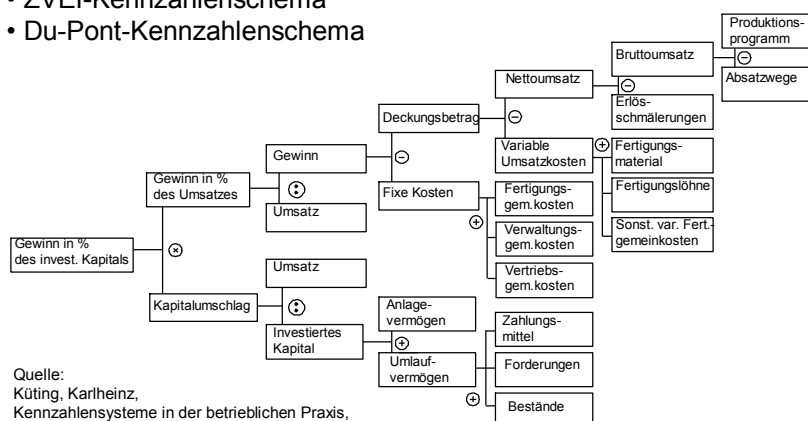
Kennzahlensysteme

- Zusammenstellung unterschiedlicher Kennzahlen zur rechentechnischen Systematisierung
- Einzelkennzahlen stehen in sachlich sinnvoller Beziehung zueinander



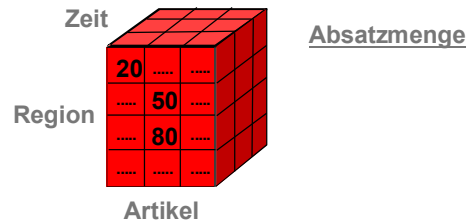
Bekannte Kennzahlensysteme

- RL-Kennzahlenschema von Reichmann/Lachnit
- ZVEI-Kennzahlenschema
- Du-Pont-Kennzahlenschema

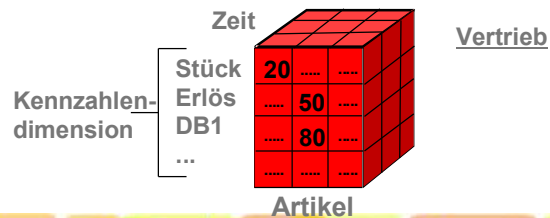


Kennzahlen als Bestandteile multidimensionaler Datenstrukturen

- Betrachtung von Einzelkennzahlen



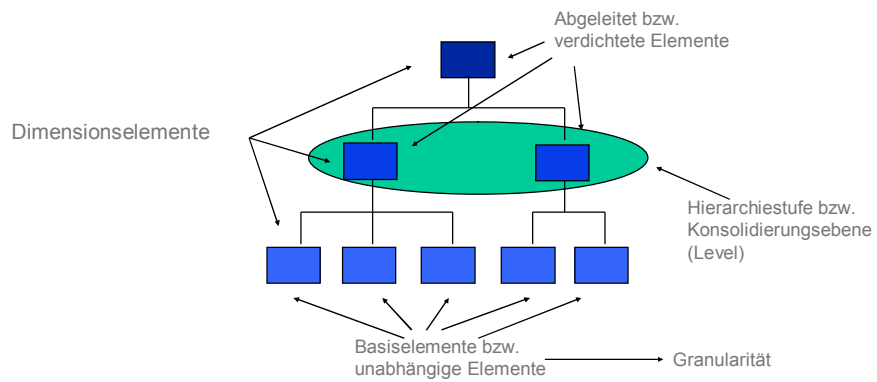
- Betrachtung mehrerer Kennzahlen



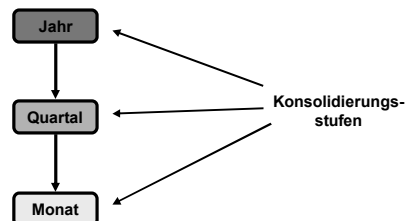
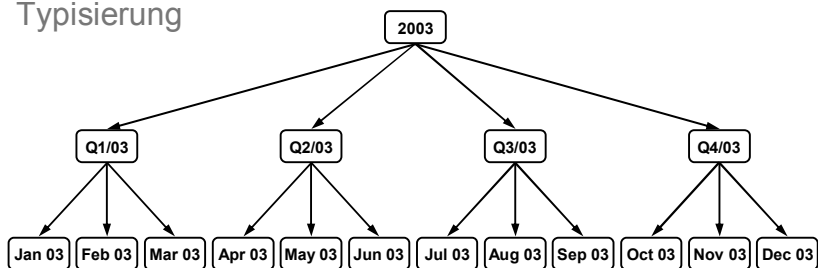
Dimensionen

- Aufgabe:
Aufgliederung / Konkretisierung von Kennzahlen bzw. Zahlenwerten
- Inhalt:
Sachlogisch zusammengehörige Umweltobjekte, gleichen Geschäftsaspekt betreffend
- Ausprägungen:
 - Einzelobjektdimensionen (Element bestimmt)
wenige, aufzählbare Dimensionselemente (z.B. Szenario- oder Kennzahlendimension)
 - Mengendimensionen (Ebenen bestimmt)
viele bis sehr viele Dimensionselemente (z.B. Kunden- oder Artikeldimension)
Beschreibung nur über Objektmengen

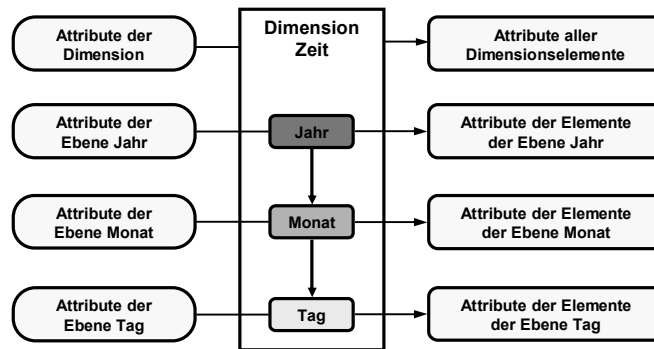
Dimensionskomponenten



Typisierung



Attribute



Dimensionstypen

- Standarddimensionen (z.B. Artikel, Kunden, Regionen)
- Kennzahlendimension (Measure Dimension)
- Szenario- bzw. Wertartdimension (Plan, Ist, Soll, ...)
- Zeitdimension (vorbestimmte Struktur)

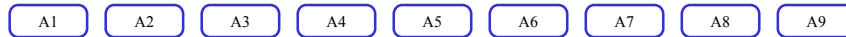
Weitere Regeln

- Würfelübergreifende Verknüpfungsregeln
→ OLAP-Join: Verbindung zwischen Würfeln (Multicube-Ansatz)
- Integritätsregeln
→ Hierarchische Integrität bei nicht dynamischer Berechnung
- Regeln für die Behandlung von Ausnahmen
→ z.B. Konsolidierungs-Sonderfälle

Agenda

- Phasen der Datenmodellierung
- Grundbestandteile mehrdimensionaler Datenstrukturen
- **Hierarchische Dimensionsstrukturen**
- Gestaltungsempfehlungen für die Modellierung

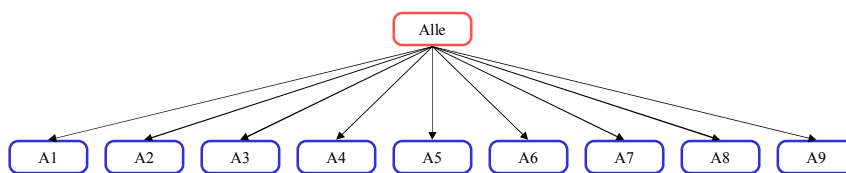
Flache Struktur



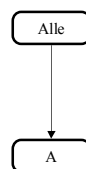
Typisierte flache Struktur



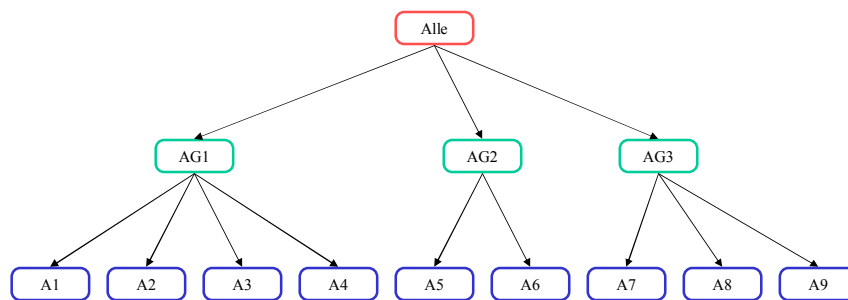
Flache Struktur mit totaler Verdichtung



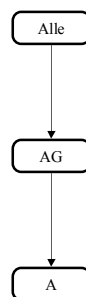
Typisierte flache Struktur mit totaler Verdichtung



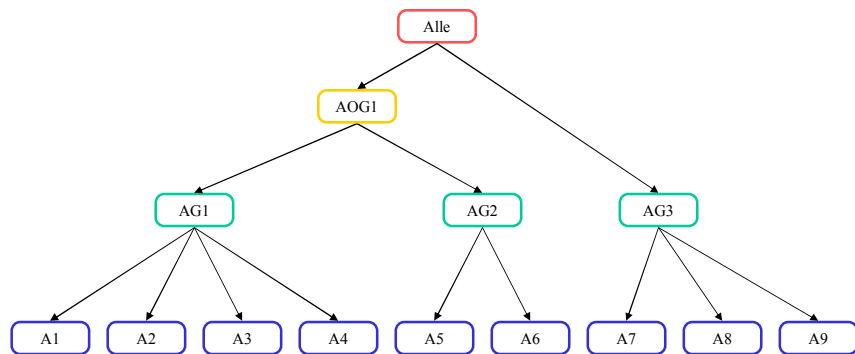
Baumstruktur



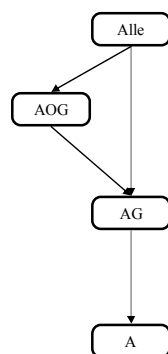
Typisierte Baumstruktur



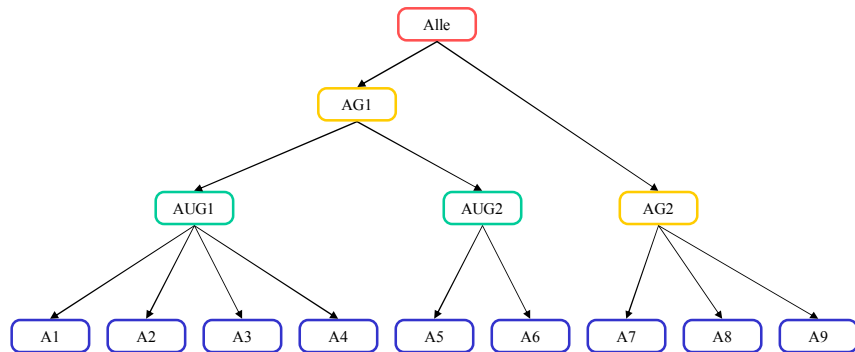
Baumstruktur mit unterschiedlichen Tiefen Variante I



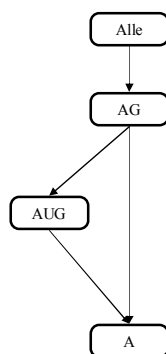
Typisierte Baumstruktur mit unterschiedlichen Tiefen Variante I



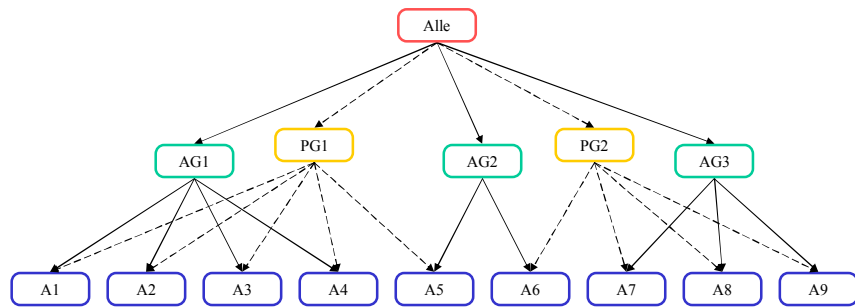
Baumstruktur mit unterschiedlichen Tiefen Variante II



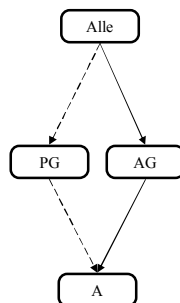
Typisierte Baumstruktur mit unterschiedl. Tiefen Variante II



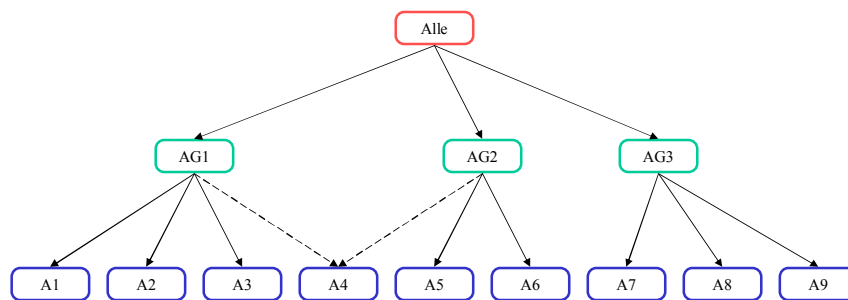
Parallele Hierarchie



Typisierte parallele Hierarchie

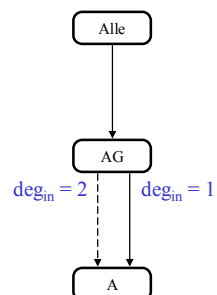


Heterarchie (m:n-Beziehungen)

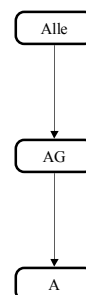


Typisierte Heterarchie

Mit Unterscheidung der
„Kardinalität“ der Kanten



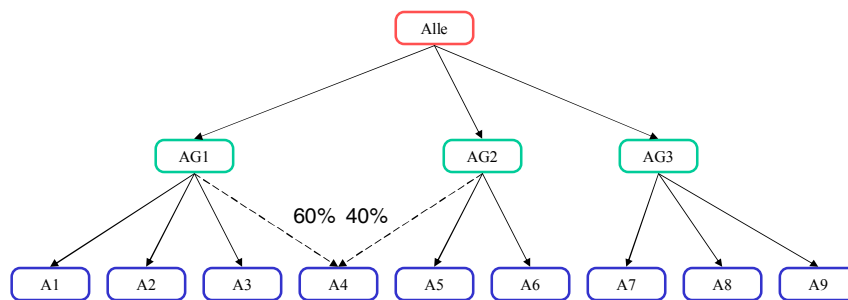
Ohne Unterscheidung der
„Kardinalität“ der Kanten



Anteilige Verrechnung

Sonderfall:

Anteilige Verrechnung



Regeln (Additivitätseigenschaften)

Ausnahmeregeln für nicht-additive Kennzahlen

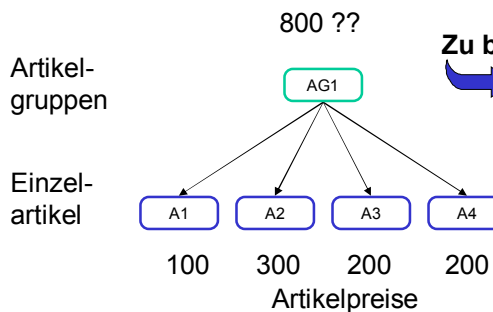
z.B. Preis, Lagerendbestand

Gegebenenfalls sinnvoll:

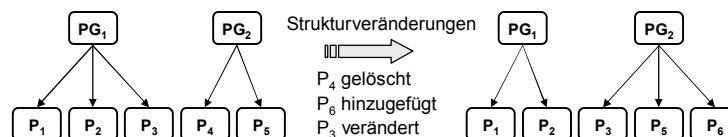
Durchschnittsberechnung oder letzter Wert

Zu beachten:

Semi-additive Größen
(Lagerendbestände lassen sich zwar sinnvoll über alle Lagerorte addieren, nicht jedoch über die Zeit)



Zeitabhängigkeit: Strukturelle Veränderungen in Dimensionen



Lösungsalternativen bei Strukturbrüchen

- **Anpassung des historischen Datenmaterials an neue Strukturen**
Vorteile: Keine Aufblähung des Datenbestandes; Datenstrukturen bleiben überschaubar
Nachteile: Alte Strukturen sind verloren; Anwender wollen u. U. Berichte mit alten Strukturen
- **Separate Speicherung des historischen Datenbestandes zusätzlich zum Komplettbestand mit neuen Strukturen**
Vorteile: Alte Auswertungen können abgerufen werden
Nachteile: Datenvolumen; aufwendige Aktualisierung, wenn Anwender auch neue Zahlen in den alten Strukturen sehen wollen; verwirrend für Endbenutzer
- **Aufbau paralleler Hierarchien mit alten/neuen Strukturen**
Vorteile: Alle Zahlen können mit beliebigen Strukturen angezeigt werden
Nachteile: Dimensionsstruktur kaum überschaubar
- **Temporale Datenbanken - Gültigkeitsstempel**
Vorteile: Alle Zahlen können mit beliebigen Strukturen angezeigt werden
Nachteile: Performance; Konzepte noch nicht ausgereift

Agenda

- Phasen der Datenmodellierung
- Grundbestandteile mehrdimensionaler Datenstrukturen
- Hierarchische Dimensionsstrukturen
- **Gestaltungsempfehlungen für die Modellierung**

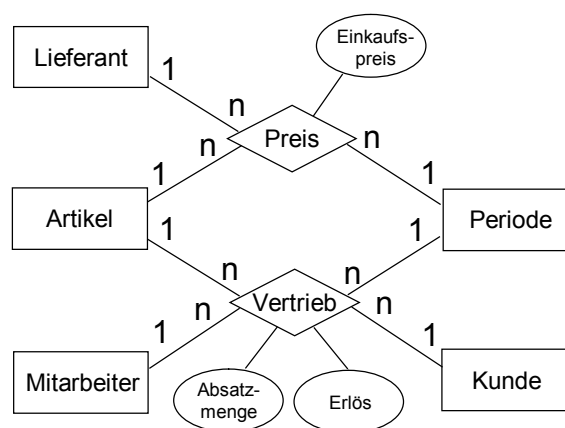
Gestaltungsüberlegungen der Dimensionsmodellierung auf semantischer Ebene

- Anzahl der Dimensionen zwischen vier und zehn (ideal zwischen sechs und acht)
- Anzahl der Hierarchiestufen (maximal sieben Hierarchiestufen)
- Anzahl der Elemente je Konsolidierungselement (maximal fünfzehn bis zwanzig Elemente)
- Bestimmung von Dimensionen (1:1-Beziehungen ungeeignet; 1:N-Beziehungen bilden Dimensionshierarchie; M:N-Beziehungen stellen zwei Dimensionen dar)

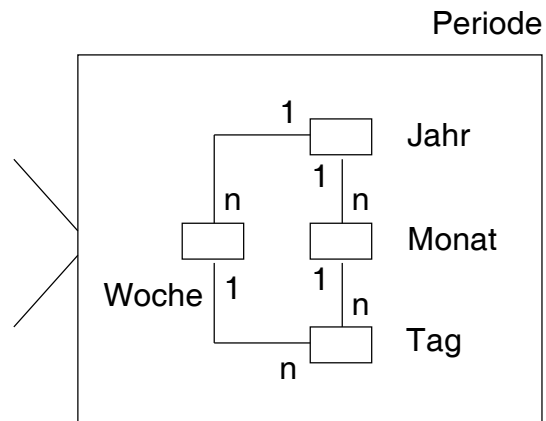
Agenda

- 1) Abgrenzung und Einordnung von CPM und BI
- 2) Architekturen und Komponenten analyseorientierter Informationssysteme
- 3) Bestandteile multidimensionaler Datenstrukturen
- 4) Methoden der semantischen Modellierung mehrdimensionaler Datenstrukturen**
- 5) Umsetzung mit multidimensionalen Datenbanksystemen
- 6) Varianten und Bestandteile des Star Schemas
- 7) Erweitertes Star Schema des SAP Business Information Warehouse
- 8) Mehrdimensionales Datenmodell des Microsoft SQL Server
- 9) Zusammenfassende Gegenüberstellung relationaler und multidimensionaler Modellierung

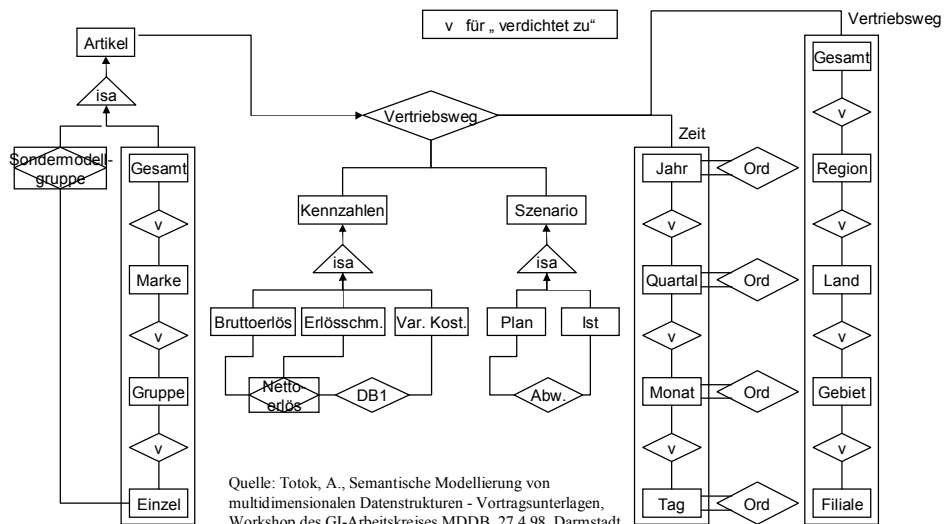
ER-Modellierung multidimensionaler Datenstrukturen



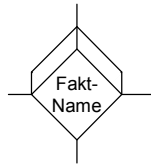
Erweiterung eER zur Darstellung von Dimensionshierarchien



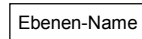
Beispiel ER-Modellierung multidimensionaler Datenstrukturen



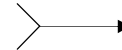
Darstellungsojekte im ME/R-Modell



**Fakt-
beziehungstyp**



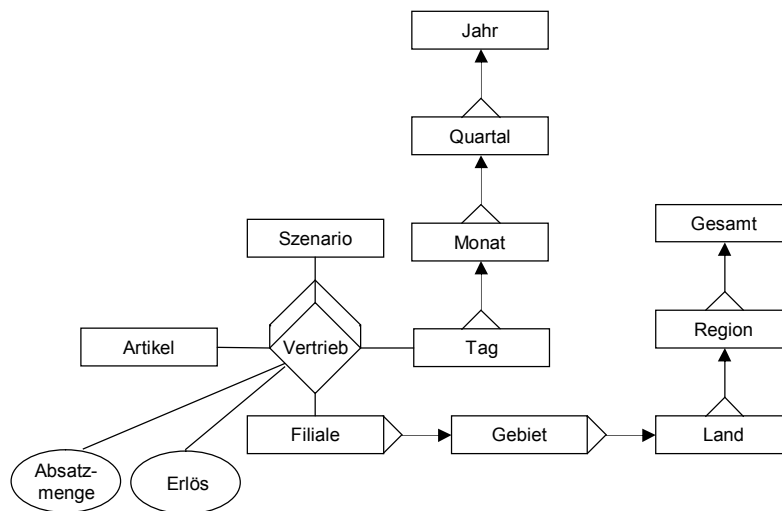
**Dimensionsebenen
Entitytyp**



**Roll-Up
Beziehungstyp**

Quelle: Sapia, Carsten; Blaschka, Markus; Höfling, Gabriele.; Dinter, Barbara: Extending the E/R Model for the Multidimensional Paradigm.
In: Proceedings of the Advances in Databases Technologies Conference, November 19-20, 1998, Singapore

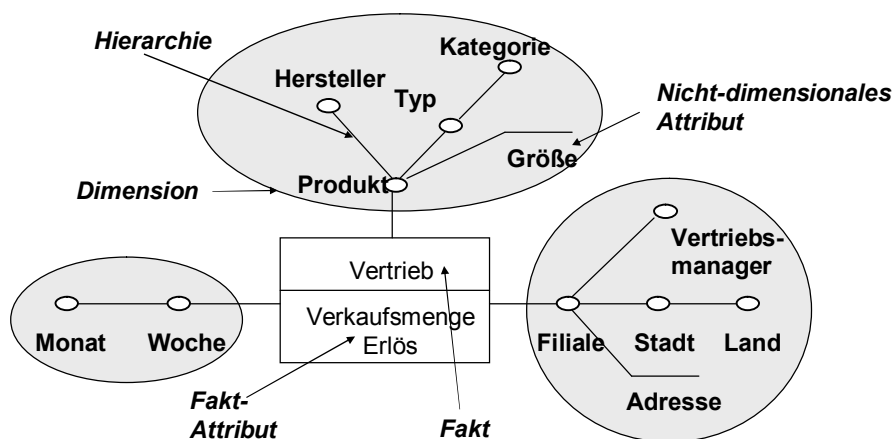
Beispiel eines ME/R-Modells



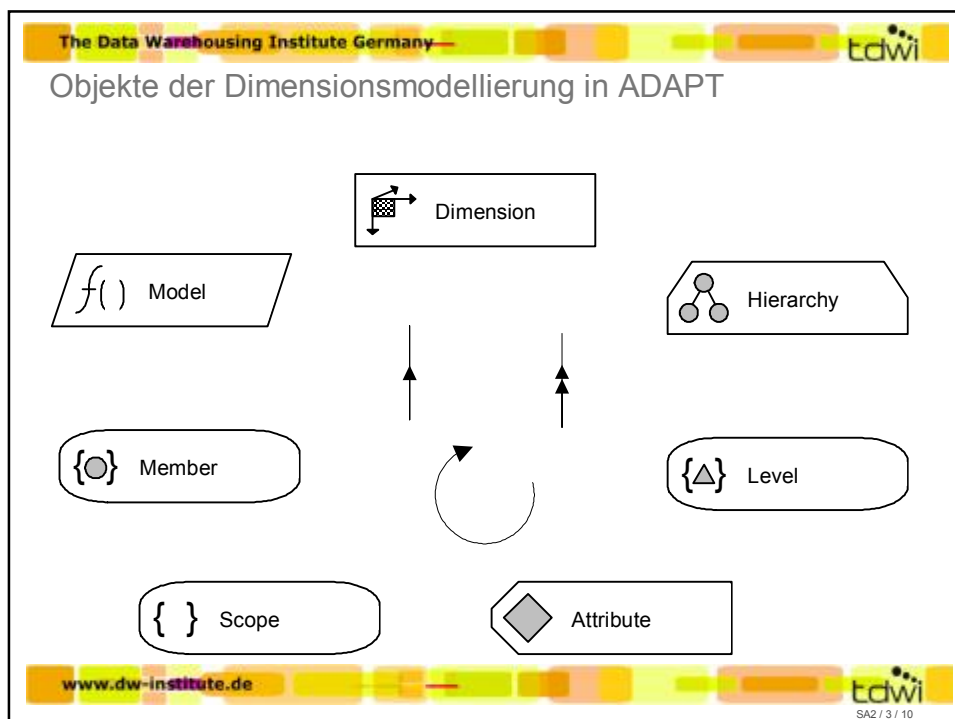
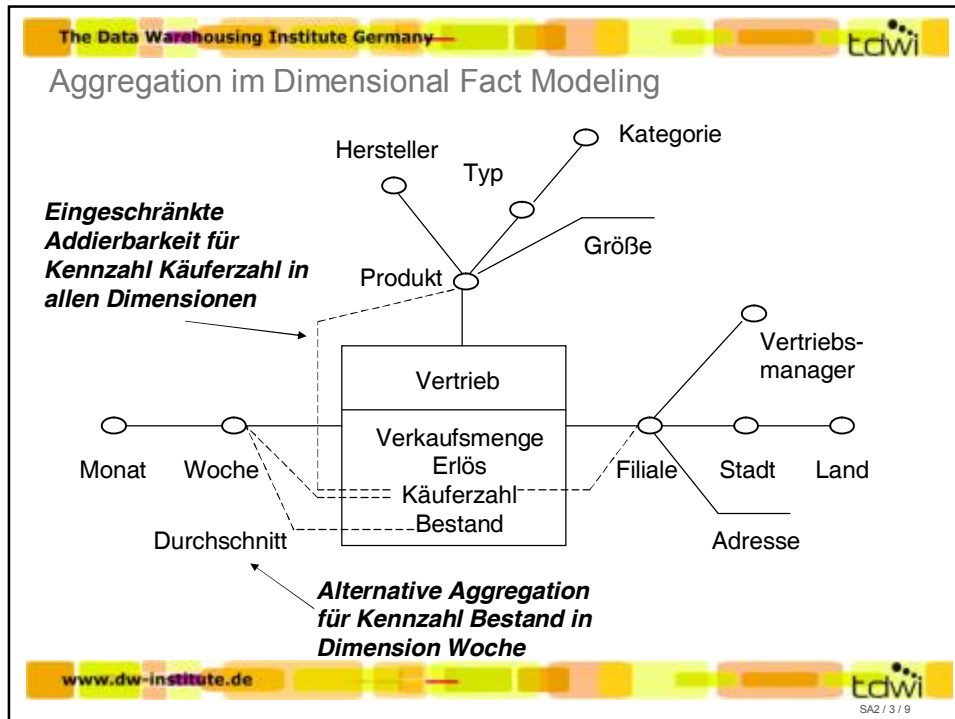
Grenzen der ER-Notation

- Unzulängliche Möglichkeiten zur Abbildung von Einzelobjekten (Kennzahlen, Szenario)
- Konzentration auf Objektklassen
- Verknüpfungen zwischen Einzelobjekten (z.B. relative Kennzahlen) wird nicht unterstützt

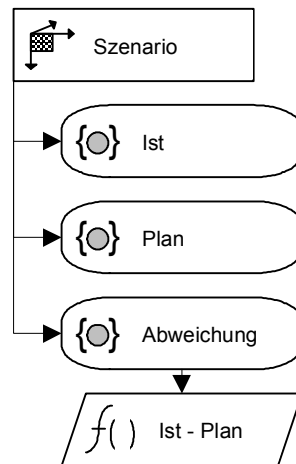
Dimensional Fact Modeling



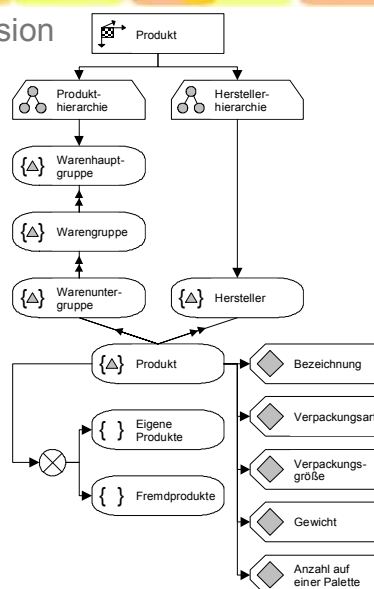
Quelle: Golfarelli, Matteo; Maio, Dario; Rizzi, Stefano: Conceptual Design of Data Warehouses from E/R Schemas.
In: Proceedings of the Hawaii International Conference On System Sciences, January 6-9, 1998, Kona, Hawaii.



Elementbestimmte Dimension



Ebenenbestimmte Dimension



Beziehungstypen in ADAPT



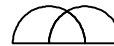
Fully exclusive



Fully overlapping

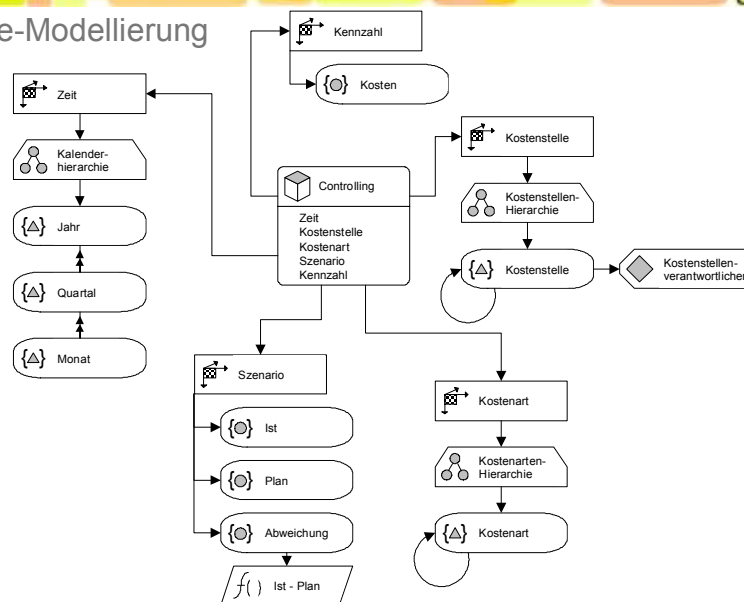


Partially exclusive

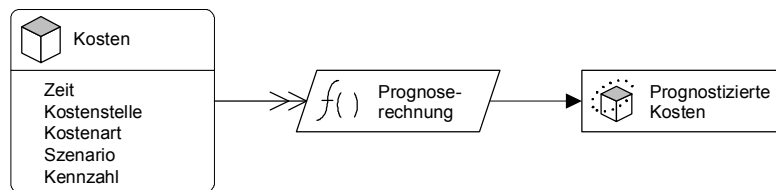


Partially overlapping

Cube-Modellierung



Modellierung einer abgeleiteten Teilsicht eines Cubes

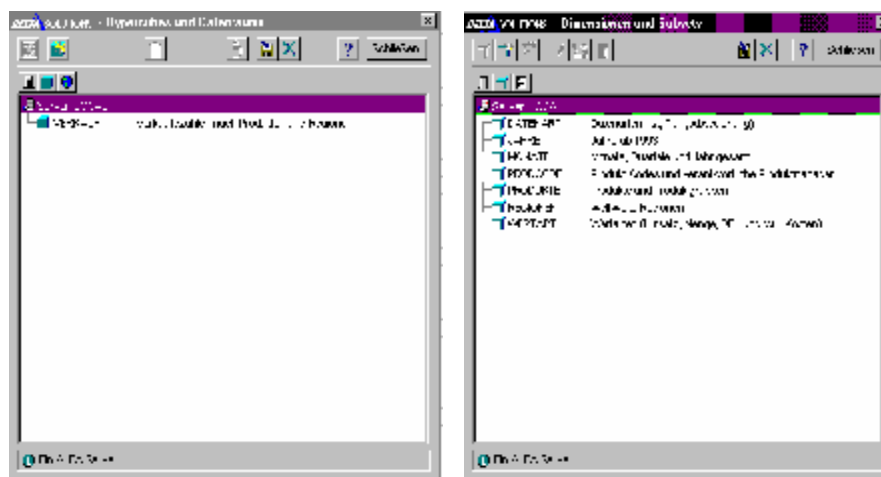
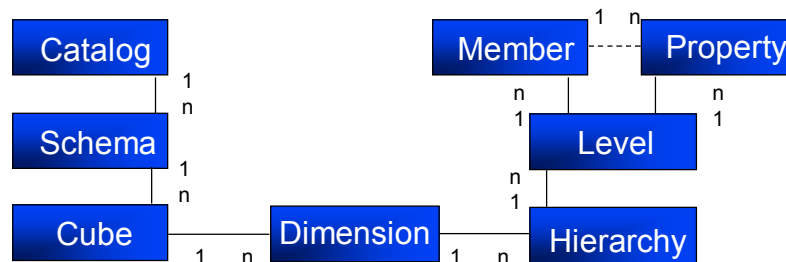


Agenda

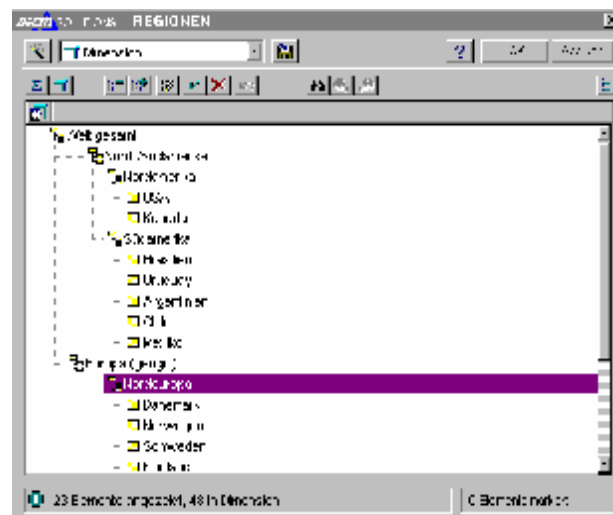
- 1) Abgrenzung und Einordnung von CPM und BI
- 2) Architekturen und Komponenten analyseorientierter Informationssysteme
- 3) Bestandteile multidimensionaler Datenstrukturen
- 4) Methoden der semantischen Modellierung mehrdimensionaler Datenstrukturen
- 5) **Umsetzung mit multidimensionalen Datenbanksystemen**
- 6) Varianten und Bestandteile des Star Schemas
- 7) Erweitertes Star Schema des SAP Business Information Warehouse
- 8) Mehrdimensionales Datenmodell des Microsoft SQL Server
- 9) Zusammenfassende Gegenüberstellung relationaler und multidimensionaler Modellierung

DV-Speicherungskonzept mit multidimensionalen Datenbanken

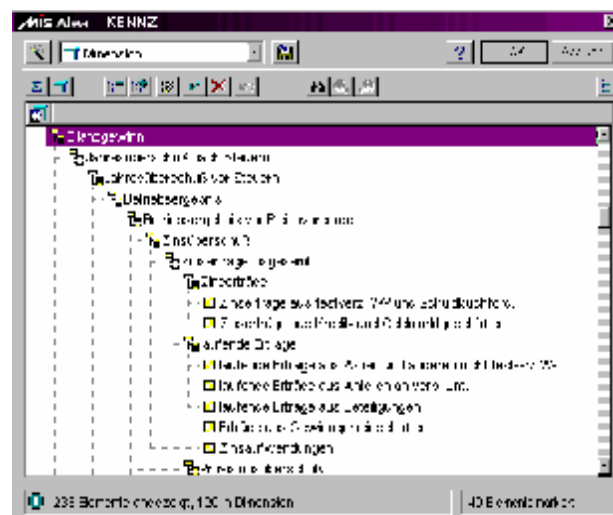
- Spezifische Ausrichtung auf multidimensionale Datenstrukturen ermöglicht i.d.R. 1:1-Umsetzung des semantischen Konzeptes
 - Hypercube- vs. Multicube-Ansätze
 - Komplexe Rule-Programmierung
 - Keine Standards (ggf. durch die Microsoft-API OLE DB for OLAP)
- ↓
- Keine einheitlichen Schema-Vorgaben



Dimensionsstrukturen in Delta Alea



Dimensionsstrukturen in Delta Alea



The Data Warehousing Institute Germany — edwi

Dimensionsstrukturen in Hyperion Essbase

www.dw-institute.de edwi

SA2 / 3 / 7

The Data Warehousing Institute Germany — edwi

Regeln

Verknüpfungs- und Ableitungsregeln für Dimensionselemente

Hyperion Essbase

MIS Alea

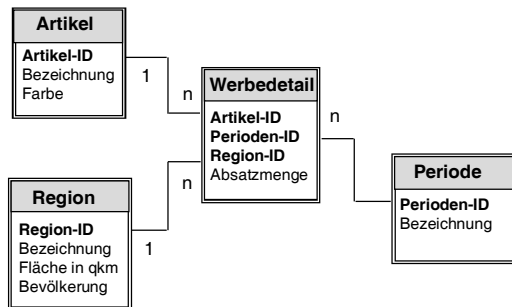
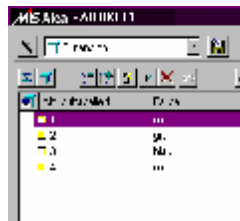
www.dw-institute.de edwi

SA2 / 3 / 8

Freiheitsgrade beim Design

Beispiel Farbe

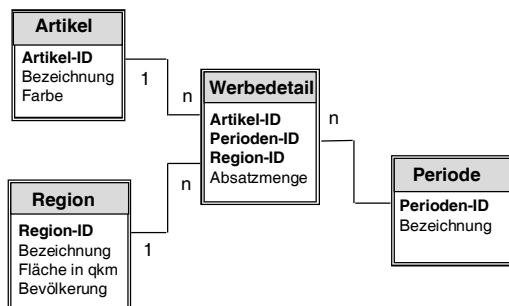
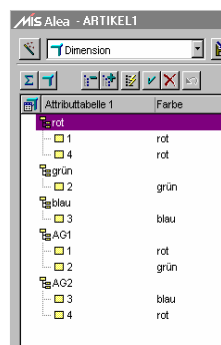
1.) Farbe als Attribut



Freiheitsgrade beim Design

Beispiel Farbe

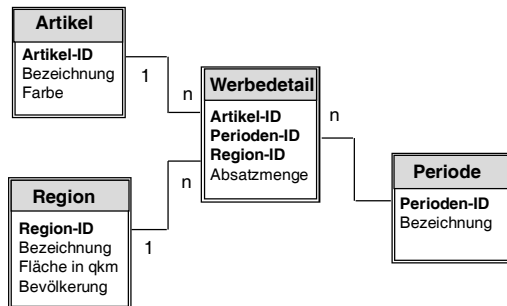
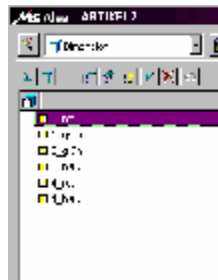
2.) Farbe als Hierarchie



Freiheitsgrade beim Design

Beispiel Farbe

3.) Farbe als Teil der Artikel-ID



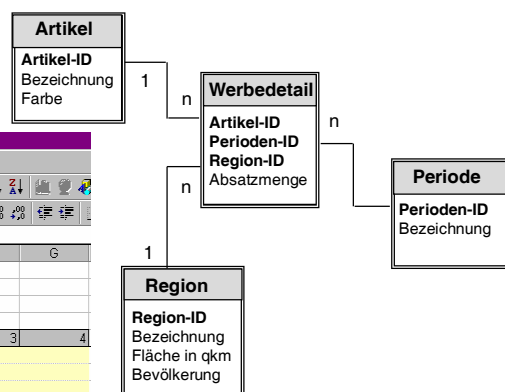
Freiheitsgrade beim Design

Beispiel Farbe

4.) Farbe als Dimension

Microsoft Excel - [Allea Ansicht] - [MIS1]Tabelle1

	B	C	D	E	F	G
19	LOCAL					
20	TEST					
21						
22		ARTIKEL ...				
23	FARBE	Alle Artikel	1	2	3	4
24	farbig					
25	rot					
26	grün					
27	blau					
28	mono					
29	weiß					
30	schwarz					



Agenda

- 1) Abgrenzung und Einordnung von CPM und BI
- 2) Architekturen und Komponenten analyseorientierter Informationssysteme
- 3) Bestandteile multidimensionaler Datenstrukturen
- 4) Methoden der semantischen Modellierung mehrdimensionaler Datenstrukturen
- 5) Umsetzung mit multidimensionalen Datenbanksystemen
- 6) Varianten und Bestandteile des Star Schemas**
- 7) Erweitertes Star Schema des SAP Business Information Warehouse
- 8) Mehrdimensionales Datenmodell des Microsoft SQL Server
- 9) Zusammenfassende Gegenüberstellung relationaler und multidimensionaler Modellierung

Agenda

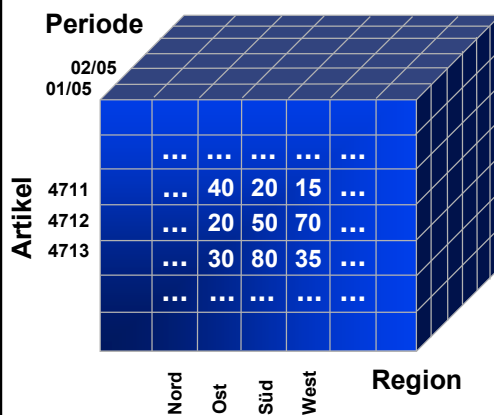
- Grundform des Star Schemas
- Modellierung von Dimensionshierarchien und Attributen
- Normalisierung von Dimensionen
- Kennzahlen und Kennzahlensysteme
- Aggregate
- Modellierungsvarianten

Agenda

- Grundform des Star Schemas
- Modellierung von Dimensionshierarchien und Attributen
- Normalisierung von Dimensionen
- Kennzahlen und Kennzahlensysteme
- Aggregate
- Modellierungsvarianten

Übergang vom Würfel zur Relation

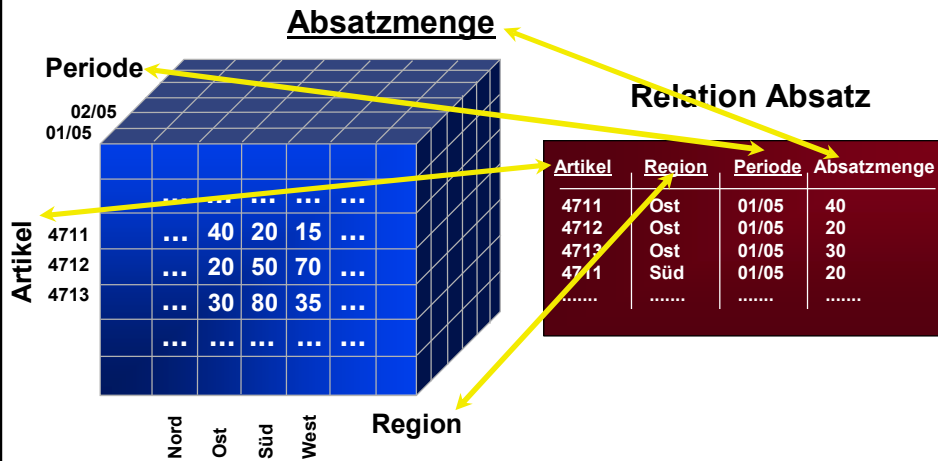
Absatzmenge



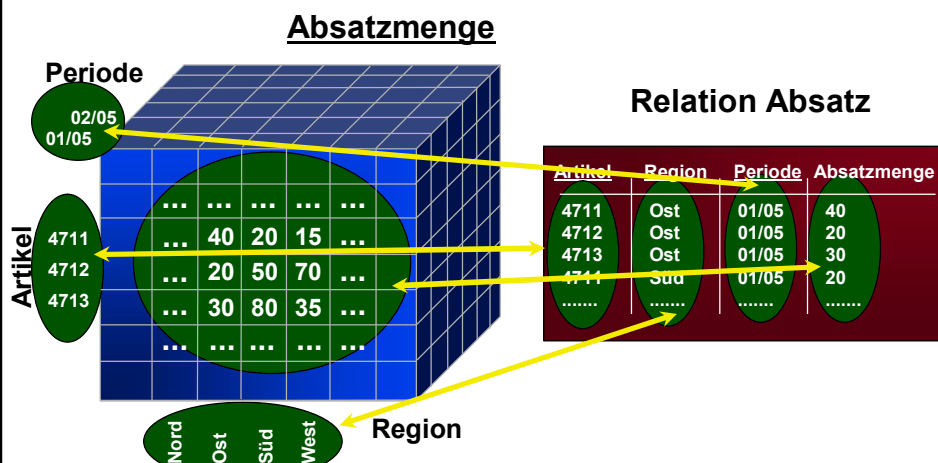
Relation Absatz

Artikel	Region	Periode	Absatzmenge
4711	Ost	01/05	40
4712	Ost	01/05	20
4713	Ost	01/05	30
4711	Süd	01/05	20
.....			

Übergang vom Würfel zur Relation

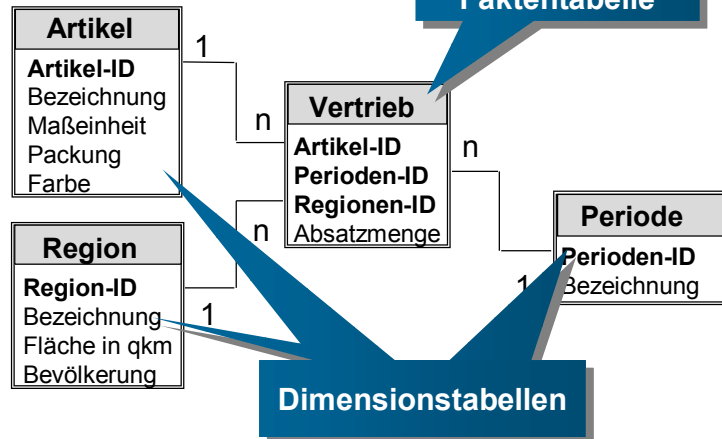


Übergang vom Würfel zur Relation

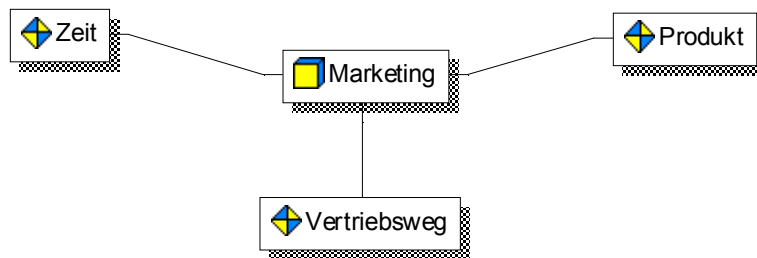


Relationale Mehrdimensionalität

Star-Schema



Einfaches Star Schema



Agenda

- Grundform des Star Schemas
- **Modellierung von Dimensionshierarchien und Attributen**
- Normalisierung von Dimensionen
- Kennzahlen und Kennzahlensysteme
- Aggregate
- Modellierungsvarianten

Flache Struktur: Dimensionstabelle Vertriebsweg

Vertriebsweg_Id	Vertriebsweg
1	Partner
2	Katalog
3	E-Shop

Balancierte Baum/Waldstruktur: Dimensionstabelle Zeit

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

Unbalanciert: Dimensionstabelle Produkt (Ausschnitt)

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

Parallele Hierarchie: Kalender und Geschäftsjahressicht

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

Anteilige Verrechnung und Heterarchien

- Aufhebung der 1:n-Beziehung zwischen Ebenen der Konsolidierungshierarchie impliziert
 - entweder Summation als Verdichtung (Heterarchie)
 - oder allgemeine Verdichtung (anteilige Verrechnung)
- Realisierung über zwei unterschiedliche Dimensionen
- Für anteilige Verrechnung sind Aggregate notwendig

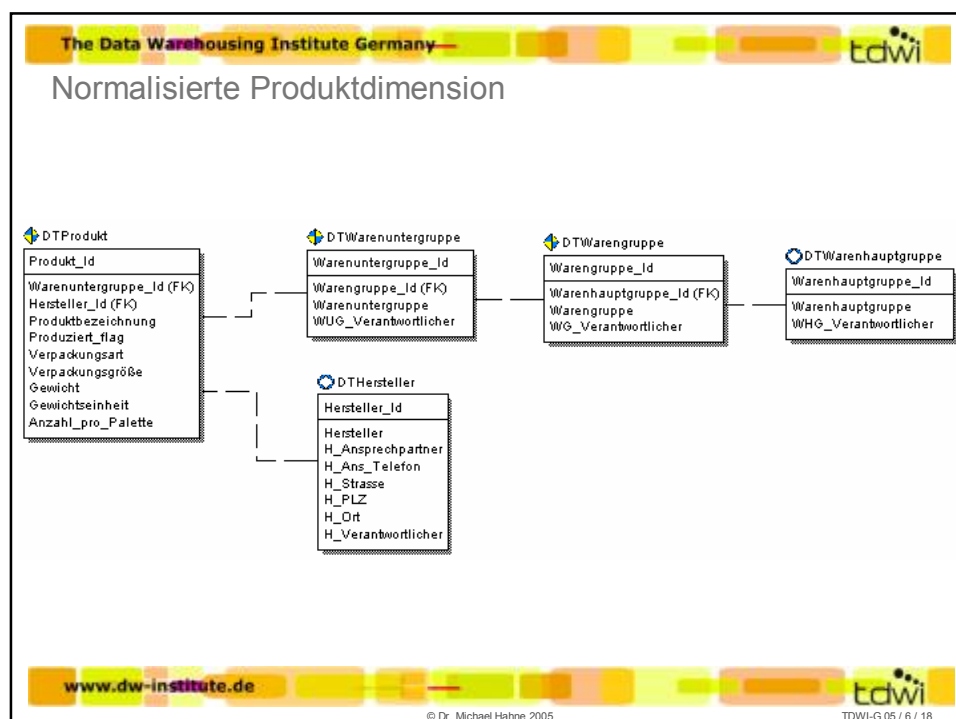
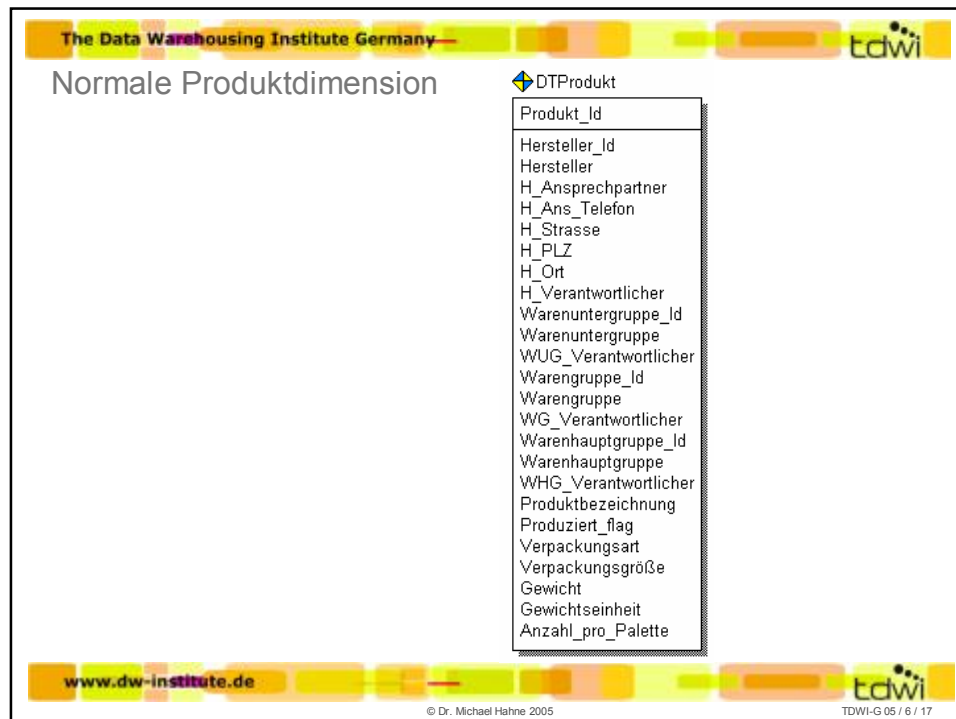
Attribute in Dimensionen

DTZeit

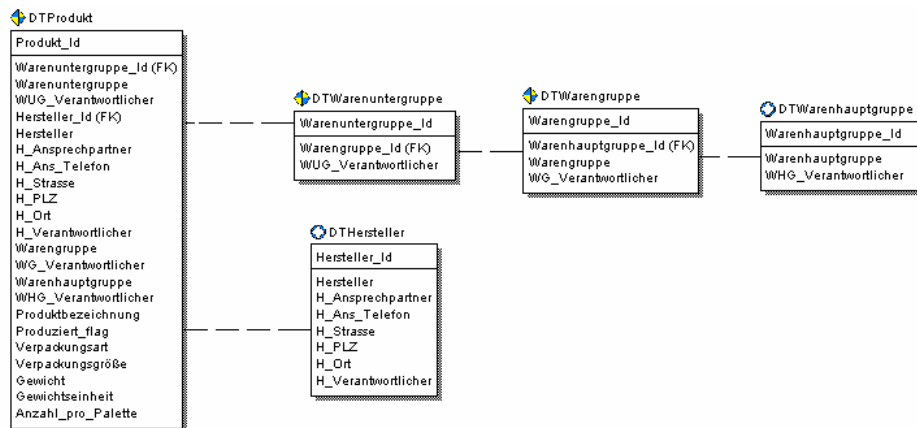
Zeit_Id
Tag
Wochentag
Monat
Quartal
Jahr
Geschäftsjahr
Werktag_flag
Ferien_flag
Jahreszeit
Ereignis
Letzter_Tag_im_Monat

Agenda

- Grundform des Star Schemas
- Modellierung von Dimensionshierarchien und Attributen
- **Normalisierung von Dimensionen**
- Kennzahlen und Kennzahlensysteme
- Aggregate
- Modellierungsvarianten



Partitionierte Produktdimension



Agenda

- Grundform des Star Schemas
- Modellierung von Dimensionshierarchien und Attributen
- Normalisierung von Dimensionen
- **Kennzahlen und Kennzahlensysteme**
- Aggregate
- Modellierungsvarianten

Kennzahlen und Kennzahlensysteme

- Speicherung von Kennzahlen als Wertspalten in der Faktentabelle
 - Alle Kennzahlen sind isoliert und unabhängig voneinander
 - Berechnung ist Teil des ETL-Prozesses
- Alternativ Kennzahlen in Dimension
 - Künstliche Faktenspalte für Kennzahlenwert in Faktentabelle
 - Bei nicht additiven Abhängigkeiten mit gespeicherten Aggregaten

FTMarketing

Zeit_Id (FK)
Produkt_Id (FK)
Vertriebsweg_Id (FK)

Absatz
Umsatz

Agenda

- Grundform des Star Schemas
- Modellierung von Dimensionshierarchien und Attributen
- Normalisierung von Dimensionen
- Kennzahlen und Kennzahlensysteme
- **Aggregate**
- Modellierungsvarianten

- Dynamische Berechnung:
 - Abfrage über SQL
 - Typischerweise viele Datensätze der Faktentabelle betroffen
 - Join aller beteiligten Tabellen
- Vorberechnete Aggregate in der Faktentabelle:
 - Qualifikation der Aggregattupel über entsprechende Dimensionsschlüssel
 - Zusätzliches Level-Attribut hilfreich

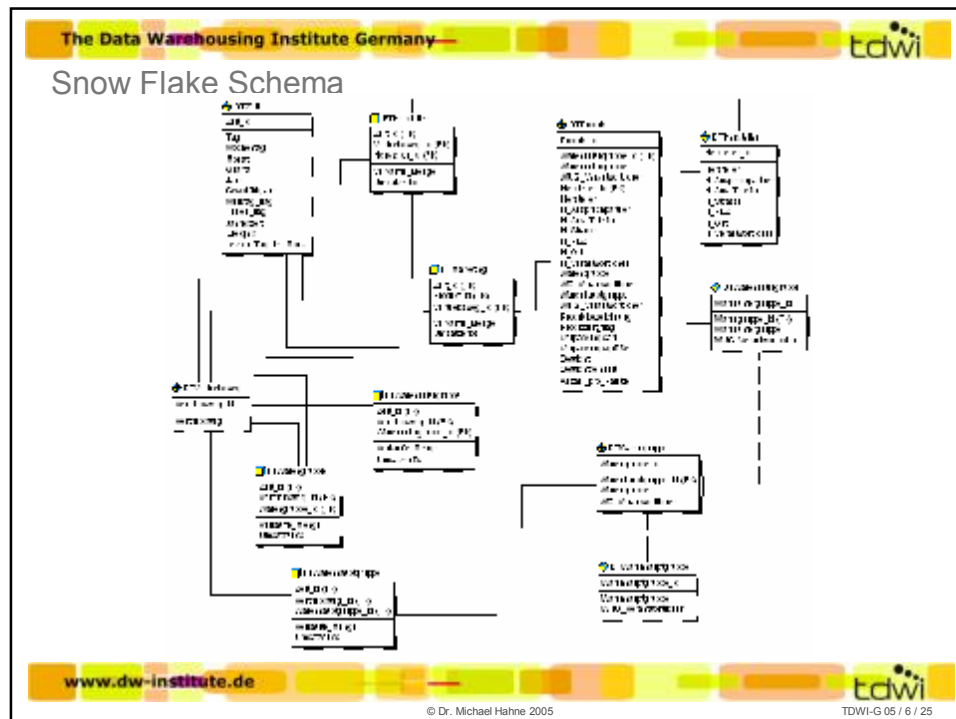
Zeit_Id	Monat	Quartal	Jahr
1	Januar	Q1	2004
2	Februar	Q1	2004
3	März	Q1	2004
4	April	Q2	2004
...	...	...	...
12	Dezember	Q4	2004
13	Januar	Q1	2005
14	Februar	Q1	2005
...	...	...	...
24	Dezember	Q4	2005
101	NULL	Q1	2004
...	...	...	...
104	NULL	Q4	2004
105	NULL	Q1	2005
...	...	...	...
108	NULL	Q4	2005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

The diagram illustrates the relationships between five tables in a database schema:

- DTZeit** (Time Data): Attributes include Zeit_Id (primary key), Tag, Wochentag, Monat, Quartal, Jahr, Geschäftsjahr, Wektag_flag, Ferien_flag, Jahreszeit, Ereignis, and Letzter_Tag_im_Monat.
- FTM_P** (Product Data): Attributes include Zeit_Id (FK), Produkt_Id (FK), Filiale_Id (FK), Verkaufte_Menge, and Umsatzerlös.
- DTProdukt** (Product Data): Attributes include Produkt_Id (primary key), Produktbezeichnung, Warenuntergruppe, Warengruppe_Id, Warengruppe, Warenhauptgruppe_Id, Warenhauptgruppe, Hersteller, Produziert_flag, Verpackungsart, Verpackungsgröße, Gewicht, Gewichtseinheit, and Anzahl_pro_Palette.
- DTFiliale** (Branch Data): Attributes include Filiale_Id (primary key), Filialname, Filiale_Straße, Filiale_Postleitzahl, Filiale_Ort, Filiale_Stadtteil, Filiale_Kreis_Id, Filiale_Kreis, Filiale_Bundesland, Filiale_Größe, Filiale_Eröffnungsdatum, and Filialkonzern.
- FTM_P_Warenggruppe** (Product Group Data): Attributes include Zeit_Id (FK), Filiale_Id (FK), Warengruppe_Id, Verkaufte_Menge, and Umsatzerlös.
- FTM_P_Warenhauptgruppe** (Product Main Group Data): Attributes include Zeit_Id (FK), Filiale_Id (FK), Warenhauptgruppe_Id, Verkaufte_Menge, and Umsatzerlös.
- FTM_P_Warenggruppe_Kreis** (Product Group Circle Data): Attributes include Zeit_Id (FK), Filiale_Kreis_Id, Warengruppe_Id, Verkaufte_Menge, and Umsatzerlös.

Relationships are indicated by lines connecting attributes between tables:

- DTZeit** to **FTM_P**: Zeit_Id to Zeit_Id (FK).
- DTZeit** to **FTM_P_Warenggruppe**: Zeit_Id to Zeit_Id (FK).
- DTZeit** to **FTM_P_Warenhauptgruppe**: Zeit_Id to Zeit_Id (FK).
- DTProdukt** to **FTM_P**: Produkt_Id to Produkt_Id (FK).
- DTProdukt** to **FTM_P_Warenhauptgruppe**: Warenhauptgruppe_Id to Warenhauptgruppe_Id (FK).
- DTFiliale** to **FTM_P**: Filiale_Id to Filiale_Id (FK).
- DTFiliale** to **FTM_P_Warenggruppe**: Filiale_Id to Filiale_Id (FK).
- DTFiliale** to **FTM_P_Warenhauptgruppe**: Filiale_Id to Filiale_Id (FK).
- FTM_P_Warenggruppe** to **FTM_P_Warenhauptgruppe**: Warengruppe_Id to Warenhauptgruppe_Id.
- FTM_P_Warenggruppe** to **FTM_P_Warenggruppe_Kreis**: Warengruppe_Id to Warengruppe_Id.



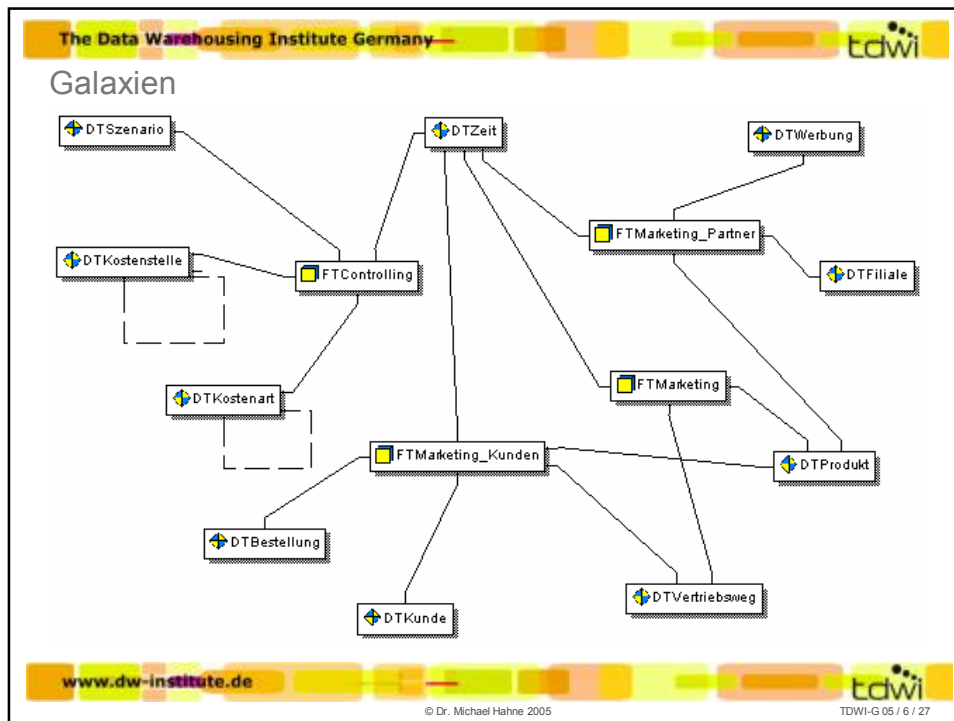
The Data Warehousing Institute Germany — **tdwi**

Agenda

- Grundform des Star Schemas
- Modellierung von Dimensionshierarchien und Attributen
- Normalisierung von Dimensionen
- Kennzahlen und Kennzahlensysteme
- Aggregate
- Modellierungsvarianten

www.dw-institute.de **tdwi**

© Dr. Michael Hahne 2005 TDWI-G 05 / 6 / 26



The Data Warehousing Institute Germany — edwi

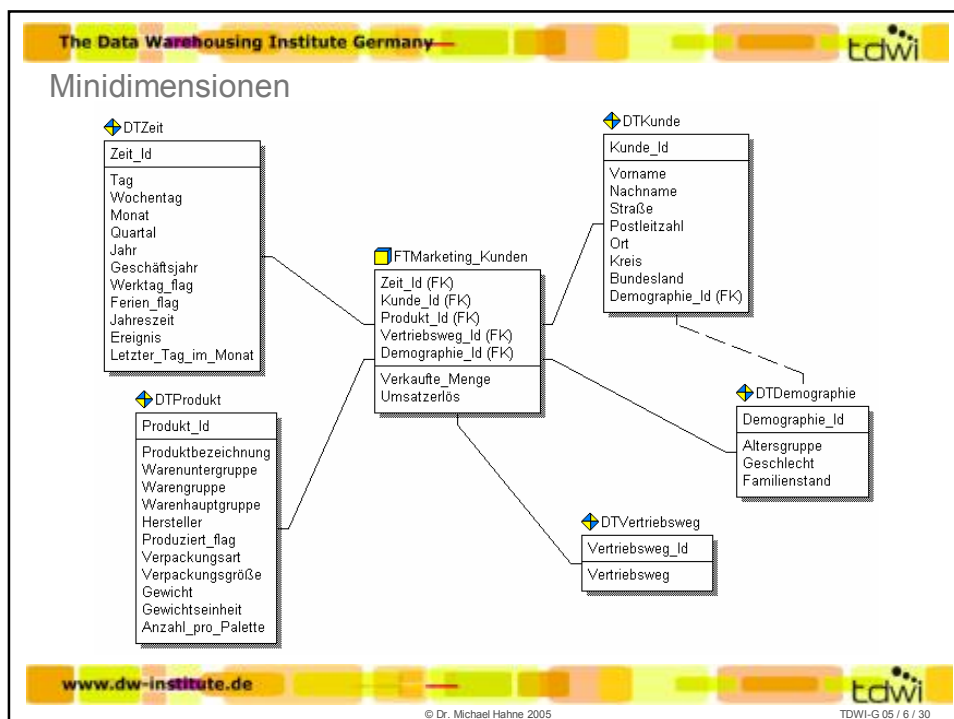
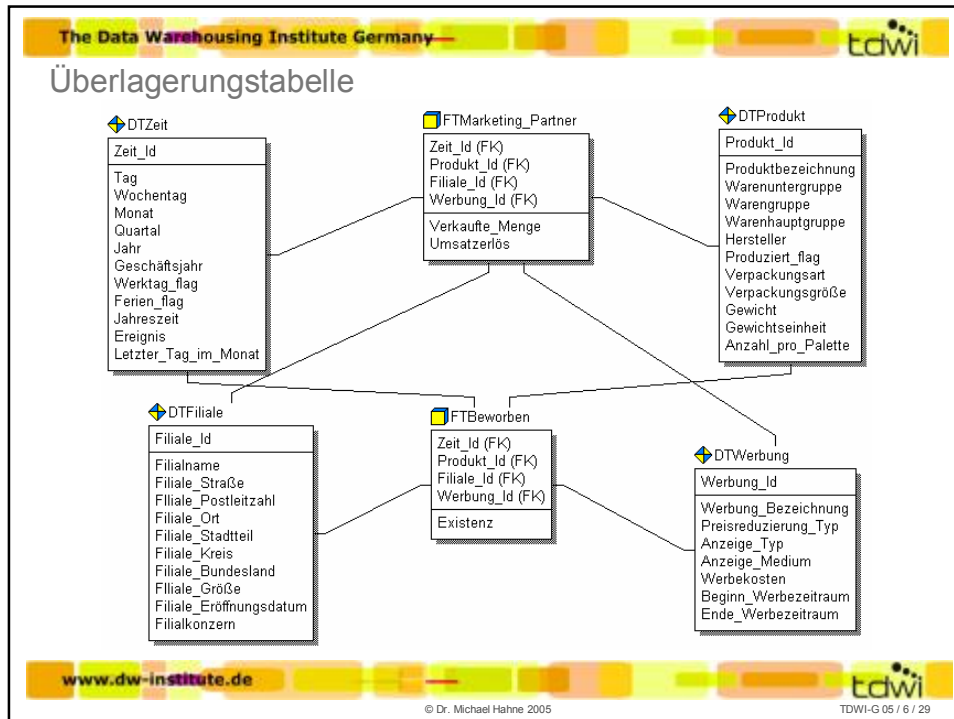
Faktenlose Faktentabellen

- Factless Fact Tables: Faktentabellen, die gar keine Kennzahl beinhalten
- Dienen der Verfolgung von Ereignissen (Vorkommen der Schlüsselkombination, oftmals ergänzt um künstliches Zählfaktum in der Faktentabelle)
- Ebenfalls nicht eingetretene Ereignisse (z.B. beworbene aber nicht verkaufte Produkte)

Beispiel: Wirksamkeit von Werbemaßnahmen

- Lift: Ist der Umsatz beworbener Produkte höher als üblich?
- Time shifting: Fallen die Umsätze der beworbenen Produkte nach der Werbung wieder auf das vorherige Niveau?
- Cannibalization: Verdrängt der Mehrumsatz eines beworbenen Produktes den Umsatz anderer Produkte?

www.dw-institute.de © Dr. Michael Hahne 2005 TDWI-G 05 / 6 / 28



Hierarchien über rekursive Beziehungen

DTKostenstelle

Kostenstelle_Id
Kostenstelle
Kostenstellenverantwortlicher
Gültig_von
Gültig_bis
Übergeordnete_Kostenstelle (FK)

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

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

Temporale Aspekte / Slowly Changing Dimensions

- Kimball:
 - Keine Historisierung
 - Einzelne Versionen
 - Erste und aktuelle Ausprägung

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

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

Agenda

- 1) Abgrenzung und Einordnung von CPM und BI
- 2) Architekturen und Komponenten analyseorientierter Informationssysteme
- 3) Bestandteile multidimensionaler Datenstrukturen
- 4) Methoden der semantischen Modellierung mehrdimensionaler Datenstrukturen
- 5) Umsetzung mit multidimensionalen Datenbanksystemen
- 6) Varianten und Bestandteile des Star Schemas
- 7) Erweitertes Star Schema des SAP Business Information Warehouse**
- 8) Mehrdimensionales Datenmodell des Microsoft SQL Server
- 9) Zusammenfassende Gegenüberstellung relationaler und multidimensionaler Modellierung

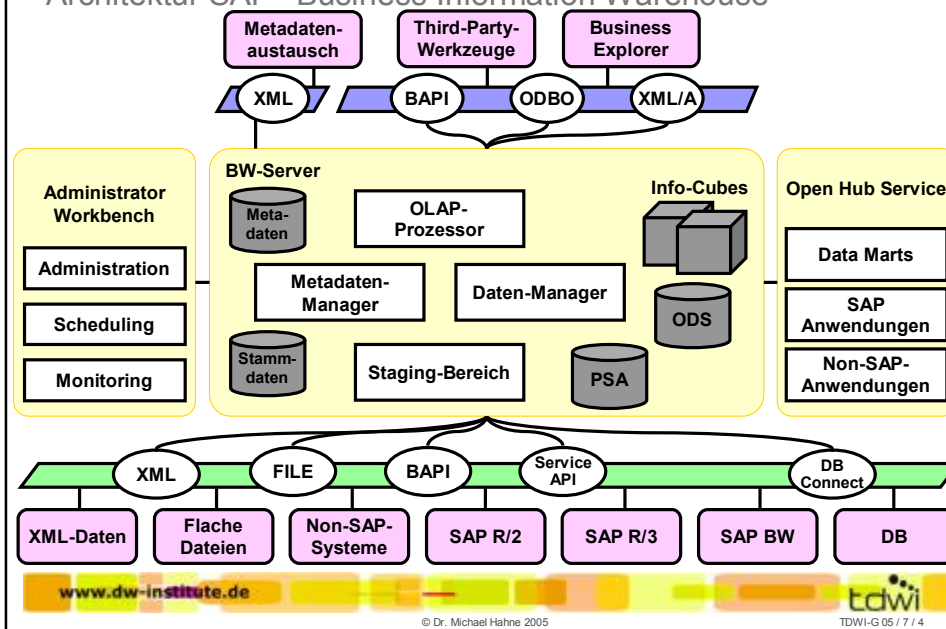
Agenda

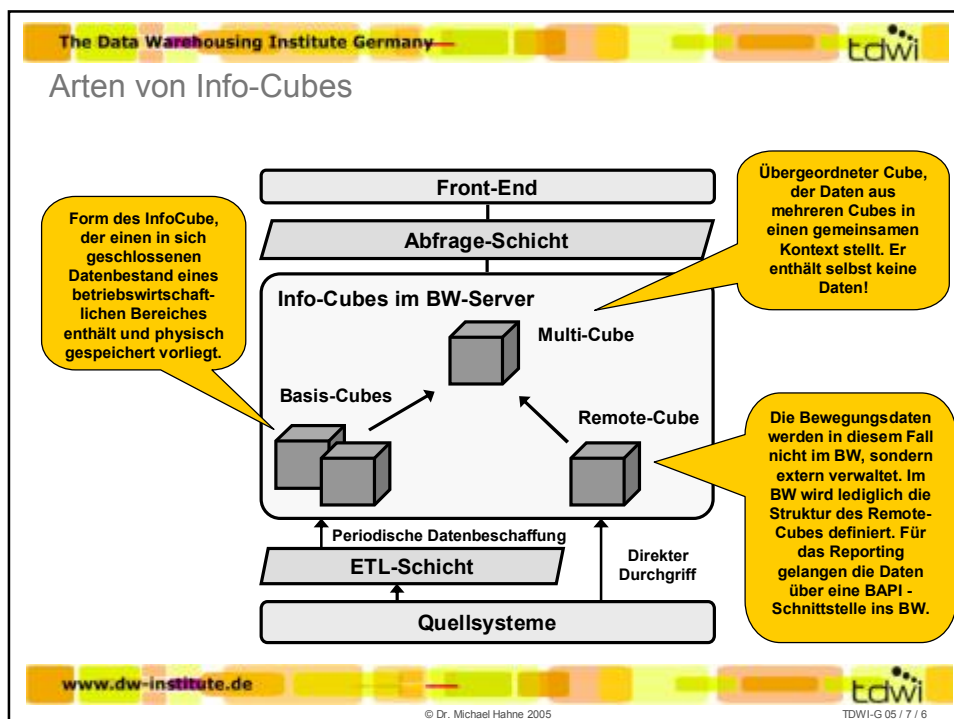
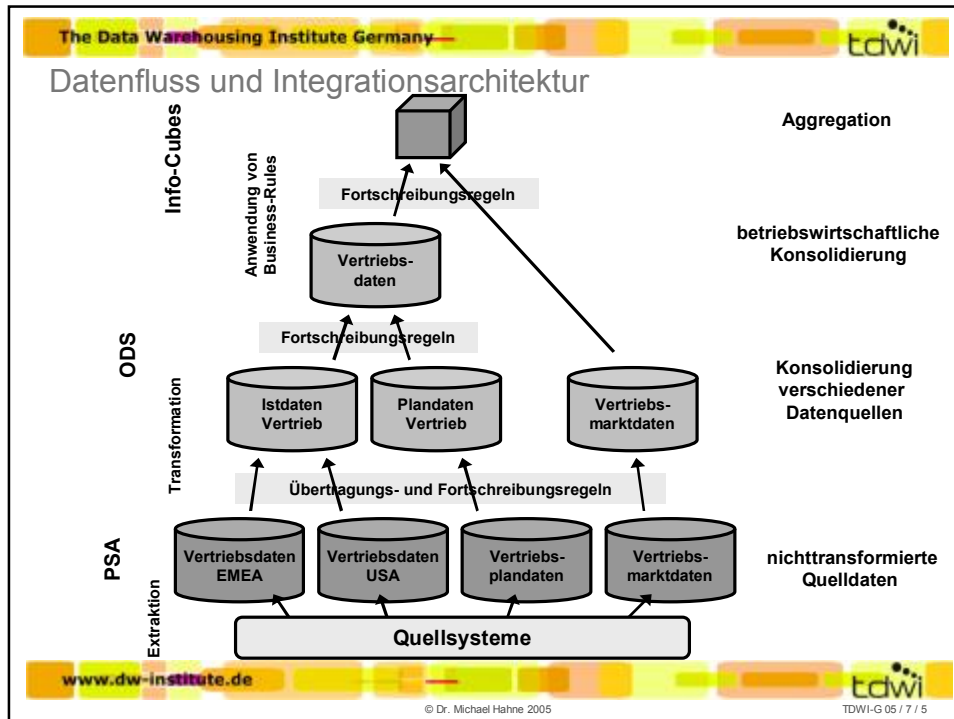
- Architektur des Business Information Warehouse der SAP AG
- Erweitertes Star Schema der SAP AG
- Modellierungsvarianten von hierarchischen Dimensionsstrukturen
- Zeitabhängigkeit und Versionierung
- Gestaltungsempfehlungen für die Modellierung
- Grafische Modellrepräsentation

Agenda

- Architektur des Business Information Warehouse der SAP AG
- Erweitertes Star Schema der SAP AG
- Modellierungsvarianten von hierarchischen Dimensionsstrukturen
- Zeitabhängigkeit und Versionierung
- Gestaltungsempfehlungen für die Modellierung
- Grafische Modellrepräsentation

Architektur SAP® Business Information Warehouse





Agenda

- Architektur des Business Information Warehouse der SAP AG
- **Erweitertes Star Schema der SAP AG**
- Modellierungsvarianten von hierarchischen Dimensionsstrukturen
- Zeitabhängigkeit und Versionierung
- Gestaltungsempfehlungen für die Modellierung
- Grafische Modellrepräsentation

Datenziele im SAP® BW

Datenziele sind Objekte, in denen Bewegungsdaten für Reporting- und Analysezwecke gehalten werden. Die Wichtigsten sind:

Info-Cubes

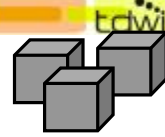


ODS-Objekte



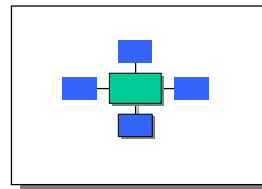
Daneben gibt es im BW noch weitere Datenziele, die u. a. ein direktes Stammdaten-Reporting ermöglichen (Info-Sets, Info-Objekte)

Definition Info-Cube



- Der InfoCube ist ein **zentraler Datenspeicher**, auf dem Berichte und Analysen im SAP® BW basieren. Er enthält einen in sich geschlossenen Datenbestand z. B. eines betriebswirtschaftlichen Bereichs oder einer Geschäftseinheit.
- InfoCubes enthalten zwei Datentypen: **Kennzahlen** und **Merkmale**.
- Der Begriff „InfoCube“ bezeichnet eine Tabellenstruktur, in der eine Anzahl relationaler Tabellen im Sinne des sog. Star Schema-Datenmodells miteinander verknüpft sind. (multi-dimensionale Datenspeicherung)
- Star Schema: Dimensionstabellen gruppieren sich sternförmig um eine zentrale Faktentabelle.

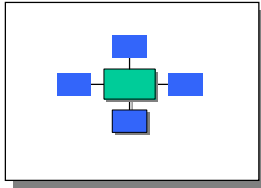
Star Schema



- Das Star Schema ist die häufigste Art der Abbildung mehrdimensionaler Datenstrukturen in relationalen Datenbanken.
- Im Star Schema werden die Fakten in einer separaten **Faktentabelle** gespeichert, während die Merkmale in **Dimensionstabellen** gruppiert werden. Die Dimensionstabellen sind über Fremdschlüssel- / Primärschlüsselbeziehungen (DIM ID) mit der Faktentabelle verbunden.
- Auf diese Weise werden alle Datensätze aus der Faktentabelle **eindeutig** durch eine Wertekombination dieser Fremdschlüssel aus den Dimensionstabellen gekennzeichnet.

Vor- und Nachteile des allgemeinen Star Schemas

- ☺ Hohe Performance bei der Analyse von Daten
- ☺ Hohe Flexibilität beim Hinzufügen von Merkmalen und Kennzahlen.

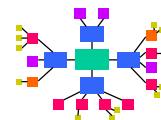


- ☹ Aufgrund der Eindeutigkeit des Schlüssels in den Dimensionstabellen können keine
 - N : M Beziehungen und
 - nur schlecht unbalancierte Hierarchien dargestellt werden.

Deshalb entschied sich die SAP AG zur Erweiterung des Star – Schemas. Beim **erweiterten Star Schema** werden Stammdaten separat und unabhängig von den InfoCubes gespeichert.

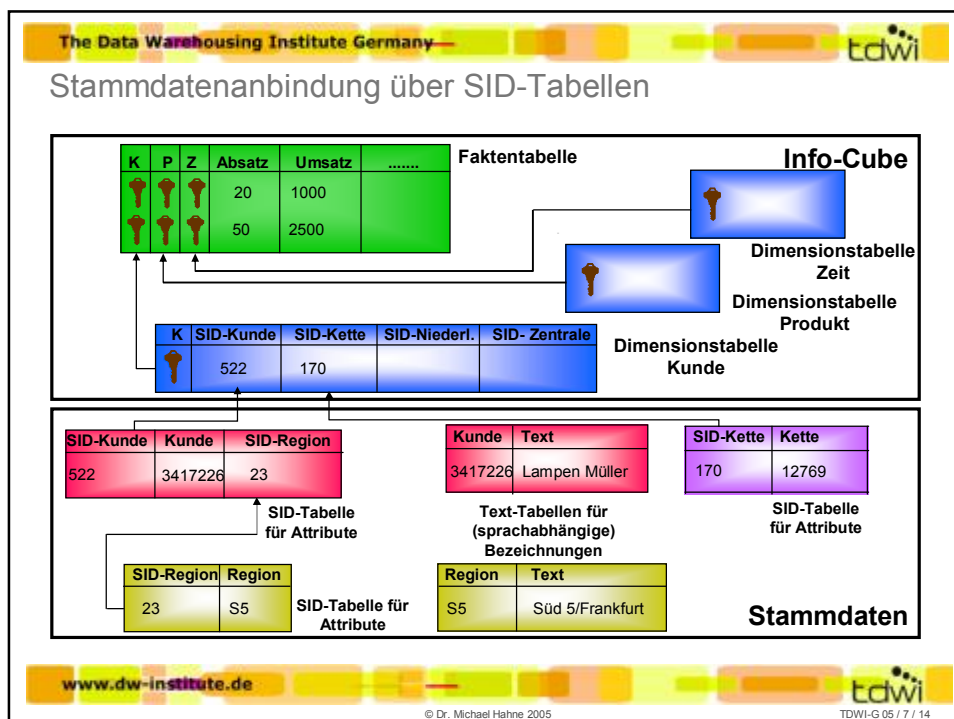
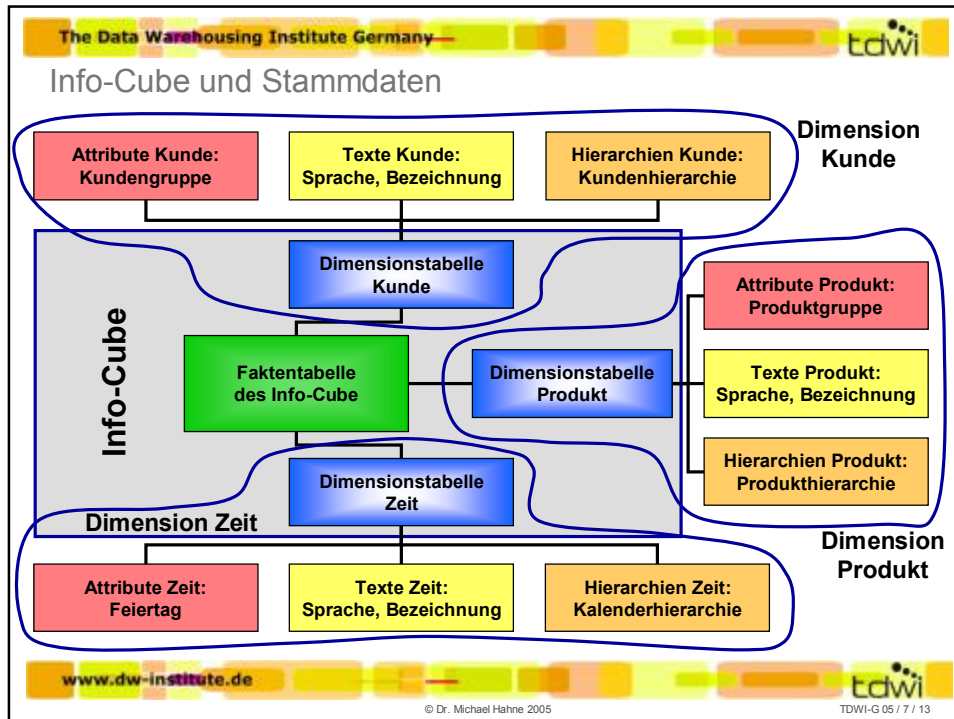
Erweitertes Star Schema

Das erweiterte Star Schema ermöglicht Zugriff auf:



- Stammdatentabellen und ihre zugehörigen Attribute
- Texttabellen mit umfassenden mehrsprachigen Beschreibungen
- Externe Hierarchietabellen für den strukturierten Datenzugriff

SID-Tabellen (Zeigertabellen) liefern die technische Verknüpfung zu den Stammdaten- und Hierarchietabellen außerhalb der Dimensionstabellen eines Star-Schemas.



Verschiedene Stammdaten-Tabellen

K	SID-Kunde	SID-Kette	SID-Niederl.	SID-Zentrale
	522	170		

Dimensionstabelle
Kunde

S	SID-Kunde	Kunde
	522	3417226

SID-Tabelle BIC/SKunde
Standard SID-Tabelle

P	Kunde	Kundenname
	3417226	Lampen Müller

Stammdaten-Tabelle BIC/PKunde
für nicht-zeitabhängige
Anzeige-Attribute

Q	Kunde	DateFrom	DateTo	Region
	3417226	...	...	Süd 5/Frankfurt

Stammdaten-Tabelle BIC/QKunde
für zeitabhängige
Anzeige-Attribute

X	SID-Kunde	Kunde	SID-Region
	522	3417226	23

SID-Tabelle BIC/XKunde
für nicht-zeitabhängige
Navigations-Attribute

Y	SID-Kunde	Kunde	DateFrom	DateTo	SID-Klasse
	522	3417226	...	...	12

SID-Tabelle BIC/YKunde
für zeitabhängige
Navigations-Attribute

Hierarchietabellen

Dim-ID	SID-Kunde	SID-Kette	SID-Niederl.	SID-Zentrale
	522	170	...	...

Dimensions-Tabelle Kunde

S	SID-Kunde	Kunde
	522	3417226

SID-Tabelle /BIC/SKunde
Standard SID-Tabelle

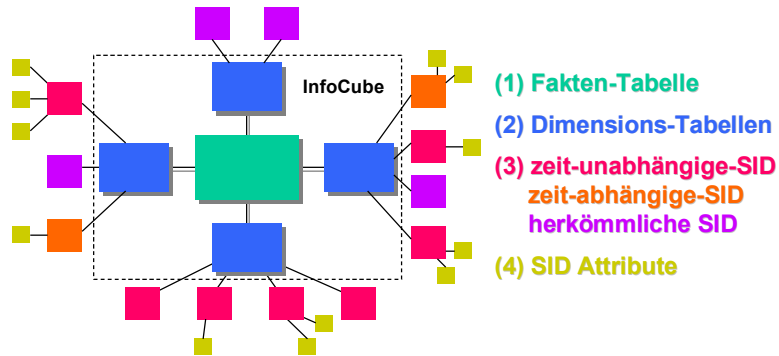
I	SID der Hierarchie	pred	succ
	234	-21	522

Hierarchietabelle /BIC/IKunde
Parent-Child-Tupel der Hierarchien

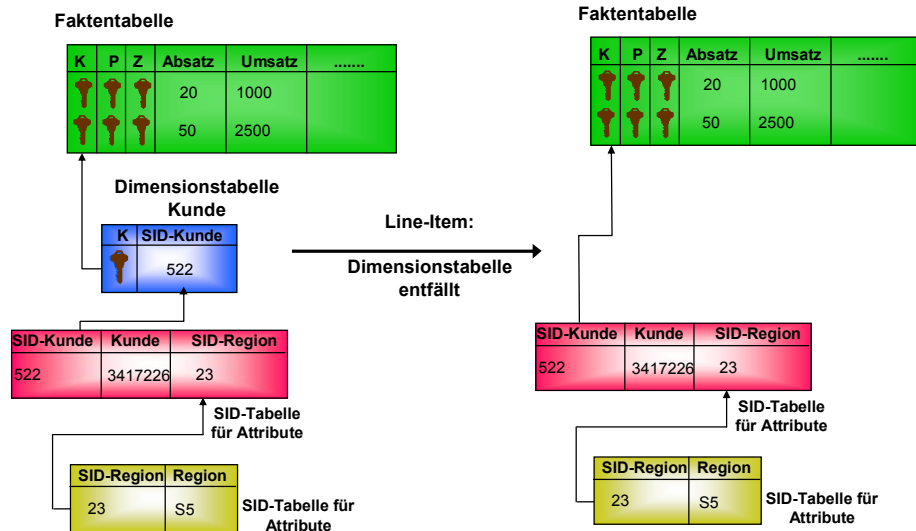
K	SID der Hierarchie	Knoten	SID
	234	Kundengruppe A	-21

Hierarchietabelle /BIC/KKunde
Textknoten der Hierarchien

Komplexität des erweiterten Star Schemas im Überblick



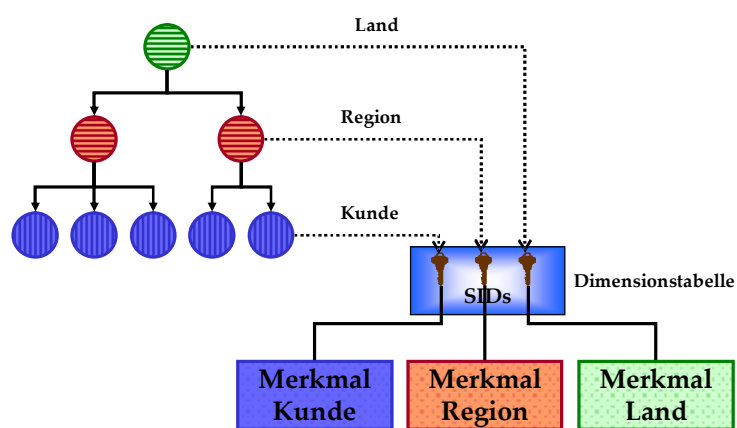
Line-Item Dimensionen



Agenda

- Architektur des Business Information Warehouse der SAP AG
- Erweitertes Star Schema der SAP AG
- **Modellierungsvarianten von hierarchischen Dimensionsstrukturen**
- Zeitabhängigkeit und Versionierung
- Gestaltungsempfehlungen für die Modellierung
- Grafische Modellrepräsentation

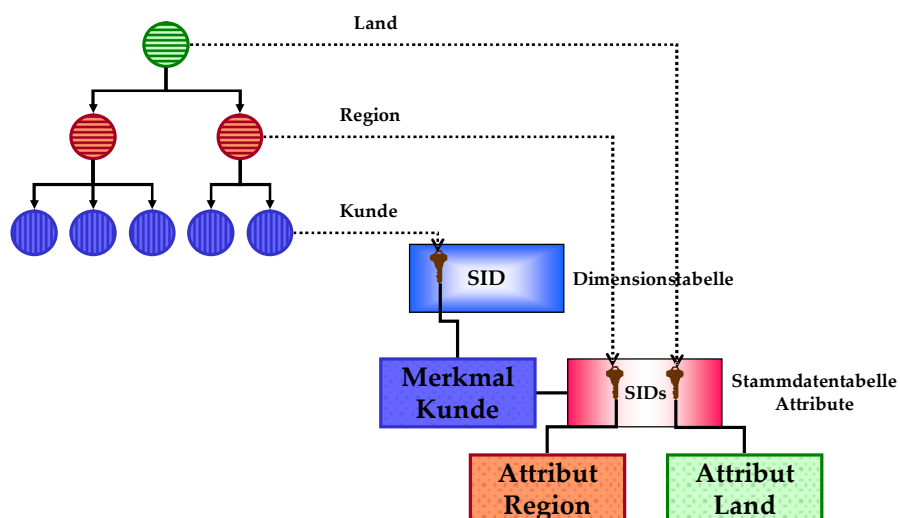
Hierarchische Strukturen über Merkmale



Hierarchien in Dimensionen / Eigenschaften

- Ein Info-Objekt je level → balancierte Struktur
- Schnellste Variante zur Abbildung von hierarchischen Strukturen
- Übergeordnete Attribute sind in Aggregate mit einzuschließen
- Keine vordefinierten Navigationspfade im BEx

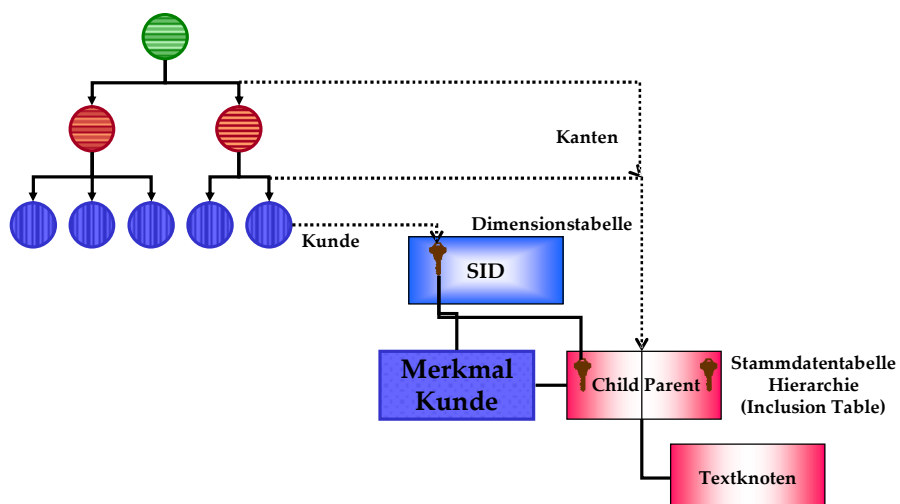
Navigationsattribute als Basis für hierarchische Strukturen



Hierarchien über Attribute in Stammdaten / Eigenschaften

- Ein Info-Objekt je level → balancierte Struktur
- Übergeordnete Attribute sind in Aggregate mit einzuschließen
- Keine vordefinierten Navigationspfade im Bex
- Schlechte Performance
- Hohe Flexibilität zur Reorganisation

Externe Hierarchien im BW



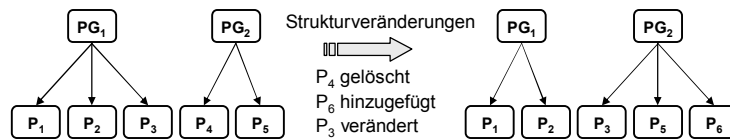
Externe Hierarchien / Eigenschaften

- Gut bei häufigen Strukturänderungen
- Unbalancierte Strukturen
- Mehrere Hierarchien pro Info-Objekt abbildbar und selektierbar
- Schlechte Performance ähnlich Navigationsattribute
- große Hierarchien problematisch
- Aggregate bei Zeitabhängigkeit nur bei gesamter Struktur zeitabhängig

Agenda

- Architektur des Business Information Warehouse der SAP AG
- Erweitertes Star Schema der SAP AG
- Modellierungsvarianten von hierarchischen Dimensionsstrukturen
- **Zeitabhängigkeit und Versionierung**
- Gestaltungsempfehlungen für die Modellierung
- Grafische Modellrepräsentation

Zeitabhängigkeit: Strukturelle Veränderungen in Dimensionen



Beispiel für "Slowly Changing Dimensions"

Produktdimension in 2003-04

Produkt	Produktgruppe
P A	PG X
P B	PG X
P C	PG Y
P D	PG Y

Produktdimension in 2003-05

Produkt	Produktgruppe
P A	PG X
P B	PG Y (geändert)
P C	PG Y
P D	PG Y
P E	PG Y (neu)

Fakten-Tabelle

Produkt	Periode	Umsatz
P A	2003-04	100
P B	2003-04	100
P C	2003-04	100
P D	2003-04	100
P A	2003-05	100
P B	2003-05	100
P C	2003-05	100
P D	2003-05	100
P E	2003-05	100

Berichtsanforderungen - Szenarien

Berichtsszenario aktuelle Struktur

Produktgruppe	Umsatz 2003-04	Umsatz 2003-05
PG X	100	100
PG Y	300	400

Berichtsszenario alte Struktur

Produktgruppe	Umsatz 2003-04	Umsatz 2003-05
PG X	200	200
PG Y	200	200

Berichtsszenario historische Wahrheit

Produktgruppe	Umsatz 2003-04	Umsatz 2003-05
PG X	200	100
PG Y	200	400

Berichtsszenario vergleichbare Resultate

Produktgruppe	Umsatz 2003-04	Umsatz 2003-05
PG X	100	100
PG Y	200	200

Szenario I : Bericht mit aktueller Struktur

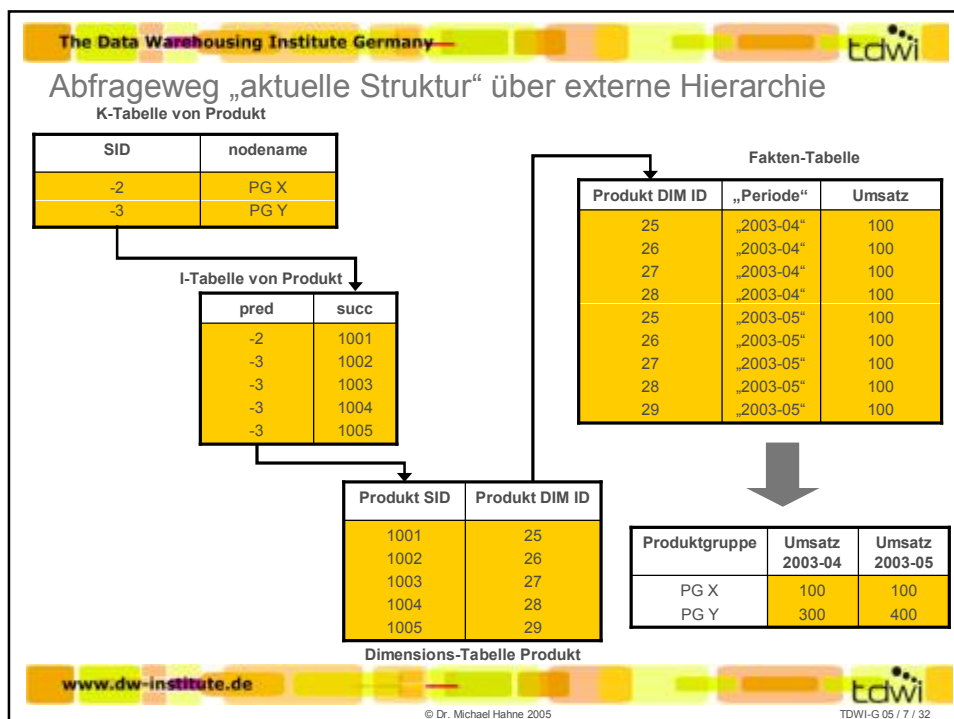
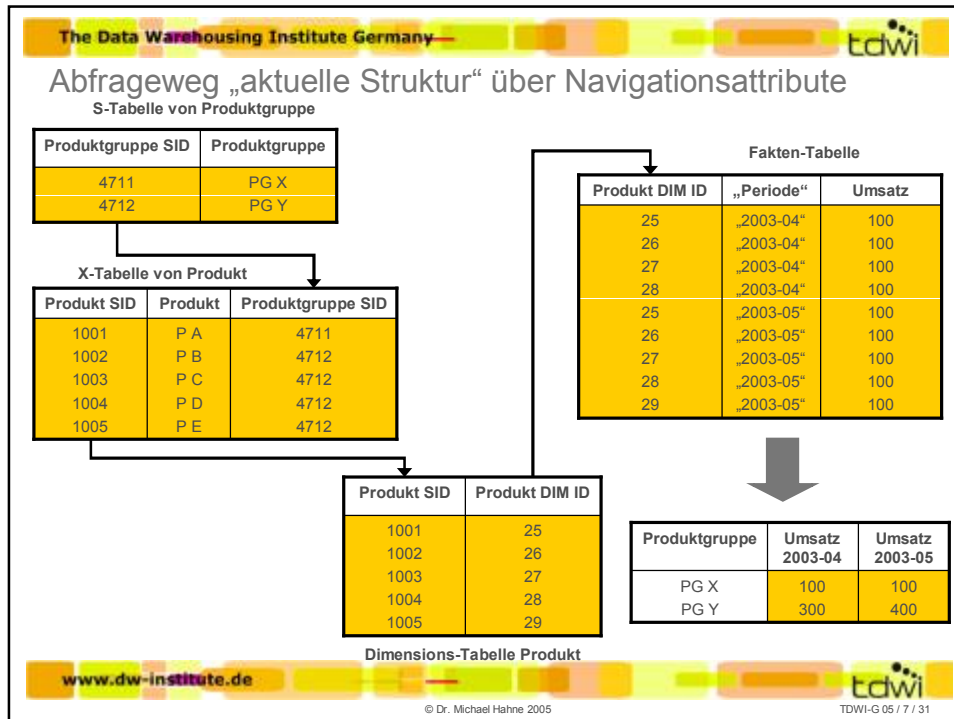
Produktdimension in 2003-05

Produkt	Produktgruppe
P A	PG X
P B	PG Y (geändert)
P C	PG Y
P D	PG Y
P E	PG Y (neu)

Fakten-Tabelle

Produkt	Periode	Umsatz
P A	2003-04	100
P B	2003-04	100
P C	2003-04	100
P D	2003-04	100
P A	2003-05	100
P B	2003-05	100
P C	2003-05	100
P D	2003-05	100
P E	2003-05	100

Produktgruppe	Umsatz 2003-04	Umsatz 2003-05
PG X	100	100
PG Y	300	400



Szenario II : Bericht mit alter Struktur

Fakten-Tabelle

Produkt	Periode	Umsatz
P A	2003-04	100
P B	2003-04	100
P C	2003-04	100
P D	2003-04	100
P A	2003-05	100
P B	2003-05	100
P C	2003-05	100
P D	2003-05	100
P E	2003-05	100

Produktdimension in 2003-04

Produkt	Produktgruppe
P A	PG X
P B	PG X
P C	PG Y
P D	PG Y

Produktgruppe	Umsatz 2003-04	Umsatz 2003-05
PG X	200	200
PG Y	200	200

Abfrageweg „alte Struktur“ über zeitabh. Navigationsattribute

S-Tabelle von Produktgruppe

Produktgruppe SID	Produktgruppe
4711	PG X
4712	PG Y

Query key date
2003-04

Fakten-Tabelle

Produkt DIM ID	„Periode“	Umsatz
25	„2003-04“	100
26	„2003-04“	100
27	„2003-04“	100
28	„2003-04“	100
25	„2003-05“	100
26	„2003-05“	100
27	„2003-05“	100
28	„2003-05“	100
29	„2003-05“	100

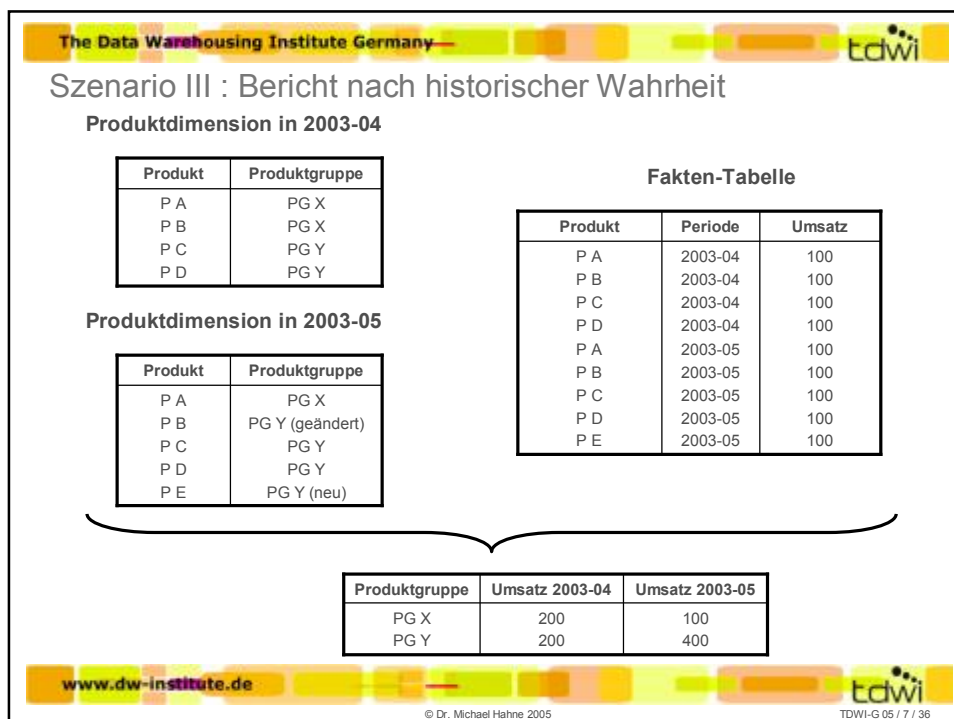
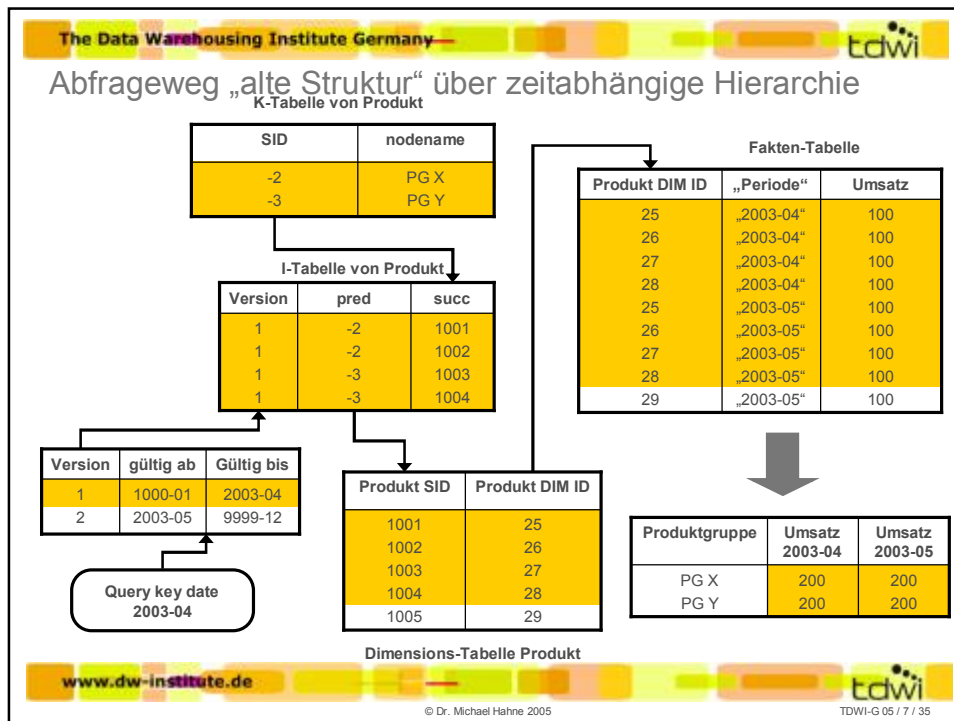
Y-Tabelle von Produkt

Produkt SID	Produkt	Produktgruppe SID	gültig ab	gültig bis
1001	P A	4711	1000-01	9999-12
1002	P B	4711	1000-01	2003-04
1002	P B	4712	2003-05	9999-12
1003	P C	4712	1000-01	9999-12
1004	P D	4712	1000-01	9999-12
1005	P E	4712	2003-05	9999-12

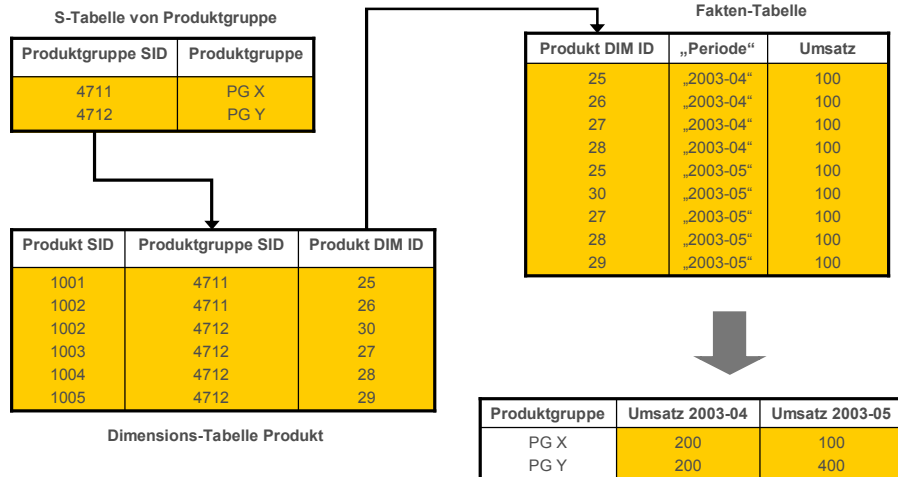
Produkt SID	Produkt DIM ID
1001	25
1002	26
1003	27
1004	28
1005	29

Dimensions-Tabelle Produkt

Produktgruppe	Umsatz 2003-04	Umsatz 2003-05
PG X	200	200
PG Y	200	200



Abfrageweg „historische Wahrheit“ über Merkmale



Szenario IV : Bericht mit vergleichbaren Resultaten

Produktdimension in 2003-04

Produkt	Produktgruppe
P A	PG X
P B	PG X
P C	PG Y
P D	PG Y

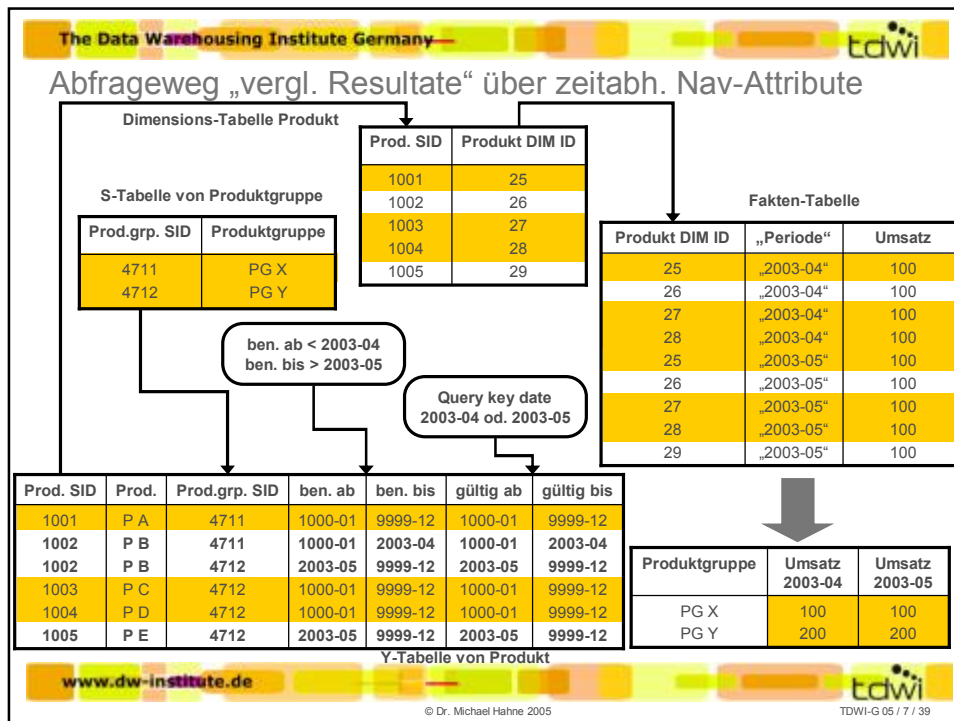
Produktdimension in 2003-05

Produkt	Produktgruppe
P A	PG X
P B	PG Y (geändert)
P C	PG Y
P D	PG Y
P E	PG Y (neu)

Fakten-Tabelle

Produkt	Periode	Umsatz
P A	2003-04	100
P B	2003-04	100
P C	2003-04	100
P D	2003-04	100
P A	2003-05	100
P B	2003-05	100
P C	2003-05	100
P D	2003-05	100
P E	2003-05	100

Produktgruppe	Umsatz 2003-04	Umsatz 2003-05
PG X	100	100
PG Y	200	200



The Data Warehousing Institute Germany — tdwi

Agenda

- Architektur des Business Information Warehouse der SAP AG
- Erweitertes Star Schema der SAP AG
- Modellierungsvarianten von hierarchischen Dimensionsstrukturen
- Zeitabhängigkeit und Versionierung
- **Gestaltungsempfehlungen für die Modellierung**
- Grafische Modellrepräsentation

www.dw-institute.de — tdwi

© Dr. Michael Hahne 2005 TDWI-G 05 / 7 / 40

Gestaltungsempfehlungen

- Modellierung von Dimensionsstrukturen
- Gestaltung von Info-Providern

Gestaltungsüberlegungen der Dimensionsmodellierung auf semantischer Ebene

- Anzahl der Dimensionen zwischen vier und zehn (ideal zwischen sechs und acht)
- Anzahl der Hierarchiestufen (maximal sieben Hierarchiestufen)
- Anzahl der Elemente je Konsolidierungselement (maximal fünfzehn bis zwanzig Elemente)
- Bestimmung von Dimensionen (1:1-Beziehungen ungeeignet; 1:N-Beziehungen bilden Dimensionshierarchie; M:N-Beziehungen stellen zwei Dimensionen dar)

Gestaltungsüberlegungen der Dimensionsmodellierung auf logischer und physischer BW-Ebene

- Merkmale mit vielen Ausprägungen als Line-Item modellieren
- Attribute, die sich häufig ändern, sollten als eigene Dimension modelliert werden (Line Item beachten!)
- Steigt die gesamte Dimensionsanzahl, sind ggf. Merkmale in eine Dimension zusammen zu fassen, die jeweils selbst wenige Ausprägungen haben
- Große Dimensionen möglichst eigene Dimension, Parent-Merkmale evtl. in eigene Dimension

Kriterien zur Entscheidungshilfe der logischen Modellierung von Dimensionsstrukturen im BW

- Historisierung
- Wirkungsbereich
- Performance
- Navigationspfade
- Unbalancierte Dimensionsstrukturen
- Blätter mit mehreren Parent-Elementen
- Strukturänderungen und Reorganisation

Hierarchie-Guideline: Historisierung

Externe Hierarchie	Hierarchie in Dimension über Merkmale	Hierarchie über Attribute
• Sicht zum Zeitpunkt des Ladens nicht vorhanden • Sichten per Version oder Schlüsseldatum	• Nur Sicht zum Zeitpunkt des Ladens möglich	• Sicht zum Zeitpunkt des Ladens nicht vorhanden • Nur die gültige Sicht der Stammdaten zum Zeitpunkt der Auswertung

Hierarchie-Guideline: Wirkungsbereich

Externe Hierarchie	Hierarchie in Dimension über Merkmale	Hierarchie über Attribute
Hängt am Info-Object und gilt daher für dieses Info-Object in allen Info-Cubes	Nur im Info-Cube	Hierarchie hängt in den Stammdaten und gilt daher für dieses Info-Object in allen Info-Cubes

Hierarchie-Guideline: Performance

Externe Hierarchie	Hierarchie in Dimension über Merkmale	Hierarchie über Attribute
Aggregate werden unbedingt benötigt für performante Abfragen	Auch ohne Aggregate recht performant	Aggregate werden unbedingt benötigt für performante Abfragen

Hierarchie-Guideline: Navigations-Pfade

Externe Hierarchie	Hierarchie in Dimension über Merkmale	Hierarchie über Attribute
Vordefinierte Navigation entlang der Hierarchie-Struktur	Ebenen können übersprungen werden, da Navigationspfad nicht vordefiniert ist	Ebenen können übersprungen werden, da Navigationspfad nicht vordefiniert ist

Hierarchie-Guideline: unbalancierte Dimensionsstrukturen

Externe Hierarchie	Hierarchie in Dimension über Merkmale	Hierarchie über Attribute
Unbalancierte Strukturen sind abbildbar	Jedes Merkmal korrespondiert mit einer festgelegten Ebene der Hierarchie, also nur balancierte Strukturen	Jedes Attribut korrespondiert mit einer festgelegten Ebene der Hierarchie, also nur balancierte Strukturen

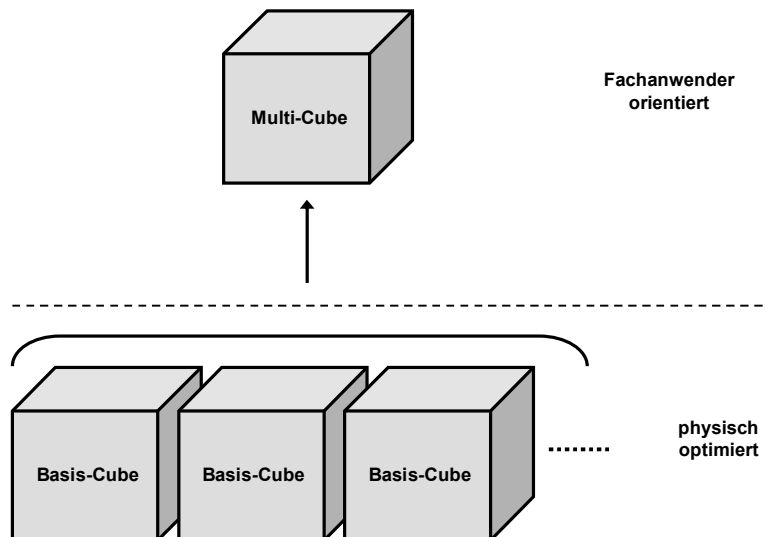
Hierarchie-Guideline: Blätter mit mehreren Parent-Elementen

Externe Hierarchie	Hierarchie in Dimension über Merkmale	Hierarchie über Attribute
m:n-Beziehungen zwischen Ebenen möglich	Nur in dem Maße, wie es in der Faktentabelle gebucht wurde	Nicht möglich

Hierarchie-Guideline: Strukturänderungen und Reorganisation

Externe Hierarchie	Hierarchie in Dimension über Merkmale	Hierarchie über Attribute
Schnelles Verändern der Hierarchie möglich	Reorganisation nicht möglich, ohne den Info-Cube neu zu laden	Reorganisation möglich durch zusätzliche Attribute

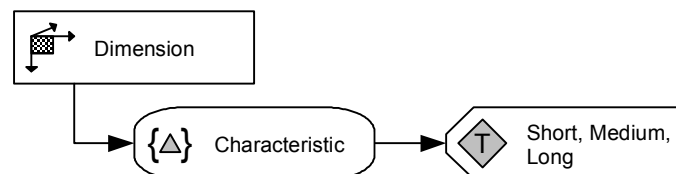
2-Layer Konzept der Cube-Modellierung



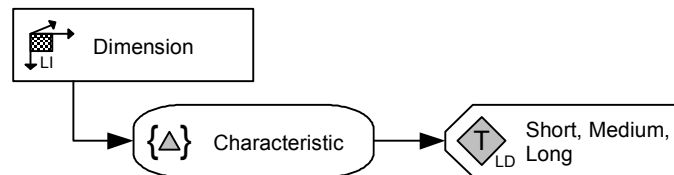
Agenda

- Architektur des Business Information Warehouse der SAP AG
- Erweitertes Star Schema der SAP AG
- Modellierungsvarianten von hierarchischen Dimensionsstrukturen
- Zeitabhängigkeit und Versionierung
- Gestaltungsempfehlungen für die Modellierung
- **Grafische Modellrepräsentation**

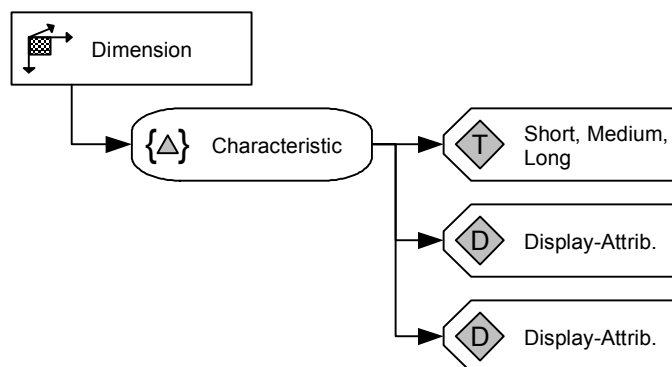
Dimension mit einem Merkmal



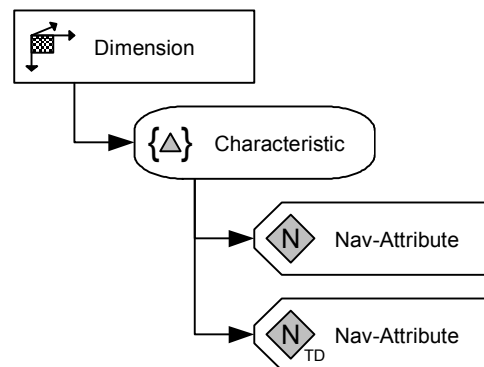
Line-Item-Dimension



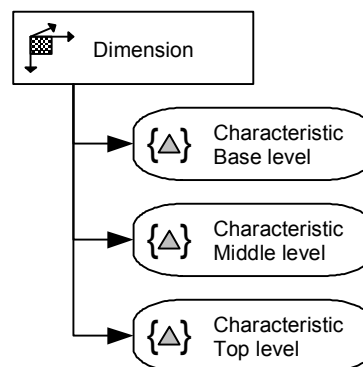
Anzeige-Attribute eines Merkmales



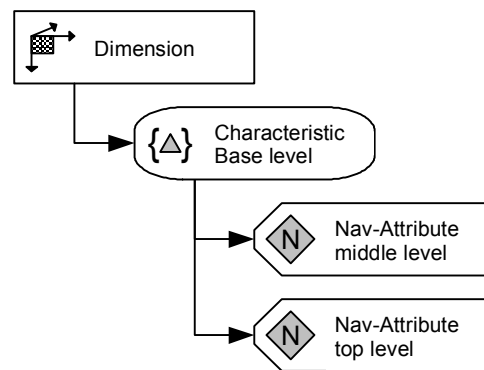
Navigationsattribute eines Merkmales



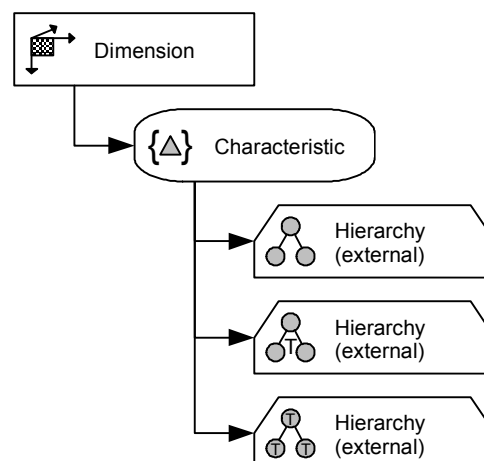
Hierarchie über Merkmale



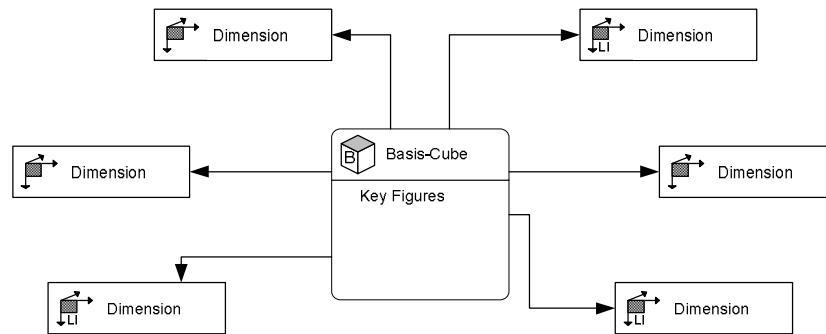
Hierarchie über Navigationsattribute



Externe Hierarchien



Modellierung von Basis-Cubes



Agenda

- 1) Abgrenzung und Einordnung von CPM und BI
- 2) Architekturen und Komponenten analyseorientierter Informationssysteme
- 3) Bestandteile multidimensionaler Datenstrukturen
- 4) Methoden der semantischen Modellierung mehrdimensionaler Datenstrukturen
- 5) Umsetzung mit multidimensionalen Datenbanksystemen
- 6) Varianten und Bestandteile des Star Schemas
- 7) Erweitertes Star Schema des SAP Business Information Warehouse
- 8) Mehrdimensionales Datenmodell des Microsoft SQL Server**
- 9) Zusammenfassende Gegenüberstellung relationaler und multidimensionaler Modellierung

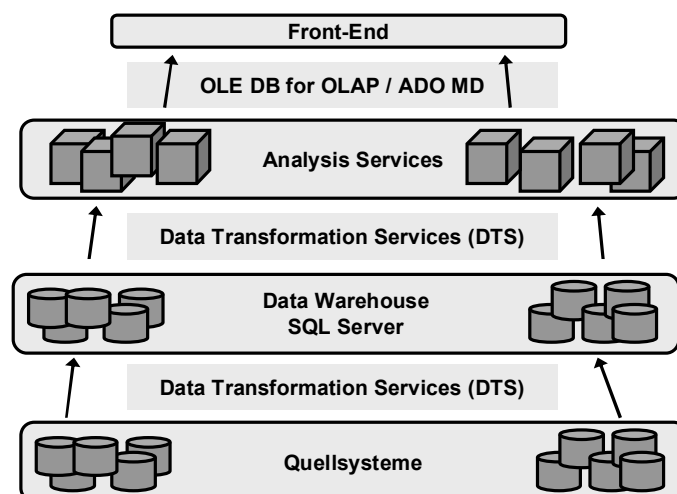
Agenda

- Microsoft SQL-Server 2000
 - Microsoft Data Warehouse Architektur
 - Dimensionsmodellierung mit den Microsoft Analysis Services
 - Attribute und Dimensionen
 - Modellierung von Cubes
 - Kennzahlen
- Microsoft SQL Server 2005
 - Unified Dimensional Model
 - Dimensionen

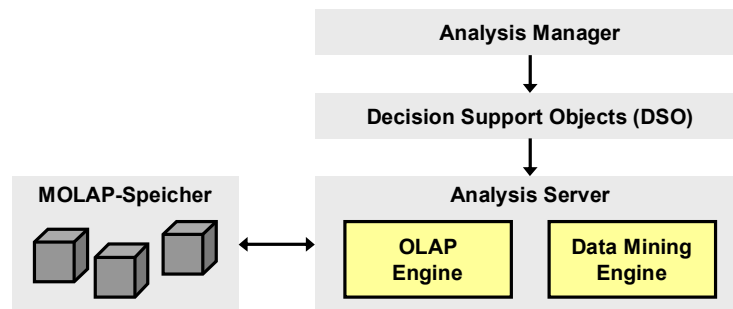
Agenda

- Microsoft SQL-Server 2000
 - **Microsoft Data Warehouse Architektur**
 - Dimensionsmodellierung mit den Microsoft Analysis Services
 - Attribute und Dimensionen
 - Modellierung von Cubes
 - Kennzahlen
- Microsoft SQL Server 2005
 - Unified Dimensional Model
 - Dimensionen

Architektur Microsoft Data Warehouse



Architektur Analysis Services

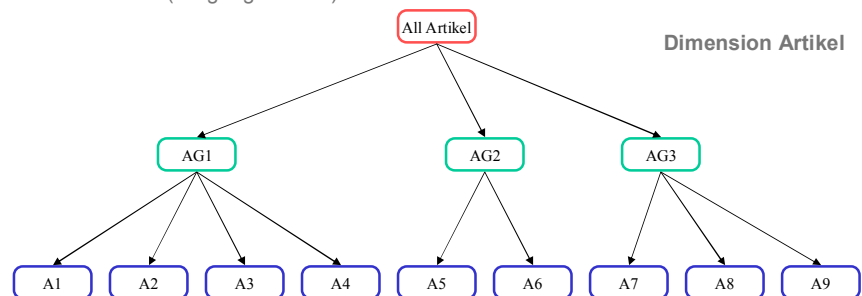


Agenda

- Microsoft SQL-Server 2000
 - Microsoft Data Warehouse Architektur
 - **Dimensionsmodellierung mit den Microsoft Analysis Services**
 - Attribute und Dimensionen
 - Modellierung von Cubes
 - Kennzahlen
- Microsoft SQL Server 2005
 - Unified Dimensional Model
 - Dimensionen

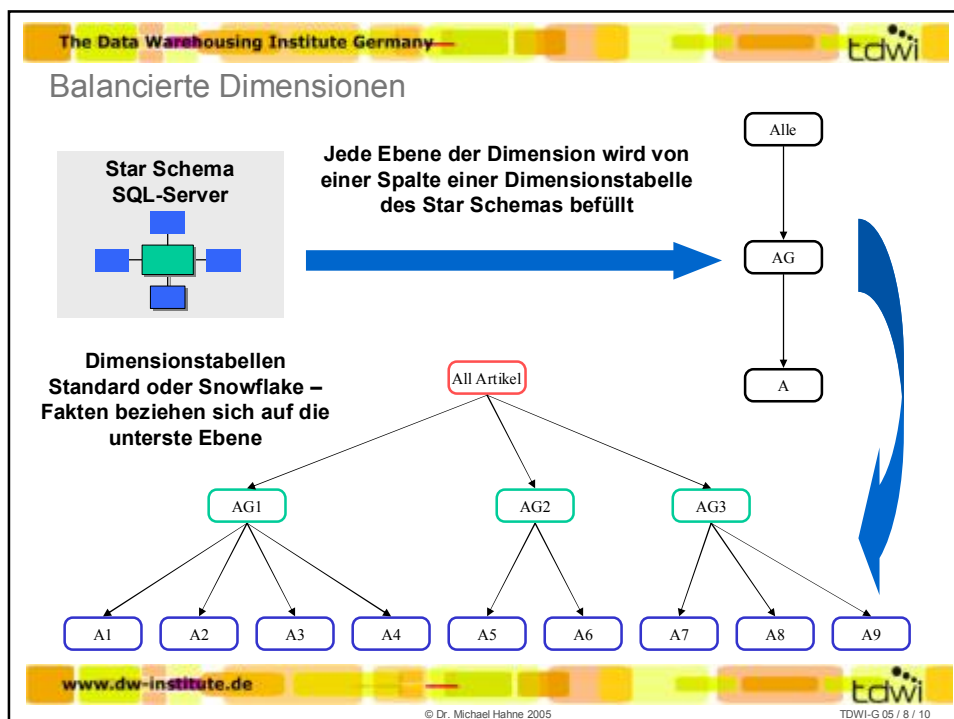
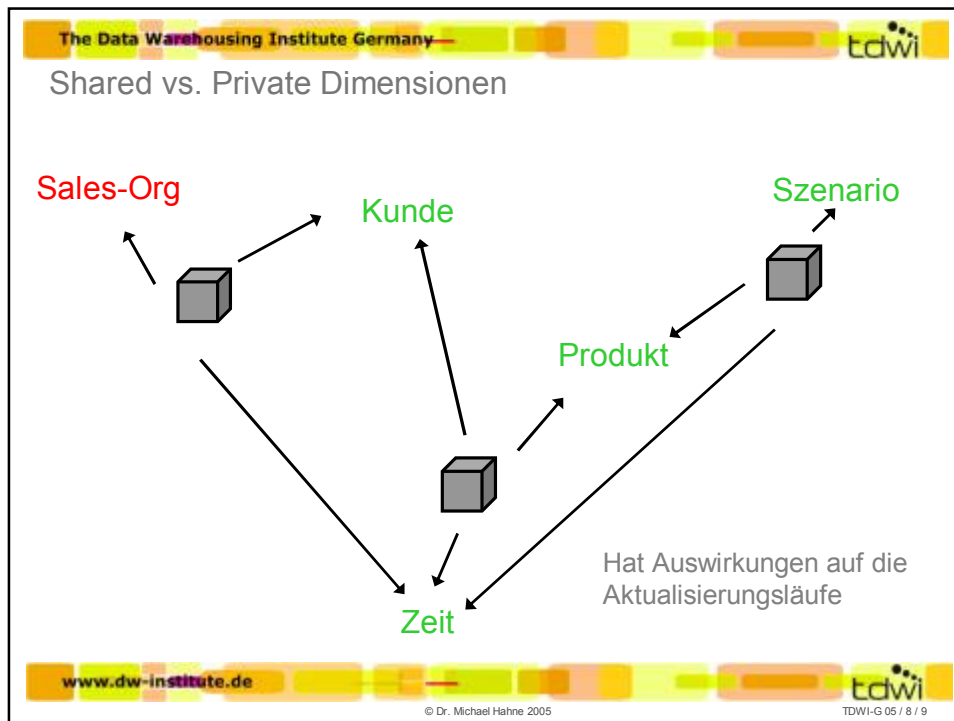
Dimensionen in den Analysis Services

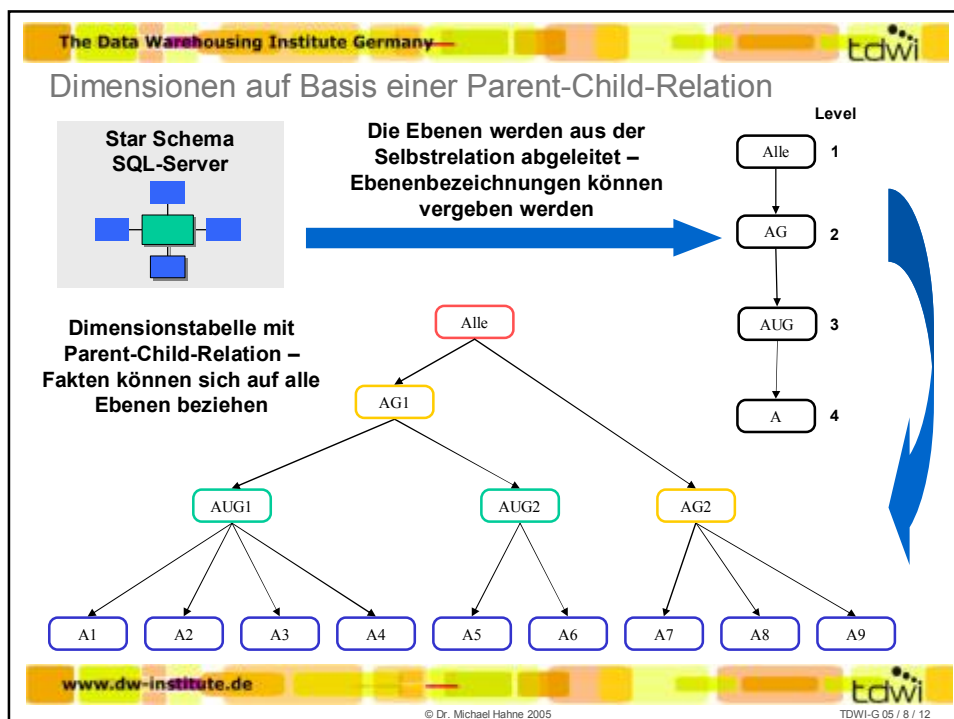
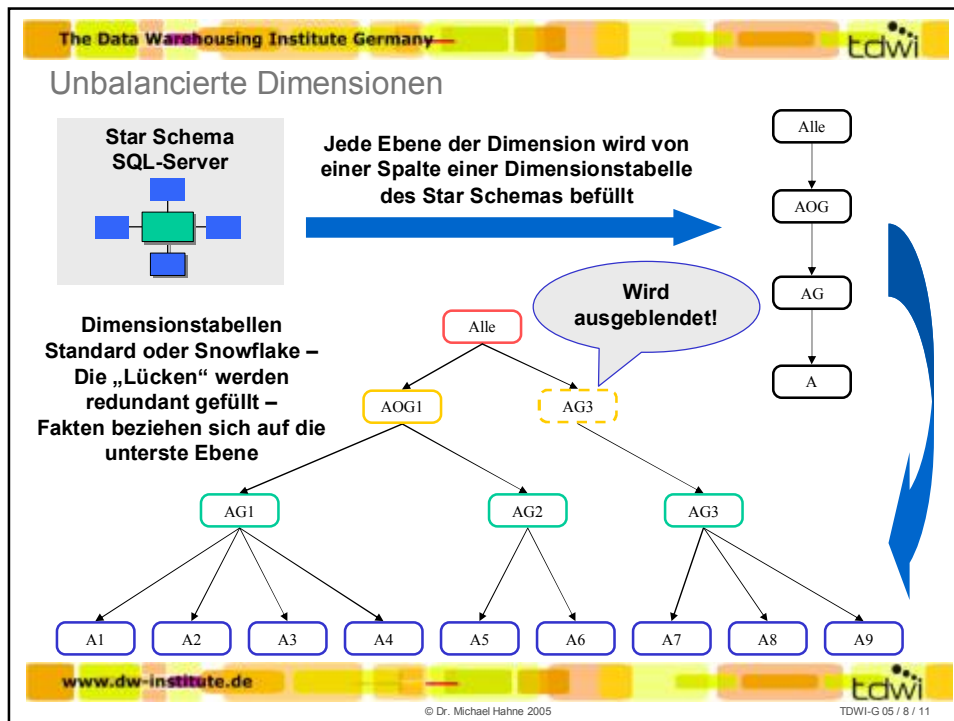
- Parents (übergeordnete Elemente)
- Siblings (Elemente mit dem gleichen übergeordneten Element)
- Children (untergeordnete Elemente)
- Cousins (Elemente einer Ebene mit verschiedenen übergeordneten Elementen)
- Descendants of x (Nachfolger von x)
- Ancestors of x (Vorgänger von x)



Technische Restriktionen

- 65535 Dimensionen pro Datenbank
- 65535 Ebenen pro Datenbank
- 128 Dimensionen pro Cube
- 256 Ebenen pro Cube
- 64 Ebenen pro Dimension
- 64000 untergeordnete Elemente je übergeordnetem Element

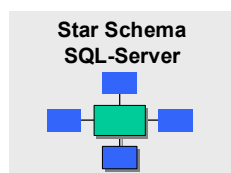




Agenda

- Microsoft SQL-Server 2000
 - Microsoft Data Warehouse Architektur
 - Dimensionsmodellierung mit den Microsoft Analysis Services
 - **Attribute und Dimensionen**
 - Modellierung von Cubes
 - Kennzahlen
- Microsoft SQL Server 2005
 - Unified Dimensional Model
 - Dimensionen

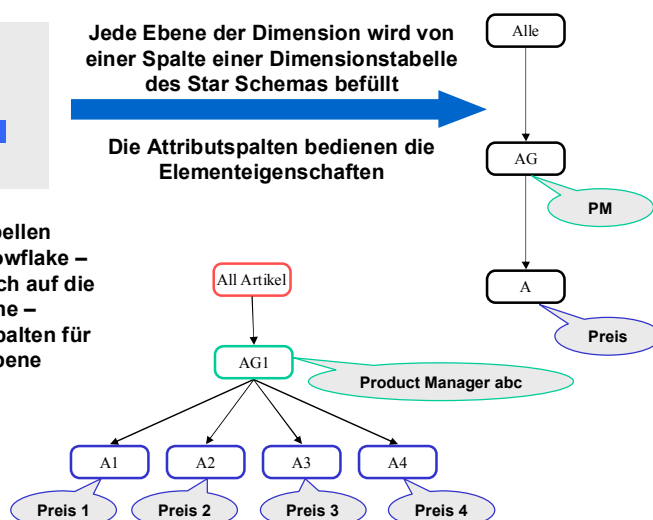
Attribute (Elementeigenschaften)



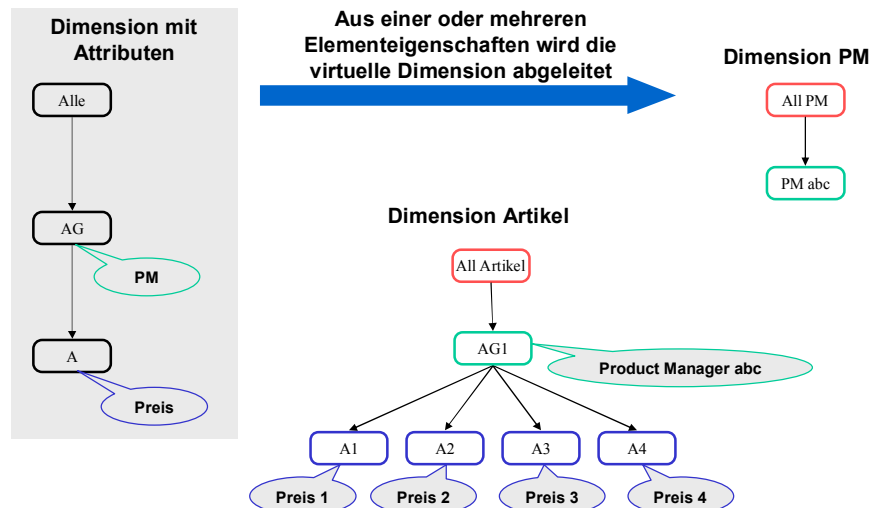
Jede Ebene der Dimension wird von einer Spalte einer Dimensionstabelle des Star Schemas befüllt

Die Attributspalten bedienen die Elementeigenschaften

Dimensionstabellen
Standard oder Snowflake –
Fakten beziehen sich auf die
unterste Ebene –
Mit zusätzlichen Spalten für
Attribute je Ebene



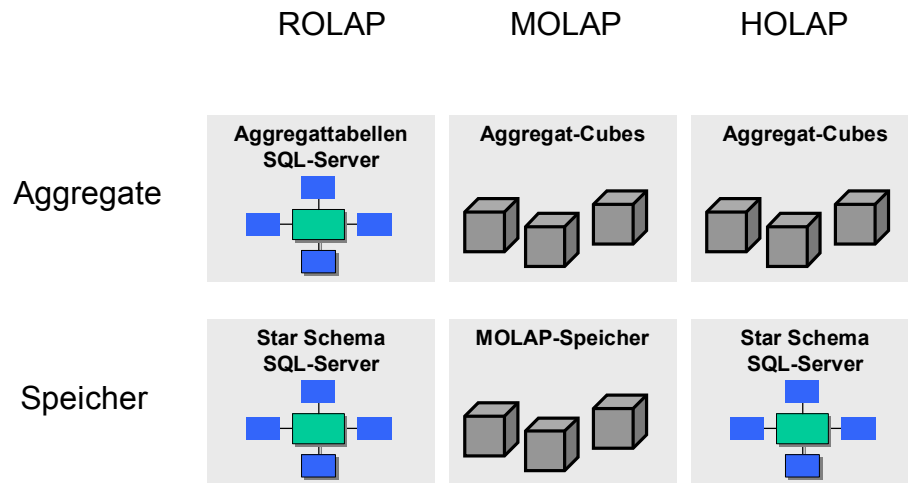
Virtuelle Dimensionen



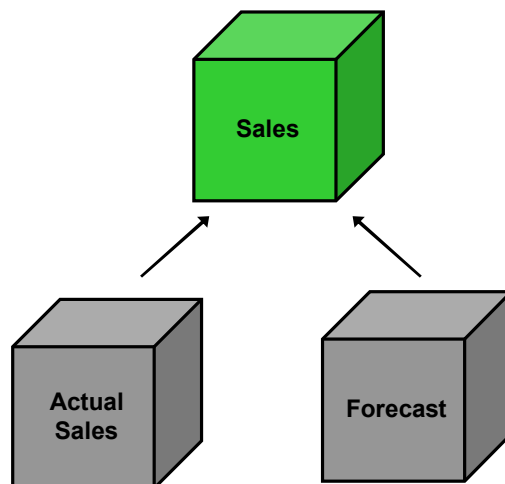
Agenda

- Microsoft SQL-Server 2000
 - Microsoft Data Warehouse Architektur
 - Dimensionsmodellierung mit den Microsoft Analysis Services
 - Attribute und Dimensionen
 - **Modellierung von Cubes**
 - Kennzahlen
- Microsoft SQL Server 2005
 - Unified Dimensional Model
 - Dimensionen

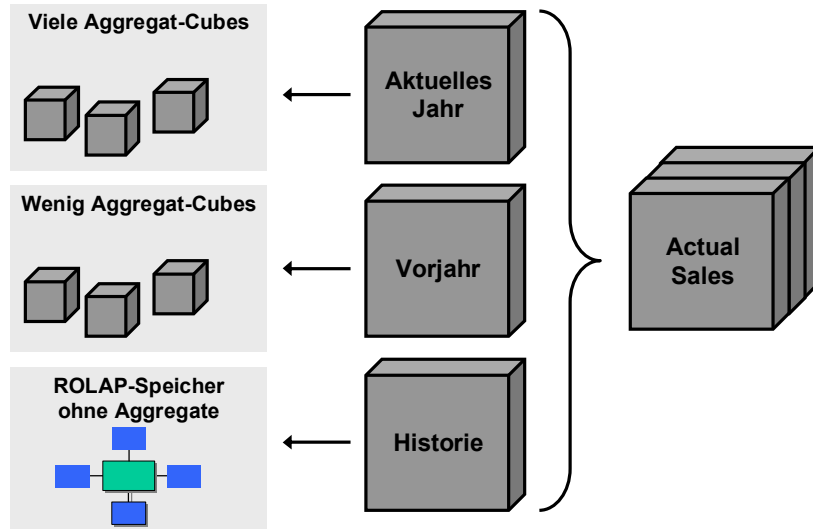
Speichervarianten



Virtuelle Cubes



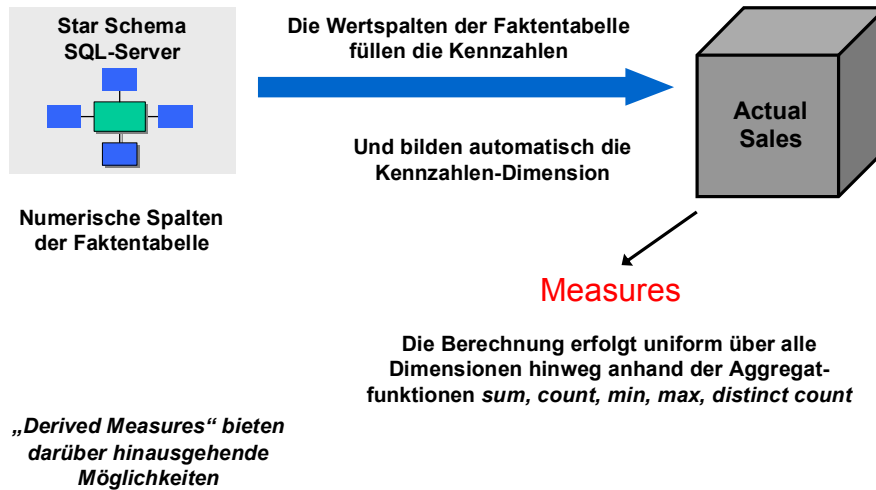
Partitionen



Agenda

- Microsoft SQL-Server 2000
 - Microsoft Data Warehouse Architektur
 - Dimensionsmodellierung mit den Microsoft Analysis Services
 - Attribute und Dimensionen
 - Modellierung von Cubes
 - **Kennzahlen**
- Microsoft SQL Server 2005
 - Unified Dimensional Model
 - Dimensionen

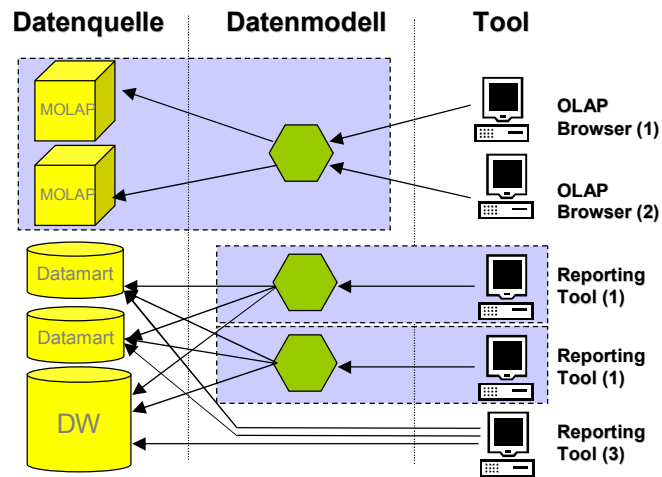
Modellierung von Kennzahlen



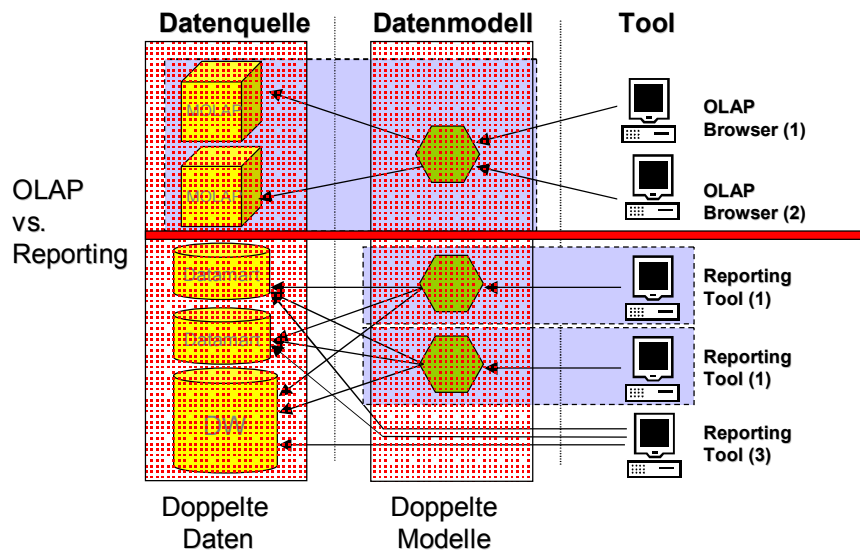
Agenda

- Microsoft SQL-Server 2000
 - Microsoft Data Warehouse Architektur
 - Dimensionsmodellierung mit den Microsoft Analysis Services
 - Attribute und Dimensionen
 - Modellierung von Cubes
 - Kennzahlen
- Microsoft SQL Server 2005
 - **Unified Dimensional Model**
 - Dimensionen

Aktueller Aufbau unternehmensweiter BI-Lösungen



Probleme unternehmensweiter BI-Lösungen



Relationales vs. OLAP-Reporting

Feature	Relational	OLAP
Flexible schema	✓	✗
Real time data access	✓	✗
Single data store	✓	✗
Simple management	✓	✗
Detail reporting	✓	✗
High performance	✗	✓
End-user oriented	✗	✓
Ease of navigation and exploration	✗	✓
Rich analytics	✗	✓
Rich semantics	✗	✓

The Unified Dimensional Model (UDM)

Relational Reporting	OLAP Cubes
• Multiple fact tables • Full richness the dimensions' attributes • Transaction level access • Star, snowflake, 3NF... • Complex relationships: Multi-grains, many-to-many, role playing, indirect... • Recursive self joins • Slowly changing dimensions	• Multidimensional navigation • Hierarchical presentation • Friendly entity names • Powerful MDX calculations • Central KPI framework • "Actions" • Language translations • Multiple perspectives • Partitions • Aggregations • Distributed sources

The Unified Dimensional Model (UDM)

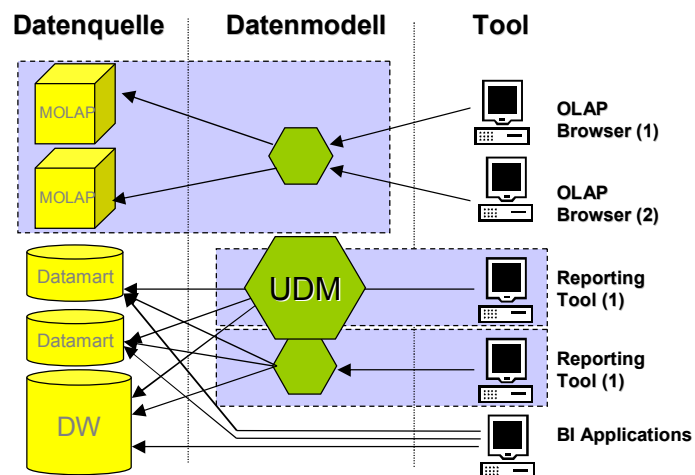
Relational Reporting

- Multiple fact tables
- Full richness the dimensions' attributes
- Transaction level access
- Star, snowflake, 3NF...
- Complex relationships: Multi-grains, many-to-many, role playing, indirect
- Recursive self joins
- Slowly changing dimensions

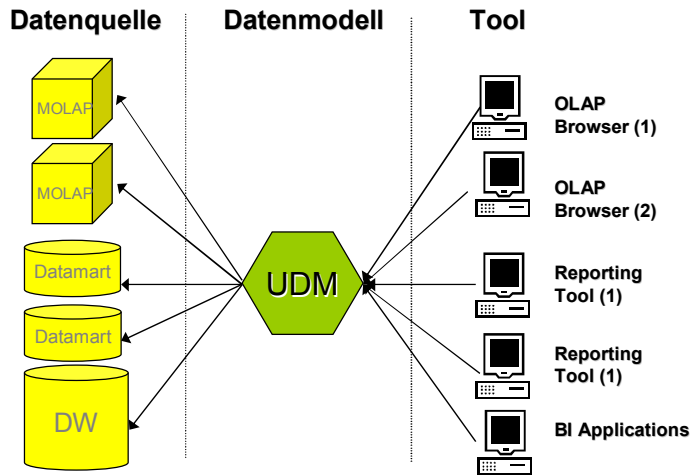
OLAP Cubes

- Multidimensional navigation
- Hierarchical presentation
- Friendly entity names
- Powerful MDX calculations
- Central KPI framework
- "Actions"
- Language translations
- Multiple perspectives
- Partitions
- Aggregations
- Distributed sources

Unternehmensweite BI-Lösungen mit dem UDM (I)

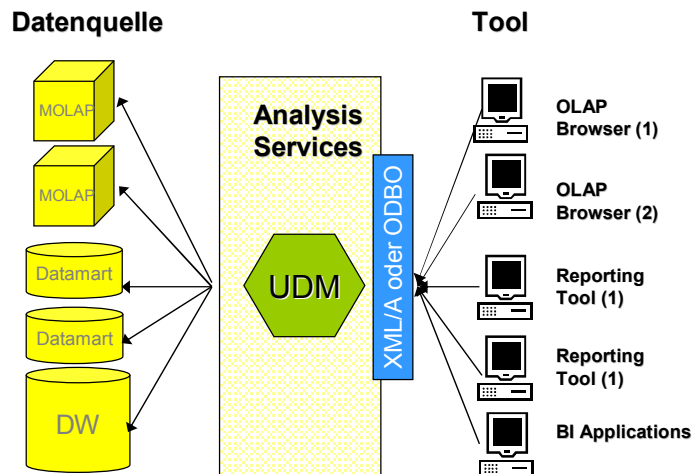


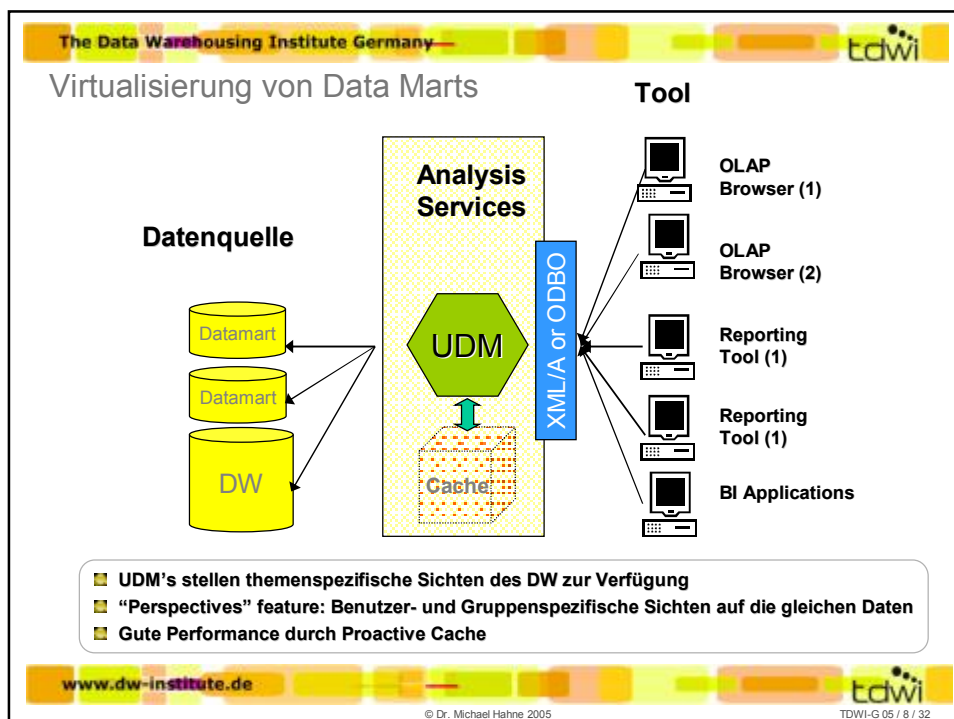
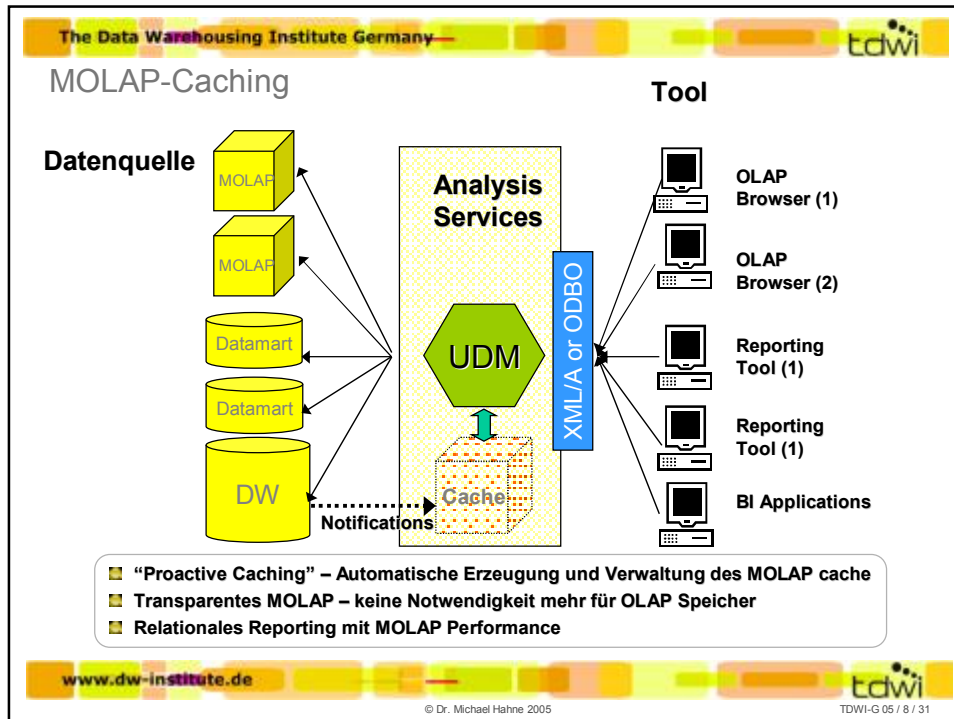
Unternehmensweite BI-Lösungen mit dem UDM (II)

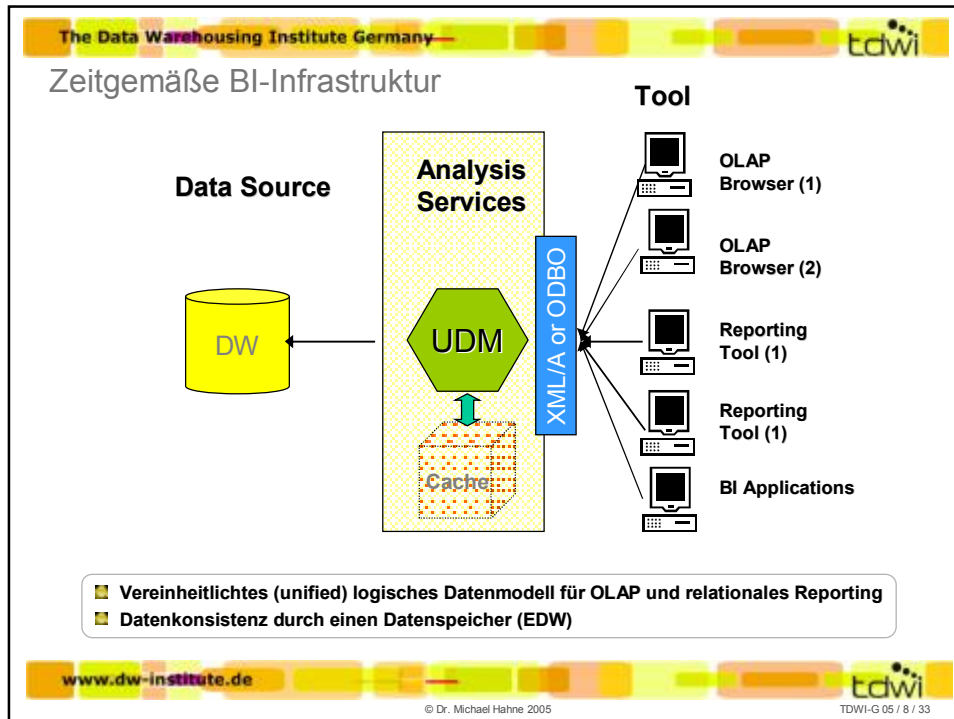


Ein dimensionales Modell für OLAP Analysen und relationales Reporting

Analysis Services – UDM-Server







- The Data Warehousing Institute Germany — edwi
- ## Agenda
- Microsoft SQL-Server 2000
 - Microsoft Data Warehouse Architektur
 - Dimensionsmodellierung mit den Microsoft Analysis Services
 - Attribute und Dimensionen
 - Modellierung von Cubes
 - Kennzahlen
 - Microsoft SQL Server 2005
 - Unified Dimensional Model
 - Dimensionen
- www.dw-institute.de
- © Dr. Michael Hahne 2005
- TDWI-G 05 / 8 / 34

Changing Dimension Support

- Drei Arten von Dimensionen, abhängig von Häufigkeit der Änderungen:
 - **Regular Dimensions:**
nicht-zeitabhängig, repräsentiert aktuelle Struktur gemäß ETL-Prozess
 - **Slowly Changing Dimensions (SCDs):**
Dimensionsänderungen treten nicht so häufig auf und müssen auswertbar sein
 - **Rapidly Changing Dimensions (RCDs):**
Häufige oder umfangreiche Änderungen, die auswertbar sein müssen

Slowly Changing Dimensions (SCDs)

- **SCD Type I:**
keine Historisierung, Attributwerte werden überschrieben, Schlüssel bleibt unverändert
- **SCD Type II:**
Einzelne Versionen, künstlicher Schlüssel, Gültigkeit über Zeitintervall abgebildet
- **SCD Type III:**
Originalwert und aktueller Wert als eigene Attribute,
auch anwendbar bei n festen Versionen

Rapidly Changing Dimensions (RCDs)

- Bei sehr großen Dimensionen und häufigen Änderungen von Attributwerten ist die Versionierung mit SCD Type II unbefriedigend
 - Konzept der RCDs
- Speicherung aller möglichen Kombinationen von Dimensionselement und Attributwert in einer Matrix → Rapidly Changing Dimension
- Zusammengesetzter Schlüssel ist Primärschlüssel der Dimensionstabelle und Fremdschlüssel der Faktentabelle, Update erfolgt in der Faktentabelle

Agenda

- 1) Abgrenzung und Einordnung von CPM und BI
- 2) Architekturen und Komponenten analyseorientierter Informationssysteme
- 3) Bestandteile multidimensionaler Datenstrukturen
- 4) Methoden der semantischen Modellierung mehrdimensionaler Datenstrukturen
- 5) Umsetzung mit multidimensionalen Datenbanksystemen
- 6) Varianten und Bestandteile des Star Schemas
- 7) Erweitertes Star Schema des SAP Business Information Warehouse
- 8) Mehrdimensionales Datenmodell des Microsoft SQL Server
- 9) **Zusammenfassende Gegenüberstellung relationaler und multidimensionaler Modellierung**

Speicherkomponenten und -technologien analytischer Informationssysteme



Integrierte Detaildaten
ggf. bis auf Belegebene
für operatives Reporting

Relationale
Speichertechnologie

Klassische relationale
Datenablage
(3. Normalform)

Verdichtete Daten (z. B.
Tagesbasis)

Relationale
Speichertechnologie

Star-Schema-Modelle
Denormalisiert

Stark verdichtete Daten mit
thematischer Ausrichtung

Relationale oder
multidimensionale
Speichertechnologie

Star-Schema oder physisch
multidimensionale Ablage

Vergleich relationales und multidimensionales Speicherkonzept

Relationales Speicherkonzept

- Skalierbarkeit
- Häufig Read-Only-Lösungen
- Beliebige Datenmengen
- Zahlreiche Attribute in jeder Dimension möglich
- Mäßige Performance
- Zusatzoptionen (Spezielle Indextechniken)

Multidimensionales Speicherkonzept

- Plattformabhängig
- Datenmodifikation möglich (What If)
- Aufblähung des Datenbestandes
- Eingeschränkte Attributanzahl je Dimension
- Gute Performance
- Black Box