
Data Warehousing und Data Mining

Eine Einführung in entscheidungsunterstützende Systeme

Interaktive Folien zu Kapitel 5
Data Warehouses



› Entscheidungen unterstützen heisst Daten

- sammeln
- aufbereiten
- abfragen
- auswerten
- präsentieren



Business Intelligence mit Data Warehousing und Data Mining

Einordnung




Nutzwertanalyse am Beispiel von AHP

- ✓ Kioskstandort  , Personalauswahl 

Was-Wenn-Analyse

- ✓ Erfolgsrechnung 
- ✓ Anzeigenplanung , Produktionsplanung 

Regelbasierte Systeme

- ✓ Spesen , Betriebskredit 
- ✓ Regelverkettung  

⇒ **Data Warehouses**

- ⇒ **Anlageberatung**  
- ⇒ **Lieferfrist, Handel, Verkauf**  

Data Mining - Ein Überblick

- Zeitschriften , Bank 

Regelinduktion

- Spesen , Bonitätsklassifikation  

Neuronale Netze

- Bonitätsklassifikation , Bonitätsvorhersage 
- EindimPerzeptron  , ZweidimPerzeptron  
- MehrklassPerzeptron  
- MehrstufPerzeptron  

Unterrichtsmaterial

Software

- Demonstrationsversion von *Cognos PowerPlay*
(Abfragewerkzeug für mehrdimensionale Würfel)
- Demonstrationsversion von *if...Synchrony*
(Abfragewerkzeug für relationale Data Warehouses)

Beispiele und Übungen

- [Anlageberatung](#) mit *Cognos PowerPlay* 📌🖱️
- [Handel](#) mit *if...Synchrony* 📌🖱️

Produktinformation

- <http://www.cognos.com/>
- <http://www.iftime.com/synchrony.htm>

Daten- und modellgetriebene Methoden

① Modellgetriebene Methoden

- kleine Datenmengen
- bewusste Datenauswahl
- rechnerisch weniger aufwendig



- ✓ Nutzwertanalyse, insb. AHP
- ✓ Optimierung

② Datengetriebene Methoden

- grosse Datenmengen
- Stichprobenauswahl
- rechenintensiv



Data Warehousing und Data Mining

- ⇒ Data Warehousing, insb. OLAP
- ⇒ Regelbasierte Systeme
- ⇒ Regelinduktion
- ⇒ Neuronale Netze

Grundlagen

Grundlagen

- ⇒ Operative und analytische Datenbanken 9
- ⇒ Enterprise Data Warehouse und Data Marts 24
- ⇒ Mehrdimensionale Daten 38

Endbenutzerzugriff 50

- OLAP 61
- ANLAGEBERATUNG mit *PowerPlay*   87

Modellierung relationaler Data Warehouses 100

- Datenmodellierung 100
- LIEFERFRIST  121
- VERKAUF  122
- EINZELHANDEL mit *if...Synchrony*   148

Entwicklung und Betrieb 152

- Entwicklung und Betrieb 152
- Effizienz 176

Netzzugriff 214

- Client/Server-Architekturen 219
- Zugriff über Internet und Intranet 229

Data Warehouses

“Ten Years ago I could have told you how Doritos were selling west of the Mississippi. Today, not only can I tell you how Doritos sell west of the Mississippi, I can tell you how well they are selling in California, in Orange County, in the town of Irvine, in the local Von’s supermarket, in the special promotion, at the end of aisle four, on Thursdays.”

(D.W. Calloway, CEO von Pespico, an einer Aktionärsversammlung¹)

¹ zitiert in *Building a Decision Support Architecture for Data Warehousing*, ATG’s Data Warehousing Technology Guide Series,, The Applied Technologies Group 1998 (www.techguide.com), p.1

Data Warehouses und Data Mining

Business Intelligence mit EUS

Data Warehousing ist die
Gedächtnisfunktion von EUS

Data Mining ist die **Beurteilungs-**
funktion von EUS



Data Warehousing eher ein Kind langjähriger
Praxis als der Wissenschaft \Rightarrow Terminologie
und Strategien noch oft uneinheitlich

Strategische¹ Planung

Aufbau langfristiger Erfolgspotentiale



strategische Daten

Operative Planung

Ausschöpfung vorhandener Erfolgspotentiale



operative Daten

¹ Taktische Planung ist mittelfristig. Im folgenden heisst “stragisch” immer auch “taktisch”.

Produktionsdatenbank

Produktionsdatenbank :=

Datenbank, die operative Geschäftsvorgänge aufzeichnet, indem sie ...

✓ Einzeldaten in

✓ häufigen

✓ ▶ Transaktionen¹

Buchungsvorgänge an einem Bankschalter

✓ in Echtzeit

✓ laufend

✓ vollständig und

✓ redundanzarm

✓ fortschreibt

¹ Transaktionsintensive Branchen sind z.B. Telekommunikation, Handel und Banken. Sie verarbeiten bis zu 10'000-30'000 Transaktionen pro Minute.

Mängel von Produktionsdatenbanken

‣Operative Entscheidungen



Produktionsdatenbanken sind ...

☹ unübersichtlich

(originäre Detaildaten)

☹ momentbezogen

(keine historischen Periodendaten)

☹ benutzerfeindlich

(‣SQL- oder ‣3GL-Kenntnisse)



Data Warehouses (DW)



‣Strategische Entscheidungen

Eine mögliche Definition

Data Warehouse (DW) :=

Datenbank, die strategische Entscheidungen unterstützt, indem sie ...

- ✓ umfangreiche und
- ✓ regelmässige Auszüge aus
- ✓ Produktionsdatenbanken
- ✓ periodenbezogen und
- ✓ oft aggregiert¹
- ✓ Endbenutzern
- ✓ auch zur **ad hoc**²-Analyse bereit stellt

¹ zusammengefasst, konsolidiert (z.B. als Summe, Durchschnitt, Kennzahl)

² unvorbereitet (im Gegensatz zur Standardauswertung)

5.1 Operative vs. analytische Datenbanken

Schwerpunkt	Produktions- datenbank	Data Warehouse
Daten <i>operativ</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Daten <i>vollständig</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Daten <i>detailliert</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Daten <i>redundanzarm</i> (fortschreibungsfreundlich)	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Daten <i>änderungsintensiv</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Schlüssel i.d.R. <i>natürlich</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
meist <i>Auswahl</i> abfragen	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Abfragen häufig <i>ad hoc</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Daten <i>strategisch</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Daten <i>periodenbezogen</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Daten oft <i>abgeleitet</i> (v.a. zusammenfassend)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>endbenutzerorientiert</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Schlüssel i.d.R. <i>künstlich</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Typische Attribute von Telecom-Transaktionen?

Data Warehouses dienen ...

- ⇒ Managern zur Entscheidungsunterstützung
- ⇒ Spezialisten für das Data Mining

90% der 2000 grössten Unternehmungen hatten 1998 Data Warehousing - Projekte

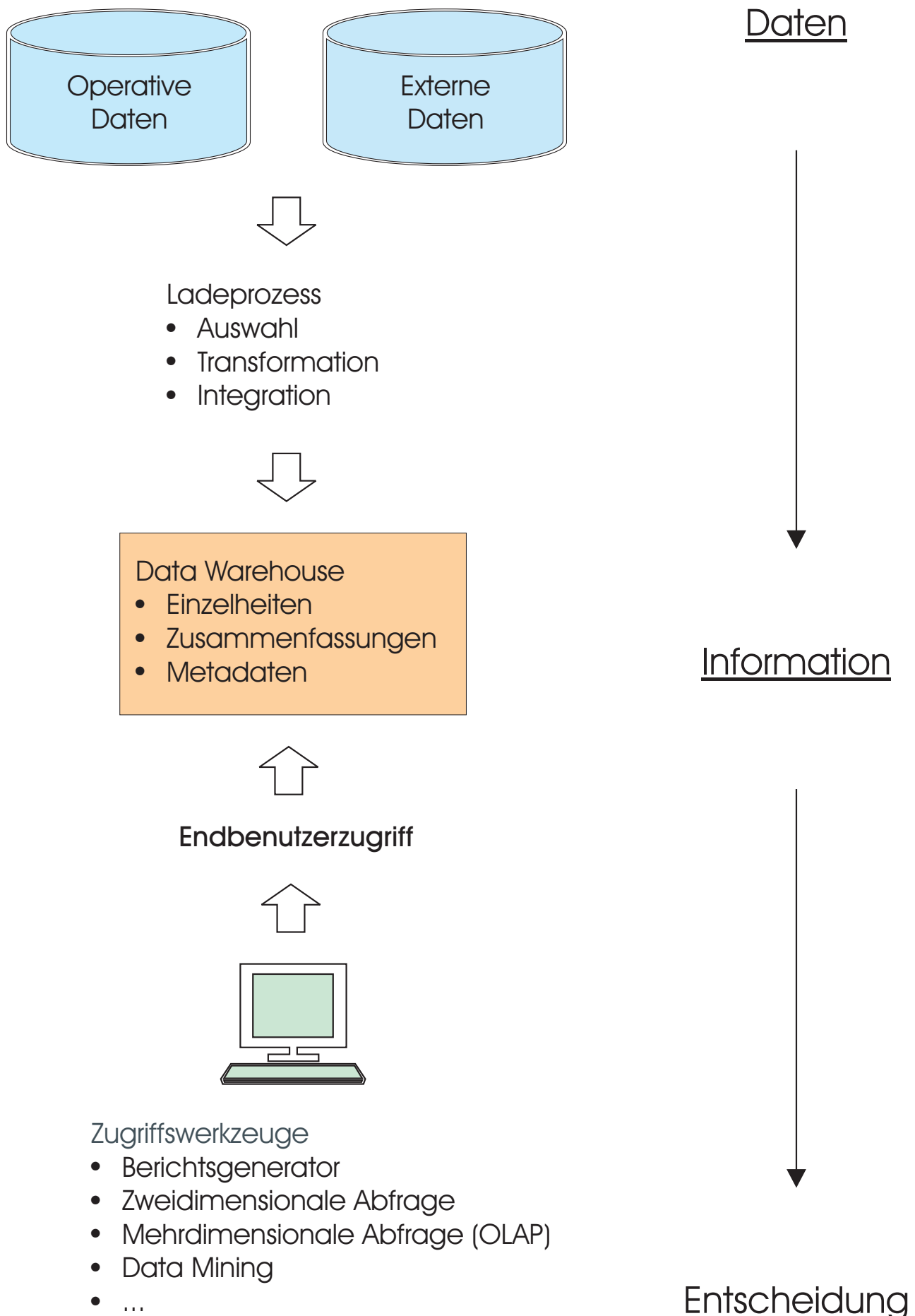
Gründe für zwei Buzzwords ...

Data Warehousing und Data Mining



- ⇒ CPU-Leistung alle zwei Jahre verdoppelt
- ⇒ Speicherplatz billiger
- ⇒ Netzbandbreite grösser
- ⇒ Integration lokaler Systeme leichter

5.3 Eine einfache Architektur



5.2 HANDEL - Fallbeispiel

Beispielunternehmung

- Einzelhandelskette ...
- mit Hunderten von Filialen, ...
- die je Tausende von Produkten führen, ...
- davon die meisten mit Bar Codes erfassen, ...
- die von POS¹-Systemen gelesen werden

Beispieldaten

- Produktdaten
- Verkaufsdaten
- Lagerdaten
- Marktforschungsdaten ²
- ...

Weshalb ein Data Warehouse?

- Daten zur Entscheidungsunterstützung integrieren
- Datenanalyse verbessern (v.a. für die Marktforschung)

Beispielabfragen

- Welche Werbeaktion hat sich prozentual am stärksten auf den Umsatz ausgewirkt?
- Wo werden Reinigungsmittel am besten aufgestellt?

¹ POS = P..... O..... S.....

² eigene Daten oder gekaufte Daten (syndicated data)

Data Warehouses sind speicheraufwendig

Ein DW integriert Daten von Bankkunden

z.B. Kreditkartenbezüge, Sparkontoeinlagen, Aktienkäufe, ...



- Kundenzahl

1 Million

- Informationseinheiten pro Kunde

1000 pro Kunde pro Monat

- Aufbewahrungszeit

5 Jahre



Informationsflut

Gesamtzahl der Werte =

$$1000 \cdot 5 \cdot 12 \cdot 10^6 = 60 \text{ Mia.}$$



Entwicklungsprobleme !

Entwicklung

Operative Daten verwalten

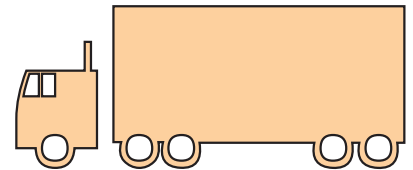
- *Bestellungen*
- *Rechnungen*
- *Produkte*
- ...



Daten laden

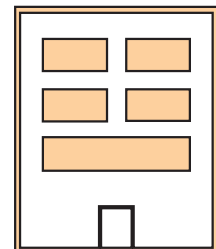
Operative Daten ...

- *zeitlich synchron auswählen*
- *temporär speichern*
- *transformieren*
- *integrieren*
- ...



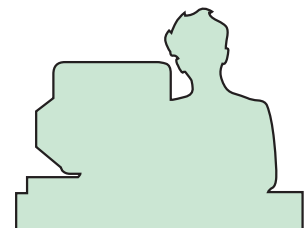
Analytische Daten verwalten

- *Fakten (Indikatoren) wie Geldumsatz speichern*
- *Dimensionen wie Zeit und Gebiet speichern*
- *Speicheraufwand und Laufzeit optimieren*
- *Daten sichern*
- *Vorberechnen, insbesondere aggregieren*
- *Nicht mehr benötigte Daten archivieren*
- ...



Zentrale Daten an lokale Systeme weiterleiten

- *lokale Data Warehouses*
- *Tabellenkalkulationswerkzeuge*
- *lokale Datenbanksysteme*
- ...



Endbenutzer-Abfragen beantworten

"Wie wirksam waren die Direct Mail-Kampagnen der letzten 3 Jahre ?"

...

Backend
Data Warehouse
Frontend

5.4 HANDEL - Eine Entwicklungsskizze

Werkzeuge

- **Hardware und Systemsoftware**
 - Client/Server-System
- **Datenmodell**
 - *Logisches* -
 - Sternschema auf einer ▸ relationalen Datenbank →
 - *Physisches* -
 - Datendefinitionen, Schlüssel, ▸ Integritätsregeln, Optimierungen, ...
- **Data Mining**
 - Werkzeuge vom Unternehmungsberater
 - Relationales ▸ OLAP-Frontend →
 - Konventionelle statistische Datenanalyse →
 - Induktion von ▸ Entscheidungsbäumen →
 - Visualisierung →

Projektmanagement

- Externer Workshop für die Entwickler
- Regelmässiger Kontakt mit den Endbenutzern
- Unterstützung durch das höhere Management
- ...

Data Warehouses verschiedener Grösse

Tabellenblatt

bis 1 **Gigabyte** (10^9 Byte)

Data Mart (lokales DW)

bis **Dutzende** von Gigabytes

Unternehmensweites DW (globales DW)

bis **Terabytes** (10^{12} Byte)



›Data Marts und ›Enterprise Data Warehouse

Data Mart (DM) :=

lokales Data Warehouse,
das sich auf die Daten ...

- ✓ eines Funktionsbereichs,
- ✓ einer Abteilung,
- ✓ einer Arbeitsgruppe
- ✓ oder einer einzelnen Person

beschränkt

Data Mart-Ziele

Intramodulare Bindung maximieren

Benutzer so homogen wählen, dass ...

⇒ Antwortzeit minimal (▷Partitionierung)

⇒ Benutzerfreundlichkeit maximal

Datenmodell für Funktionsbereich, Abteilung, Arbeitsgruppe¹

⇒ Werkzeuganpassung maximal (▷MOLAP, ▷DOLAP →)

Intermodulare Koppelung minimieren

Schnittstelle zwischen den Data Warehouses so schmal, dass ...

⇒ Data Mart-übergreifende Abfragen selten



¹ Kriterien der Verteilung von ▷DW-Kosten auf Abteilungen?

Enterprise Data Warehouse

Zentrales Data Warehouse :=

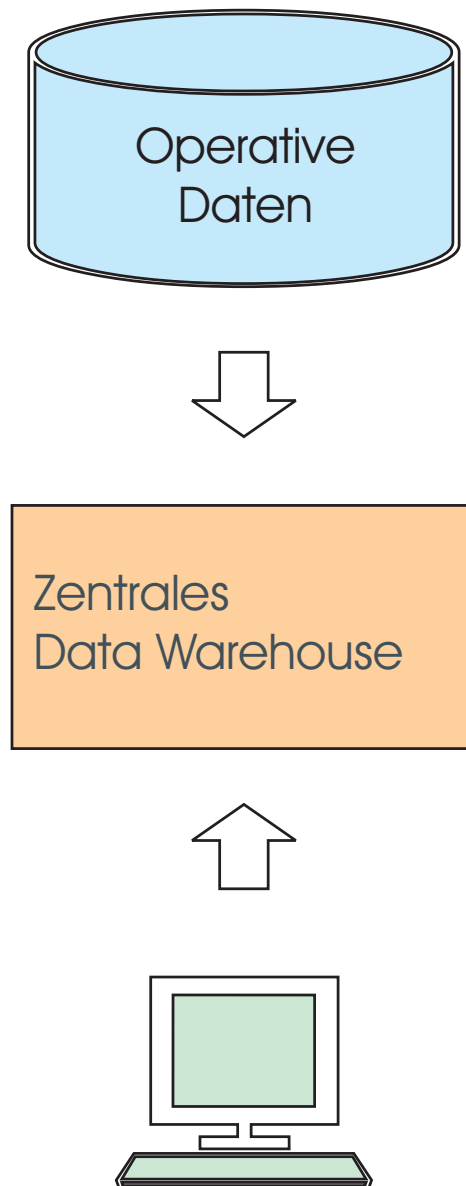
- ✓ DW, das
- ✓ Produktionsdaten
- ✓ transformiert und
- ✓ koordiniert
- ✓ Data Marts zu Verfügung stellt

Enterprise Data Warehouse (EDW) :=

- ✓ **zentrales** Data Warehouse¹ oder
- ✓ **Vereinigung** von Data Marts, das/die
- ✓ **unternehmensweite** Information verteilt

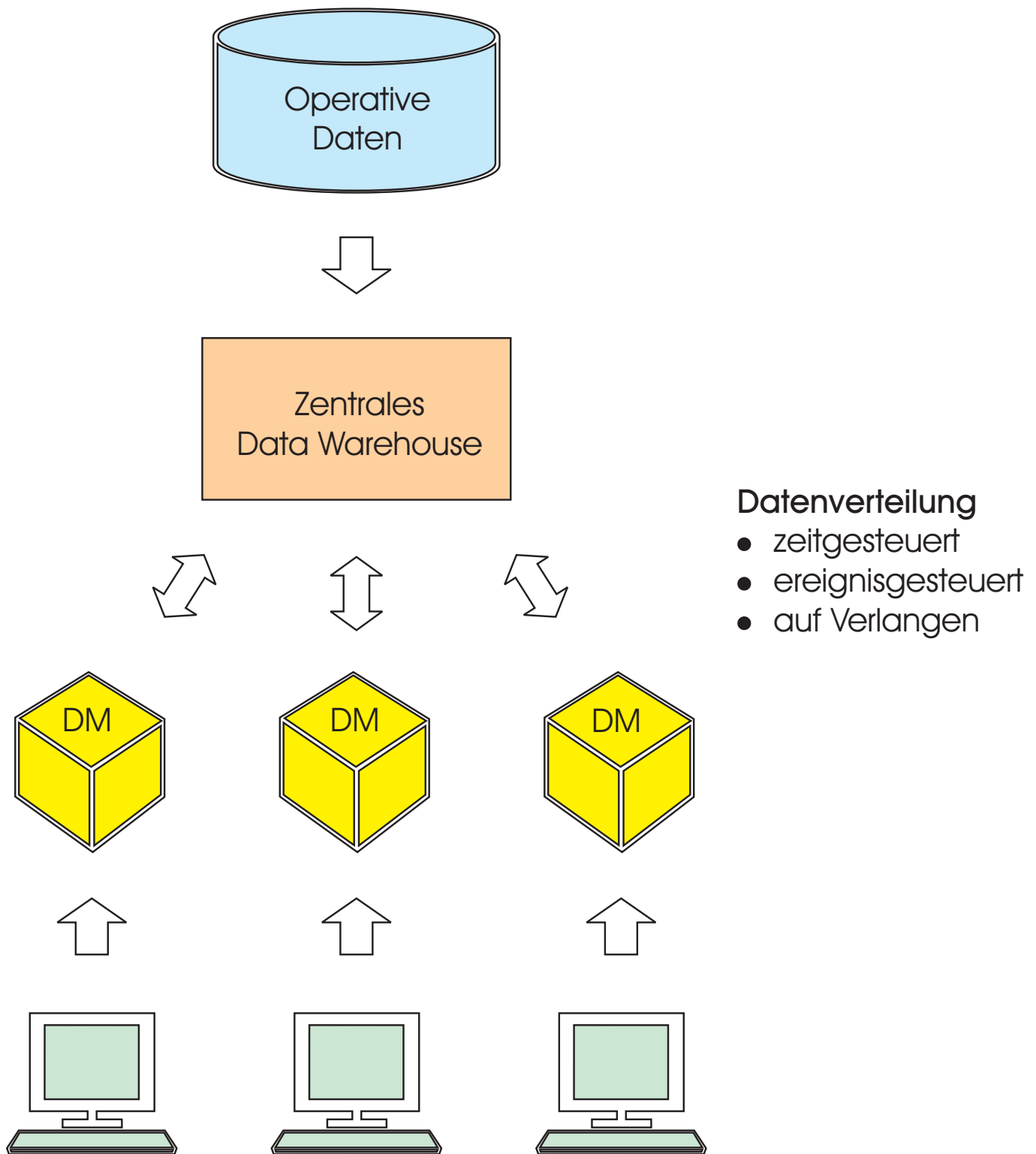
¹ Zentral kann auch ein DW für die Data Mart einer Niederlassung, eines Funktionsbereichs oder einer grossen Abteilung sein.

5.6 Zentralisierte EDW-Architektur¹



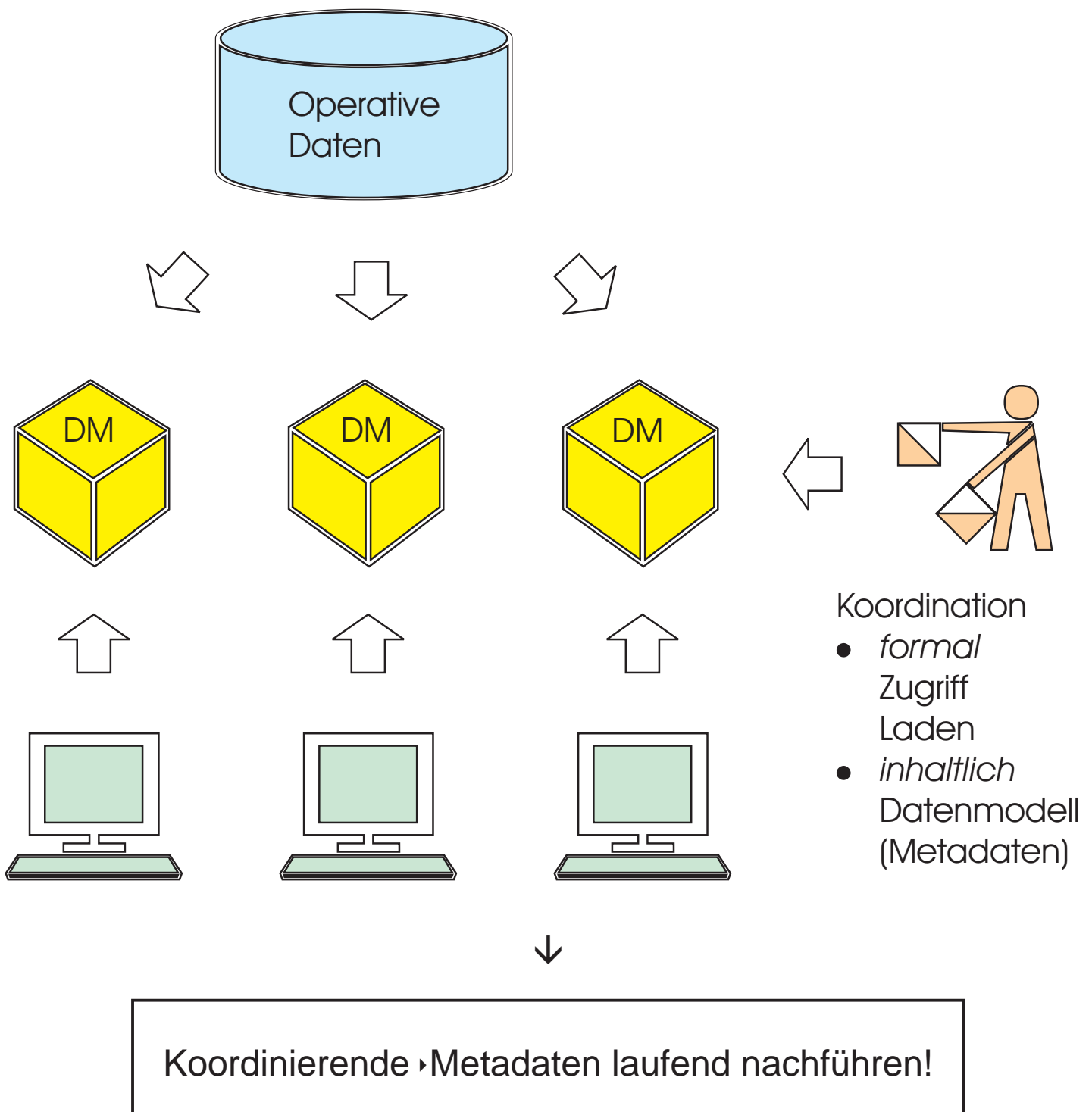
¹ Die Entwicklung eines zentralisierten DW ist für die meisten Unternehmungen zu komplex

5.7 Hierarchische EDW-Architektur (HEDW)



› Attribute mit der gleichen Bedeutung verwenden in allen Data Marts
die gleichen Bezeichner, › Datentypen, Schlüssel und Datenquellen.

5.10 Koordinierte EDW-Architektur (KEDW)¹



¹ Vereinigung aller Data Marts. »Attribute mit der gleichen Bedeutung verwenden in allen Data Marts die gleichen Bezeichner, »Datentypen, Schlüssel und Datenquellen.

Architekturen

Architektur	Zentrales DW	Data Marts	Koordinierte DM
zentralisiert	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
hierarchisch	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
koordiniert	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
unkoordiniert	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Eine *unkoordinierte* Architektur führt zu ...

- ☹ unkontrollierter **Redundanz**
- ☹ **Inkonsistenz**
- ☹ hohem **Entwicklungsaufwand**
- ☹ ungenügender **Skalierbarkeit**

Data Mart-Koordination - Voraussetzungen

› Dimensionen oder › Fakten ...

① gleicher Bedeutung ...

- ✓ gleich benennen
(Synonyme vermeiden!)
- ✓ mit den gleichen › Attributen beschreiben

② unterschiedlicher Bedeutung ...

- ✓ verschieden benennen
(Homonyme vermeiden!)

5.9 Data Mart-Koordination - Abfragebeispiel

Produkt sei eine **koordinierte Dimension** der ›Sternschemata

① Produktion

② Lagerhaltung

③ Verkauf



Eine **sortierende ›Abfrage** nach Produkten ergibt für jedes Sternschema eine Ergebnistabelle



Mischen der drei Tabellen ergibt die folgende Tabelle:

Produkt	① Produziert	② Gelagert	③ Verkauft
A	2000	200	1800
B	4000	1000	3000
C	3000	400	3600

Anforderungen an Attribute

Attribute sind im Idealfall ...

- ✓ koordiniert
- ✓ verständlich
- ✓ vollständig
- ✓ korrekt
- ✓ dokumentiert

Aufgabe Anforderungen an Attribute

Entwerfen Sie nach den folgenden Regeln eine möglichst breit verwendbare Adressstruktur:

- a) Strukturieren Sie eine Adresse so, dass die Suche nach ihren Elementen möglichst leicht wird
- b) Vermeiden Sie Abkürzungen
- c) Schliessen Sie unterschiedliche Kommunikationsarten ein (Post, Telefon, ...)
- d) Beschränken Sie sich auf natürliche Personen (Adressen juristischer Personen stellen andere Anforderungen).



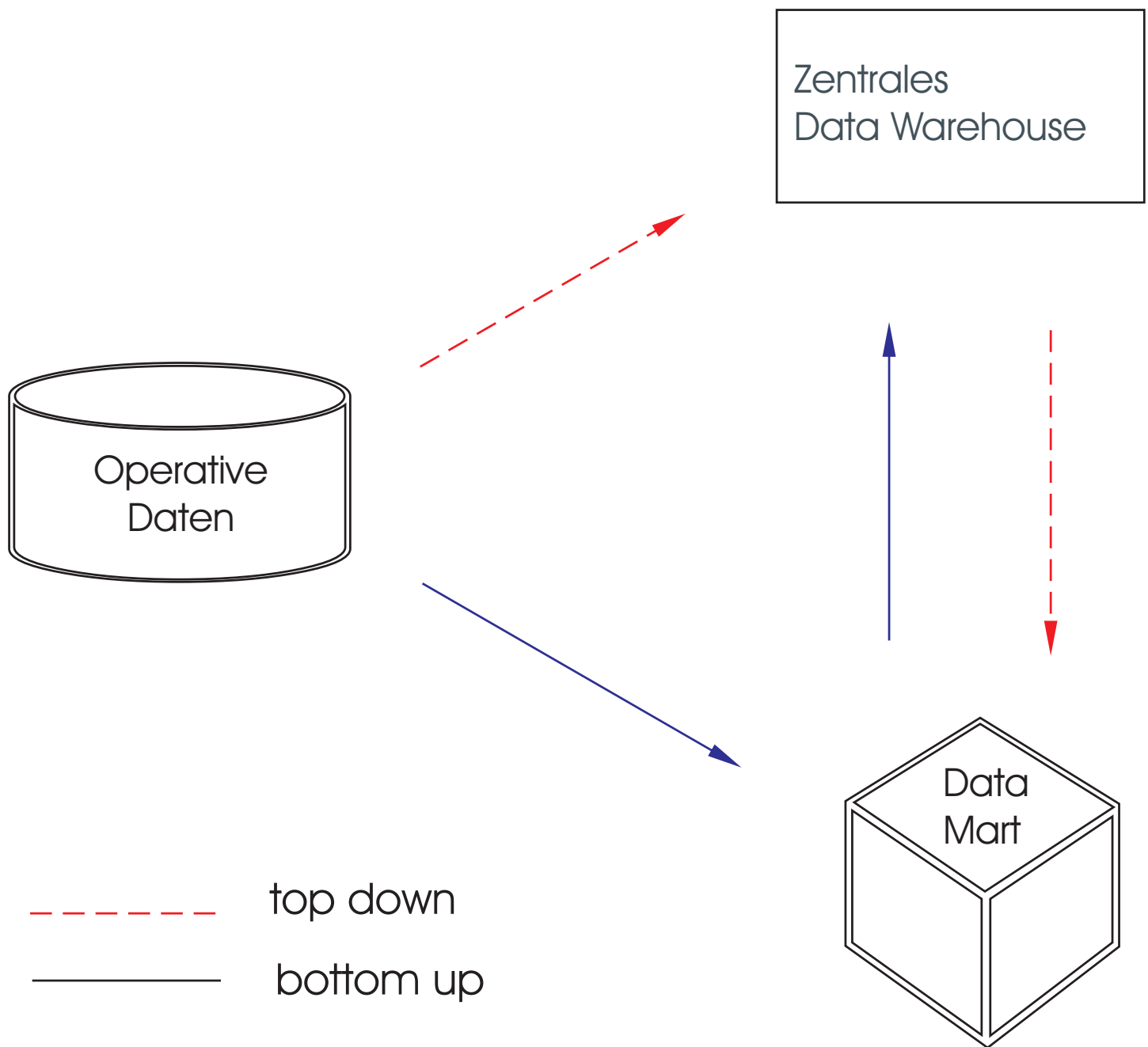
5.8 Data Mart vs. Enterprise Data Warehouse

	Data Mart (DM)	EDW
<i>Ähnliche Begriffe</i>	lokales ¹ DW	globales DW
<i>Hauptziel</i>	Modularisierung lokaler Daten und Abfragen	redundanzarmes Unternehmungsdatenmodell für ›EUS
<i>Domäne</i>	Abteilung oder Anwendung	Unternehmung
<i>Entwicklung / Betrieb</i>	Fachabteilung	IT-Zentrale
<i>Grössenordnung</i>	bis 100 Gb (10 ⁹ Byte)	bis Terabytes (10 ¹² Byte, oft grösser als Produktionsdaten)
<i>Datenquelle</i>	EDW oder Produktions-DB	Produktions-DB
<i>Granularität (Datendetail)</i>	oft gross (eher aus ›Aggregaten)	meist klein (viele Transaktionsdetails)
<i>Normalisierungsgrad</i>	klein	gross
<i>Hauptzugriffsart</i>	›OLAP (meist ad hoc)	›SQL (oft vordefiniert)
<i>Entwicklung</i>	gut überblickbar	aufwendig
<i>Typische Plattform</i>	mehrdimens. oder ›relationaler Datenbankserver	›RDBMS auf Unix-, Windows NT- oder Grossrechner
<i>Produktbeispiel</i>	<u>Sybase IQ</u>	<u>Sybase MPP</u> ²

1 EDW und DM sind verteilt oder koexistieren auf einer einzigen Hardware- und Software-Plattform.

2 RDBMS für Massively Parallel Processors

5.5 Data Mart-Entwicklung



Top Down - Entwicklung

Bottom Up - Entwicklung

Parallele Entwicklung

EDW und Data Mart - Ein Kontinuum

	<i>Operative DB</i>	<i>Enterprise DW</i>	<i>Abteilungs- Data Mart</i>	<i>Mini- Data Mart</i>
<i>Redundanz</i>	minimal	kontrolliert	kontrolliert	kontrolliert
<i>Normalisiert</i>	vollständig	teilweise	kaum	kaum
<i>Granularität</i>	laufendes Detail	historisches Detail	aggregiert	aggregiert
<i>Optimierungs- schwerpunkte</i>	Speicher Zugriff Integrität	Speicher Integrität	Zugriff	Zugriff

① Mini-Data Mart

Ein Benutzer analysiert Kundendaten nach der Dimension Herkunft und den Kategorien Land, Region und Ort und identifiziert Regionen mit unterdurchschnittlichem Umsatz.

② Abteilungs-Data Mart

Unter Umständen analysiert der gleiche Benutzer diese Regionen näher und sammelt z.B. Daten über die umsatzschwächsten Orte.

③ Enterprise Data Warehouse

Für ausgewählte Kunden untersucht er vielleicht sogar die letzten fünf Jahre.

④ Operative Datenbank

Falls die Käufe eines Kunden in dieser Periode konsistent abgenommen haben, sendet er die laufende Adresse des Kunden einem Direct Mailing-Mitarbeiter.

Beurteilung von Data Marts

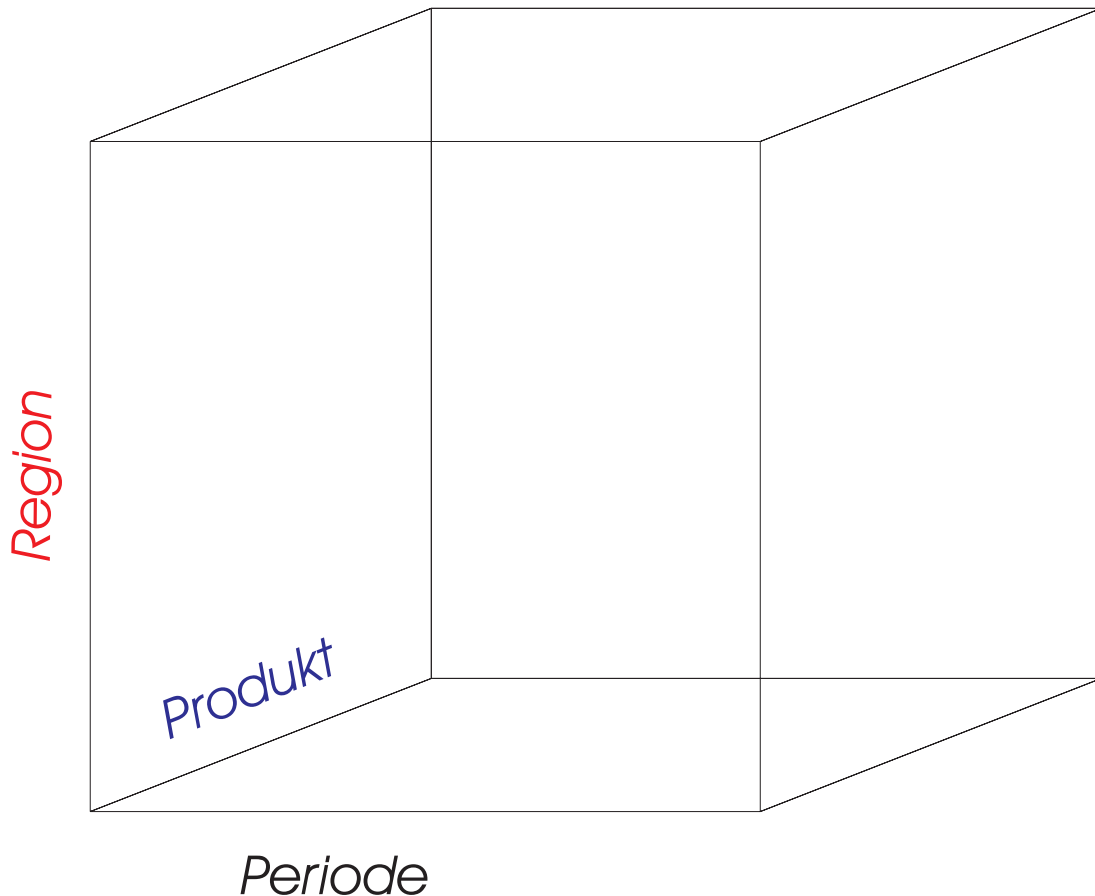
Vorteile

- 😊 Entwicklungskomplexität kleiner
- 😊 Endbenutzerbeteiligung wahrscheinlicher
- 😊 Flexibilität grösser
- 😊 Datenmodell einfacher
- 😊 Abfrageeffizienz grösser

Probleme

- 😞 Skalierbar?
- 😞 Integrierbar in Enterprise Data Warehouse?
- 😞 Redundanz kontrollierbar? →
- 😞 Data Marts untereinander konsistent?
- 😞 Entwicklungs- und Wartungskosten tragbar?

5.12 Betriebsnähe durch Indikator u. Dimensionen



“Welches ist der *UMSATZ* nach *Periode* und *Produkt* und *Region*?”

- Würfel veranschaulicht die drei **Dimensionen** des Indikators
- Am Schnittpunkt der drei Dimensionen ist der Umsatz.
UMSATZ ist ein **Indikator** der betrieblichen Leistung

- 1 “Würfel” mit mehr als drei Dimensionen heissen Hyperwürfel (hypercubes)
- 2 Suchen Sie nach weiteren Dimensionen des Umsatzes!

Würfel - Begriffe

Würfel (engl. cube bzw. hypercube) :=

- ✓ mehrdimensionale Datenstruktur, welche die
- ✓ gleichzeitige Analyse mehrerer
- ✓ Indikatoren und
- ✓ Dimensionen ermöglicht

Indikator (Fakt, engl. business measure) :=

- ✓ aggregierbares
- ✓ meist numerisches und kontinuierliches Attribut, das die
- ✓ mehrdimensionale Messung
- ✓ eines betrieblichen Erfolgskriteriums erlaubt

Dimension :=

- ✓ meist symbolisches und diskretes Attribut, das die
- ✓ Auswahl, Zusammenfassung und Navigation eines
- ✓ Indikators erlaubt

Einen Indikator nach seinen Dimensionen messen

Fakten (Erfolgsindikatoren) wie ...

- *Umsatz*

Dimensionen wie ...

- *Zeitperiode*
- *Produkt*

Mehrdimensionale Abfragen wie ...

Umsatz nach Zeit nach Produkt

Januar			
Produkt A	Produkt B	Produkt A	Produkt B
10,5 Mio.	2,8 Mio.	9,4 Mio.	5 Mio.

Die Dimension **Zeit** (-punkt oder/und -periode) kommt in jedem Data Warehouse vor !

HANDEL - Fakten nach Dimensionen auswerten

Fakten

Mengenumsatz

Geldumsatz

Kundenzahl

...

Dimensionen

Periode

Produkt

Filiale

Werbeaktion

...

Abfragen

Umsatz **nach** Produkt

Umsatz **nach** Produkt **nach** Monat

Umsatzzunahme **nach** Aktion

Kundenzahl **nach** Filiale

Indikator versus Dimensionen

	<i>Indikator</i>	<i>Dimension</i>
<i>Synonyme</i>	Fakt performance measure, key business measure	constraint
<i>Beispiele</i>	“Umsatz” nach ...	“Produkt”, “Region”
<i>Zweck</i>	Erfolg nach mindestens drei Dimensionen messen	Indikatorwerte wählen, zusammenfassen und durchlaufen
<i>Datentyp</i> (Datenformat)	i.d.R. numerisch und ›kontinuierlich (aggregierbar)	i.d.R. ›symbolisch und ›diskret
<i>Schlüssel</i>	›Primärschlüssel aus den ›Fremdschlüsseln der Dimensionen	Primärschlüssel
<i>Datenvolumen</i>	gross (i.d.R. etwa 70% des Data Warehouse)	klein

- 1 Einteilung in Indikatoren oder Dimensionen nicht immer eindeutig
- 2 Ein Attribut kann als *stetiges* Attribut Indikator (Bsp. *Börsenwert*) und als *diskretes* Attribut Dimension sein (Bsp. Gewinn eines Unternehmens nach der Dimension *Börsenwert-Klasse* analysieren)

Aufgabe *Fakt oder Dimension* (A 5.1)

- ›POS-Transaktion (›Marktkorbanalyse)
- Region
- Zeit
- Produkt
- Kontenbewegung
- Gewinn
- Niederlassung
- Mitarbeiterzahl
- Lieferant
- ›Funktionsbereich
- Branche
- Börsenwert
- ›Eigenkapitalrentabilität
- ...

Mehrdimensionale Daten speichern

Speicherung mehrdimensionaler Daten in einem ...

① › relationalen › physischen Modell

- als zweidimensionale Objektdaten
- mehrdimensional verbunden durch Metadaten
- in einem relationalen Server-Datenbanksystem

② mehrdimensionalen physischen Modell

- als mehrdimensionale Objektdaten
- in einem proprietären mehrdimensionalen Server-DBMS

Beide physischen Datenmodelle speichern mehrdimensionale Daten meist aggregiert und oft komprimiert

Relationale versus mehrdimensionale DB

	Relationale DB ¹	Mehrdimens. DB
Abfragefreundlichkeit	-	+
Abfrageeffizienz	-	+ ²
Speichereffizienz	+	-
Skalierbarkeit	+	-
Kompatibilität	+ ³	-



In der Regel werden **Enterprise Data Warehouses** durch spezialisierte RDBMS und **Data Marts** durch relationale oder mehrdimensionale DBMS verwaltet

- 1 ›RDBMS mit ›mehrdimensionalem Datenmodell heissen **multirelational**
- 2 Grund: Mehrdimensionale ›DB speichern Objektdaten mehrdimensional
- 3 Mehrdimensionale DB sind ›proprietär, relationale stützen sich hingegen auf ein gemeinsames Datenmodell, eine standardisierte Datenbanksprache (›SQL), eine Vielfalt von Werkzeugen und langjährige Erfahrungen

Anforderungen an RDBMS

Oberkriterium	Kriterium
Modelltreue	
	Sind Objekt- und Metadaten in Tabellenform, und lassen sich darauf Tabellenoperationen ausführen? (Data Dictionary)
	Lassen sich fehlende Werte einheitlich als Nullwerte darstellen?
	Unterstützt die Datenbanksprache die relationalen Grundoperationen?
	Abhängigkeit von Änderungen der physischen und logischen Datenorganisation? (zum Beispiel nach Optimierungen oder Änderungen von Basistabellen)
Integrität	
deklarative -	Welche Integritätsbedingungen lassen sich mit einer deklarativen Datenbeschreibungssprache im Katalog (Data Dictionary) festhalten?
	Entitäts- und Beziehungsintegrität automatisch?
prozedurale -	Programmierung von Integritätsbedingungen als gespeicherte Prozeduren und Trigger?
Sprachen	
	Benutzerfreundlichere Art der deklarativen Datendefinition, -abfrage, -manipulation und -kontrolle als SQL? (zum Beispiel QBE)
	Wie umfangreich sind DDL, DML und DCL?
	Zusammengesetzte Schlüssel definierbar?
	Operationen auf Views?

Oberkriterium	Kriterium
	Wieviele ›Tabellen und Schachtelungen unterstützten die Abfragesprachen?
	Für welche Objekte und Eigenschaften lassen sich Synonyme vergeben?
	Mehrere Versionen von ›Datenbankobjekten? (engl. versioning)
Entwicklung	
Werkzeug- umfang	Autonome Entwicklung ganzer Anwendungen?
	Werden alle Entwicklungsphasen unterstützt? (insbesondere auch der Datenentwurf)
	Strukturierte Programmierung? (z.B. durch Steuerkonstrukte, vollständige ›Datentypen, lokale und globale Objekte)
	Objektorientierte Anwendungsentwicklung? (Kapselung, Vererbung und Polymorphie)
	Programmierungsumgebung? (zum Beispiel Editor, Debugger, Browser und Versioning)
	Wie vollständig und programmierfreundlich lassen sich die Datenbankfunktionen aus externen ›3GL-Sprachen aufrufen? (z.B. ›Cursors)
Werkzeug- komfort	Kann der Entwickler die Benutzerschnittstelle ›deklarativ gestalten? (›Formular- und ›Berichtsgeneratoren)
	Lassen sich Standardanwendungen deklarativ entwickeln? (z.B. durch ›Applikationsgeneratoren)
Kompatibilität	
Portabilität	Welche Betriebssysteme und Hardwareware-Plattformen unterstützt das ›DBMS?

Oberkriterium	Kriterium
<i>Kommunikation</i>	Werden ›Mehrprozessorsysteme unterstützt?
	Anwendungsübergreifend? (zum Beispiel über Import/Export, ›COM oder ›DDE)
	Welche ›Datentypen unterstützt das ›DBMS? (zum Beispiel ›BLOBs, Memofelder, COM)
<i>Standards</i>	Portabilität des ›SQL-Dialekts?
	Wie nahe kommt die eingebaute ›prozedurale Programmiersprache an eine verbreitete ›3GL- oder ›4GL-Sprache?
<i>Integration</i>	Data Warehousing? Data Mining? Text Mining?
Datensicherheit	
<i>Verfügbarkeit</i>	Lassen sich z.B. Schreiboperationen aus Sicherheitsgründen gleichzeitig auf mehreren Geräten speichern? (weitere Massnahmen: Stand by-Server, On line-Verwaltung)
<i>Backup</i>	Backup während des laufenden Betriebs?
<i>Recovery</i>	Automatischer und manueller ›Recovery?
	Komfort der Erstellung von ›Log-Dateien?
<i>Transaktionssteuerung</i>	Welches ›Sperrprotokoll steuert den konkurrierenden Zugriff mehrerer ›Transaktionen auf gemeinsame Objekte?
	Welche ›Isolierungsgrade werden unterstützt?
	Welches ist die Granularität der ›Sperrungen? (Datenbank, Tabelle, Page, Satz, Feld)
	›Deadlocks automatisch erkannt und aufgelöst?
Datenschutz	
	Datenschutzmassnahmen (z.B. Passwortschutz, Verschlüsselung, Zugriffsprotokolle)

Oberkriterium	Kriterium
	Granularität der Datenschutzmassnahmen? (System, Datenbank, Tabelle, Page, Satz, Feld)
	Hiding von ›Datenbankobjekten?
	Lassen sich Benutzergruppen bilden?
	Lassen sich Berechtigungen vererben?
Netzfähigkeit	
	Welche Netzwerkprotokolle werden unterstützt?
	›Verteilte Architekturen? (›Fileserver-, ›Client/Server-, ›verteilte Datenbanken)
	Erlaubt das ›DBMS heterogen verteilte Daten?
	Wie gut wird ›Replikation unterstützt?
	Qualität der ›World Wide Web-Einbindung? (z.B. E-Commerce)
Physische Grenzen	
	z.B. Speicherbedarf; Anzahl Netzknoten; Länge von Feld, Schlüssel, Satz, Namen, Datei, Datenbank und Programm?
	Skalierbarkeit
Effizienz	
Speicher-	Wie flexibel lässt sich der Cache (›Puffer) verwalten?
	Lassen sich Sätze physisch zusammenfassen und als Cluster speichern?
	Lassen sich die Daten komprimieren?
Laufzeit-	Wie kann der Zugriff optimiert werden? (z.B. durch ›B-Bäume und ›Hashfunktionen)
	Werden Suchabfragen automatisch optimiert? Lokal und global? (in verteilten Datenbanken)

Oberkriterium	Kriterium
	Erzeugt der Übersetzer der ›prozeduralen Sprache (v.a. ›gespeicherter Prozeduren) Maschinen- oder Zwischencode?
Komfort	
	Bietet das System vollständige Einführungs- und Referenzhilfen an? (on line, schriftlich, Schulung)
	Wie gut ist die laufende Unterstützung durch den Anbieter? (technischer Support)
	Existieren Benutzervereinigungen?
Anbieter- und Produktqualität	
	Produktumstände wie Reifegrad oder Häufigkeit und Aufwand von Versionswechseln?
	Herstellerfirma? (zum Beispiel ihre wirtschaftliche Lage, die lokale und globale Marktpräsenz und die eingeholten Referenzen)
	In welches Abhängigkeitsverhältnis gegenüber dem Anbieter begeben wir uns?

Grundlagen

Grundlagen 2

- ✓ Operative und analytische Datenbanken 9
- ✓ Enterprise Data Warehouse und Data Marts 24
- ✓ Mehrdimensionale Daten 38

Endbenutzerzugriff

- ⇒ OLAP 61
- ⇒ ANLAGEBERATUNG mit *PowerPlay*   87

Modellierung relationaler Data Warehouses 100

- Datenmodellierung 100
- LIEFERFRIST  121
- VERKAUF  122
- EINZELHANDEL mit *Synchrony*   148

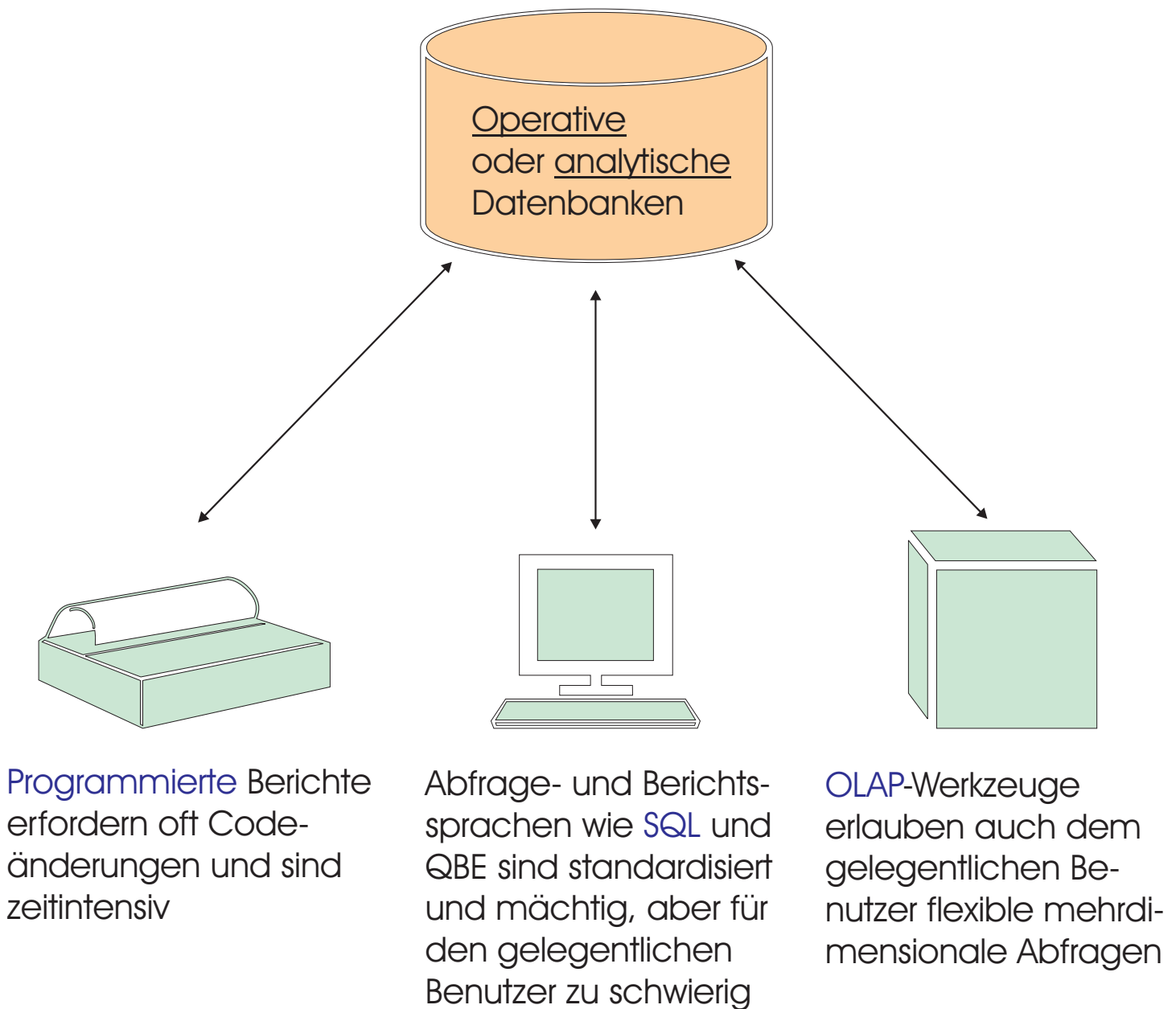
Entwicklung und Betrieb 152

- Entwicklung und Betrieb 152
- Effizienz 176

Netzzugriff 214

- Client/Server-Architekturen 219
- Zugriff über Internet und Intranet 229

5.15 Zugriff auf operative und analytische Daten



Benutzerfreundlichere Abfragen mit wenig ...

- Papier
- Entwicklungsaufwand
- Auswertungsaufwand (vgl. Spreadsheets)

Datenbankabfragen

Abfragen i.w.S. sind ...

① **Auswahl**abfragen (Abfragen i.e.S.)

✓ Suchabfragen

② **Aktions**abfragen

✓ Anfüge-

✓ Lösch-,

✓ Aktualisierungs-

✓ Tabellenerstellungs-

Abfragesprachen

Prozedurale - sind flexibler

- ✓ Navigation meist satzweise
- ✓ Anweisungsreihenfolge wichtig

Visual Basic, Cobol, dBASE

Nichtprozedurale - (deklarative) sind komfortabler

- ✓ Navigation tabellenweise
- ✓ einzige verbale oder grafische Anweisung

QBE, SQL

5.16 Prozedurale und deklarative Abfragen

① Umgangssprachliche -

Welche Eigenschaften haben die Produkte der Abteilung “Farben”?

② Formalsprachliche -

▸ prozedurale - (dBASE)

```
use PRODUKTE
% -- temporär zwischenspeichern
copy to TMP
use TMP
delete for Produktart <> 'Farben'
% -- Duplikate eliminieren
total on PRODUKTE to ERGEBNIS
display all
```

▸ deklarative - (SQL)

```
select *
from PRODUKTE
where Produktart = "Farben"
```

5.17 Query by Example (QBE)

“Wieviele **Stunden** unterrichtet Dozent **Meier** im **Sommersemester 1993** ?”

ORGANISATION

Name	Semester	Dozent	Raumnummer
x	SS93	Meier	

VERANSTALTUNG

Name	Stunden
x	√

ANTWORT

Stunden
2

Tabellenskelette

Ausgabespalten

Ausgabezeilen

Verbundattribute

Abfragenname

SQL - Grobübersicht

SQL := Structured Query Language

- ▶ deklarativ (WAS statt WAS UND WIE)
- interaktiv oder in ein Programm eingebettet

Operationen

▶ Datenbankobjekte definieren

- `create`

Datenbankobjekte manipulieren

- `insert`
- `update`
- `delete`

Datenbanken abfragen

- `select`

Datenbanken schützen

- `grant`

5.18 SQL - Teilsprachen

SQL-Kategorie	Zweck	Anweisungen
Data Definition Language (DDL)	Definition von Datenbankobjekten	create, alter, drop table, view, index
Data Manipulation Language (DML)	Abfrage Änderung	select insert, delete, update
Data Control Language (DCL)	Transaktionen Schutz / Sicherheit	lock, commit, rollback grant, revoke

Übersicht 6.10: Kategorien von SQL-Anweisungen

SQL - Vorteile

- ☺ portabel (▷ANSI / ▷ISO-Standard)
- ☺ für ▷verteilte Datenbanken geeignet
- ☺ in ▷prozedurale Programme einbettbar
- ☺ anders als ▷QBE terminalunabhängig

Mängel von SQL

Mehrdimensionale SQL-Abfragen sind benutzer~~un~~freundlich, weil oft ...

- ☹ Anweisungen **zahlreich**
- ☹ Anweisungen **komplex**
(Mehrtabellenverbund, Aggregation, Sortieren, Views)
- ☹ **Zeitreihenanalyse** aufwendig
- ☹ **Rechenfunktionen** unzulänglich
(z.B. finanzmathematische, statistische)



Benutzerfreundliche › mehrdimensionale Abfragen mit OLAP

5.19 Abfragearten

	<i>Hauptzweck</i>	<i>Flexibilität</i>	<i>Komfort</i>	<i>Interaktivität</i>
SQL	allgemein	nein	-	ja
Bericht	Abfrage breit	ja	+ ¹	nein ²
Menü	Abfrage eng	ja	+	ja
QBE	Abfrage breit	nein	∅ ³	ja
OLAP	Abfrage und Analyse	nein	+	ja



Integration von ›Berichtsgeneratoren,
›OLAP-Werkzeugen und ›Data Mining Tools

- Business Objects voll integriert
Bericht + Menüabfrage + OLAP
(*Business Miner* getrennt erhältlich)
- Cognos Improptu, PowerPlay, Scenario als Suite
Bericht, OLAP, Data Mining

1 aus Benutzersicht (Entwicklung mit ›3GL, ›4GL oder EIS Tools aufwendig)

2 Desktop Report Writers für den Entwickler beschränkt interaktiv

3 ›QBE syntaktisch einfach, verlangt aber ›logisches Datenmodell



Endbenutzerzugriff mit OLAP - Definition

OLAP (On Line Analytical Processing) :=

Abfragemethode, die ...

- ✓ Endbenutzern einen
- ✓ mehrdimensionalen
- ✓ schnellen Zugriff und eine
- ✓ benutzerfreundliche interaktive Analyse
(wie mit mehrdimensionalen Tabellenkalkulationspaketen)
- ✓ auf Data Marts ermöglicht

E.F. Codd führte 1993 das Konzept ein

5.20 OLAP versus OLTP

	OLTP	OLAP
<i>Abfragedaten</i>	operative Transaktionsdaten	managementkritische Analysedaten
<i>Granularität</i> (Detaillierungsgrad)	mikroskopisch (originäre Daten)	makroskopisch (oft abgeleitete Daten)
<i>Aktualität</i>	vollständiger Istzustand	Sequenzen historischer Snapshots
<i>Hauptoperationen</i>	fortschreiben (read / write)	abfragen, berichten und berechnen (meist read only)
<i>Werkzeuge</i>	›3GL, ›4GL (v.a.SQL)	›proprietär
<i>Abfragefreundlichkeit</i>	gering	gross
<i>Speichereffizienz</i>	gross	gering
<i>Optimierungsstrategie</i>	›transaktionsorientiert	datenorientiert

OLTP

On Line **Transaction** Processing

OLAP

On Line **Analytical** Processing

Anwendungen

Finanz- und Rechnungswesen

- Kurzfristige Erfolgsrechnung
- Jahresabschlussanalyse
- Cash Flow - Analyse
- Kennziffernanalyse (insb. ROI-Analyse)

Absatz

- Soll-/Ist-Vergleiche
- Produkt- und Kundenvergleiche
- Qualität des Kundendienstes

Beschaffung

- Bestandsanalysen
- Lieferfristenüberwachung

Produktion

- Kapazitätsanalysen
- Qualitätskontrolle

Personalwesen

- Personalverwaltung
- Mitarbeiterqualifikationen

...

Definitionselemente

OLAP ist ...

✓ **mehrdimensional**, weil es ...

Indikatoren nach ihren Dimensionen analysiert

✓ **endbenutzerfreundlich**, weil es ...

① direkte Manipulation ermöglicht¹

② zusammenfasst und nur auf Verlangen detailliert

‣ Drilling Up and Down navigiert entlang einer Dimension

③ analysiert und synthetisiert

‣ Slicing addiert und *Dicing* subtrahiert Dimensionen

④ Ergebnisse einfach darstellt, insbesondere visualisiert

✓ **schnell**, weil es ...

• ‣ ad hoc-Abfragen in Sekunden beantwortet

✓ **Data Mart** - orientiert, weil es ...

• weniger Details als das zentrale ‣ EDW unterstützt

• i.d.R. nicht direkt auf ‣ Produktionsdaten zugreift

¹ Vgl. mit der Komplexität einer generierten SQL-Abfrage von *if...Synchrony*

Datenbasis *NASDAQ*

NASDAQ (National Association of Security Dealers Automated Quotation System) ist ein Bildschirmkommunikationssystem für den Handel mit ausserbörslich kotierten Wertpapieren von etwa 5000 Unternehmungen

Datenmodell *Anlageberatung*

Würfel mit ...

25 **Indikatoren** (Fakten)

- jährlicher Geldumsatz
- jährlicher Nettoerfolg

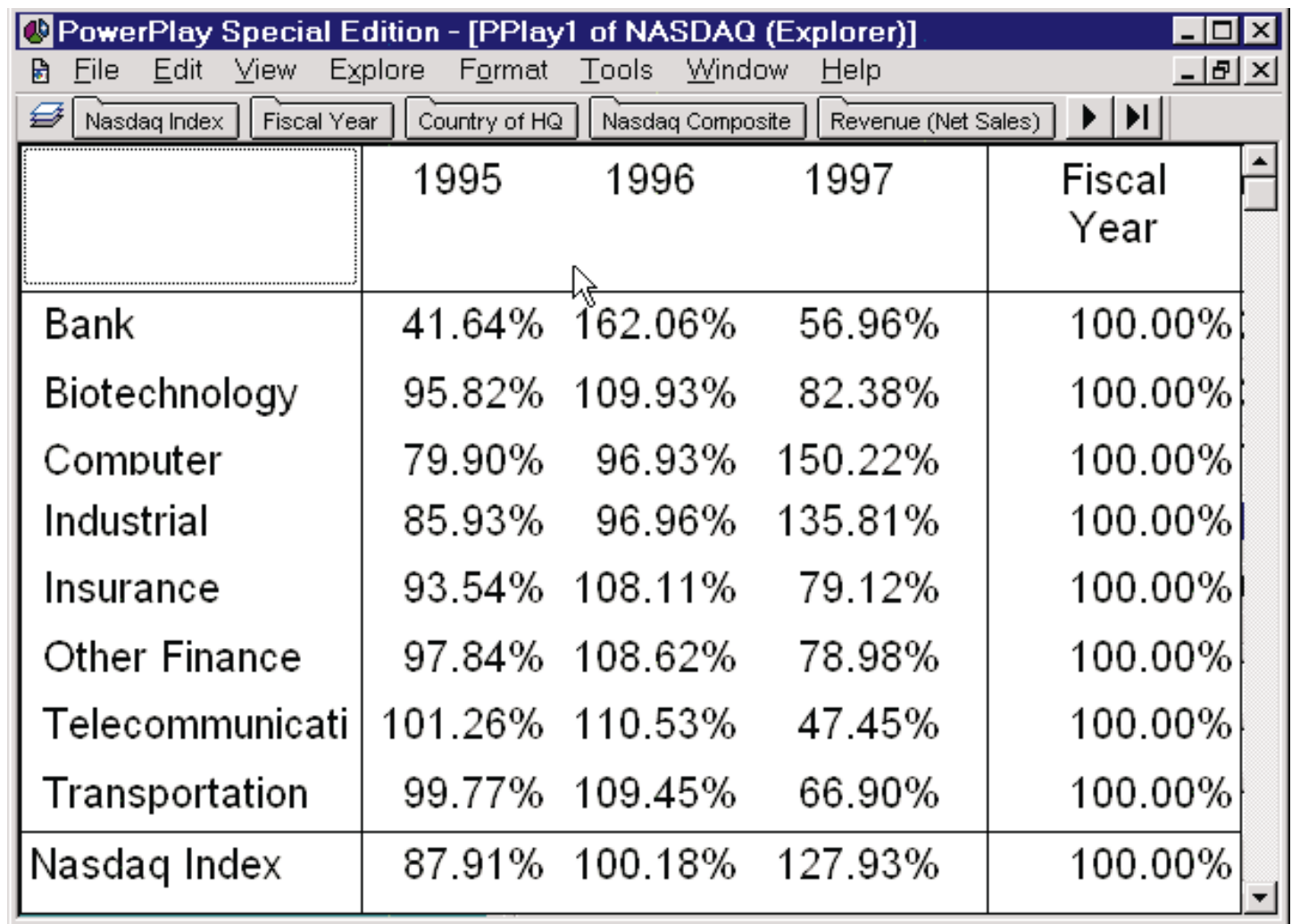
...

15 **Dimensionen**

- Branche
- Firmensitz
- Fiskaljahr
- Börsenwert (Klassenzugehörigkeit)
- jährliche Nettoerfolgsänderung in % (Klassenzugehörigkeit)
- Eigenkapitalrendite (Klassenzugehörigkeit)
- ...

 NASDAQ.mdc (multidimensional cube)

📌 Einfache Darstellung von Ergebnissen



The screenshot shows the 'PowerPlay Special Edition - [PPlay1 of NASDAQ (Explorer)]' window. The menu bar includes File, Edit, View, Explore, Format, Tools, Window, and Help. The toolbar has buttons for 'Nasdaq Index', 'Fiscal Year', 'Country of HQ', 'Nasdaq Composite', and 'Revenue (Net Sales)', along with navigation arrows. The main data table is as follows:

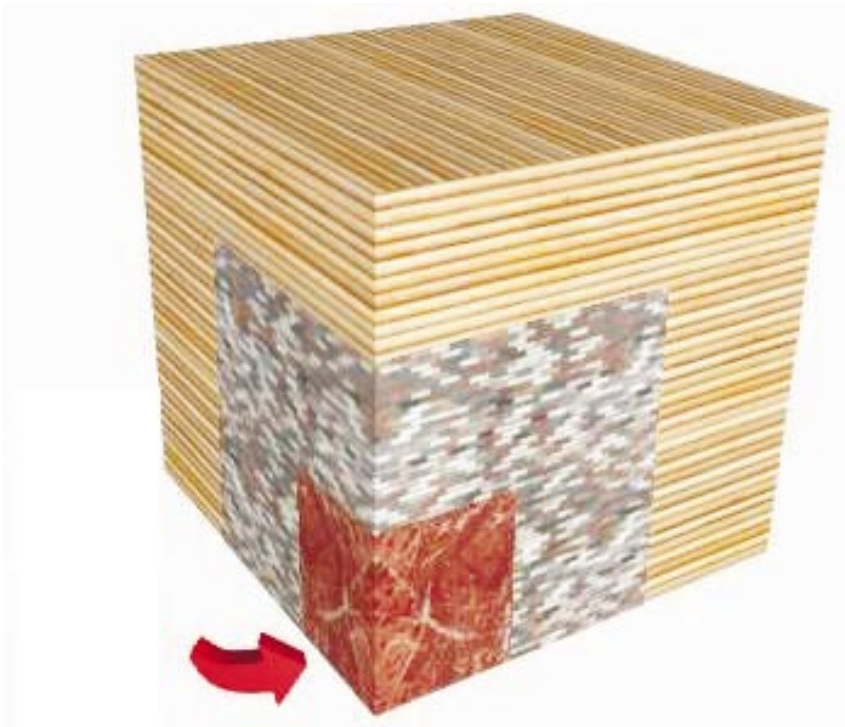
	1995	1996	1997	Fiscal Year
Bank	41.64%	162.06%	56.96%	100.00%
Biotechnology	95.82%	109.93%	82.38%	100.00%
Computer	79.90%	96.93%	150.22%	100.00%
Industrial	85.93%	96.96%	135.81%	100.00%
Insurance	93.54%	108.11%	79.12%	100.00%
Other Finance	97.84%	108.62%	78.98%	100.00%
Telecommunicati	101.26%	110.53%	47.45%	100.00%
Transportation	99.77%	109.45%	66.90%	100.00%
Nasdaq Index	87.91%	100.18%	127.93%	100.00%

Beispiel: Der Menüpunkt *View/Show Values As ...* wechselt zwischen absoluter und relativer Darstellung

› Indikator **in Prozent von ...**

- Zeilentotal
- Spaltentotal
- Ebene (engl. layer)
- Gesamttotal

Importierte und abgeleitete Daten



1. **Importierte** Daten

v.a. aus operativen Datenbanken und EDW

2. **Abgeleitete** Daten

z.B. mit gespeicherten Prozeduren

3. **Ad hoc abgeleitete** Daten

z.B. mit generiertem SQL

Je mehr permanent abgeleitete Daten, desto ...

- kleiner die Antwortzeit
- grösser die Ladezeit
- grösser der Speicherbedarf

Zusammenfassungen und Vergleiche

Zusammenfassung (Aggregation)

- ✓ *Quartalerträge* des letzten Jahres

Vergleich

- ✓ *Absoluter* Vergleich der diesjährigen mit letztjährigen Gewinnen
- ✓ *Prozentualer* Beitrag aller Produkte an den Gesamtumsatz
- ✓ Abschlüsse der *besten* 10% US-Vertreter

...



- Trendanalysen
- Soll- / Istvergleiche
- Benchmarkvergleiche
- . . .

- ① SQL sucht nach Details
 - ② OLAP fasst zusammen und vergleicht

Zusammenfassung - Ein Beispiel

StarTracker Demo

File Edit **Aggregates** Sequences Comparisons Help

Run Report

PERIODEN:
 Sum
 Count
 Count Distinct
 Avg
 Min
 Max
 AvgPeriodSum
 AvgPeriodCount
 Tagesanzahl
 Wochenanzahl

PRODUKTE: Alle Produkte
 Browse
 Expand
 Produkt#
 Name
 Gewicht
 Lieferant
 Unterkategorie
 Kategorie

Verkäufe
 Periode#
 Produkt#
 Aktion#
 Filiale#
 Geldumsatz
 Mengenumsatz
 Kosten
 Kundenzahl
 Durchschnittlich verkaufte Stückzahl*
 Durchschnittliche Auftragsgrösse*
 Durchschnittskosten*
 Durchschnittspreis*
 Gewinn*
 Kostenquote*
 Report Columns*

FILIALEN: Alle Filialen
 Browse
 Expand
 Filiale#
 Strasse
 Ort
 Kanton
 PLZ
 Geschäftszeiten

AKTIONEN: Alle Aktionen
 Browse
 Expand
 Aktion#
 Name
 Preisreduktion
 Anzeige
 Display
 Anzeigenmedium

Kostenvergleich
 for 4. Quartal '98

	A	B	C	D	E	F	G	H	I	J	K
	Ort	Veränderung in %	Kosten 4. Quartal '98	Kosten 4. Quartal '97							
1	Aarau	2.51%	15898.66	15508.86							
2	Basel	-13.63%	17456.13	20209.79							
3	Bellinzona	-18.30%	15501.94	18975.36							
4	Bern	8.35%	17459.52	16113.75							
5	Biel	-16.48%	15068.73	18042.93							
6	Brig	-1.08%	17070.44	17257.63							
7	Chur	7.32%	19072.19	17771.85							
8	Freiburg	-1.62%	17100.58	17381.68							
9	Genf	21.93%	17656.07	14480.07							
10	Laufen	7.74%	17182.35	15947.90							
11	Lausanne	16.87%	20263.64	17338.48							
12	Liestal	-3.82%	17476.27	18169.48							
13	Locarno	-3.39%	15828.11	16383.57							
14	Luzern	-14.94%	16680.32	19609.66							
15	Martigny	1.28%	16880.94	16667.19							
16	Rheinfelden	12.73%	17751.06	15746.94							
17											

Family: Handelsunternehmen 16.10.97 14:04

OLAP-Werkzeuge generieren im Hintergrund oft SQL-Abfragen

Zusammenfassungen können verfälschen

Eine **Aggregation** (Zusammenfassung) kann zu Informationsverlusten führen

Beispiel

”Gründe für tiefe Umsätze ? “



Konsolidierung von Tagesumsätzen zu Wochenumsätzen verdeckt den Einfluss einzelner Tagesumsätze

Browsing

Browsing :=

benutzerfreundliche Abfrage mit ...

✓ **Drilling Down and Up**

Zusammenfassung und Detaillierung entlang einer Dimension

✓ **Drilling Through**

Detailzugriff auf operative Datenbanken
(meist über vordefinierte SQL-Anweisungen)

✓ **Drilling Across**

Zugriff auf mehrere Data Marts

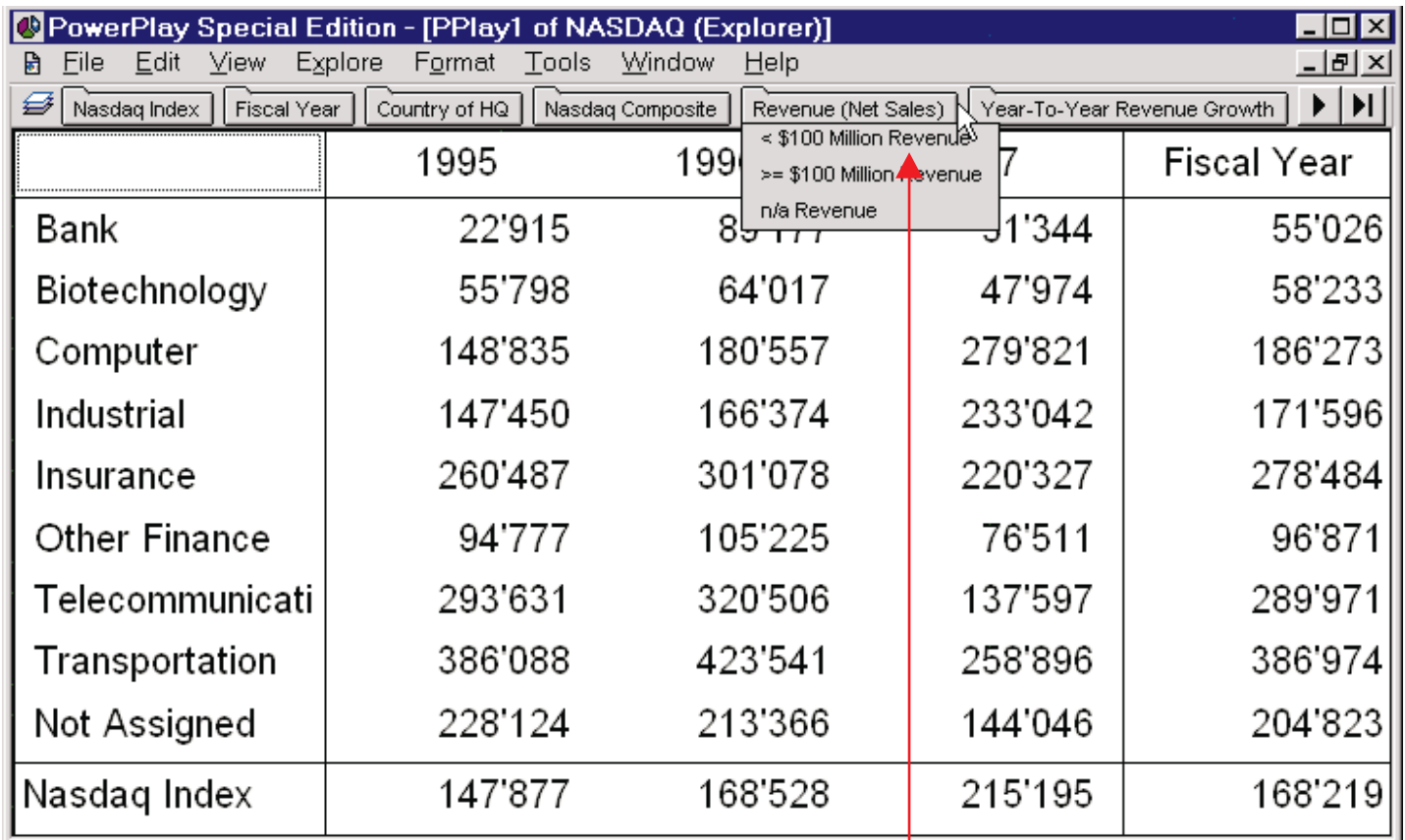
✓ **Filtern**

Werteauswahl

✓ **Slicing and Dicing**

Dimensionsauswahl

5.21 📌 Filtern beschränkt die angezeigten Werte



The screenshot shows the PowerPlay Special Edition interface. The 'Revenue (Net Sales)' dimension is selected in the horizontal dimension bar, and a filter menu is open showing the criteria: '< \$100 Million Revenue', '>= \$100 Million Revenue', and 'n/a Revenue'. The '>= \$100 Million Revenue' criterion is selected, and a red vertical line indicates the filter is applied to the data table below.

	1995	1996	1997	Fiscal Year
Bank	22'915	85'177	51'344	55'026
Biotechnology	55'798	64'017	47'974	58'233
Computer	148'835	180'557	279'821	186'273
Industrial	147'450	166'374	233'042	171'596
Insurance	260'487	301'078	220'327	278'484
Other Finance	94'777	105'225	76'511	96'871
Telecommunicati	293'631	320'506	137'597	289'971
Transportation	386'088	423'541	258'896	386'974
Not Assigned	228'124	213'366	144'046	204'823
Nasdaq Index	147'877	168'528	215'195	168'219

Filtern (Werteauswahl) :

Cursor auf eine Dimension der horizontalen Dimensionenleiste bewegen und auf ein Filterkriterium klicken

Beispiel: Unternehmungen mit einem Jahresertrag von ≥ 100 Mio

Drilling Up and Down

Drilling Up (zusammenfassend)
Drilling Down (detaillierend)

Umsatzzahlen für ...

ganz Europa

Mitteleuropa

Deutschland

9

5.21 📌 Drilling Down - Ausgangslage

PowerPlay Special Edition - [PPlay2 of NASDAQ (Explorer)]

File Edit View Explore Format Tools Window Help

Nasdaq Index Fiscal Year Country of HQ Nasdaq Composite Revenue (Net Sales)

		1996	1997	Fiscal Year
Bank	915	89'177	31'344	55'026
Biotechnology	55'798	64'017	47'974	58'233
Computer	148'835	180'557	279'821	186'273
Industrial	147'450	166'374	233'042	171'596
Insurance	260'487	301'078	220'327	278'484
Other Finance	94'777	105'225	76'511	96'871
Telecommunications	293'631	320'506	137'597	289'971
Transportation	386'088	423'541	258'896	386'974
Nasdaq Index	147'877	168'528	215'195	168'219

① Indikator

Beispiel: *Jahresertrag*

② Dimension :

Cursor über einem Dimensionsreiter der Dimensionszeile

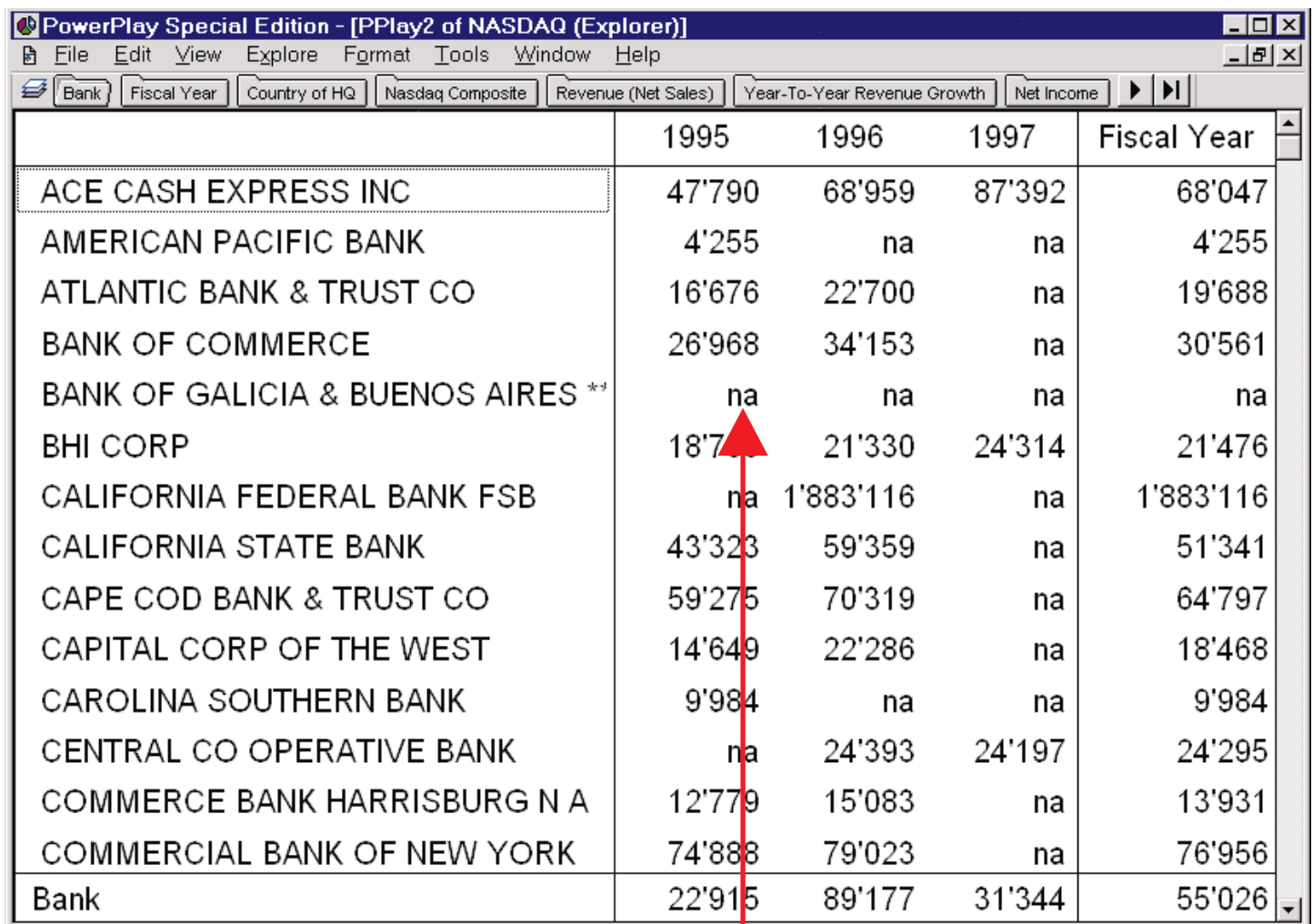
Beispiel: Dimensionen *Branche* und *Zeit*

③ Drilling Down or Up →

Down: Doppelklick auf Zellenwert. Up: rechte Maustaste

Beispiel: Jahreserträge der einzelnen *Bankunternehmungen*

5.22 📌 Drilling Down - Ergebnis



The screenshot shows a software window titled "PowerPlay Special Edition - [PPlay2 of NASDAQ (Explorer)]". It has a menu bar (File, Edit, View, Explore, Format, Tools, Window, Help) and a toolbar with buttons for "Bank", "Fiscal Year", "Country of HQ", "Nasdaq Composite", "Revenue (Net Sales)", "Year-To-Year Revenue Growth", and "Net Income". Below the toolbar is a table with the following data:

	1995	1996	1997	Fiscal Year
ACE CASH EXPRESS INC	47'790	68'959	87'392	68'047
AMERICAN PACIFIC BANK	4'255	na	na	4'255
ATLANTIC BANK & TRUST CO	16'676	22'700	na	19'688
BANK OF COMMERCE	26'968	34'153	na	30'561
BANK OF GALICIA & BUENOS AIRES **	na	na	na	na
BHI CORP	18'7	21'330	24'314	21'476
CALIFORNIA FEDERAL BANK FSB	na	1'883'116	na	1'883'116
CALIFORNIA STATE BANK	43'323	59'359	na	51'341
CAPE COD BANK & TRUST CO	59'275	70'319	na	64'797
CAPITAL CORP OF THE WEST	14'649	22'286	na	18'468
CAROLINA SOUTHERN BANK	9'984	na	na	9'984
CENTRAL CO OPERATIVE BANK	na	24'393	24'197	24'295
COMMERCE BANK HARRISBURG N A	12'779	15'083	na	13'931
COMMERCIAL BANK OF NEW YORK	74'888	79'023	na	76'956
Bank	22'915	89'177	31'344	55'026

A red arrow points from the "na" value in the 1995 column for "BHI CORP" to the text "na (not available)" below the table.

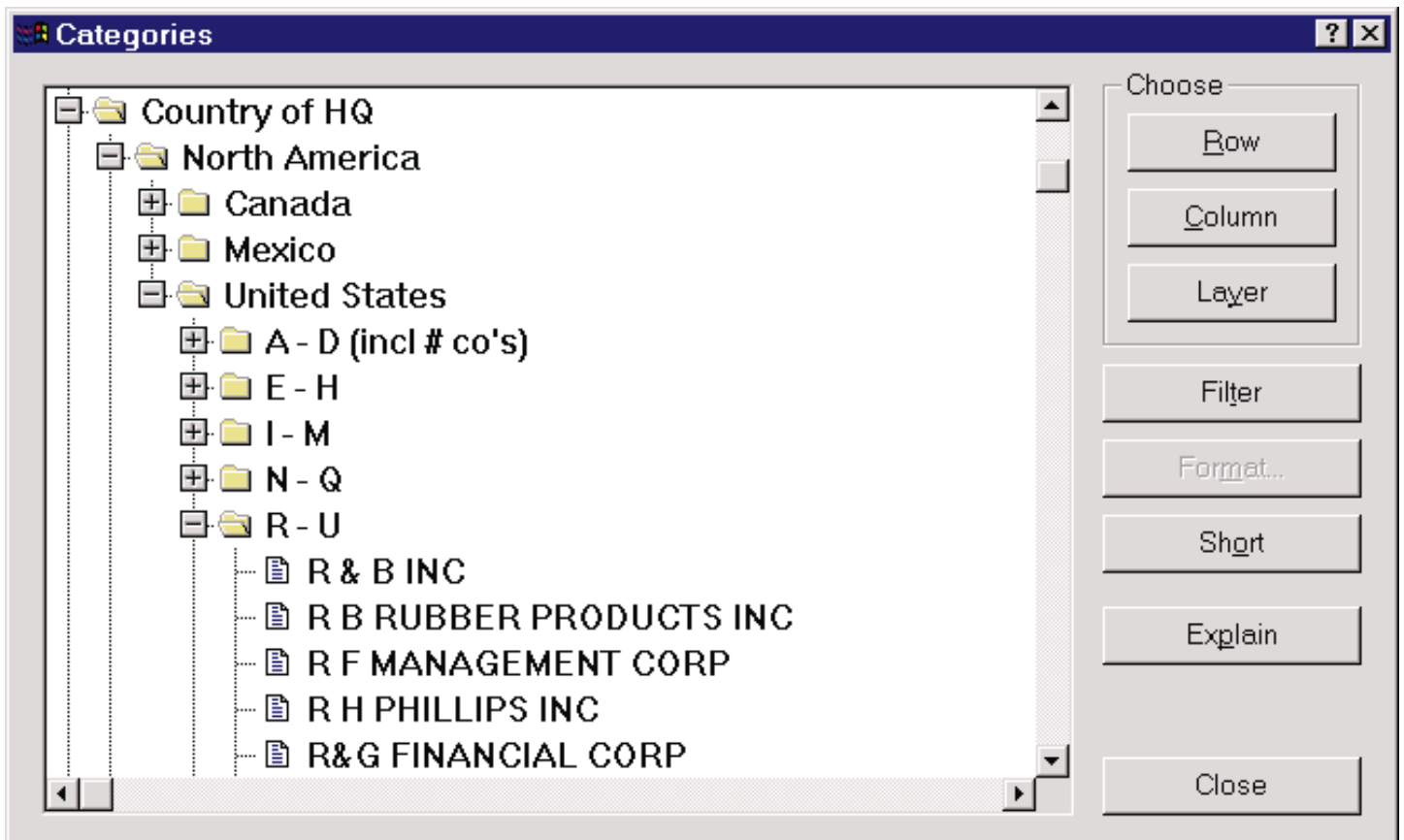
na (not available)

- Je grösser die **Granularität**, desto mehr überflüssiger Speicher
- Je mehr **fehlende Werte**, desto mehr Speicherineffizienz



Wie werden **dünn besetzte Matrizen** am besten gespeichert?

📌 Drilling navigiert in Kategorienhierarchien



Kategorienhierarchie einer Dimension

Die Dimension FIRMENSITZ ("Country of HeadQuarter") ordnet ihre Kategorien (Wertebereiche) **hierarchisch**:

Dimension FIRMENSITZ

Kategorie Ländergruppe (Vaterknoten)

Kategorie Land (Kindknoten von Ländergruppe)

Kategorie alphabetische Gruppe

Kategorie Einzelunternehmung

Dimensionen mit Kategorien beschreiben

① Hierarchische Kategorien

Bsp. Kategorien der Dimension *Periode*

Perioden

1995

1. Quartal

Januar

...

2. Quartal

...

...

1996

...

Die Kategorienhierarchie der Dimension *Periode* ordnet die Kategorien **Jahr**, **Quartal** und **Monat** hierarchisch

② Flache Kategorien

Bsp. Kategorien der Dimension *Produkt*

Name, Farbe, Grösse, Art ...

Ein Indikator oder eine Dimension kann *mehrere* flache oder hierarchische Kategorien haben

Drilling Down kann verfälschen

Zu kleine **Fallzahlen** können bei mehrdimensionalen Analysen zu Informationsverlusten führen

Beispiel

“Gründe für tiefe Umsätze ?“



Die Abfrage “Umsatz nach Produkt, Filiale und Abteilung” reduziert die Stichprobe zu stark

5.23 † Slicing und Dicing analysiert / synthetisiert

Slicing and Dicing (Dimensionswahl):

Änderung der angezeigten Dimension mit Drag and Drop einer Dimension der horizontalen Leiste über dem Anzeigefenster

Beispiel: Drag and Drop der neuen Dimension *Firmensitz* (Country of HQ) in die horizontale oder vertikale Titelleiste des Anzeigebereichs ersetzt die alte Dimension *Fiskaljahr*

	Nasdaq Index	Fiscal Year	Country of HQ	Nasdaq Composite	Revenue (Net Sales)
		1995	1996		
Bank		22'915	89'177		
Biotechnology		55'798	64'017		
Computer		148'835	180'557		

	Nasdaq Index	Fiscal Year	Country of HQ	Nasdaq Composite	Revenue (Net Sales)
		North America	Europe		
Bank		56'023	na		
Biotechnology		54'825	26'932		
Computer		191'503	80'836		

📌 Pivoting tauscht Zeilen und Spalten

PowerPlay Special Edition - [PPlay1 of NASDAQ (Explorer)]

File Edit View Explore Format Tools Window Help

Nasdaq Index Fiscal Year Country of HQ Nasdaq Composite Revenue (Net Sales)

	Bank	Biotechnology	Computer	Industrial
North America	56'023	54'825	191'503	172'717
Europe	na	26'932	80'836	149'610
Asia Pacific	na	920	794'566	90'836
Other	21'476	314'846	52'088	93'256
Country of HQ	55'026	58'233	186'273	171'596


Pivoting :
Zeilen und Spalten tauschen¹

Beispiel: Tausch der Dimensionen *Branche* (Bank, ...) und *Firmensitz*

¹ to pivot heisst drehen

5.24 † Eine **Ebene** ist eine Zusatzdimension

Ebenen: Für jeden Wert einer Zusatzdimension eine zusätzliche Ebene (**Drag and Drop** von *Country of HQ* auf das Ebenensymbol)



	Country of HQ	Nasdaq Composite	Revenue (Net Sales)	Year-To-Year Revenue Growth
	1995	1996	1997	Fiscal Year
Bank	22'915	89'177	31'344	55'026
Biotechnology	55'798	64'017	47'974	58'233
Computer	148'835	180'557	279'821	186'273

Beispiel

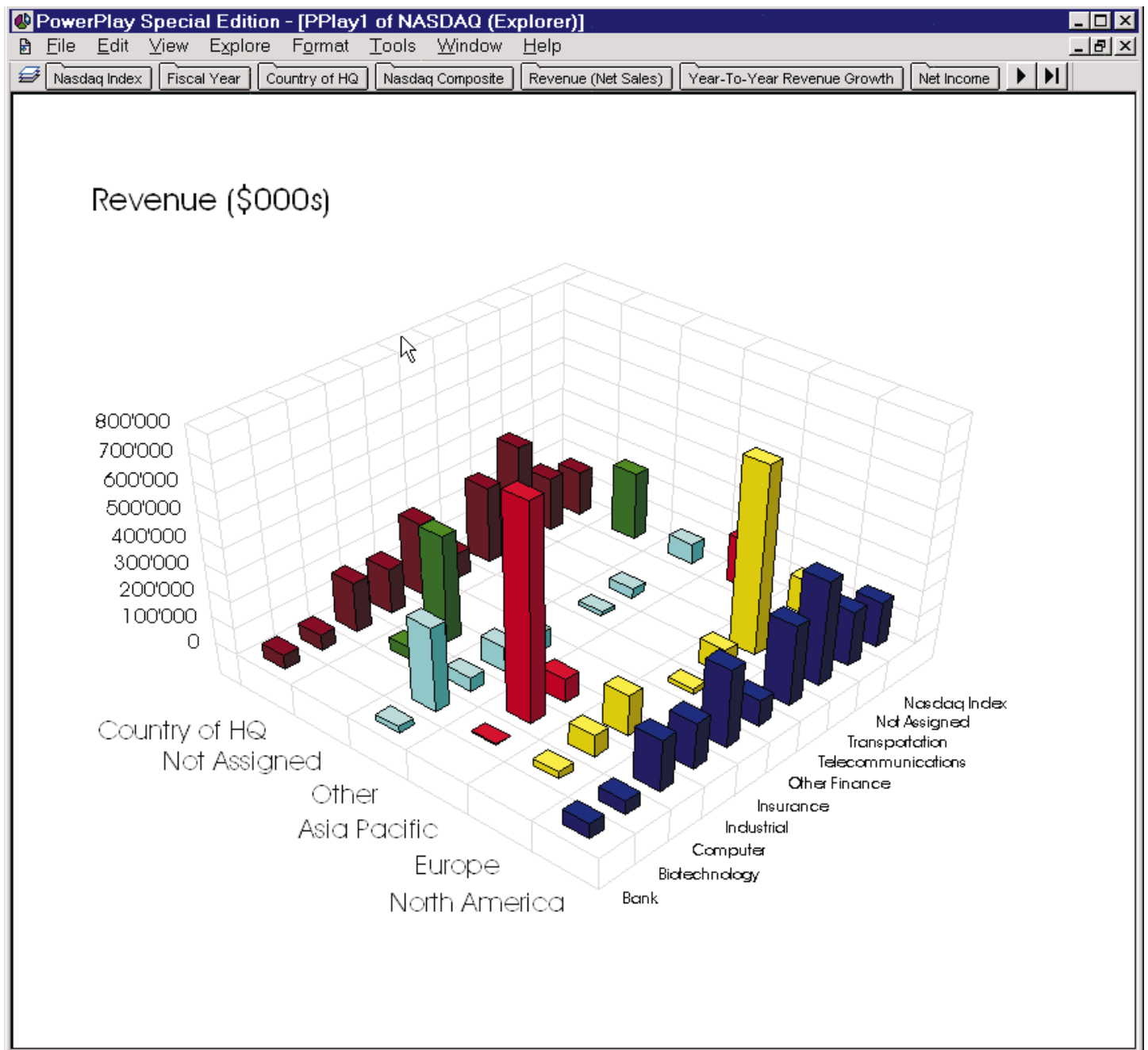
North America als Wert der Zusatzdimension *Firmensitz*:

Unterschied zum Slicing and Dicing

Anzeige von 3 (statt nur 2) Dimensionen

North America	Layer 1 of 6			
	1995	1996	1997	Fiscal Year
Bank	23'011	90'590	32'047	56'023
Biotechnology	51'968	59'625	48'826	54'825
Computer	153'948	186'081	279'691	191'503
Industrial	148'311	167'423	234'267	172'717
Insurance	270'406	312'268	220'327	288'479
Other Finance	94'996	105'454	76'847	97'115
Telecommunicati	303'266	332'067	137'597	298'746
Nasdaq Index	149'300	170'102	216'055	169'744

5.25 📌 Die Visualisierung hebt hervor



Dreidimensionale Visualisierung :

Dreidimensionale grafische Veranschaulichung eines Indikators und zweier Dimensionen

Beispiel: Indikator *Jahresertrag*, Dimensionen *Branche*, *Firmensitz*

OLAP-Arten - Definitionen

ROLAP (engl. Relational OLAP) :=

- ✓ OLAP
- ✓ mit mehrdimensionalen Metadaten
- ✓ in einem relationalen Server-Datenbanksystem

MOLAP (engl. Multidimensional OLAP) :=

- ✓ OLAP
- ✓ mit mehrdimensionalen Objektdaten
- ✓ in einem proprietären Server-Datenbanksystem

DOLAP (engl. Desktop OLAP) :=

- ✓ MOLAP oder ROLAP
- ✓ auf einem Client (i.d.R. einem PC)

OLAP-Arten - Vergleich

	ROLAP	MOLAP	DOLAP
<i>Abkürzung</i>	<u>R</u> elationales OLAP	<u>M</u> ehrdimen- sionales -	<u>D</u> esktop -
<i>EDW-Daten i.d.R.</i>	›relational	relational	relational,/ multdimen- sional
<i>Mehrdimensiona- lität in den ...</i>	Metadaten (Direktzugriff auf EDW möglich)	Objektdaten	›Metadaten/ Objektdaten
<i>Datenumfang</i>	gross	gross	klein
<i>Datenort</i>	Server	Server	Client (PC)
<i>Datenmodell</i>	standardisiert	›proprietär	›proprietär
<i>Abfrageeffizienz</i>	-	+	+
<i>Flexibilität</i>	+	-	+/-
<i>Skalierbarkeit</i>	+	Ø	-
<i>Integrierbarkeit</i>	+	-	+/-
<i>Vorberechnung</i>	geringer	umfangreich	umfangreich
<i>Beispielanbieter</i>	<u>Oracle</u> <u>Essbase</u>	<u>Arbor</u> , <u>Oracle</u>	<u>Cognos</u> , <u>Brio</u>

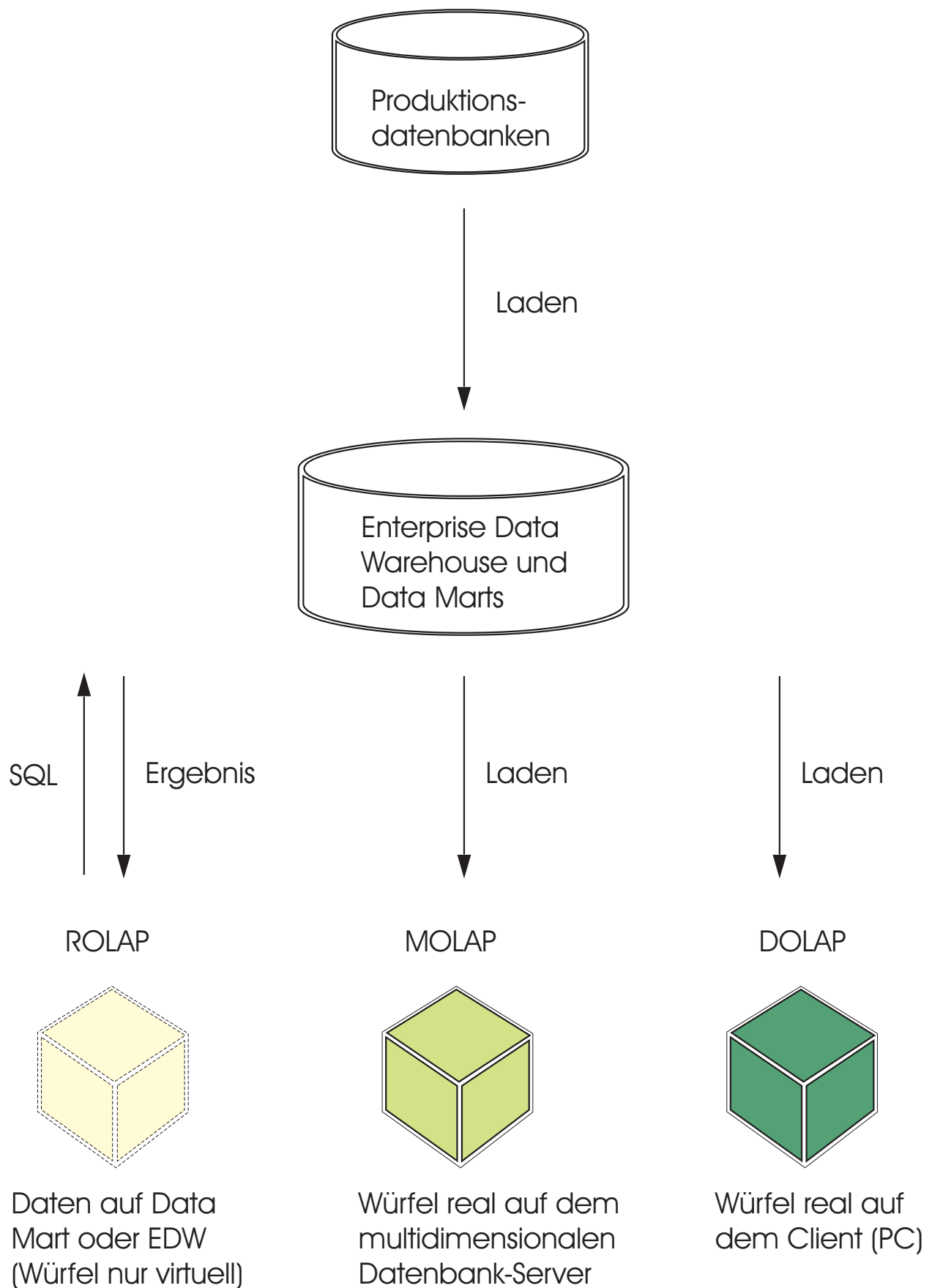
Hybride OLAP-Systeme

- MOLAP auf einem mehrdimensionalen Würfel
- ROLAP durch generiertes SQL auf einem relationalen DW

Bsp. Pilot Decision Support Suite

15

5.28 OLAP-Architekturen



5.29 OLAP-Architekturen und Produkte

	Speicherung in RDBMS	Speicherung im MDBMS	Speicherung im Client
Verarbeitung mit RDBMS	DSS Agent Server		
Verarbeitung mit MDBMS	Oracle Express	Hyperion Essbase Oracle Express	
Verarbeitung im Client	PowerPlay Server	Hyperion MBA	BrioQuery PowerPlay Client
<u>Legende</u>	ROLAP	MOLAP	DOLAP
	light client		
	fat client		

Produkte und Hersteller

- [DSS Agent](#) (Server) von MicroStrategy
- [Express](#) (MOLAP und ROLAP) von Oracle
- [PowerPlay](#) (Server) von Cognos
- [Hyperion Essbase](#) von Hyperion Software
- [Hyperion Enterprise](#) von Hyperion Software (spezialisiert für Finanzanwendungen)
- [PowerPlay](#) (Client) von Cognos (populäres DOLAP-Produkt, Data Mining-Ergänzung Cognos [Scenario](#))
- [BrioQuery](#) von Brio Technology

PowerPlay ist . . .

❶ ein **DOLAP** - Frontend ...

Desktop - OLAP-Frontend auf dem

- Client oder
- Server

❷ für benutzerfreundliche **Abfragen** ...

- ad hoc statt geplant (vordefiniert)
- direktmanipulierend statt befehlsorientiert (wie SQL)
- explorativ statt off line (Drilling, Slicing/Dicing, ...)

❸ auf einem **Würfel** mit proprietär Datenstruktur

PowerPlay setzt einen *vordefinierten* Würfel
aus Data Warehouse-Daten voraus

5.30 Entwicklung eines DOLAP-Würfels

Metadatenverwalter des Data Warehouse

(z.B. Sybase PowerDesigner Warehouse Architect)

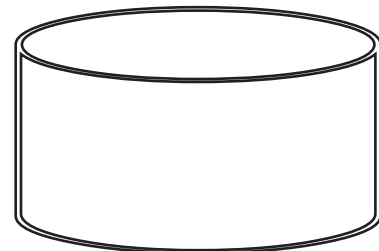
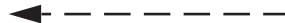
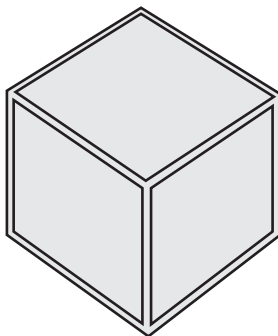


Grobes physisches Modell der multidimensionalen Daten



Metadatenverwalter des DOLAP-Würfels

(z.B. Cognos PowerPlay Administrator)

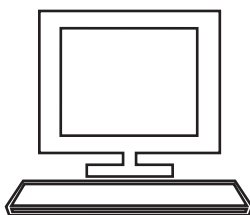


Multidimensionaler Würfel

(z.B. PowerPlay PowerCube)

Data Warehouse

(z.B. Sybase - RDBMS)



DOLAP - Viewer


(z.B. Cognos PowerPlay)

Cognos PowerPlay und ein Würfel aus NASDAQ-Daten sollen einen Anlageberater bei Portfolioentscheidungen unterstützen. Im ersten Teil der folgenden Aufgabe lernen Sie PowerPlay geleitet kennen. In einem zweiten Teil können Sie das Werkzeug an einem Anlageberatungsbeispiel selbständig erkunden.

Lernziele

- ⇒ Das Grundmodell der mehrdimensionale Analyse erkennen
- ⇒ OLAP-Abfragekonzepte anwenden

1 PowerPlay kennen lernen

- Laden Sie den Würfel  [NASDAQ.mdc](#).
- Mit den Menüpunkten *View* und *File/Preferences* können Sie die Benutzeroberfläche konfigurieren.
- Wenn Sie den Cursor auf ein Icon der Symbolleiste positionieren, erhalten Sie eine Kurzbeschreibung. Ausführliche Hilfe zur Symbol- und Menüleiste finden Sie unter dem Menüpunkt *Help/Inhalt/Reference Information/ Shortcuts*. Die Shortcuts Ctrl-Z und Ctrl-Y stehen für undo bzw. redo.
- Informationen zum Inhalt des NASDAQ-Würfels erhalten Sie mit einem Rechtsklick auf eine Zelle oder auf ein Register (engl. tab) der Dimensionenleiste. Der Kontextmenüpunkt “Explain” erklärt dann die Bedeutung des Indikators, der Dimension oder des Wertes.

a) Dimensionen und Indikatoren

- Nach dem Start erscheint die voreingestellte Ansicht “3-D-Bar” (Falls keine dreidimensionale Balkengrafik erscheint, klicken Sie auf das Icon 3-D-Bar). Identifizieren Sie den Indikator, die Dimensionen und deren Wertebereiche (Kategorien).

- Ein Klick auf das Icon “Crosstab” wechselt zur tabellarischen Ansicht. Unter der Symbolleiste finden Sie mehrere Register zu den Dimensionen und Indikatoren des Würfels. Den Indikator “Revenue” (Ertrag in 1000\$) finden Sie am rechten Ende der Registerleiste. Wenn Sie den Mauszeiger darauf positionieren, erscheinen weitere mögliche Indikatoren.
- Ein Klick auf ein Register zeigt die Kategorien der Dimension bzw. die Measures in einem Pop Up-Fenster. Schauen Sie sich die Kategorienhierarchien der Dimensionen *Country of HQ* und *Net Income* an.

b) Filter

Die Ausgangstabelle enthält den Indikator “Durchschnittlicher Unternehmungsertrag” nach den Dimensionen Branche (“Nasdaq Index”) und Steuerjahr (“Fiscal Year”). Verwenden Sie einen Filter, der nur Unternehmen mit einer Eigenkapitalrentabilität von mehr als 20% zeigt. Gehen Sie dazu mit dem Mauszeiger zur Dimension “Return on Equity”. Sobald Sie die Maus nicht mehr bewegen, erscheint ein Menü mit den Filterkriterien. Wählen Sie “>=20% ROE” (engl. Return On Equity). Experimentieren Sie mit den anderen Kriterien.

c) Pivoting

Wechseln Sie die Spalten und Zeilen mit *Explore/Swap/Rows and Columns*.

d) Drilling Down and Up

Ein Doppelklick auf eine Tabellenzelle führt eine Kategorie tiefer (Drilling Down). Eine Kategorie höher kommen Sie mit einem Rechtsklick auf eine Zelle (Drilling Up).

e) Nullwerte

“na” (not available) bezeichnet Zellen ohne Wert. *Explore/Suppress Zeros/Rows and Columns* zeigt Zeilen oder Spalten, die nur aus Nullwerten bestehen, nicht an.

f) Ansichten (engl. views)

View/Show Values As zeigt Ergebnisse prozentual statt absolut an, und *Explore/Hide Selected Categories* blendet markierte Zeilen und Spalten aus.

g) Sortierung (engl. rank)

Sortieren Sie markierte Spalten oder Zeilen mit dem Menüpunkt *Explore/Rank*.

h) Ebenen (engl. layers)

Tabellen können nur zwei Dimensionen gleichzeitig darstellen. Um eine dritte Dimension zu zeigen, können Sie jeden ihrer Werte als zusätzliche Tabelle (Ebene) anzeigen lassen. Ziehen Sie dazu die gewünschte dritte Dimension mit der Maus auf das Stapelsymbol am linken Ende der Dimensionsleiste.

i) Visualisierung

Wechseln Sie mit *View/Change Display* von der Tabellenansicht (crosstab) zur Würfelsicht (3D-Bar). Wenden Sie Filter, Slice and Dice, Drill Down and Up und Sortierung auch auf diese Sicht an.

2 PowerPlay selbständig anwenden

a) Zeigen Sie die TopTen der Banken mit den höchsten Eigenkapitalrenditen (ROE) des Jahres 1997 an.

b) Welchen Marktwert hatte "Microsoft" 1997?

c) Veranschaulichen Sie die Risiko-Ertragslage in einem Multiline-Diagramm, das die Eigenkapitalrendite (ROE) und den Beta-Koeffizient branchenweise darstellt. Welche Branche weist das ungünstigste Verhältnis aus? Welche das beste? Diskutieren Sie zunächst, welche Beta-Werte günstig oder ungünstig sind. Nehmen Sie zur Lösung der Aufgabe an, dass ein möglichst kleiner (positiver) Beta-Wert ideal sei. Warum kommen Sie u.U. zur falschen Lösung, wenn Sie den ROE durch Beta dividieren und

das höchste Resultat als das beste und das tiefste als das schlechteste Verhältnis ansehen?

d) Reporter

Sie möchten ein Aktienportefeuille nach verschiedenen Kriterien erstellen. Da Sie aber nicht alles auf eine Karte setzen wollen, verteilen Sie ihr Geld nach den folgenden Kriterien auf fünf Aktienpakete:

- Es kommen nur Unternehmen in Frage, die 1997 einen Nettogewinn ("net income") von mindestens 250 Mio \$ erzielt haben.
- Wählen Sie die Firmen mit dem höchsten ROE/Beta-Verhältnis.
- Beteiligen Sie sich proportional zur Mitarbeiterzahl.

3 Möglichkeiten und Grenzen erkennen

- a) Kritisieren Sie die Portfoliostrategie des Anlageberaters.
- b) Stellen Sie sich eigene Aufgaben. Suchen Sie dabei nach neuen Funktionen und Grenzen von *PowerPlay*.
- c) Welche Vor- und Nachteile hat ein OLAP-Tool wie *PowerPlay* im Vergleich zu QBE, SQL oder Tabellenkalkulationsprogrammen?
- d) Implementieren Sie die Aufgabe 2d in QBE und SQL.

Funktionalität von *PowerPlay*

Eingabe

- ✓ Würfeldaten real oder als Teil eines ›RDBMS speichern
- ✓ Realen Würfel auf einem PC oder einem Server speichern
- ✓ Benutzerfreundliche ›DOLAP-Abfragen definieren

Verarbeitung

- ✓ Slicing and Dicing
- ✓ Drilling Down and Up
- ✓ Drilling Through mit Cognos *Impromptu*
- ✓ Automatische PowerCube-Generierung
- ✓ Zugriff auf relationale und mehrdimensionale DBMS
- ✓ ›API für MS *Visual Basic*, MS *Excel* u.a. Pakete

Ausgabe

- ✓ zweidimensionale Tabellen
- ✓ dritte Dimension durch Ebenen
- ✓ Point and Click - Grafik

① Ausgewählte Kriterien für OLAP-Clients

Beispiele von OLAP-Clients

- Brio Enterprise von [Brio Technology](#)
- Business Objects [Business Objects](#)
- PowerPlay von [Cognos](#)
- Express von [Oracle](#)
- Seagate Info von [Seagate](#)

Funktionalität

- ✓ ›Drill Down, Up und Through
 - textuell (z.B. durch Doppelklick)
 - grafisch (z.B. durch Klick auf ein grafisches Element)
- ✓ ›Slice and Dice
- ✓ ›Filterung
- ✓ Sortierung
- ✓ ›Ebenen
(dreidimensionale Anzeige)
- ✓ Verwaltung mehrerer Würfel bzw. Sternschemata
- ✓ ›Views
- ✓ Ad hoc-Berechnung temporärer Dimensionen
- ✓ ›Pivoting
- ✓ Point and Click Graphing
- ✓ Automatische Unterdrückung von ›Nullwerten

✓ Dynamische Berichte

✓ Agenten

(automatische Aktionen bei Vorliegen einer benutzerdefinierten Bedingung, zum Beispiel automatischer Versand von EMail, Fax oder Bestellungen)

Integration

✓ SQL-Schnittstelle

✓ Integration von Software zu Tabellenkalkulation, Data Mining, Gruppenarbeit, GIS (Geographical Information Systems), ...

✓ Integration von Technologien wie ›COM und WWW

✓ Erweiterungsfähigkeit

(z.B. durch ›Visual Basic oder C++)

Benutzerfreundlichkeit

✓ Wahl von ›Indikatoren, ›Dimensionen und ›Kategorien durch Auswahl (zum Beispiel in einem Listenfeld) *und* Namensnennung

✓ einfache Navigation in Kategorienhierarchien

✓ Transparenz (Implementationsdetails für Benutzer nicht erkennbar)

✓ ...

② Ausgewählte Kriterien für OLAP-Server

Beispiele von OLAP-Servern

- Microsoft SQL Server 7 OLAP Services
- Hyperion Essbase OLAP Server von [Hyperion Software](#)
- IBM DB2 OLAP Server
- MetaCube ROLAP Option von [Informix](#)

Architektur

- ✓ ein-, zwei- oder mehrstufig
- ✓ relationale oder mehrdimensionale Objektdaten
- ✓ ein einziger Würfel oder mehrere kombinierbare Würfel

Funktionalität

- ✓ Zahl und Art der unterstützten Frontends
- ✓ Zahl und Art der unterstützten RDBMS-Quellen
- ✓ verteilte Datenhaltung
- ✓ Definition von Attributen zur Beschreibung von Indikatoren, Dimensionen, Kategorien und Werten möglich
- ✓ Vielfalt der Datentypen (numerische und nichtnumerische, einfache und zusammengesetzte)
- ✓ Strukturen und Operationen auf der Zeitdimension
- ✓ ad hoc- und vorberechnete Zusammenfassungen
- ✓ Aggregationen (insb. arithmetische Operationen)
- ✓ Verknüpfung von Würfeln oder Sternschemata
- ✓ Drill Through (Zugriff auf die originären Daten operativer Datenbanken über vordefinierte SQL-Anweisungen)

- ✓ Drill Across (Zugriff über mehrere Data Marts)
- ✓ Sicherheitskomfort (v.a. Zugriffsberechtigungen)
- ✓ Kompatibilität zu OLAP Council-Empfehlungen

Wartbarkeit

- ✓ Ladeaufwand (z.B. inkrementelles Laden)
- ✓ Modellierungsaufwand (z.B. grafischer Datenbankentwurf)
- ✓ Optimierungsaufwand

Effizienz

- ✓ Ladezeit (zur Initialisierung und periodischen Nachführung)
- ✓ Antwortzeit
- ✓ Speicheraufwand (Höchstzahl von Zellen, Vorberechnungsgrad, Verwaltung dünn besetzter Matrizen)

Preis

- ✓ Software
- ✓ Schulung
- ✓ Beratung
- ✓ Wartung

OLAP-Produkte und Hersteller

- Brio Enterprise von [Brio Technology](#)
- Business Objects [Business Objects](#)
- Business Information Warehouse von [SAP](#)
- Decision Support Suite von [Pilot](#)
- DSS Agent von [Micro Strategy](#)
- Express von [Oracle](#)
- Holos von [Seagate](#)
- Hyperion Essbase von [Hyperion Software](#)
- Hyperion Enterprise von [Hyperion Software](#)
(spezialisiert für Finanzanwendungen)
- Seagate Info von [Seagate](#)
- InfoBeacon von [Platinum](#)
- MetaCube von [Informix](#)
- PowerPlay von [Cognos](#)
- Produktpalette von [SAS](#)

5.31 Methode im Vergleich

Kriterium	AHP	Optimierung	OLAP	Regelbasierte Systeme	Induktion	Neuronale Netze	Regression
Methode breit anwendbar	Ø	–	+	Ø	Ø	Ø	–
Automatisierungsgrad	–	+	1	Ø	+	+	+
Ergebnis genau	–	+	+	Ø	+	+	+
Unabhäng. Var. gewichtbar	–	–	–	–	Ø	–	+
Lösungsweg begründbar	Ø	–	Ø	+	+	–	–
Methode plausibel	+	Ø	+	+	Ø	–	Ø
Ergebnis einbettbar	–	+	Ø	Ø	+	Ø	+
Entwicklungsaufwand	+	+	–	–	+	Ø	Ø
Rechnerbelastung	+	+	–	Ø	Ø	–	+

Kriterien

1 Hauptnachteil

Modellierung relationaler Data Warehouses

Grundlagen 2

- ✓ Operative und analytische Datenbanken 9
- ✓ Enterprise Data Warehouse und Data Marts 24
- ✓ Mehrdimensionale Daten 38

Endbenutzerzugriff 50

- ✓ OLAP 61
- ✓ ANLAGEBERATUNG mit *PowerPlay*   87

Modellierung relationaler Data Warehouses

- ⇒ Datenmodellierung 101
- ⇒ LIEFERFRIST  121
- ⇒ VERKAUF  122
- ⇒ EINZELHANDEL mit *Synchrony*   148

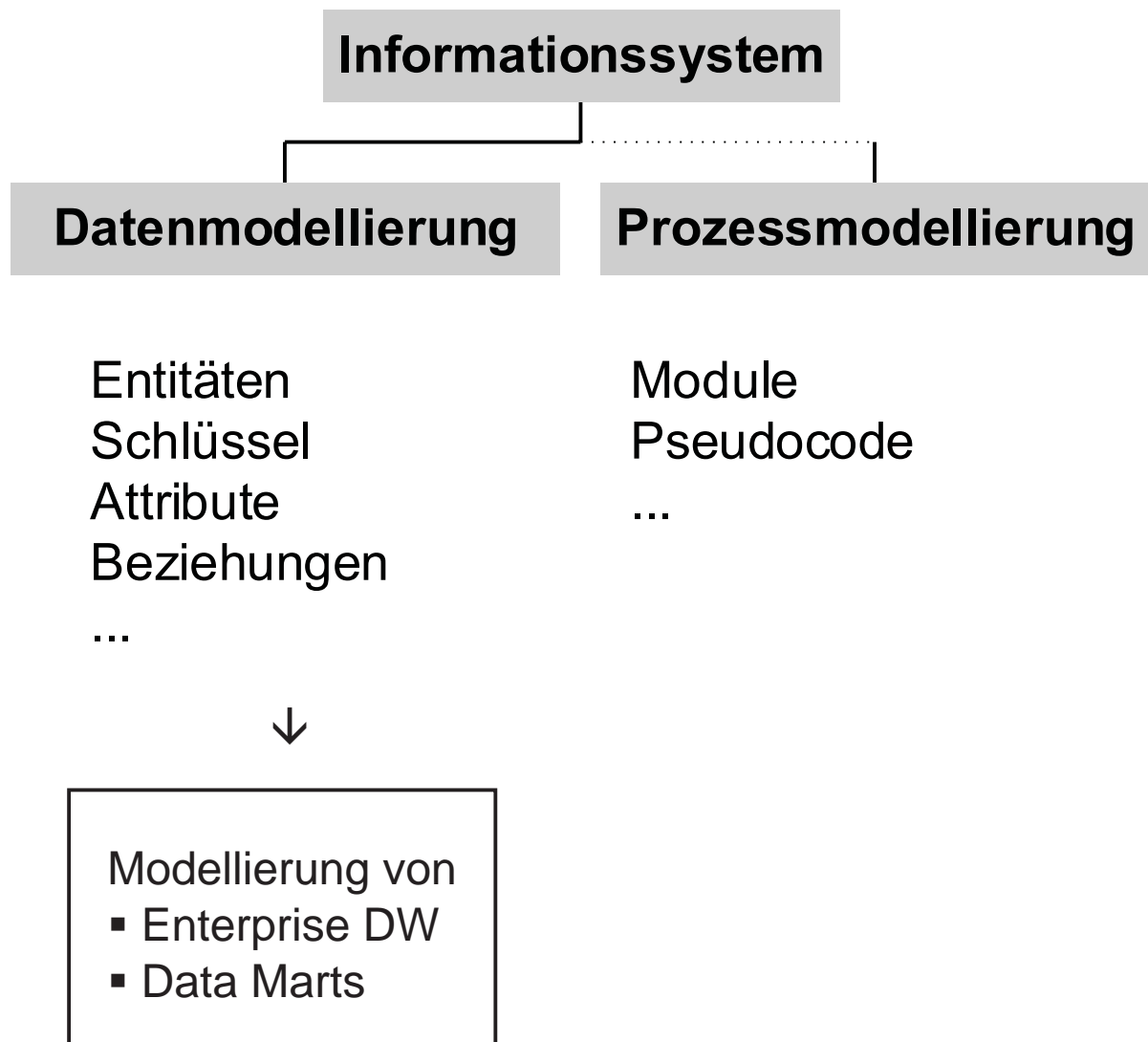
Entwicklung und Betrieb 152

- ⇒ Entwicklung und Betrieb 152
- Effizienz 176

Netzzugriff 214

- Client/Server-Architekturen 219
- Zugriff über Internet und Intranet 229

Modellierung von Informationssystemen



Operative und analytische Datenmodelle¹

Operatives Unternehmungsdatenmodell²

logische Verbindung grober Bereichsdatenmodelle



Operative Bereichsdatenmodelle

detaillierte logische ›Entity-Relationship-Diagramme



Enterprise Data Warehouse-Modell

globale Datenmodelle mit einigen ›physischen Festlegungen



Data Mart - Sternschemata und -Würfel

lokale Datenmodelle mit vielen physischen Festlegungen
(v.a. stark ›denormalisiert)

¹ Diagramme verändert nach Silverston et al. 1997

² Die Erstellung von Unternehmungsdatenmodellen ist aufwendig. Viele Unternehmungen verzichten deshalb auf ein übergreifendes globales Datenmodell

Unwünschte und wünschbare Redundanz

Produktionsdatenbanken

Einfüge-, Lösch- und Fortschreibungsanomalien minimieren



Redundanz *minimieren*

Data Warehouses

Benutzerfreundlichkeit und Abfrageeffizienz maximieren



Redundanz *introduzieren*

Redundanzquellen?

Data Warehouses führen zu Redundanz

<i>Redundanzquellen</i>	<i>Gründe der Einführung von Redundanz</i>
Analytische Daten entstehen aus operativen Daten	benutzernahe Datenmodellierung Zugriffseffizienz
Lokale analytische Daten entstehen aus zentralen analytischen Daten	gemeinsame Quelle für Data Marts benutzernahe Datenmodellierung Zugriffseffizienz
Denormalisierte entstehen aus normalisierten Tabellen	benutzernahe Datenmodellierung Zugriffseffizienz
Abgeleitete Daten wiederholen Rohdaten	Zugriffseffizienz ((Bsp. ▶Aggregate) benutzernahe Datenmodellierung (Bsp. Zeitreihen)

Ziele der Datenmodellierung

Produktionsdatenbanken

Transaktionsverarbeitung durch hohe Granularität



Fortschreibungsfreundlichkeit durch Normalisierung



Speichereffizienz durch Normalisierung

Data Warehouses

Benutzerfreundlichkeit durch geringere Granularität

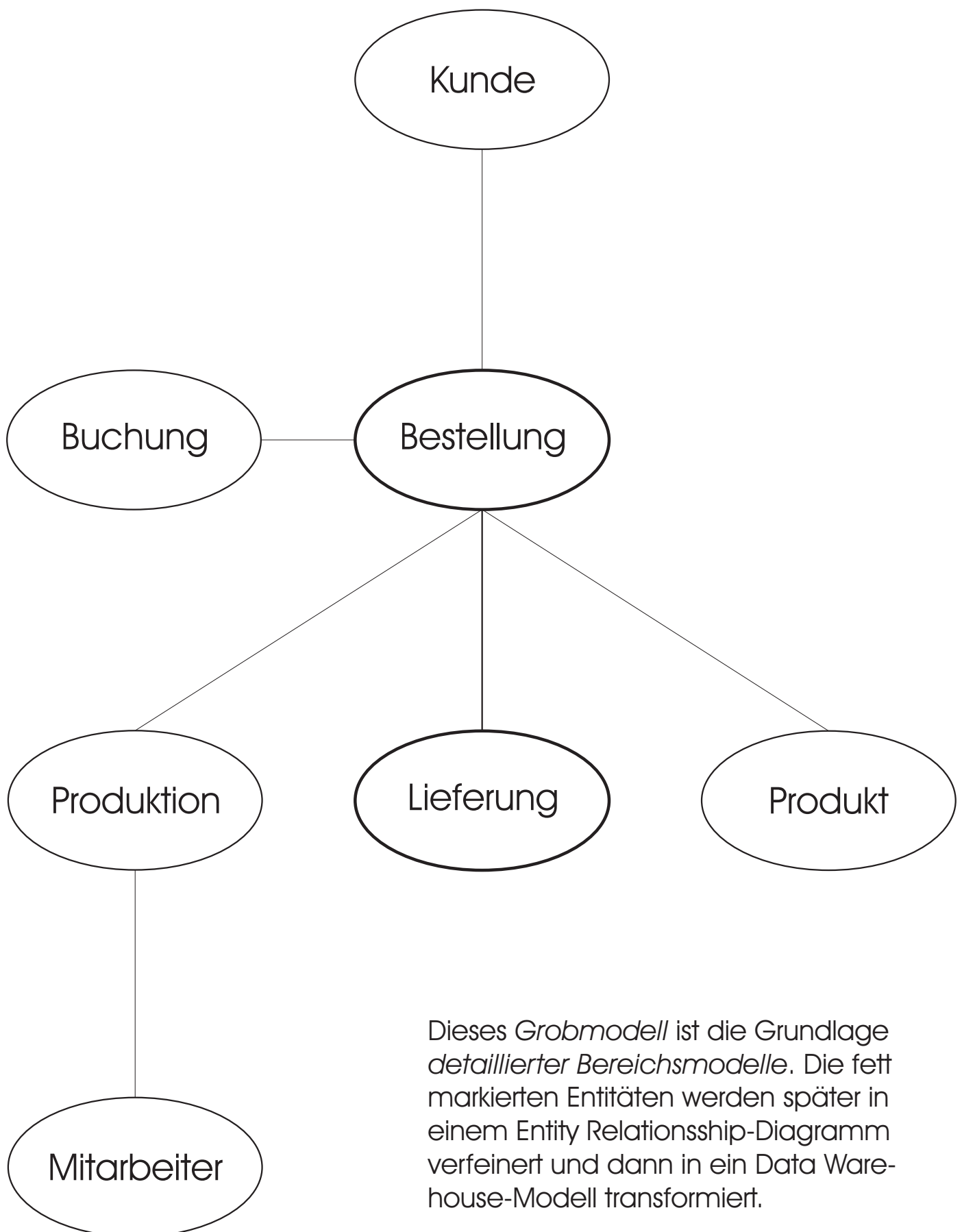


Abfrageeffizienz durch Denormalisierung

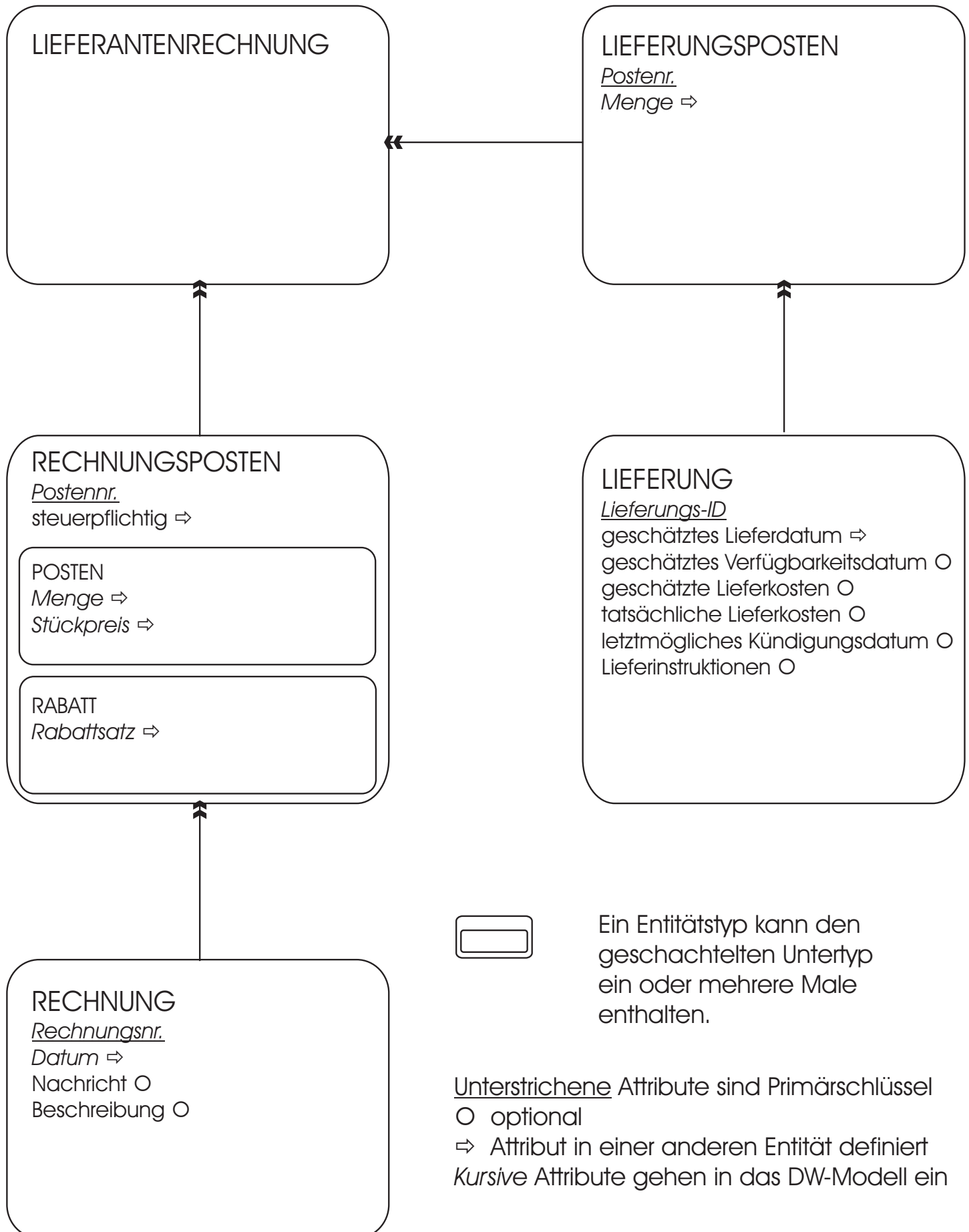


Betriebsnähe durch Indikatoren und ihre Dimensionen

5.32 Ein operatives Unternehmungsdatenmodell



5.34 Ein operatives Bereichsmodell *Fakturierung*



Produktionsdatenbanken



Data Warehouses

① **Logische Modellierung**

Ziel sind **abbildungstreue** *logische* Schemata

‣ **Zentrales Data Warehouse**

Anforderungsspezifikation und logisches Schema

⇒ ‣ Entity-Relationship-Diagramme

Data Marts

Anforderungsspezifikation und logisches Schema

⇒ ‣ Sternschemata für relationale DW

⇒ ‣ Würfel für mehrdimensionale DW



② **Physische Modellierung**

Ziel sind **effiziente** *physische* Schemata

Transformationen des operativen Modells

Künftiger **Nutzen** eines Attributs
oder einer Beziehung

versus

Einbusse an **Effizienz und Wartbarkeit**
von Data Warehouses

① **Operative** Daten weglassen

Rechnungsstatus (z.B. offen, bezahlt, Zahl der Mahnungen)

② **Zeitdimension** integrieren

Ladedatum, Zeitkategorien

③ **Attribute** berechnen

einfache Vorberechnungen wie *Betrag = Menge * Stückpreis*

Verdichtungen (z.B. Aggregationshierarchien)

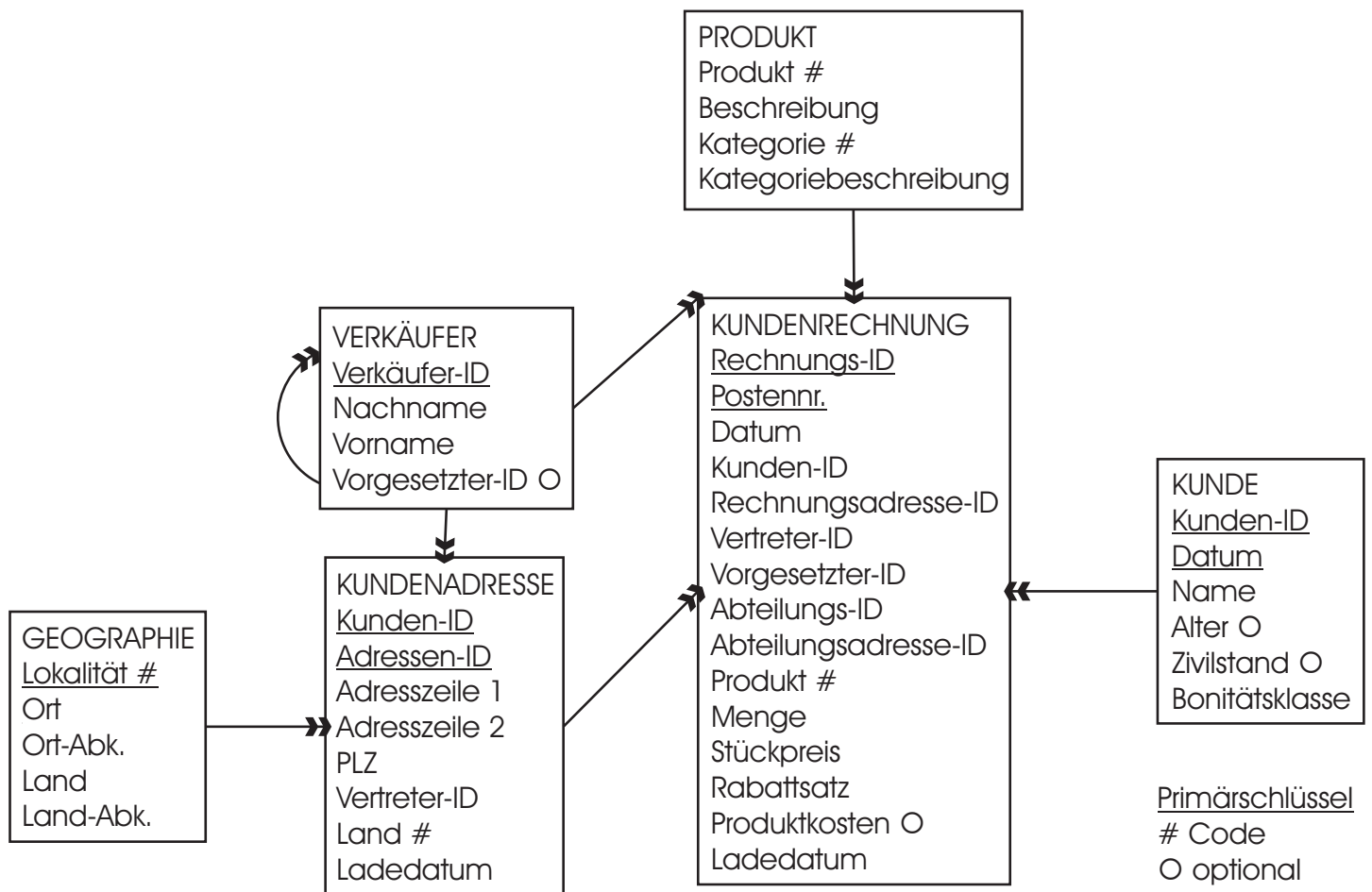
Umgruppierungen

④ **Verbundoperationen** verringern

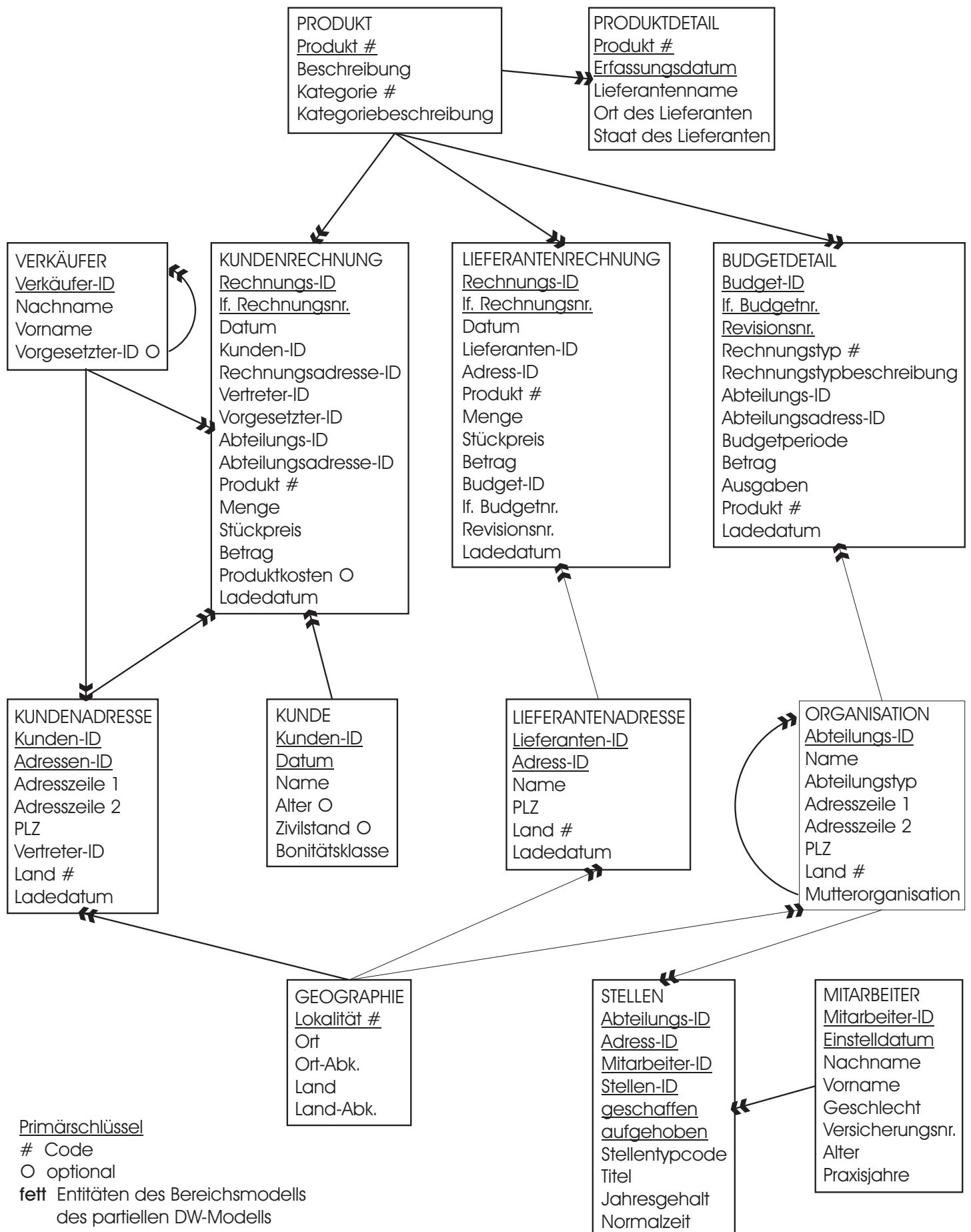
z.B. in LIEFERUNG und LIEFERUNGSPOSTEN

...

5.35 Operatives Bereichs- → partielles DW-Modell



5.36 Vom partiellen DW-Modell zum EDW-Modell



5.37 Logische Modellierung im Vergleich

	<i>Operative Daten</i>	<i>Analytische Daten</i>
<i>Ziel</i>	Redundanz-minimierung	Abbildungstreue und Abfragekomfort
<i>Logisches Datenmodell</i>	▸ normalisiert	▸ denormalisiert und mehrdimensional
<i>Ergebnis</i>	viele änderungs-freundliche Tabellen mit je wenigen Attributen	grosse <u>Faktentabellen</u> aus meist stetigen numerischen Attributen kleine <u>Dimensionstabellen</u> aus meist symbolischen Attributen

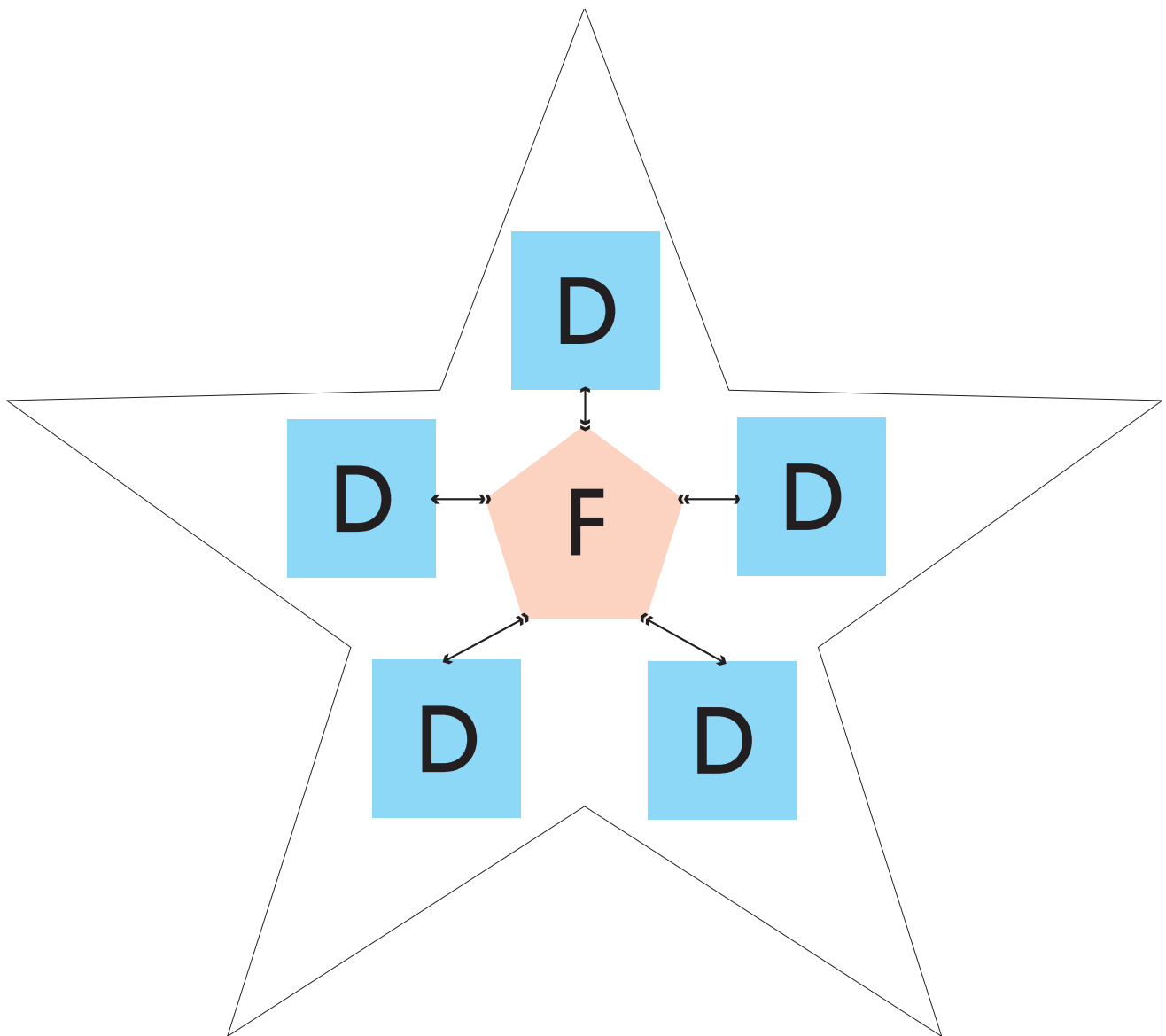


- **Logische** Datenmodelle ermöglichen ...
 - **abbildungstreue** Tabellenstrukturen
 - benutzerfreundliche **Abfragen**



Würfel für mehrdimensionale Data Marts
Sternschemata für *relationale* Marts →

5.38 Sternschemata für relationalen Data Mart



Sternschema := logisches Datenbankschema, das die **Dimensionstabellen** (D) eines *relationalen* DW abfragefreundlich um eine **Faktentabelle** (F) ordnet

Dutzende von Sternschemata für ein einzige EDW

Tabelleninhalte



Verfügbare Daten

(aus operativen Modellen und Interviews mit Datenbankadministratoren)

Beispiele	Attribute
POS-Tagesabschlüsse und andere operative Daten	Mengenumsatz Geldumsatz Kundenzahl



Wünschbare Auswertungen

(aus Interviews mit Endbenutzern)

Beispiele	Fakt pro Dimension
Zeitvergleich	VERKÄUFE pro PERIODE
Produktvergleich	VERKÄUFE pro PRODUKT
Lieferantenvergleich	VERKÄUFE pro LIEFERANT
Marktkorbanalyse	VERKÄUFE pro KUNDE

5.40/5.41 Anforderungsdiagramm

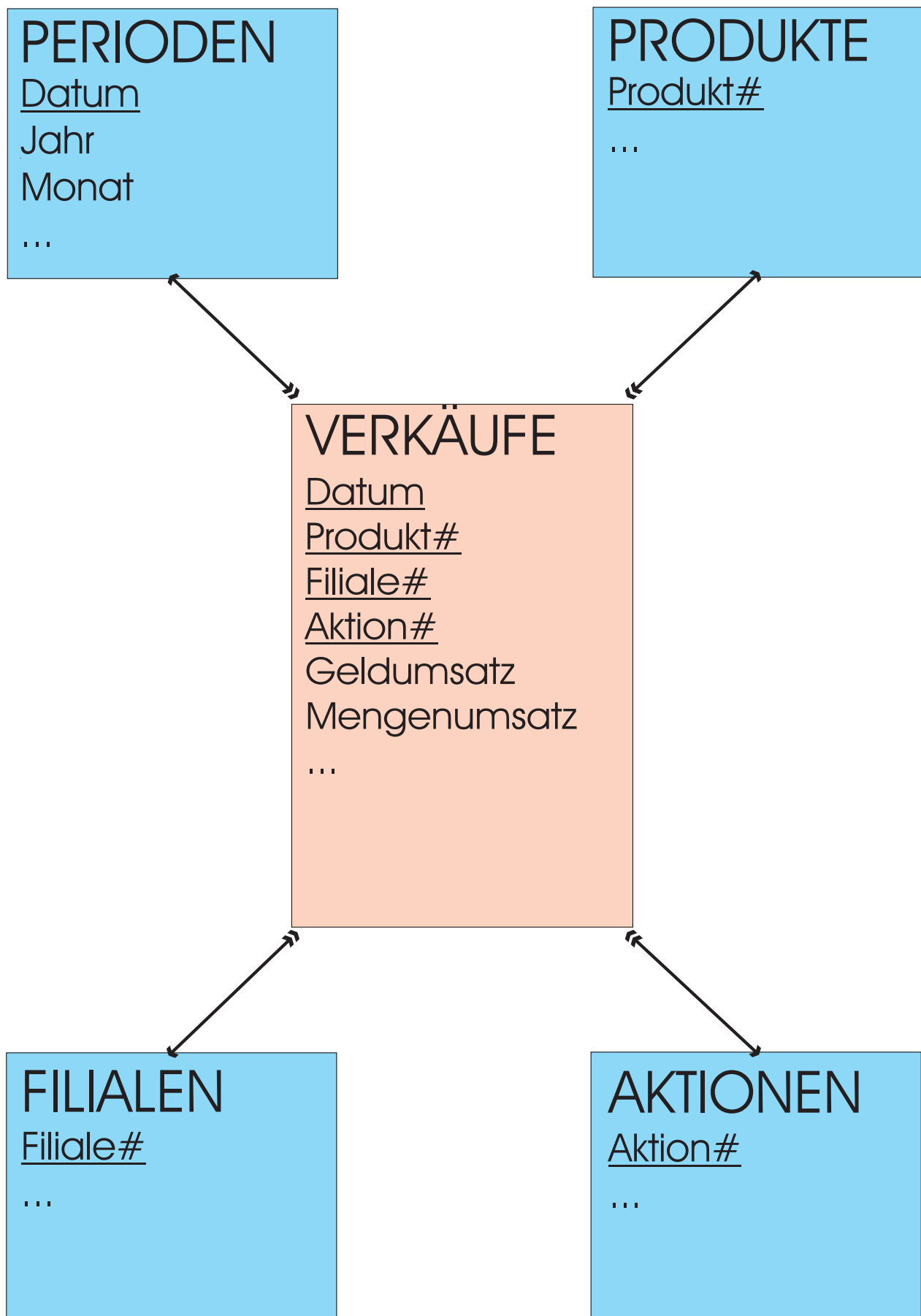
Beispiel

Mengenumsatz, ...			
Produkt	Periode	Ort	...
Produktklasse	Jahr	Land	
Produktgruppe	Quartal	Region	
Einzelprodukt	Monat	Filiale	
...	

Erläuterung

<i>Komponente</i>	<i>Beispielfragen</i>	<i>Definition</i>
① Indikatoren	Wie gut?	Attribute, die Unternehmungs- ergebnisse bewerten
② Dimension	Was?, Wann?, Wo?	Attribut, entlang welchem ein Fakt gemessen werden kann
③ Kategorie	Wie genau ?	Wertebereiche der Dimension

5.42 Ein Sternschema für HANDEL



5.43 Mehrere Faktzeilen pro Dimensionszeile

Mehrere VERKÄUFEzeilen

<u>Datum</u>	<u>Produkt#</u>	<u>Filial#</u>	<u>Aktion#</u>	<u>Mengenumsatz/Tag</u>	...
2.3.98	231	14	3	1'863	...
3.3.98	231	14	4	533	...
...



Eine PRODUKTE-Zeile

<u>Produkt#</u>	<u>Produktkategorie</u>	<u>Produktname</u>	<u>Farbe</u>	...
231	Mountain Bike	"Colorado"	Farbe grün	...
...

Produkt #231 kommt in *mehreren* VERKÄUFEzeilen vor

① VERKÄUFE

- Schlüssel aus vier Fremdschlüsseln vermittelt 1:n-Beziehungen zu den Dimensionstabellen

② ZEIT (bzw. PERIODE)

- Dimensionstabelle kann berechenbare und nicht berechenbare Attribute enthalten

Werktag, Feiertag, Saisonzeit, Steuerperiode

- Dimension enthält oft mehrere Kategorien

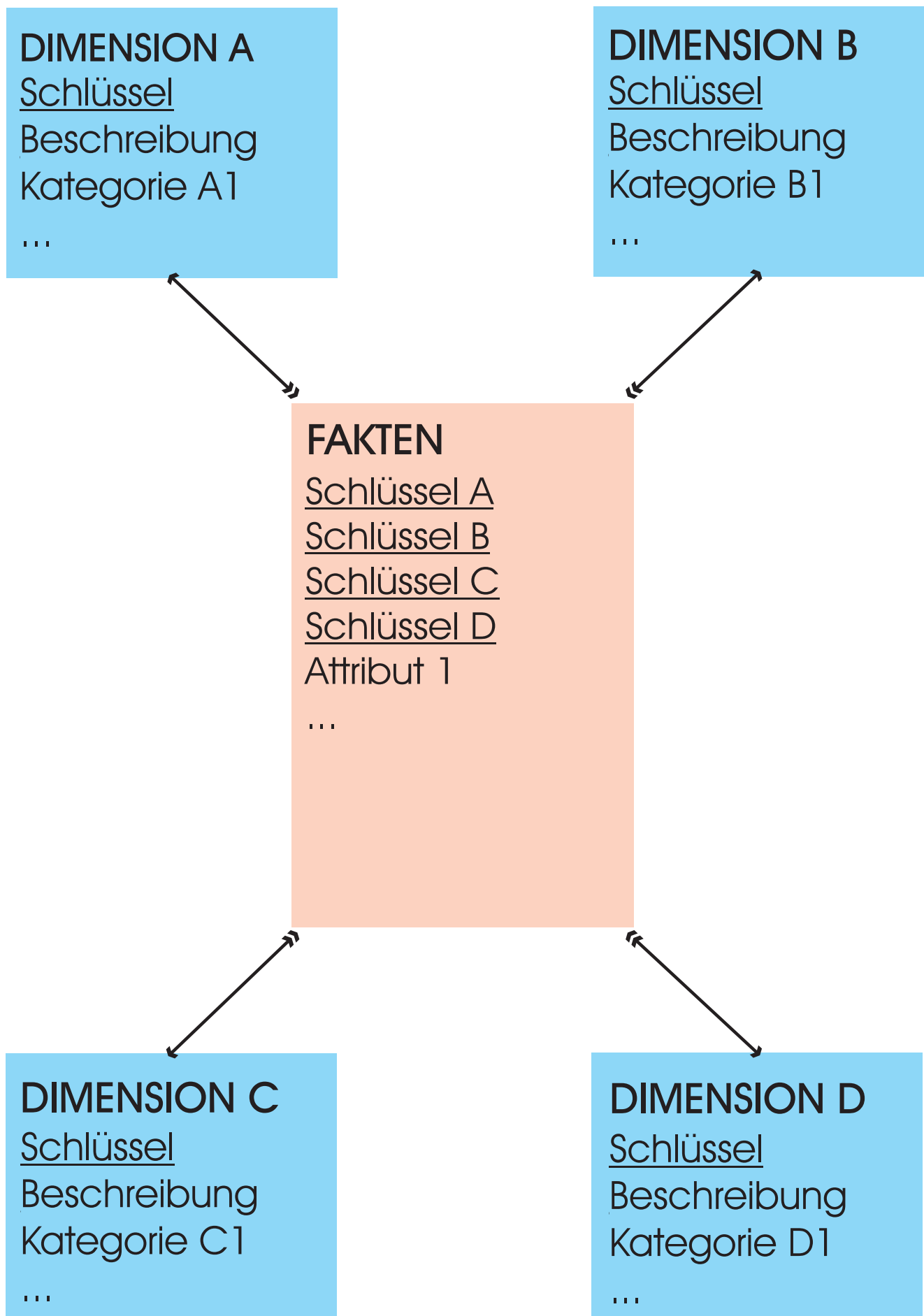
Jahr, Quartal, Monat, Woche und Tag

③ PRODUKTE

④ FILIALEN

⑤ WerbeAKTIONEN

5.44 Sternschema allgemein



5.45 Sternschemata filtern Abfragen ...

1. *Dimensionstabelle* PRODUKTE

Produkt#	Name	Gewicht	...
1	A	5	...
2	B	4	...
3	C	7	...
...



2. *Einschränkung durch Selektion (Filterung)* (Produkt = 2 oder 3) und (Gewicht = 4)



3. *Eingeschränkte Dimensionstabelle* PRODUKTE

Produkt#	Name	Gewicht	...
2	B	4	...



4. *Abfrage auf einer Faktentabelle*

LIEFERFRIST - Sternschema (A 5.3)

Die *Kundenzufriedenheit* hängt nicht zuletzt von der Einhaltung der *Lieferfristen* ab. Ihre Überwachung gehört deshalb zu den wichtigsten Aufgaben eines Data Warehouse. Aus der sorgfältigen Diagnose nicht eingehaltener Fristen kann sich zum Beispiel die folgende Therapie ergeben :

- Erhöhung der Lagerbestände
- Änderung der Transportmittel und -wege
- zuverlässigere Berechnung der Lieferfristen.

Eine Unternehmung mit mehreren Produktionsstandorten möchte die *Diagnose von Lieferverzögerungen* in ihre Data Warehouse-Lösung integrieren.

- a) Erstellen Sie ein *Anforderungsdiagramm* mit mindestens fünf Dimensionen und mehreren importierten und berechneten Fakten. (Am besten versetzen Sie sich in die Unternehmungsleitung und fragen nach Gründen von Lieferverzögerungen)
- b) Zeichnen Sie ein *Sternschema*

VERKAUF - EDW-Modell und Sternschema (A 5.4)

Betrachten Sie das Entity-Relationship-Diagramm 5.47. Es enthält einen Ausschnitt aus dem Datenmodell eines Enterprise Data Warehouse.

1. Datenmodell des Enterprise Data Warehouse

- a) Woran erkennen Sie, dass der EDW-Ausschnitt von Bild 5.47 nicht aus dem Datenmodell einer Produktionsdatenbank stammt?
- b) Welche Schritte durchläuft die Transformation eines operativen Datenmodells in ein EDW-Datenmodell?
- c) Definieren Sie anhand des EDW-Datenmodells eine SQL-Abfrage, welche die Produkte und die jeweiligen Mengen auflistet, die Personen mit einem Wohnsitz in der Schweiz im Jahre 1998 gekauft haben.

2. Sternschema des Funktionsbereichs VERKAUF

- d) Erstellen Sie aus dem Datenmodell ein Sternschema. Der Verkaufsleiter will Zeit-, Produkt-, Kunden-, Verkäufer- und Regionalvergleiche erstellen. Ausserdem möchte er wissen, ob das Kreditlimit, das Alter des Kunden oder der Zivilstand den Absatz beeinflussen.
 - Tragen Sie die Fakten, Dimensionen und Kategorien in ein Anforderungsdiagramm.
 - Ordnen Sie die Dimensionstabellen um die Faktentabelle an und definieren Sie deren Beziehungen.
- e) Definieren Sie die SQL-Abfrage aus Aufgabe 1 c) auf dem Sternschema. Beschreiben Sie die Unterschiede zwischen den beiden Abfragen. Weshalb wäre die gleiche Abfrage mit einem OLAP-Werkzeug wie PowerPlay benutzerfreundlicher?

3. Anwendungsbereich “Strategische Planung”

Eine Analyse der Absatzmengen eines Produkts nach der Dimension Zeit stellt Daten zur Lückenplanung (engl. gap analysis) bereit.

- f) Genügen die in unserem Sternschema vorhandenen Daten auch für weitere strategische Planungsinstrumente? Wenn nicht, welche Daten fehlen?
- g) Beschreiben Sie die Dimensionen und Fakten zur Erstellung eines Boston Consulting Group-Portfolio.

Relationales OLAP erfordert Sternschemata

The screenshot shows the StarTracker Demo application interface. At the top, there is a menu bar with options: File, Edit, Aggregates, Sequences, Comparisons, and Help. Below the menu bar is a toolbar with a 'Run Report' button. The main area displays a star schema diagram with four fact tables connected to a central dimension table. The fact tables are: PERIODEN: Alle Perioden, PRODUKTE: Alle Produkte, FILIALEN: Alle Filialen, and AKTIONEN: Alle Aktionen. The central dimension table is 'Verkäufe'. The 'Verkäufe' table lists fields: Periode#, Produkt#, Aktion#, Filiale#, Geldumsatz, Mengenumsatz, Kosten, Kundenzahl, Durchschnittlich verkaufte Stückzahl*, Durchschnittliche Auftragsgrösse*, Durchschnittskosten*, Durchschnittspreis*, Gewinn*, Kostenquote*, and Report Columns*. The 'PERIODEN' table lists fields: Periode#, Datum, Tag, Tagesnr_im_Mt, Tagesnr_total, and Wochenim Jahr. The 'PRODUKTE' table lists fields: Produkt#, Name, Gewicht, Lieferant, Unterkategorie, and Kategorie. The 'FILIALEN' table lists fields: Filiale#, Strasse, Ort, Kanton, PLZ, and Geschäftsgebiet. The 'AKTIONEN' table lists fields: Aktion#, Name, Preisreduktion, Anzeige, Display, and Anzeigenmedium. Below the star schema diagram is a data table with columns A through N and rows 1 through 17. The table is currently empty. At the bottom of the application window, there is a status bar showing 'Family: Handelsunternehmen', the date '16.10.97', and the time '13:58'.

Beispielwerkzeug *If StarTracker* bzw. *If Synchrony*

Family Panel

Sternschema für eine *Familie* von Abfragen (ein Sternschema)

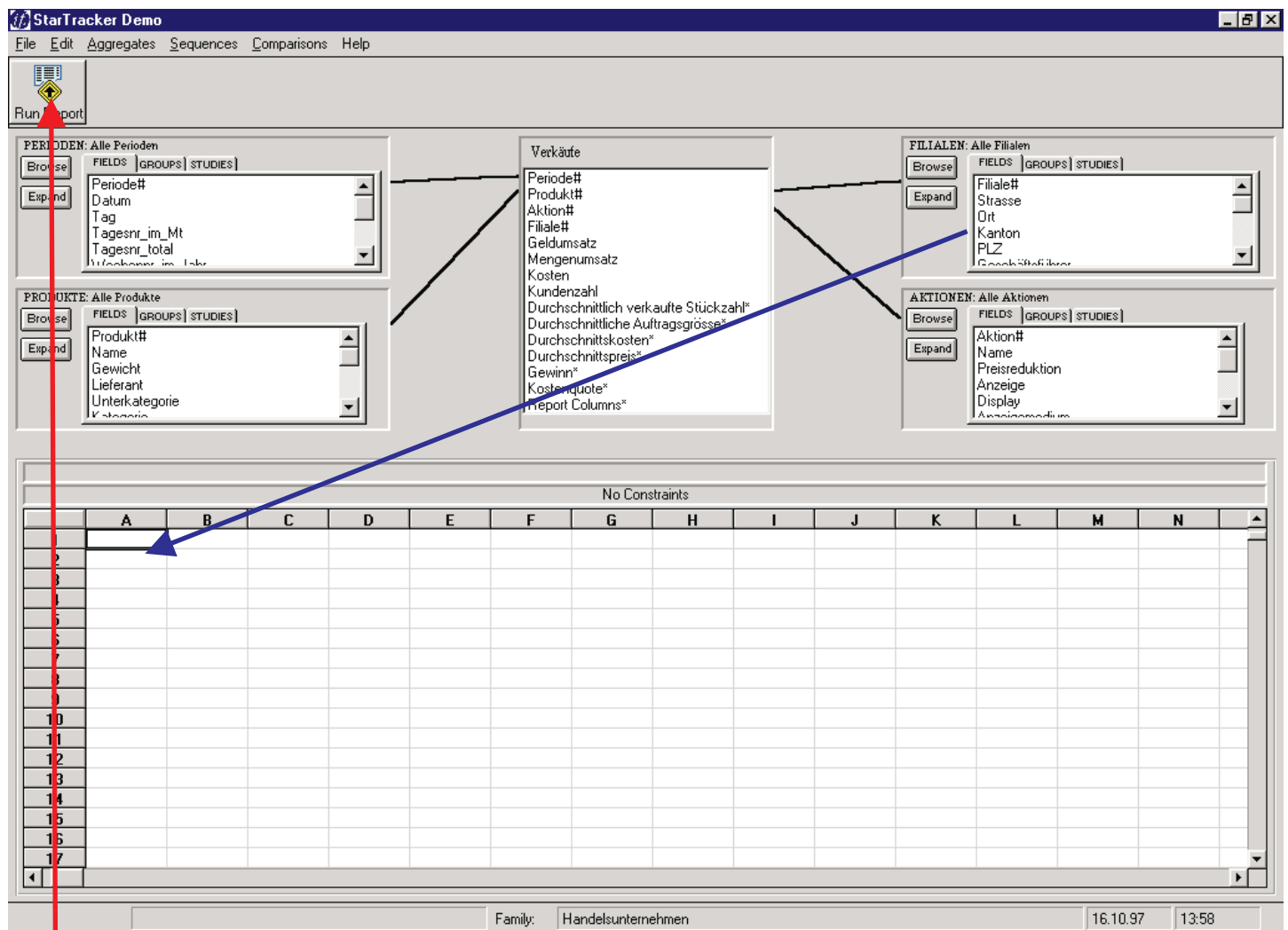
Report Panel

Ergebnisse einer *Abfrage* (hier engl. report)

 *if... Synchrony*

◀18 ▶19

OLAP-Abfragen auf einem Sternschema



- 1 **Attribute** vom *Family Panel* in den *Report Panel* **ziehen**
- 2 Knopf "Run Report" **klicken**

Eine SQL-Abfrage auf einem Sternschema

```
SELECT      PRODUKT.Name,  
            SUM ( VERKÄUFE.Geldumsatz )  
FROM        VERKÄUFE, PRODUKTE, PERIODEN  
WHERE       PERIODEN.Monat = VERKÄUFE.Monat  
GROUP BY    PRODUKT.Name  
ORDER BY    PRODUKT.Name1
```



```
SELECT      Fakt- oder Dimensionsattribut  
FROM        Fakt- oder Dimensionstabellen  
WHERE       Bedingung  
GROUP BY    Fakt- oder Dimensionsattribut  
ORDER BY    Fakt- oder Dimensionsattribut
```

SQL ist für den Endbenutzer *schlecht* geeignet



Benutzerfreundlichere Abfragen durch
Sternschemata und SQL-Generierung

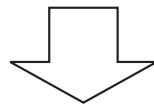
¹ Beschreiben Sie die SQL-Abfrage umgangssprachlich

5.48 Sternschemata durch Denormalisierung



Denormalisierung von Produktionsdaten führt zu

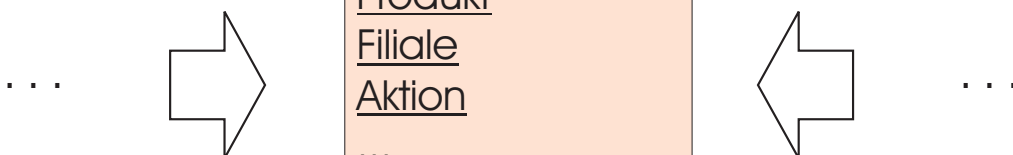
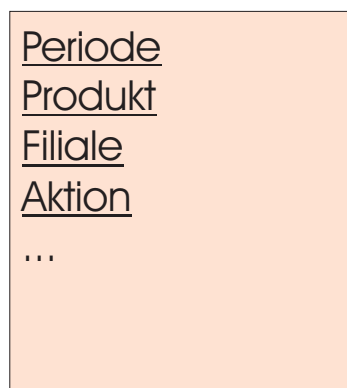
- mehr *Redundanz*, aber
- weniger Verbundoperationen



Dimension FILIALE



Fakten VERKÄUFE



Stern- versus normalisiertes Schema I

Normalisiertes Schema

Ziele

- Redundanzminimierung
- Effiziente Transaktionsverarbeitung
(Zugriff auf den einzelnen ›Satz)

Ergebnis

- Komplexes, globales und idiosynkratisches Schema
- Viele Entitäten und Beziehungen
(Hunderte für eine grössere Unternehmung)

Sternschema

Ziel

- Benutzerfreundliche Abfrage
(Aggregatzugriff, verständlicheres Datenmodell)

Ergebnis

- Einfaches, lokales und standardisiertes Datenmodell
(10-25 Sternschemata für grössere Unternehmungen)
- 1 Faktentabelle und wenige Dimensionstabellen
(i.d.R. 5-15)

Stern- versus normalisierte Schemata II

Sternschemata sind meist ...

😊 abfragefreundlicher

- Weniger Tabellen \Rightarrow einfachere *Abfragen*
- Fakten und Dimensionen \Rightarrow Betriebsnähe

😊 effizienter

- Weniger Tabellen \Rightarrow weniger *Verbunde*

aber ...

😞 unvollständig

bilden zum Beispiel nicht alle Beziehungen ab

😞 wartungsaufwendig

insbesondere änderungsfeindlich

Vom normalisierten zum Sternschema

- ① Ein Sternschema für jeden **Geschäftsprozess**
- ② Fakttabellen aus den **aggregierbaren** Attributen mit **m:n**-Beziehungen
- ③ Übrige Attribute in Dimensionstabellen denormalisieren und **direkt** mit der Faktentabelle verbinden

Erweiterte Sternschemata

- ① Mehrere **Fak**entabellen, insb.
m:n-Verbindungstabellen
und / oder
- ② **Normalisierte** Unterdimensionen
und / oder
- ③ **Kategorientabellen**

Erweiterte Sternschemata heissen auch **Schneeflockenschema**

Beurteilung erweiterter Sternschemata

Vorteile

- 😊 Sonst nicht darstellbare Attribute und Beziehungen
- 😊 Update (weniger Redundanz)
- 😊 Speicherplatz

Nachteile

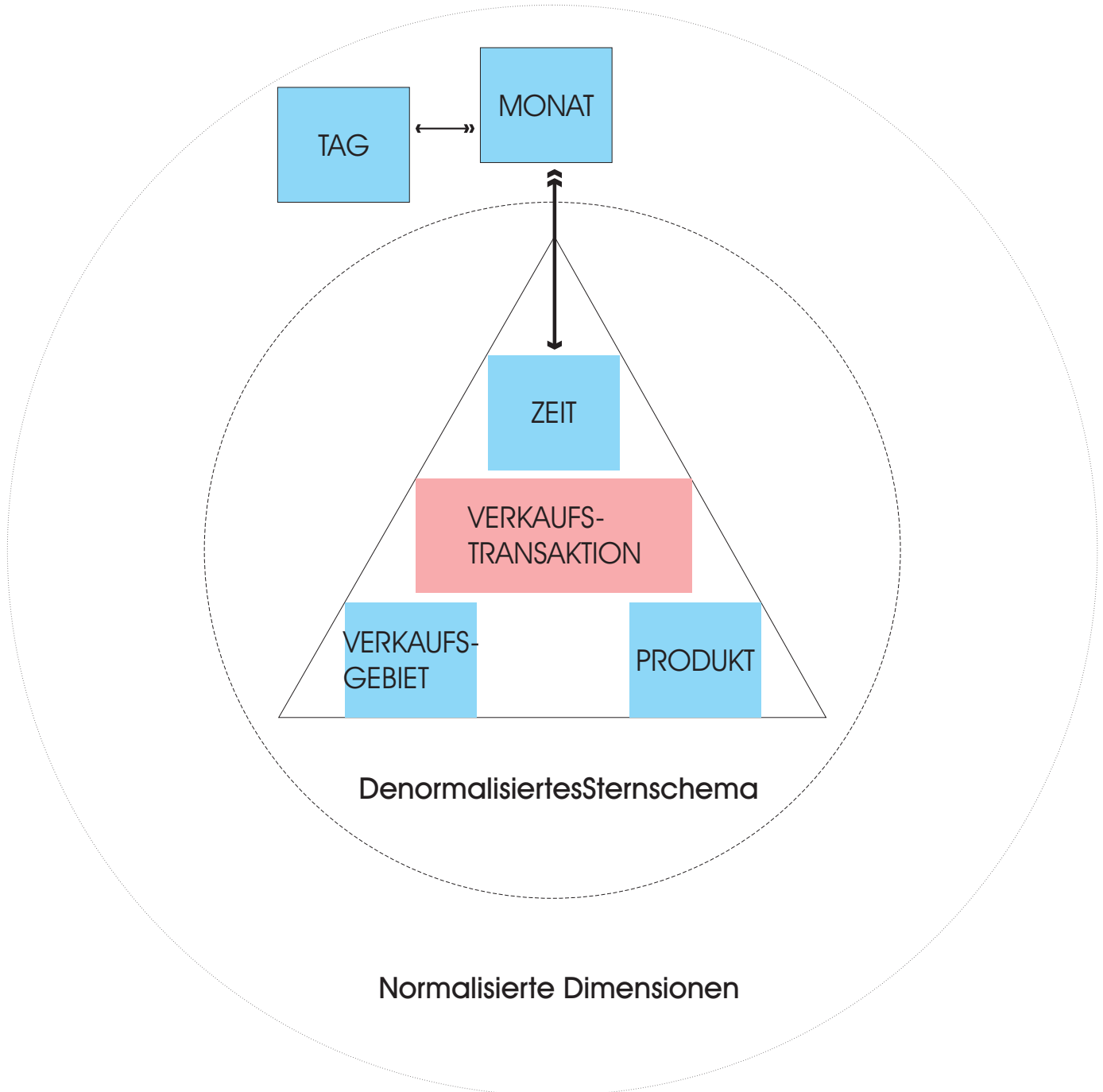
- 😞 Benutzerfreundlichkeit
- 😞 Abfrageeffizienz



Sparsam einsetzen!

Bsp. Unterdimension *Zeit* →

5.49 Erweitertes Sternschema - Normalisierung



Denormalisierte ZEIT

ZEIT (Jahreszahl, Monatszahl, Tageszahl, Monatsname, Tagesname, Werktag?)

Normalisierte ZEIT

ZEIT (Jahreszahl, Monatszahl, Tageszahl, Werktag?)

MONAT (Monatszahl, Monatsname)

TAG (Tageszahl, Tagesname)

Erweitertes Sternschema - m:n-Beziehung (A 5.6)

Bild 5.51 bildet die Rechnungsstellung eines Spitals auf ein Sternschema ab. Es sieht pro Patient nur eine einzige Diagnose vor, weil der zusammengesetzte Schlüssel von RECHNUNG nur den Fremdschlüssel #Diagnose enthält. Schlagen Sie ein Schema vor, dass mehrere Diagnosen (sogenannte Diagnosegruppen) erlaubt.

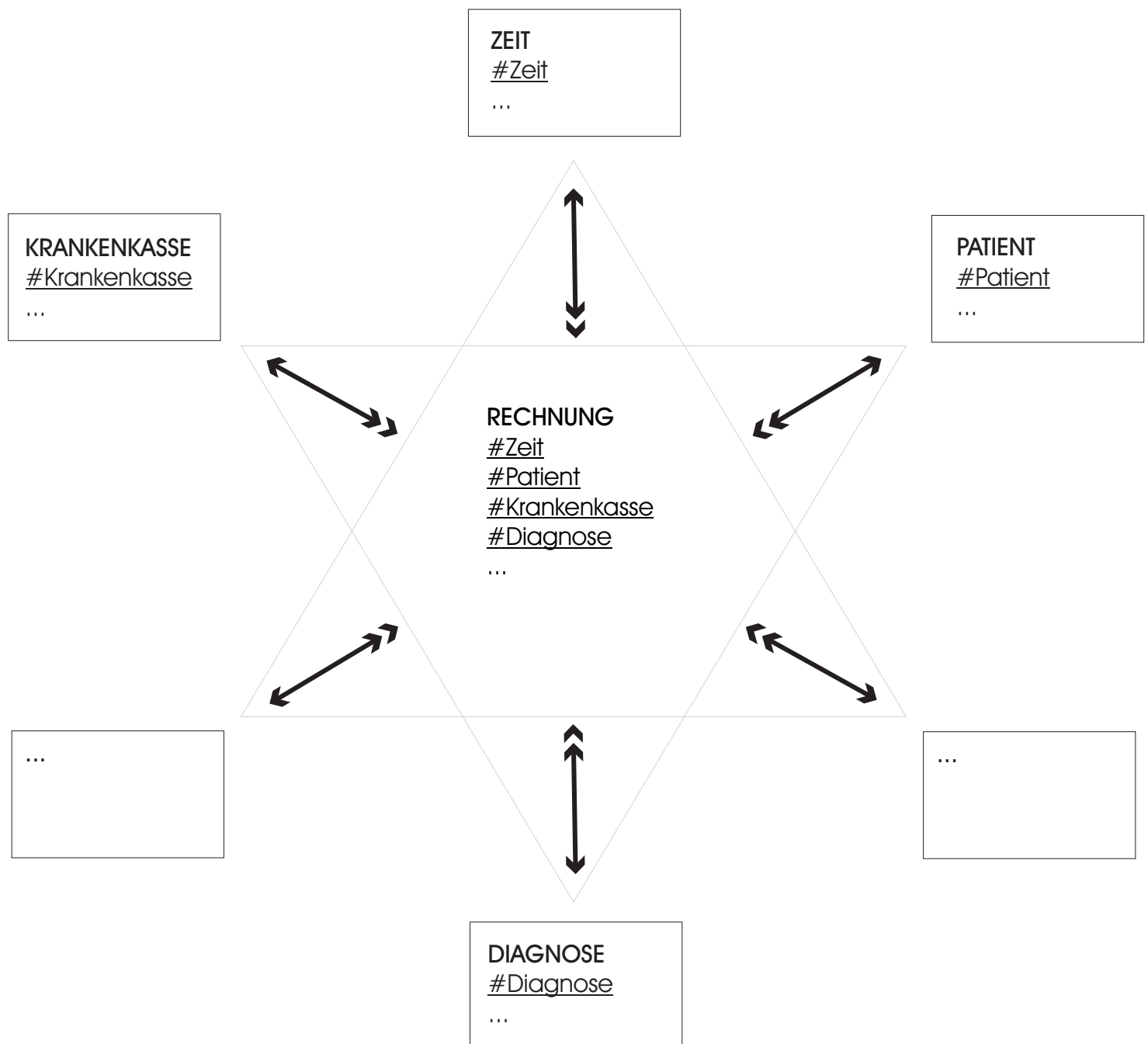


Bild 5.51: Sternschema einer Fakturierung in einem Spital

KONTO - Auflösung einer m:n-Beziehung (A 5.7)

Gegeben seien eine Faktentabelle KONTENSTAND und eine Dimensionstabelle KUNDE. Der Kontenstand wird monatlich erfasst.

- a) Begründen Sie, weshalb zwischen den beiden Tabellen in der Regel eine m:n-Beziehung besteht.
- b) Zeichnen Sie den Ausschnitt des Sternschemas, welcher die beiden Tabellen und ihre Beziehung abbildet.

Metadaten

Metadaten :=

- ✓ Daten
- ✓ über
- ✓ die “eentlichen” Daten (Objektdaten) und
- ✓ die “eentlichen” Prozesse (Objektprozesse)
- ✓ der Informationssysteme einer Unternehmung

Wie kann ich ▶ Metadaten über *verschiedenste* Objektdaten und -prozesse in einer *einheitlichen* Notation und *einfach* zugreifbar verwalten?

<i>Metadatenklasse</i>	<i>Beispiele</i>	<i>Anwendungen</i>
Bezeichner	Tabellen, Attribute, Abfragen	▸ logische Datenmodellierung
Datentypen	▸ ASCII/▸ EBCDIC, Genauigkeit numerischer Typen	Ladeprozess
Beziehungen	1:1, 1:n, m:n	▸ logische Datenmodellierung
Volumen	Datenmenge, Fortschreibungshäufigkeit, ▸ Partitions Grenzen	▸ physische Datenmodellierung
Transformationen	Aggregationsregeln	Ladeprozess
Quelle und Ziel	Quellen operativer Daten	Ladeprozess
Verantwortliche	Datenbankadministrator eines Data Mart	Entwicklung und Betrieb
Benutzersichten und Zugriffsberechtigungen	Zugriff auf die Personaldaten	Administration



Vorkommen von Metadaten

Interaktive benutzerfreundliche Abfragen



Lokale
▸ Datenmodelle

Unternehmens-
Datenmodelle



Einzelrechner-Daten



Client-Daten



Server-Daten



Programmdaten



▸ CASE-Daten



...

Data Dictionary

Data Dictionary (DD) :=

- ✓ Datenbank, die
- ✓ Metadaten
- ✓ zu Form und Inhalt von
- ✓ Applikationen enthält

Synonyme: Katalog, Repository

Data Dictionary-System

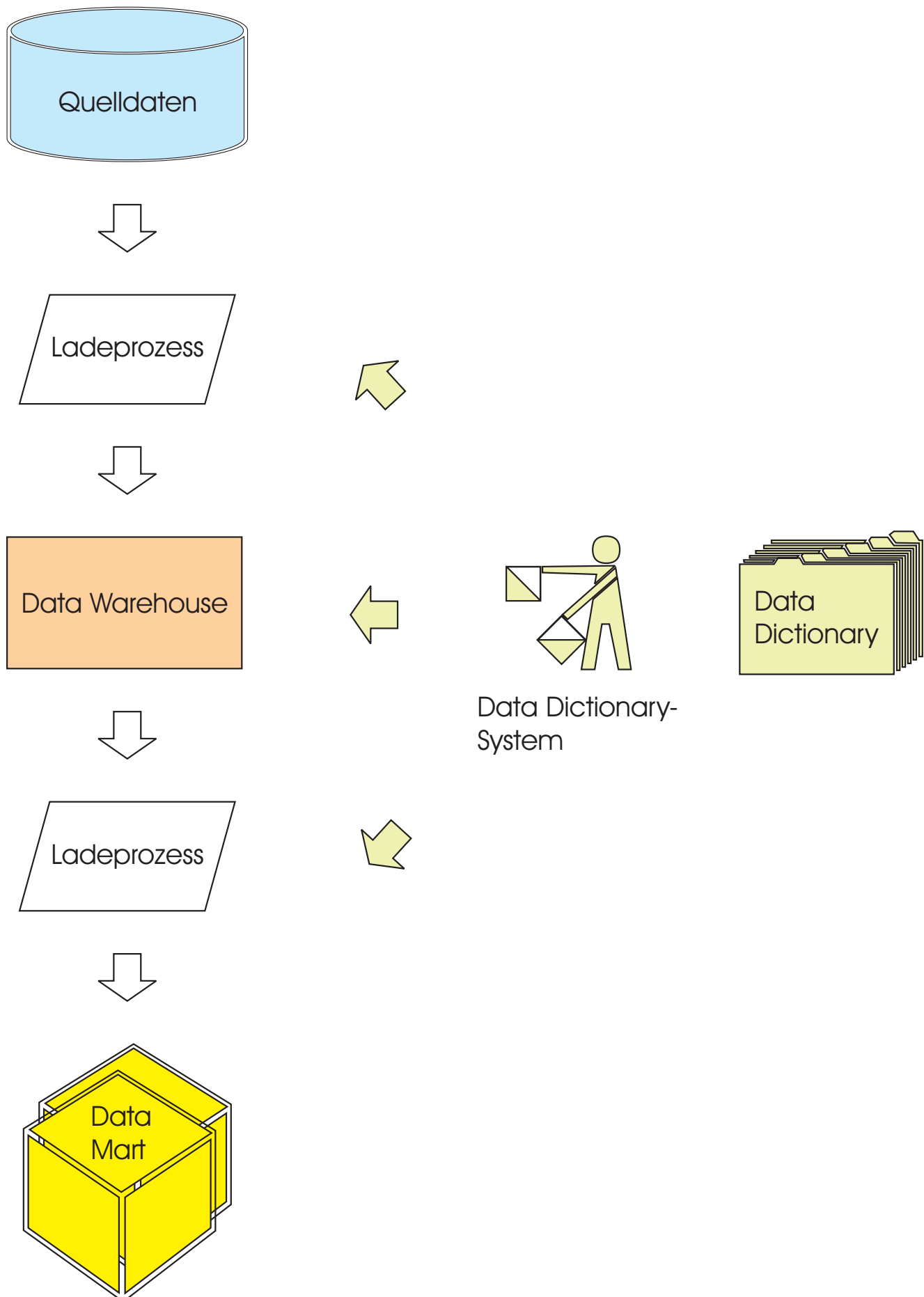
Data Dictionary-System :=

- ✓ Informationssystem, das
- ✓ Metadaten
- ✓ in einem Data Dictionary
- ✓ so verwaltet,
- ✓ dass es Softwareentwicklung und -wartung möglichst gut koordiniert

Bsp. Ladeprozesse zwischen ...

- operativen Datenbanken und Data Warehouse
- Enterprise Data Warehouse und Data Marts

5.53 Data Dictionary-System



Anforderungen an DD-Systeme

Klassen von Data Dictionary-Systemen

- *Allgemeiner* DD (für mehrere Werkzeuge und Hersteller)
- *Dedizierter* DD (für ein Werkzeug oder einen Hersteller)
z.B. Metadaten des ROLAP-Werkzeugs *Synchrony*
- *Gateway* zur Verbindung lokaler DDs

Anforderungen

- ✓ Metamodell standardisiert
z.B. *Open Information Model* auf Initiative von Microsoft,
Metadata Interchange Initiative von führenden DW-Anbietern
- ✓ Metamodell erweiterbar (z.B. ▶Entity Relationship Model)
- ✓ Impact Analysis
Rückwirkungen von Änderungen?
- ✓ Versioning
gleichzeitige kontrollierte Arbeit an mehreren Versionen
- ✓ benutzerfreundliche Abfragen und Berichte
z.B. plattformunabhängig aus einem Webbrowser
- ✓ ▶API
- ✓ ...

Trends

- Integration externer (gekaufter) Daten
- neue Datentypen (Multimedia, Webseiten, Zeit, Raum, ...)

Beispiele von DD-Systemen

Prism Warehouse Directory

- Data Warehouse-Lösung
- Austausch mit anderen DD-Systemen
- Webzugriff
- “Prism Warehouse Manager” automatisiert Ladeprozesse
- “Prism Change Manager” automatisiert DW-Wartung

(2001: Prism Solutions wurde von Ardent und anschliessend von Ascential übernommen. Der ETL-Bereich von Ardent wurde seinerseits von Informix gekauft)

Platinum Repository von Platinum Technology

- unternehmensweite Lösung
- *Data Shopper* für den GUI-Zugriff

Synchrony ist . . .

❶ ein einfaches **ROLAP** - Frontend ...

Relationales OLAP-Frontend zu einem ODBC-RDBMS
wie *MS Access*

❷ für **Abfragen** ...

- ad hoc statt vordefiniert
- benutzerfreundliche Direktmanipulation statt SQL
- explorativ statt off line (v.a. Drilling Down and Up →)

❸ auf einem **Sternschema**

if...Synchrony setzt ein nach dem Sternschema
organisiertes relationales Data Warehouse voraus

Lernziele

⇒ Sternschema

⇒ Abfragen (hier Reports), insbesondere Drilling Down and Up

⇒ Aggregation

 [Grocery.mdb](#) (MS Access Database)

(*Synchrony* selbst starten Sie vom Desktop oder aus Explorer)

Architektur von *Synchrony*

Datenbank

⇒ relational

Bsp.  Grocery.mdb

Metadaten

⇒ Sternschema oder Schneeflockenschema

Bsp.  Metadaten zu Grocery.mdb

Sicherheitsdefinitionen

Bsp.  ...\Synchrony\ifstar.mdw

Abfragen

⇒ benutzerdefinierte Abfragen

a) vordefinierte Abfragen der Metadatenbank

Bsp.  Metadaten zu Grocery.mdb

b) exportierte Abfragen

Bsp.  ...\Synchrony\Reports*.qry

Wichtige Metadaten von *Synchrony*

ifDatamart

Namen der Data Marts

ifFamily

Name und Version des Data Mart-Sternschemas, Name und Beschreibung der Fakten und Dimensionen, ...

ifMeasure

Name, Beschreibung und Schlüssel der Fakttable, ...

ifDimension

Name, Beschreibung und Schlüssel der Dimensionstabellen, ...

ifAttribute

Name, Beschreibung und Datentyp (Datenformat) der Fakt- und Dimensionsattribute, Formeln der berechneten Attribute, ...

ifReport

Definition und Beschreibung der Abfragen, ...

ifAggregate

Struktur einer Aggregatsdefinition

...

 [Metadaten zu Grocery.mdb](#)

Funktionalität von *Synchrony*

Eingabe


- ✓ Relationale Daten in einem Sternschema darstellen
- ✓ ›ODBC-Verbindung zu einem ›RDBMS erstellen
- ✓ Benutzerfreundliche ›ROLAP-Abfragen definieren

Verarbeitung

- ✓ ROLAP-Abfragen in SQL übersetzen
- ✓ SQL-Abfragen über ODBC an ein relationales System senden

Ausgabe

- ✓ ODBC-Ergebnisse in Berichtsform präsentieren

Eine Einzelhandelskette besteht aus Hunderten von Filialen und verkauft Tausende von Produkten. Die relationale MS Access-Datenbank  [Grocery.mdb](#) enthält die Objektdaten. Sie stellen einen *Ausschnitt* der Verkaufsdaten ausgewählter Filialen und Produkte für zwei Jahre dar. Die folgende Dimensionen strukturieren diese Daten :

- Perioden (Tabelle TIME)
- Produkte (Tabelle PRODUCT)
- Filialen (Tabelle STORE)
- Aktionen (Tabelle .PROMOTION).

Das Sternschema ist definiert in  [Sample Meta.mdb](#). Die Sicherheitsdefinitionen befinden sich unter  ...\\Synchrony\\ifstar.mdw.

① ***Synchrony* kennen lernen**

Benutzeroberfläche

a) Starten Sie *Synchrony* vom Desktop oder Explorer aus.

Eine Einführung erhalten Sie unter *Help/Books Online* und Hilfe unter *Help/Synchrony Help*.

b) Wählen Sie den Menüpunkt *File/New* und klicken Sie doppelt auf das Sternschema “Grocery”. Achten Sie darauf, dass die Sternschema-Sicht eingeschaltet ist (*View/Star*).

Sternschema

c) Wozu dienen die obere und die untere Hälfte des Bildschirms? Welche Synonyme verwendet der Buchtext für die *Synchrony*-Termini “Family”, “Report”, “Measures” und “Fields”?

d) Machen Sie sich mit dem Sternschema “Grocery” vertraut : Wo sind die Fakten und Dimensionen? Welches sind Schlüsselattribute? Erklären Sie die Bedeutung des Symbols, das *Synchrony* vor Faktattribute stellt.

Einfache Abfragen

- e) Formulieren Sie die Abfrage “Welchen Geldumsatz erzielt jedes Produkt?” und beschreiben Sie das Ergebnis.

(Ziehen Sie die folgenden Attribute (engl. fields) in das Report-Fenster: “Full Description” der Dimensionstabelle PRODUCT und “Sum of Dollar Sales” der Faktentabelle SALES. Klicken Sie dann das Symbol ► (Run Report))

Filter

Sie löschen eine Ergebnisspalte, indem Sie sie markieren und die Taste Del drücken.

- f) Die Groups- und Studies-Tabs der Dimensionstabellen enthalten Filter. Beantworten Sie die folgende Filterabfrage: “Welchen Geldumsatz erzielten die Produkte der Marke Squeezable Inc.?”

(Ziehen Sie den Filter “Brand = Squeezable Inc.” in das Report-Fenster und klicken Sie auf ►).

- g) Wie hoch war der Geldumsatz der Marke “Squeezable Inc.” 1995?
- h) Wechseln Sie in der Dimensionstabelle PRODUCT in die Browse-Ansicht (Rechtsklick) und inspizieren Sie die Werte einzelner Attribute (Doppelklick auf ein Attribut).

Lassen Sie sich die ganze Tabelle PRODUCT anzeigen (*View/Table*). Finden Sie heraus, wie sich Filter definieren lassen.

- i) Erklären Sie den SQL-Code, der für die Abfrage “Wie gross ist der Gesamtumsatz?” generiert wird (*View/SQL*).
- j) Interpretieren Sie den SQL-Code einer komplexeren Abfrage.

Aggregate


- k) Aggregatfunktionen fassen alle Werte eines bestimmten Fakts, zum Beispiel “Dollar Sales” zusammen. Finden Sie die Aggregat-

funktionen von Synchrony heraus. Weshalb werden gerade *diese* Aggregatfunktionen angeboten?

- l) Führen Sie die folgende Abfrage aus: “Wieviele Einheiten der Marke “American Corn” wurden nach der Werbeaktion *POS Grabbers Paper* im Vergleich zur Aktion *Shelf Talkers* durchschnittlich verkauft?”. Gehen Sie dabei wie folgt vor:
- Erstellen Sie eine erste Abfrage, indem Sie das Field “Description” aus PRODUCT in das Report-Fenster ziehen, Machen Sie dasselb mit dem Faktattribut “Unit Sales”. Schränken Sie dann “Unit Sales” durch den Filter “Brand = American Corn” aus PRODUCT ein. Damit Sie die Abfrage weiter verfeinern können, klicken Sie auf ►.
 - Nach einem Rechtsklick auf die Faktspalte wählen Sie *Properties/Aggregate* und dann die Aggregatfunktion *Average*.
 - Erstellen Sie eine weitere identische Faktspalte.
 - Wählen Sie den Filter *Promotion Name = POS Grabbers* für die erste Faktspalte und *Promotion Name = Shelf Talkers* für die zweite Faktspalte. Beide Filter finden Sie unter dem Groups-Tab der Dimensionstabelle PROMOTION.
 - Führen Sie die Abfrage aus.

② **Synchrony selbständig anwenden**

- m) Vergleichen Sie die Geldumsätze der Filialen und ihre prozentuale Änderungen von 1994 bis 1995. Umsatzrückgänge sollen rot gekennzeichnet werden.
- n) 1994 wurde in allen Filialen eine Diät-Produktreihe eingeführt. Untersuchen Sie die Umsatzentwicklung von 1994 und 1995 und empfehlen Sie, in welchen Filialen die Diätprodukte wieder aus dem Sortiment genommen werden sollen.
- o) Welche Produkt-Untergruppe verkauft sich besonders gut. Wie hoch ist die Gewinnspanne bei den einzelnen Untergruppen?

- p) Vergleichen Sie die Zahlen für die Mittwochskunden mit jenen der Samstagskunden. Welche Produkte werden vor allem an Samstagen nachgefragt?
- q) Erstellen Sie eine der erstellten Abfragen mit QBE. Öffnen Sie dazu die MS Access-Datenbank  [Grocery.mdb](#).

③ Möglichkeiten und Grenzen erkennen

- r) Worin liegen die Vor- und Nachteile eines ROLAP-Werkzeugs wie *Synchrony*?
- s) Vergleichen Sie *Synchrony* mit *MS Access*, *MS Excel* und *Cognos PowerPlay*.

Entwicklung und Betrieb

Grundlagen	<u>2</u>
✓ Operative und analytische Datenbanken	<u>9</u>
✓ Enterprise Data Warehouse und Data Marts	<u>24</u>
✓ Mehrdimensionale Daten	<u>25</u>
Endbenutzerzugriff	<u>53</u>
✓ OLAP	<u>61</u>
✓ ANLAGEBERATUNG mit <i>PowerPlay</i> 📌🖱️	<u>87</u>
Modellierung relationaler Data Warehouses	<u>101</u>
✓ Datenmodellierung	<u>101</u>
✓ LIEFERFRIST 🖱️	<u>121</u>
✓ VERKAUF 🖱️	<u>122</u>
✓ HANDEL mit <i>Synchrony</i> 📌🖱️	<u>144</u>
Entwicklung und Betrieb	
⇒ Entwicklung und Betrieb	<u>153</u>
⇒ Performance	<u>176</u>
Netzzugriff	<u>214</u>
• Client/Server-Architekturen	<u>219</u>
• Zugriff über Internet und Intranet	<u>229</u>

Von Produktionsdatenbanken zum EDW

- ① Produktionsdatenbank mit
Drittgenerations-Sprachen
wie COBOL oder C



- ② Produktionsdatenbank mit
Viertgenerations-Sprachen
wie ›SQL/ ›QBE



- ③ Endbenutzersichten (Views) auf
eine Produktionsdatenbank mit
SQL/QBE und ›Berichtsgeneratoren



- ④ Views auf aufbereiteten Daten



- ⑤ Enterprise Data Warehouse
und Data Marts mit OLAP

Entwicklung von Data Warehouses

Data Warehouse-Entwicklungszyklus

① Spezifikation (Anforderungsdefinition)

Was ist das Ziel?

- ✓ Ist-Analyse
- ✓ Soll-Analyse

② Realisierung

Wie soll das Ziel erreicht werden?

① Entwurf

- ✓ Logische Datenmodellierung
- ✓ Physische Datenmodellierung
- ✓ Dialogentwurf
- ✓ Modularisierung

② Implementation

- ✓ › Deklarative Implementation (Daten, Abfragen)
- ✓ › Prozedurale Implementation (Prozeduren)
- ✓ Modul- und Systemtest

③ Betrieb

- ✓ Installation
- ✓ Wartung

Phasenunabhängige Tätigkeiten

Planung und Kontrolle der ...

- ✓ Zwischen- und Endprodukte
- ✓ Ressourcen
 - Zeit
 - Geld
 - Personal
 - Hardware/Software

Kontrollmethoden

- ✓ Spezifikations-Walk Through
- ✓ Entwurfs-Walk Through
- ✓ Implementations-Walk Through

5.54 Data Warehouse-Ziele

<i>Kriterium</i>	<i>Beispiele</i>
Abbildungstreue	einfache Abbildung betrieblicher Fakten und Dimensionen
Benutzerfreundlichkeit	Mehrdimensionalität (Würfel oder Sternschema), Drilling, Slicing and Dicing, Pivoting, Visualisierung, Views
Flexibilität	Ausbaubarkeit von Clients, Servern und Netz, vordefinierte <i>und</i> ad hoc-Abfragen
Performance	›Durchsatz, Laufzeiteffizienz, Speichereffizienz
Verfügbarkeit	Multiprozessoren, ›RAID, Festplattenspiegelung, Stand by-Server, On line-Verwaltung ohne Shut Down von Systemkomponenten
Wartbarkeit	Synchronisation mit ›OLTP, Metadatenverwaltung
Kompatibilität	›Data Mining Tools, Produktionsdatenbanken, Anwendungssoftware

Endbenutzerbefragung - Ablauf

Befragungsteam zusammenstellen

Vorbereitungsmaterial sammeln

- Jahresbericht
- Planungshandbuch
- Marketingunterlagen
- Frühere “Data Warehousing”-Initiativen

Adressaten wählen

- Linie
- Stab (v.a. “Daten”kundige)

Fragebogen zusammenstellen →

Ablauf planen

Befragung ankündigen

Rollenverteilung und Interviewtechniken reflektieren

- Fragende
- Antwortende
- Protokollführer
- Beobachter

Befragung durchführen →

- Endbenutzer-Vokabular verwenden
- Gemeinsame Terminologie verwenden
- Interview flexibel führen

Endbenutzerbefragung - Inhalte

Einführung

Projektziele
Interviewziele
Interviewer
Zeitraumen
Folgeschritte
Pflichtenhefte

Ziele

Geschäftsziele
Produkte
Erfolgsindikatoren
Geschäftsrisiken
Früherkennung
Innovationshäufigkeit

Analyseanforderungen (Ist/Soll)

Routineanalyse
Ad hoc-Analysen
Historische Auswertungen
Berichte
Quellenzugang

Schluss

Zusammenfassung
Bitte um Zusatzmaterial
Projektchancen

Feedback



Feinspezifikation

Spezifikationsdokument

Zusammenfassung für den Manager

① **Projektziel** (Grobspezifikation)

② **Anforderungen**

- ✓ Betriebswirtschaftliche Anforderungen
- ✓ Informationstechnologische Anforderungen

③ **Datenquellen**

You Can't Eat an Elephant in One Byte !

☹ Anforderungsdefinition

- Koordination von Fachabteilungen

☹ Laden operativer Daten

- Datenextraktion, -transformation und -integration

☹ Komplexität und Redundanz der Datenmodelle

- Definition und Wartung komplexer Tabellenstrukturen
- Einführung von Redundanz

☹ Speicher- und Zugriffseffizienz

- durch mehrere Fakttabellen und grosse Dimensionstabellen
- durch umfangreiche Indizes und Vorberechnungen

☹ Datenverantwortung und -sicherheit

- Datenhoheit der Fachabteilungen

☹ Evaluation und Schulung

- Tabellenmodell (z.B. Fremdschlüssel-Begriff)
- Boolesche Abfragelogik für Endbenutzer

☹ Projektmanagement

- zu lange Projektdauer (Jahre statt Monate)
- mangelnde Erfahrung

Rollen der DW-Entwicklung und -Verwaltung

Entwicklung eines Data Warehouse	
Rolle	Aufgabe
<i>Data Warehouse</i> Administration	Entwicklung und Verwaltung in allen Phasen leiten
<i>Business</i> Requirements Analysis	Anforderungen der Endbenutzer spezifizieren und Entwurf begleiten
<i>Data Warehouse</i> Architecture	Data Warehouse-Architektur entwerfen, insbesondere Daten modellieren
<i>Data</i> Acquisition <i>Development</i>	Daten extrahieren, transformieren und integrieren
<i>Data</i> Access <i>Development</i>	Freien Zugriff für Experten und vordefinierte Pfade für gelegentliche Benutzer definieren
Programming and <i>Maintenance</i>	Programmierung, Überwachung und Sicherheitsverwaltung
Wartung eines Data Warehouse	
<i>Data Warehouse</i> Administration	Entwicklung und Verwaltung in allen Phasen leiten
<i>Organisational</i> Change Management	Betriebliche Auswirkungen von Data Warehouse-Entwicklungen begleiten
<i>Database</i> Administration	Physische Daten verwalten
Metadata Management	Metadaten verwalten

Rollen der DW-Entwicklung (A 5.8)

Während der Entwicklung eines Data Warehouse üben Mitarbeiter die unterschiedlichsten Rollen aus. Ordnen Sie den folgenden Rollen passende Tätigkeiten und Kompetenzen zu.

- a) Projektdirektor
- b) Projektmanager
- c) Analytiker
- d) Data Warehouse-Architekt
- e) Datenbankadministrator
- f) Systemadministrator
- g) Migrationsspezialist
- h) Data Mart-Entwickler
- i) Qualitätskontrolleur
- j) Legacy-System-Spezialist
- k) Ausbildner
- l) Handbuch-Autor
- m) Kommunikationsspezialist
- n) Sponsor
- o) Help Desk-Verantwortlicher
- p) Werkzeugspezialist
- q) Web Master
- r) externer Berater

Software für Data Warehouses

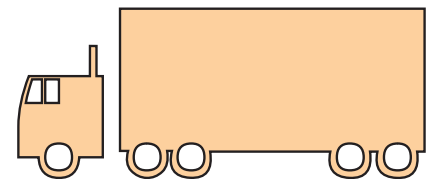
Produktionsdaten

- Relationale DBMS
(z.B. von *Oracle*)
- Nichtrelationale DBMS
- Einzeldateien



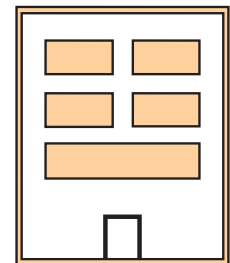
Laden

- DBMS-Funktionen wie Replikation und Snapshot
- Dedizierte Standardsoftware
(z.B. *ETI-EXTRACT* Suite von Evolutionary Technologies, *Prism Warehouse Manager* von Prism Solutions)
- Selbst erstellte Software
(z.B. gespeicherte Prozeduren, SQL, 3GL)



Data Warehouse - Verwaltung

- Spezialisierte relationale DBMS
(z.B. *Sybase IQ*, *Red Brick Data Warehouse*)
- Multidimensionale Datenbanksysteme
(z.B. *Arbor Essbase*)
- Werkzeuge zur Metadatenverwaltung
(z.B. *Sybase PowerDesigner WarehouseArchitect*)
- Werkzeuge zur Performanceoptimierung
(z.B. Analyse der Abfrageprofile durch ein OLAP-Werkzeug)
- Datensicherung und -archivierung
(z.B. Systemsoftware)



...



Data Warehouse - Zugriff

- Abfragewerkzeuge und Berichtsgeneratoren
(z.B. OLAP-Werkzeuge wie *Cognos Powerplay*)
- Internetwerkzeuge
(z.B. *Sybase PowerDynamo* oder *MS InterDev*)
- Data Mining-Werkzeuge
(z.B. *Cognos Impromptu*)
- Allgemeine 4GL-Entwicklungsumgebungen
(z.B. *MS Visual Basic* oder *PowerBuilder*)



...

IBM Information Warehouse

Hardware und Betriebssysteme

✓ › OS/2

✓ › AIX

✓ › MVS

RDBMS

✓ › DB2

Data Warehouse Management

✓ Metadatenverwaltung

✓ Datenverteilung

✓ Zugriffswerkzeuge

✓ › Data Mining-Werkzeuge

Implementation und Betrieb

Spezifikation und Entwurf

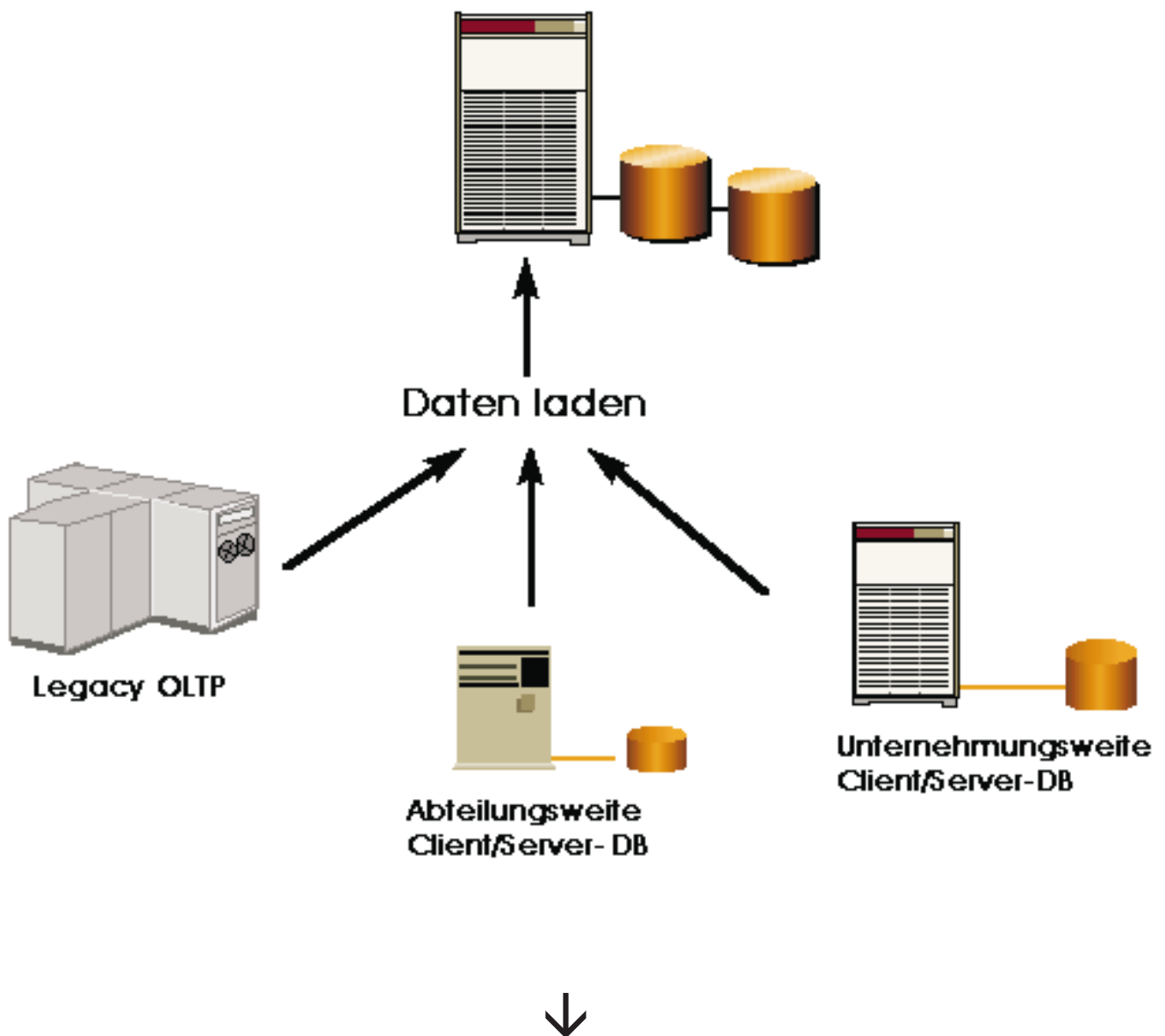


- ① Daten aus operativen Datenbanken **kopieren**
- ② Gekaufte und neu erhobene Daten **hinzufügen**
- ③ Daten **transformieren**
- ④ Daten im zentralen Data Warehouse **speichern**
- ⑤ Daten auf allfällige Data **Marts** replizieren
- ⑥ Daten **abfragen**
- ⑦ **Data Mining**
- ⑧ Ergebnisse **interpretieren**



Illustration am Beispiel ZEITSCHRIFTEN

5.58 Operative Daten laden



Daten laden

- ①-② extrahieren
- ② transformiert
- ④-⑤ integrieren

Legacy OLTP: On Line Transaction Processing durch Grossrechner

ZEITSCHRIFTEN¹ - 1. Fallbeispiel

Verlag

Zeitschriften “Auto”, “Wohnen”, “Sport”, “Musik” und “Comics”

Marktforschungsfragen

- Profil der Leser von “Wohnen”?
- Zusammenhänge zwischen “Sport”- und “Auto”-Lesern?
- ...

Mögliche Ergebnisse

- Hypothekarverschuldete Kunden sind mit einer Wahrscheinlichkeit von 0.8 “Wohnen”-Abonnenten
- “Sport”-Abonnent zwischen 20 und 30 subskribiert mit einer Wahrscheinlichkeit von 0.4 zwischen 30 und 40 auch “Auto”
- ...



Phasen des Data Warehousing und Data Mining ① - ⑧

¹ Fallbeispiel nach Adriaans/Zantinge, Data Mining

2. Daten auswählen

① Ausgewählte Produktionsdaten **kopieren**

⇒ Ausgewählte Attribute der Produktionsdatenbank in die Tabelle SUBSKRIPTIONEN kopieren

<i>Kundennr.</i>	<i>Name</i>	<i>Adresse</i>	<i>Subskription</i>	<i>Zeitschrift</i>
14005	Müller Dani	Hofstrasse 1	20.3.1996	“Auto”
18004	Meier Fritz	Stadtstrasse 5	28.2.1901	“Wohnen”
18004	Ott Hugo	Bergweg 3	6.5.1996	“Musik”
...
14005	Miller Dani	Hofstrasse 1	4.5.1997	“Sport”

② Gekaufte und getrennt erhobene Daten **hinzufügen**

⇒ Marktforschungsdaten kaufen

Geburtsdatum

steuerbares Einkommen

Privatkreditvolumen

Autoeigentum

...

⇒ Zusätzliche Daten per Telefoninterview erheben

Wohneigentum

...

3. Daten transformieren

③ Daten mit SQL transformieren

a) Duplikate und irrelevante Attribute entfernen

- ⇒ Duplikate "Müller Dani ..." und "Miller Dani ..." entfernen
(Kleine Namensunterschiede in sonst identischen Tabellenzeilen)
- ⇒ Irrelevantes Attribut *Name* weglassen

b) Nicht plausible Werte berichtigen oder weglassen

- ⇒ Zeile mit dem unplausiblen Subskriptionsjahr "1901" weglassen

<i>Nr</i>	<i>Adresse</i>	<i>Subskriptionsdatum</i>	<i>Zeitschrift</i>
18004	Stadtstrasse 5	28.2.1901	"Wohnen"

c) Attribute redefinieren

- ⇒ Zu detaillierte Attribute redefinieren

Detailliertes Attribut	Redefiniertes Attribut
Geburtsdatum	<i>Altersklasse</i>
Adresse	<i>Wohnregion</i>
Subskriptionsdatum	<i>fortlaufende Monatszahl seit 1990</i>
Einkommen	<i>Einkommensklasse</i>

D Daten transformieren und analysieren

d) Zeilen **aggregieren** und **Formate vereinheitlichen**

⇒ Zeilen mit gleicher Kundennummer gruppieren

⇒ Binäre Attribute mit den Werten 1 und 0 codieren

ja / nein und abonniert / nicht abonniert werden zu 1 / 0

<i>Kundennummer</i>	<i>Altersklasse</i>	<i>Einkommensklasse</i>	<i>Kreditklasse</i>	<i>Autoeigentümer</i>	<i>Hauseigentümer</i>	<i>Region</i>	<i>"Auto"</i>	<i>"Wohnen"</i>	<i>"Sport"</i>	<i>"Musik"</i>	<i>"Comics"</i>
18004	20	39	10	0	1	1	0	0	1	1	0
...

④ Transformierte Daten in das Warehouse **integrieren**

⑤ Data **Marts** "beliefern"

⑥ Daten **konventionell** auswerten

⑦  **Data Mining** (Teil II des Fallbeispiels *Zeitschriften*)

⑧ Ergebnisse **interpretieren**

Funktionalität von Ladeoperationen

Allgemeine Anforderungen

- ✓ Auf ein zentrales Data Dictionary zugreifen
- ✓ Stapelweise und inkrementell laden
- ✓ Auf Verlangen und automatisch laden

① Extraktion

- ✓ Aus verschiedenen Quellen laden, vor allem aus ...
 - flachen und indizierten Dateien (v.a. VSAM)
 - relationalen und anderen Datenbanken (v.a. IMS)

② Transformation

- ✓ Validität prüfen (Vollständigkeit und Konsistenz)
 - Syntaktische Fehler erkennen
 - Semantische Fehler erkennen
- ✓ Datentypen konvertieren, zum Beispiel ...
 - vom EBCDIC- zum ASCII-Code
 - vom julianischen zum gregorianischen Datum
 - von einer Nullwertcodierung zu einer anderen
- ✓ Attribute redefinieren
 - z.B. Einzelwerte in Klassen zusammenfassen
- ✓ Attribute neu einfügen
 - z.B. Zeitdimension

- ✓ Werte vorberechnen
 - z.B. Tagesumsätze in Wochenumsätze aggregieren
 - z.B. statistische Kennzahlen berechnen

③ Integration

- ✓ Geladene Daten ...
 - direkt in Warehouse-Tabellen einfügen
 - in einer sogenannten Staging Area zwischenspeichern
- ✓ Datenmodell vorbereiten
 - mehrdimensionaler Würfel
 - relationales Sternschema
- ✓ Geprüfte und transformierte Daten integrieren
- ✓ Indizes neu aufbauen oder fortschreiben
- ✓ Metadaten nachführen

Beispiel Validitätsprüfungen

Syntaktische Prüfungen

✓ Namenskonventionen

*Punkt nach Abkürzungen
ausgeschriebener Vorname*

✓ Fehlende Werte

Andrea Braun ohne Geschlechtswert

Semantische Prüfungen

✓ Wertebereiche

Monatszahl 13

✓ Homonyme

*L.R. Brown
St. Peter's Church
1st St.
St. Petersburg*

✓ Synonyme

<i>PC</i>	<i>Gestell 237</i>
<i>Personalcomputer</i>	<i>237</i>
<i>Microcomputer</i>	<i>No. 237</i>

✓ Inkonsistenzen

1997er-Transaktion mit einem erst 1998 eingeführten Produkt

Ladesoftware

Entwicklung

- Selbstentwickelte Software

zum Beispiel mit ›SQL, ›gespeicherten Prozeduren, COBOL, ›MVS ›JCL, Unix Scripts

- Standardsoftware →

Spezialisierung

- Allgemeine DBMS-Werkzeuge

zum Beispiel zur ›Replikation

Problem: meist beschränkt auf RDBMS-Quellen

- Dedizierte Ladewerkzeuge →

Technologie

- Generatoren

Problem: Verwaltung der generierten Programme

- Selbständige Systeme

Quellen

- ›flache Dateien
- indizierte Dateien
- ›hierarchische DBMS
- ›Netzwerk-DBMS
- RDBMS

Einige Ladeprodukte

Prism Warehouse Manager von *Prism Solutions*

- Auswahl, Transformation und Integration
- Generator

PASSPORT von *Carleton (von Oracle übernommen)*

- Auswahl, Transformation und Integration
- COBOL-Generierung
- dreistufige Architektur

Integrity von *Vality*

- Ergänzung von Extraktionswerkzeugen mit Integritätsprüfung

EDA Suite von *Information Builders*

- Auswahl, Transformation und Integration
- SQL-Generierung

ETI-EXTRACT Tool Suite von *Evolutionary Technologies*

- Auswahl, Transformation und Integration
- Generator

PowerMart Suite von *Informatica*

- Auswahl, Transformation und Integration
- Data Mart Builder
- inkrementelles Laden

Physische Modellierung

① Logische Modellierung

abbildungstreues und
benutzerfreundliches Datenmodell !

- Denormalisierung
- Sternschema



③ ▶ Physische Modellierung

effizientes *physisches* Schema !

- Tabellen definieren
- Schlüssel physisch definieren
- Integritätsprüfungen physisch definieren
- ...
- Performance optimieren

Performanceprobleme

Steigende **Verarbeitungszeiten**

- ☹️ **Laden**
- ☹️ **Antworten**
- ☹️ **Backup**
- ☹️ **Recovery**
- ☹️ ...

+

Sinkende **Verfügbarkeit**

- ☹️ **Shutdowns durch Laden**
- ☹️ **Shutdowns durch Backup**
- ☹️ ...

↓

Laden, Abfragen, Aggregieren

Performanceproblem Aggregation

Berechnungen brauchen Zeit



Aggregationsdaten belegen bis zu 2/3 eines DW



Vordefinierte oder Ad hoc-Aggregation?

- Betriebliche Bedeutung
- Häufigkeit
- Prozessor- und Speicherbelastung

Performanceproblem Sternschema

Mehrdimensionalität →

Dimensionenzahl und Kardinalitäten ↗



Grösse der Faktentabelle ↗

Faktenschlüssel

Dimensionskategorien im Faktenschlüssel ↗
(Jahr, Monat, Woche, Tag)



künstlicher Faktenschlüssel

Verbund von Dimensions- und Faktentabellen

Siebentabellenverbund erfordert $7! = 5040$ Paarverbunde



Faktentabelle an den Anfang eines Mehrtabellenverbunds

5.59 Performanceproblem Mehrdimensionalität I

Kardinalitäten der Dimensionen

<i>Dimensionen</i>	<i>Kategorien</i>	<i>Kardinalität</i>
Filialen	Zahl	20
Produkte	Zahl	50
Periode	Jahr / Monat / Tag	10 / 12 / 30

Zeilen des Fakts *Jahresumsatz* nach Dimensionen

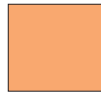
<i>Berechnung</i>	<i>Zeilen der Fakttabelle</i>
10 Jahre	10
10 Jahre, 20 Filialen	$10 \cdot 20 = \mathbf{200}$
10 Jahre, 20 Filialen, 50 Produkte	$200 \cdot 50 = \mathbf{10'000}$
120 Monate, 20 Filialen, 50 Produkte	$10'000 \cdot 12 = \mathbf{120'000}$
3'600 Tage, 20 Filialen, 50 Produkte	$120'000 \cdot 30 = \mathbf{3'600'000}$



Speicher-, Zugriffs- und Rechenaufwand optimieren

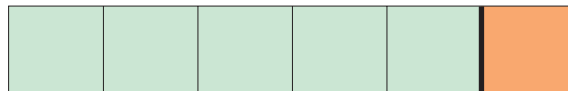
Performanceproblem Mehrdimensionalität II

Umsatz **total**



Umsatz nach 5 **Produkten**

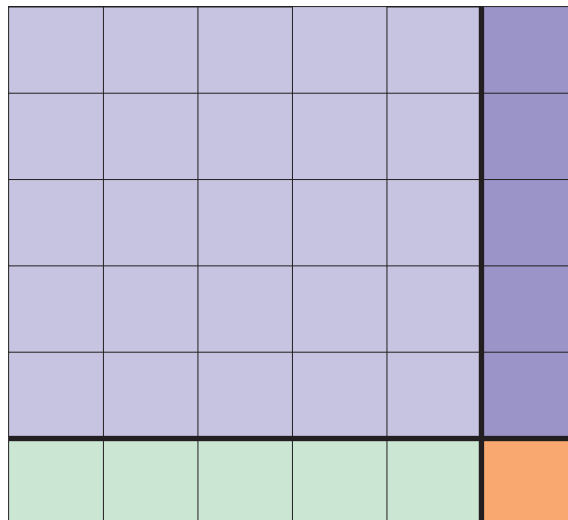
(1 Dimension und 1 Aggregat)



Umsatz nach 5 **Produkten**

und **Regionen**

(2 Dimensionen und 11 Aggregate)



kombinatorische Explosion



Speicheraufwand optimieren

Data Warehouse-Performance

Data Warehouse-Ziele

- Abbildungstreue
- Benutzerfreundlichkeit
- Flexibilität
- Verfügbarkeit
- Kompatibilität
- **Performance**



- ⇒ Speicherkapazität und -effizienz
- ⇒ Laufzeiteffizienz, insb. Antwortzeit
- ⇒ Zahl der Benutzer



Hardwaredeterminanten

- Prozessor
- Externspeicher
- Internspeicher
- Kommunikation

Datenmodell-Determinanten

- Physisches Datenmodell
- Logisches Datenmodell



Verbesserung?

Datenmodell-Performance verbessern

Logisches Datenbankmodell

Minimierung der Objekte und Beziehungen

Modularisierung nach Data Marts

Sternschema und Denormalisierung →

- Vorberechnung
- Spaltenreplikation
- Vorverbund

Physisches Datenbankmodell

Indexierung

- ›B-Baum
- Bitmuster-Index →
- ›Hashfunktion
- Künstliche Faktenschlüssel

Speicherung

- Partitionierung →
- Nullwerte
- Kompression

Abfragen vordefinieren und überwachen

- Ablaufplan
- Off line-Kandidaten aussondern
- Überwachung mit Warteschlangen und Query Blockers

Parallelisierung →

Optimierungsmassnahme Denormalisierung

Vorbereitung

Bsp. Verkäufe nach Jahr, Monat und Woche zusammenfassen

⇒ 1 Detailtabelle, 3 Aggregationstabellen

Spaltenreplikation

Verbund durch Integration fremder Spalten vermeiden

⇒ redundante Spalten

Vorverbund

Verbund durch Integration der Verbundtabelle vermeiden

⇒ redundante Tabelle



Kosten und Nutzen der Denormalisierung abwägen
(Zeitgewinn versus Speicher- und Zeitaufwand)

5.61 Normalisierung verlangsamte Abfragen

VERANSTALTUNG

<u>Name</u>	<u>Stunden</u>
<u>Recht</u>	<u>2</u>
BWL	4
VWL	3

ORGANISATION

<u>Name</u>	<u>Semester</u>	<u>Dozent</u>	<u>Raumnummer</u>
Recht	SS93	Meier	111
Recht	SS94	Meier	112
BWL	WS94	Schmid	111
BWL	SS95	Schmid	111
...

RAUM

<u>Raumnummer</u>	<u>Plätze</u>
111	200
112	150

Abfragen auf **einer** und **mehreren** Tabellen

a) Stundenzahl der Veranstaltung Recht ?

b) **Plätze** der Räume für **BWL**-Vorlesungen von **Schmid** ?

5.62 Denormalisierung beschleunigt

Plätze der Räume für **BWL**-Vorlesungen von **Schmid** ?

ORGANISATION

<u>Name</u>	<u>Semester</u>	<u>Dozent</u>	<u>Raumnummer</u>
Recht	SS87	Meier	111
Recht	SS88	Meier	112
BWL	WS88	Schmid	111
BWL	SS89	Schmid	111
...

+

RAUM

<u>Raumnummer</u>	<u>Plätze</u>
111	200
112	150

VERBUND

<u>Name</u>	<u>Semester</u>	<u>Dozent</u>	<u>Raumnummer</u>	<u>Plätze</u>
Recht	SS87	Meier	111	200
Recht	SS88	Meier	112	150
BWL	WS88	Schmid	111	200
BWL	SS89	Schmid	111	200
...

Optimierungsmassnahme Partitionierung

Begriff

Tabellenspaltung

Zweck

- Zugriffszeit optimieren
- Speichereffizienz optimieren
- Partitionsweise Lösch- und Updateoperationen erleichtern

Arten

- **Horizontale** Partitionierung (z.B. nach dem Transaktionsdatum)
- **Vertikale** Partitionierung (z.B. nach der Normalisierung)

Partitionsschlüssel

- Reihenfolge der Zeilen bzw. Spalten
- statische und dynamische Daten
- häufig zusammen verwendete Daten
- Grösse von Eingabe-/Ausgabeblocks
- ...

Partitionierungsbeispiele

Zeilenpartitionierung für Transaktionsdatenbanken

<i>Name</i>	<i>Abteilung</i>	...
Müller	Personalwesen	...
...
...
Studer	Finanzwesen	...
...

Spaltenpartitionierung für Data Warehouses

<i>Name</i>	<i>Abteilung</i>	...
Müller	Verkauf	...
Meier	Beschaffung	...
Schmid	Rechnungswesen	...
...

5.63 Spaltenpartitionierung

I/O-Blöcke so speichern, dass ...

zusammen verwendete Daten *hintereinander* stehen

+

Anwendungen verwenden in der Regel ...

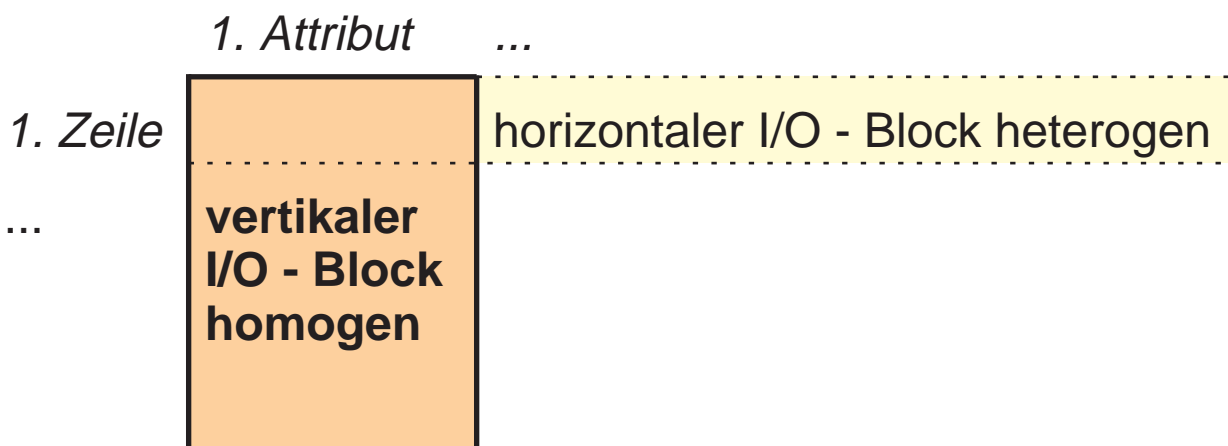
- *operative* Daten satzweise (zeilenweise)
- *analytische* Daten attributweise (spaltenweise)

+

Die **Kompression** *homogener* Daten ist leichter

⇓

Data Warehouse - Daten so speichern, dass Werte des gleichen *Attributs* im gleichen I/O-Block stehen



Optimierungsmassnahme Indexierung

Nutzen

- ☺ Zeitgewinn durch Direktzugriff

Kosten

- ☹ Speicheraufwand
- ☹ Erzeugung der Indizes zur Ladezeit
- ☹ u.U. Zeitverlust gegenüber sequentieller Suche

5.64 Optimierungsmassnahme *Bitmuster*

Datenbasis

<i>Wirtschaftssektor</i>	<i>Region</i>	<i>ISDN-Anschluss?</i>
Dienstleistung	Oberrhein	ja
Landwirtschaft	Mittelland	nein
Dienstleistung	Oberrhein	ja
Dienstleistung	Genfersee	nein

Abfrage

Wieviele *Dienstleistungs*firmen verfügen über einen *ISDN*-Anschluss und haben ihren Sitz in der Region *Oberrhein*?

<i>Dienstleistung</i>	<i>Oberrhein</i>	<i>ISDN</i>
1 (ja)	1 (ja)	1 (ja)

Bitmuster-Index der Datenbasis (Abfrageergebnis eingerahmt)

<i>Dienstleistung?</i>	<i>Oberrhein?</i>	<i>ISDN-Anschluss?</i>
1	1	1
0	0	0
1	1	1
1	0	0

Abfrageergebnis

2

Komprimierte Bitmusterindizes

Mehrwertiges Attribut

z.B. *Wirtschaftssektor* [Dienstleistung, Industrie, Landwirtschaft]



in mehrere **zweiwertige Attribute** transformieren

z.B. in [Dienstleistung, nicht Dienstleistung] und [Industrie, nicht Industrie] und [Landwirtschaft, nicht Landwirtschaft]



Mehrere Bitmusterindizes

z.B. drei Indizes statt ein Index



Bitketten der Bitmusterindizes komprimieren

Bsp. Ein *komprimierter Bitmusterindex* für eine halbe Million Sätze kann in einem einzigen Block von 4096 Bytes gelesen werden

Kompression und Effizienz bitweiser
Operationen erleichtern **Ad hoc** - Abfragen

Wann Bitmuster-Indizes?

✓ Daten **statisch**

- read only oder
- Update selten

✓ Daten **geringer Kardinalität**

- wenig Attributwerte pro Spalte
im Vergleich zur Zeilenzahl
Bsp. Ländercode

Kosten und Nutzen der Bitmuster-Indexierung

Nutzen

- 😊 bei kleiner Kardinalität kleiner als B-Baum
- 😊 kurze Abfragezeit
(kleiner Index und bitweise Operationen)

Kosten

- 😞 Speicheraufwand
- 😞 Erzeugung der Indizes zur Ladezeit
- 😞 u.U. Zeitverlust gegenüber sequentieller Suche

Performancedeterminanten Hardware

Prozessorleistung

- Mehrprozessor-Durchsatz
- Einzelprozessor-Durchsatz

Internspeicherleistung

- Kapazität
- Zugriffszeit

Externspeicherleistung

- Kapazität =
Rohdatenvolumen
+ Overhead (Indizes, Temporärspeicher)
(3 bis 10 mal das Rohdatenvolumen)
- Durchsatz

Kommunikation

- Systembus
- Netz

Hardwareperformance verbessern

① Einprozessor-Systeme

⇒ Komponenten optimieren



Flaschenhals-Problem

② Mehrprozessor-Systeme (MPS)

⇒ Divide et impera!

- Scale Up (Verbesserung der Kapazität)
- Speed Up (Verbesserung der Geschwindigkeit)



⇒ Programmierbarkeits-Problem

Komponentenperformance verbessern

Prozessor

- Prozessorarchitektur wechseln (z.B. von CISC nach RISC)
- Takt erhöhen (z.B. von 200 auf 300 MHz)
- Bus verbreitern (z.B. von 32 bit zu 64 bit)

Internspeicher

- Hauptspeicher vergrößern und beschleunigen
- Cache vergrößern und beschleunigen

Peripherie beschleunigen

- RAID (engl. Redundant Array of Inexpensive Disks)
 - viele kleine Plattenspeicher als Ganzes an den Systembus angeschlossen
 - kleinere Festplatten sicherer
 - Wiederherstellbarkeit einer ausgefallenen Platte wegen redundanter Datenspeicherung

Optimierungsmassnahme Parallelisierung

Parallelverarbeitung :=

- ✓ gleichzeitige Verarbeitung
- ✓ mehrerer Aufgaben
- ✓ auf mehreren Prozessoren

Aufgabe (engl. task) :=

- ✓ *Anweisungsteil*
- ✓ *Anweisung*
- ✓ *Prozess* aus mehreren Anweisungen
- ✓ *Job* aus mehreren Prozessen
- ✓ unabhängiges *Programm*

MPS parallelisieren ...

⇒ **Auswahl**abfragen (Abfragen i.e.S.)

- durch die Datenbankmaschine
- durch OLAP-Werkzeuge auf Clients und Servern

⇒ **Aktions**abfragen

- Extraktion, Transformation und Integration
- Sicherung und Wiederherstellung
- Indexierung

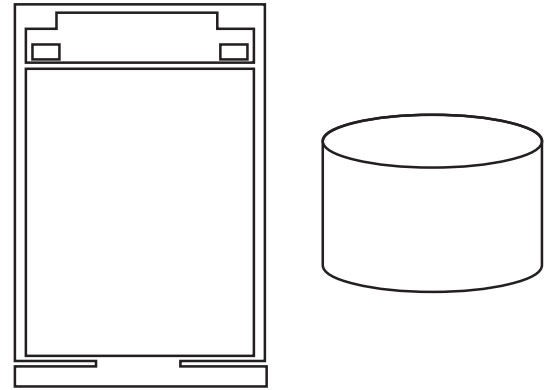


☺ Speed Up und Scale Up

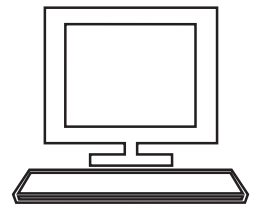
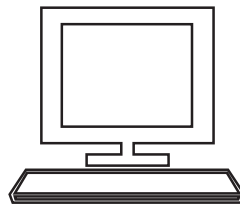
☹ Verwaltungsaufwand

5.65 Beispiel eines Data Warehouse-MPS

SMP-Server
8 Prozessoren
8 Gigabyte RAM
32 Gb Externspeicher
paralleles DBMS
(z.B. Oracle)



MS Windows NT
OLAP-Software
(z.B. Oracle Express)
Data Mining-Software
(z.B. SAS)



Arten von Mehrprozessorsystemen

① Mehrere Prozessoren **eng** koppeln

- SMP-Systeme

② Mehrere Prozessoren **lose** koppeln

- MPP-Systeme
- Clusters aus Mehrprozessor-Systemen

5.66 Symmetrisches¹ MPS (SMP)

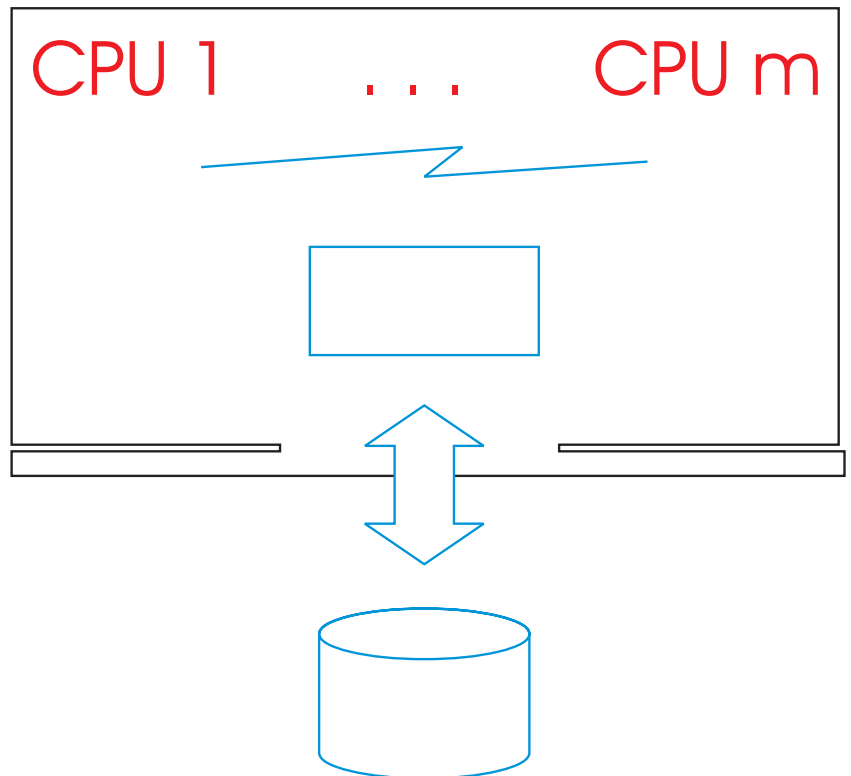
Wenige Prozessoren

Gemeinsamer Bus

Gemeinsamer Interspeicher

Gemeinsame Disk Controller

Gemeinsame Laufwerke



😊 einfach programmierbar

😞 begrenzt skalierbar
Flaschenhals Systembus

¹ symmetrisch, weil Prozessoren gleich mächtig und gleichberechtigt

Eine *mittlere* SMP-Konfiguration (1998)

Prozessoren

8 Intel Pentium Pro

Internspeicher (engl. memory)

256 Mb bis 8 Gb

Bus

PCI

Externspeicher (engl. storage)

- interne Festplatten
zusammen 1 bis 2 Tb (Terabyte)
- externe Festplatten
bis mehrere Dutzend Tb

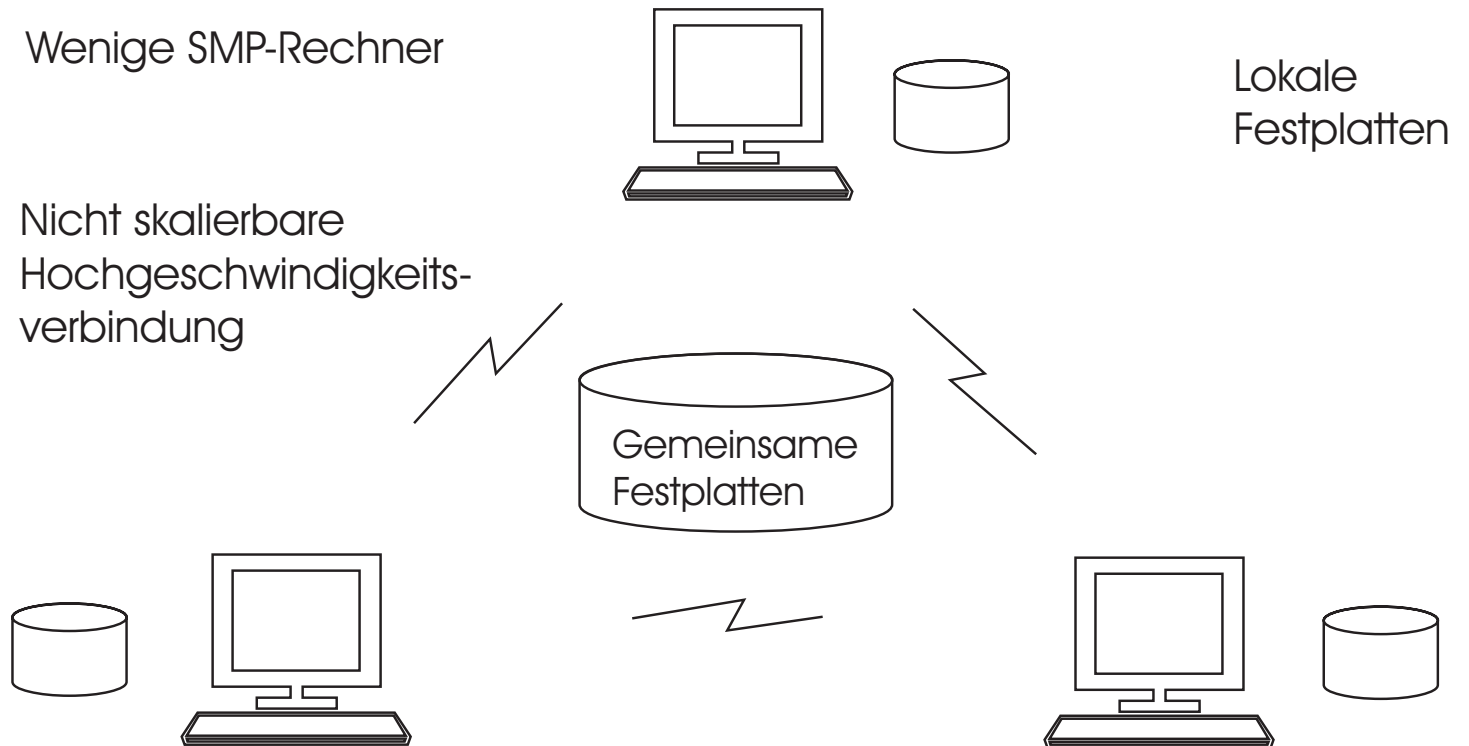
Betriebssystem

Bsp. MS Windows NT Enterprise Edition

Datenbanksystem

Bsp. MS SQL Server

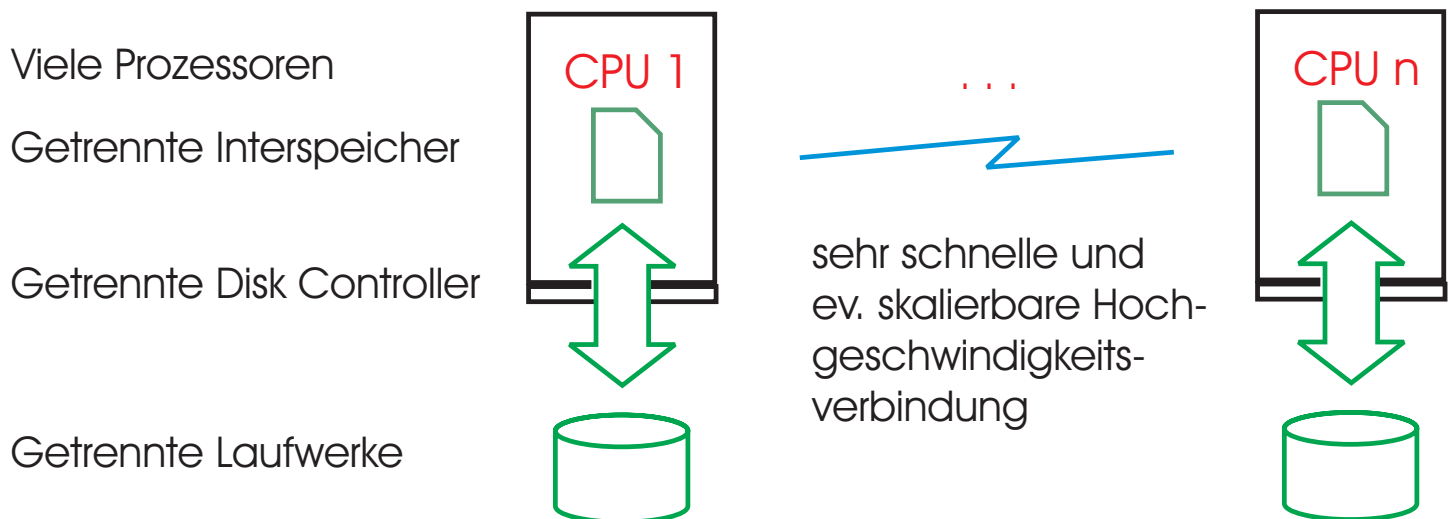
5.67 Clustering



- 😊 besser skalierbar
- 😊 fehlertolerant
- 😞 schwerer programmierbar
(Zugriff auf andere Knoten)

5.68 Massiv paralleles MPS

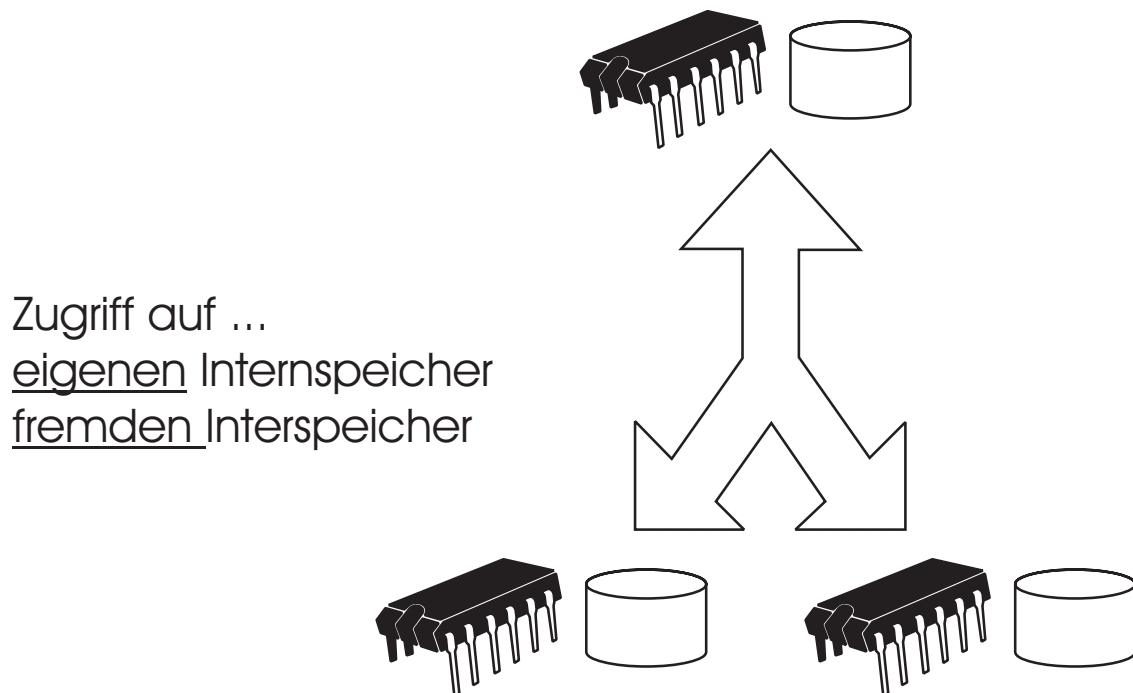
MPP := massiv paralleles Mehrprozessorsystem



- 😊 sehr gut skalierbar
- 😞 schwer programmierbar

- 1 massiv, weil mehr Prozessoren mit grösserer Bandbreite als in einem SMP
- 2 Manche Systeme greifen auch auf gemeinsame Festplatten zu
- 3 Moderne MPS bestehen auch aus SMP-Knoten (wie Clusters)

5.70 NUMA (Non Uniform Memory Access)



- ☺ gut skalierbar
- ☺ einfach programmierbar
- ☹ Internspeicherzugriff nicht uniform
 - Zugriff auf eigenen Internspeicher schnell
 - Zugriff auf fremden Internspeicher weniger schnell (z.B. 1:2)

5.69 *Symmetrische* vs. *massiv parallele* MPS

	<i>Symmetrische</i> MPS	<i>Parallele</i> MPS
<i>Synonyme</i>	Symmetric Multi-processing (SMP), Shared Memory -	Massively Parallel Processing (MPP), Shared Nothing -
<i>Prozessoren (mit Cache)</i>	wenige gleichberechtigte (symmetrische) i.d.R. 4-30	viele (bis Hunderte)
<i>Verbindung</i>	Standard-Systembus	proprietär
<i>Betriebssystem</i>	gemeinsam	getrennt
<i>Speicher</i>	gemeinsam	getrennt
<i>Leistungsfähigkeit</i>	-	+
<i>Skalierbarkeit</i>	-	+
<i>Kompatibilität</i>	+	-
<i>Programmierung</i>	leicht	schwierig
<i>Parallelisierung durch ...</i>	Betriebssystem	Anwendungsprogrammierer
<i>Installat../Betrieb</i>	+	-
<i>Kosten</i>	+	-
<i>Verbreitung</i>	stark	noch gering
<i>Hauptanwendungen</i>	DBMS, Kommunikation	“number crunching”
<i>RDBMS-Beispiel</i>	Sybase SQL Server (auch für EPS)	Sybase MPP (dediziert)
<i>Hardwarebeispiel</i>	Digital AlphaServer	IBM RS/6000 SP

Parallelisierungsverantwortliche

Betriebssystem parallelisiert ...

⇒ Operationen

⇒ Prozesse

Übersetzer parallelisiert ...

⇒ Operationen

⇒ Prozesse

Programmierer parallelisiert ...

⇒ Jobs und Programme

Hilfsprogramm parallelisiert ...

⇒ Jobs und Programme, z.B. nicht parallelisierbare ...

- Drittsoftware
- Legacy-Programme

XpertRule Profiler induziert Entscheidungs­bäume mit SQL-Anweisungen, die von einem parallelen RDBMS verarbeitet werden können

Parallelisierungsvarianten

Funktionale Parallelisierung

Verschiedene **Aufgaben** für jeden Prozessor

Daten-Parallelisierung

Verschiedenen **Daten** der gleichen Aufgabe für jeden Prozessor

Pipeline-Parallelisierung

Verschiedene **Phasen** der gleichen Aufgabe für jeden Prozessor
(z.B. Abfragephasen Scan, Verbund, Sortierung, Aggregation)



Ideal ist eine Kombination aller Parallelisierungsansätze

Datenparallelisierung

Verteilung der Eingabe/Ausgabe auf Festplatten

① **Gemeinsame** Platte für mehrere Prozessoren

⇒ langsamer wegen *sequentieller* Eingabe/Ausgabe

② **Mehrere** Platten für mehrere Prozessoren

⇒ schneller wegen *paralleler* Eingabe/Ausgabe



Wie kann ein Prozessor einer physischen oder logischen Platte (**Partition**) zugewiesen werden?

Partitionierung für die Datenparallelisierung

Parallelisierende Partitionierung := Zuordnung

- ✓ einer **physischen oder logischen** Platte
- ✓ zu **einem** Prozessor
- ✓ damit jeder Prozessor **parallel** zu den anderen
- ✓ auf **eine** Platte zugreifen kann

Zuordnung durch ...

- ⇒ Hardware (RAID)
- ⇒ Betriebssystem
- ⇒ Datenbanksystem

Partitionierungskriterien

① Bereichspartitionierung

i.d.R. nach dem meist gebrauchten Schlüssel

fiktive SQL-Anweisung

```
CREATE TABLE Verkäufe(  
    Produktnummer NUM  
    Produktname CHAR  
    ...)  
PARTITION BY VALUE  
    Produkt = 'Racing Bike' IN partition_1,  
    Produkt = 'Mountain Bike' IN partition_2,  
    ...
```

② Hashpartitionierung

Partition je nach Wert einer Hashfunktion auf einem Tabellenwert

③ Round Robin-Partitionierung

automatische Partitionierung nach der Verfügbarkeit der Prozessoren

Trade Offs einiger Optimierungsmassnahmen

	<i>Vorteile</i>	<i>Nachteile</i>
Sternschemata	benutzerfreundlich relativ schnell relativ flexibel	unvollständig redundant
Vorberechnung	schnell	unvollständig speicheraufwendig unflexibel
unkomprimierte Bitmusterindizes	schnell flexibel	read only, speicheraufwendig, kleine ›Kardinalität
Komprimierte Bitmusterindizes	schnell flexibel grosse Kardinalität (bis Tausende)	read only, speicheraufwendig

Netzzugriff

Grundlagen 2

- ✓ Operative und analytische Datenbanken 9
- ✓ Enterprise Data Warehouse und Data Marts 24
- ✓ Mehrdimensionale Daten 38

Endbenutzerzugriff 50

- ✓ OLAP 61
- ✓ ANLAGEBERATUNG mit *PowerPlay*   87

Modellierung relationaler Data Warehouses 101

- ✓ Datenmodellierung 101
- ✓ LIEFERFRIST  121
- ✓ VERKAUF  122
- ✓ HANDEL mit *Synchrony*   148

Entwicklung und Betrieb 152

- ✓ Entwicklung und Betrieb 152
- ✓ Effizienz 176

Netzzugriff

- ⇒ Client/Server-Architekturen 219
- ⇒ Zugriff über Internet und Intranet 229

5.71 Vom Host zum Client/Server-System

<i>Umgebung</i>	zentrale Aufgaben (Host oder Server)	lokale Aufgaben (Terminal oder Client)
Host-	Verarbeitung, Interaktion	E/A-Funktionen dummer Terminals
Master/Slave-	Verarbeitung, Interaktion	E/A-Funktionen intelligenter Terminals
Fileserver-	Drucker und Speicher teilen, Dateiaustausch	Verarbeitung, Interaktion
Zweistufige C/S-	Daten, Verarbeitungssteile	Verarbeitungssteile, Interaktion
Mehrstufige C/S-	Daten, Verarbeitung durch dedizierte Applikationsserver	Interaktion

Rechnernetze

Rechnernetze

- **Lokale** Netze erstrecken sich in der Regel auf ein Grundstück (LAN: Local Area Network, Bsp. Novell *Netware*)
- **Verbund**netze sind örtlich verteilt (WAN: Wide Area Network, Bsp. *Internet* →)

Komponenten

- **Clients** (inkl. Netzkarte + Netzbetriebssystem) →
- **Server** (inkl. Netzkarte + Netzbetriebssystem) →
- **Leitungen** (Telefon-, Koaxial-, Glasfaserkabel)
- **Kommunikationsprotokolle** (Bsp. ▶TCP / IP) →



Die Internet-Technologie bietet eine standardisierte Infrastruktur für LAN- und WAN-Bedürfnisse

Rechnernetze teilen Ressourcen

<i>Ressource</i>	<i>Beispiele</i>
Hardware	Prozessoren Speicher Drucker Modem Fax
Kommunikation	Telefon Fax Internet
Software	›Datenbankmaschine OLAP- und Data Mining-Werkzeuge Kommunikationssoftware (z.B. E-Mail)
Daten	›Fileserver-Dateien Produktionsdatenbank Data Warehouse E-Mail-Nachrichten

Rechnernetze für Data Warehouses

- ⇒ **Verteilte** Datenbanken
- ⇒ Laden aus den **Produktions**datenbanken
- ⇒ Laden aus dem **Enterprise** Data Warehouse
- ⇒ Endbenutzerzugriff auf Data **Marts**
- ⇒ . . .



Netzwerk-Bandbreite!

Client/Server-Systeme

Client/Server-System :=

- ✓ Verbund von Hardware und Software, der
- ✓ Aufträge von **Clients** (Arbeitsplatzrechnern)
- ✓ durch **Server** (Dienstleistungsrechner und -programme) ausführen lässt

Eine mögliche Aufgabenteilung

Client	Server
Benutzereingaben lesen, prüfen und senden	Abfragen empfangen und Ergebnisse senden
Ergebnisse empfangen, formatieren und ausgeben	Datenbankzugriffe durchführen
	Integrität, Sicherheit und Datenschutz gewährleisten

5.72 Alternative Modelle der Aufgabenteilung

<i>serverzentriert</i>		<i>verteilt</i>	<i>clientzentriert</i>	
<i>Präsentation verteilt</i>	<i>Präsentation entfernt</i>	<i>Anwendung verteilt</i>	<i>Daten entfernt</i>	<i>Daten verteilt</i>
Daten	Daten	Daten	Daten	Daten
Anwendung	Anwendung	Anwendung	Netzwerk	Netzwerk
Präsentation	Netzwerk	Netzwerk	Anwendung	Daten
Netzwerk	Präsentation	Anwendung	Anwendung	Anwendung
Präsentation	Präsentation	Präsentation	Präsentation	Präsentation

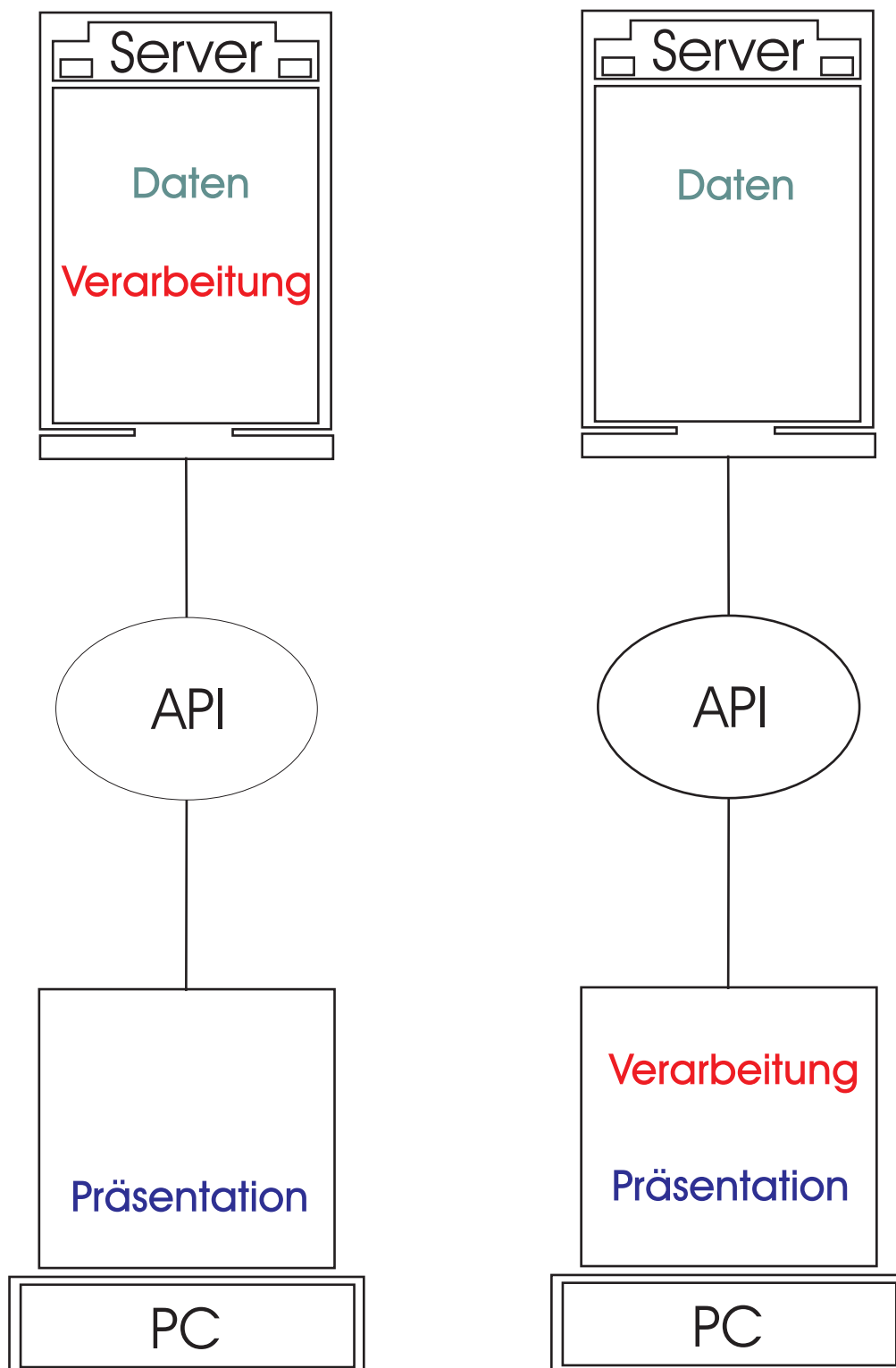
Server(s)

Client

5.73 Client/Server-Architekturen

<i>Architektur</i>	<i>Erstrechner (Client)</i>	<i>Zweitreechner (Server)</i>	<i>Dritt- (Server)</i>	<i>Netz- verkehr</i>
einstufig (one-tiered)	Daten Verarbeitung Präsentation			-
zweistufig (two-tiered, "thin client")	Daten	Verarbeitung Präsentation		Abfrage, Antwort
zweistufig (two-tiered, "fat client")	Daten Verarbeitung	Präsentation		Abfrage, Zwischen- ergebnis
dreistufig (three-tiered)	Daten	Verarbeitung	Präsen- tation	Abfrage, Zwischen- ergebnis, Antwort

Zweistufige Client/Server-Architekturen

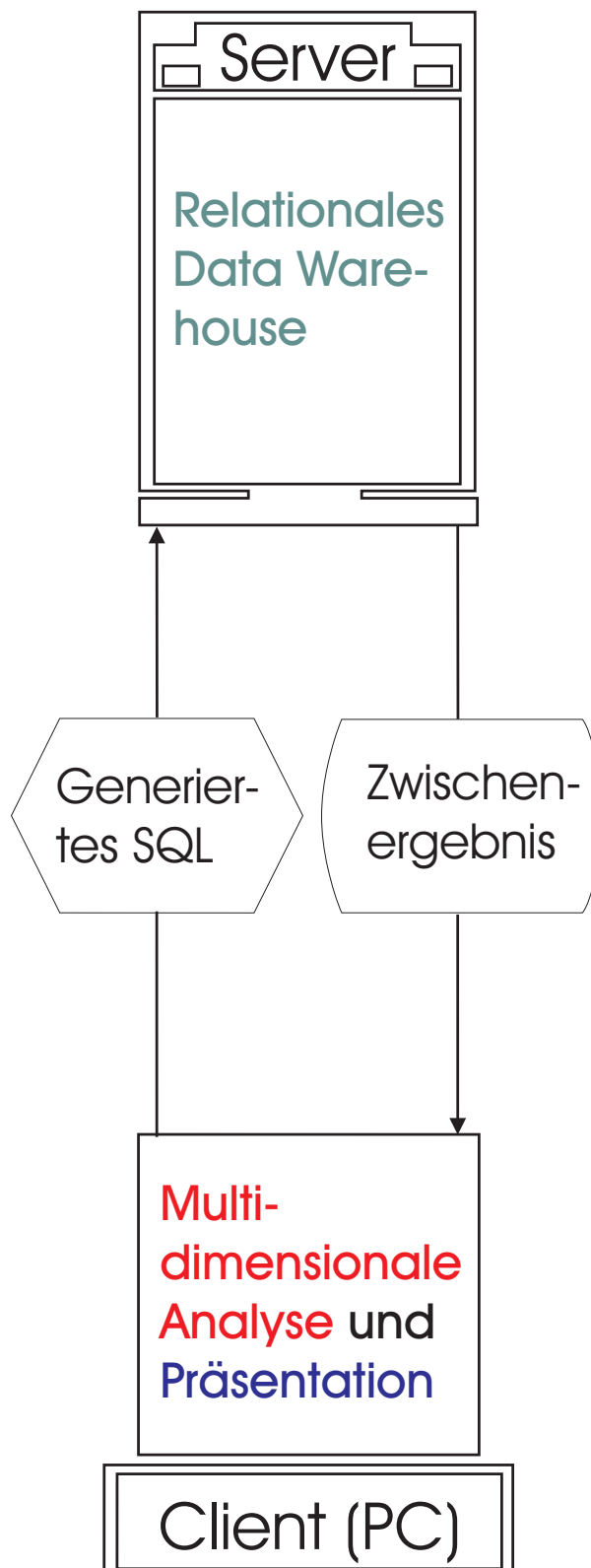


“thin client”

“*fat* client”

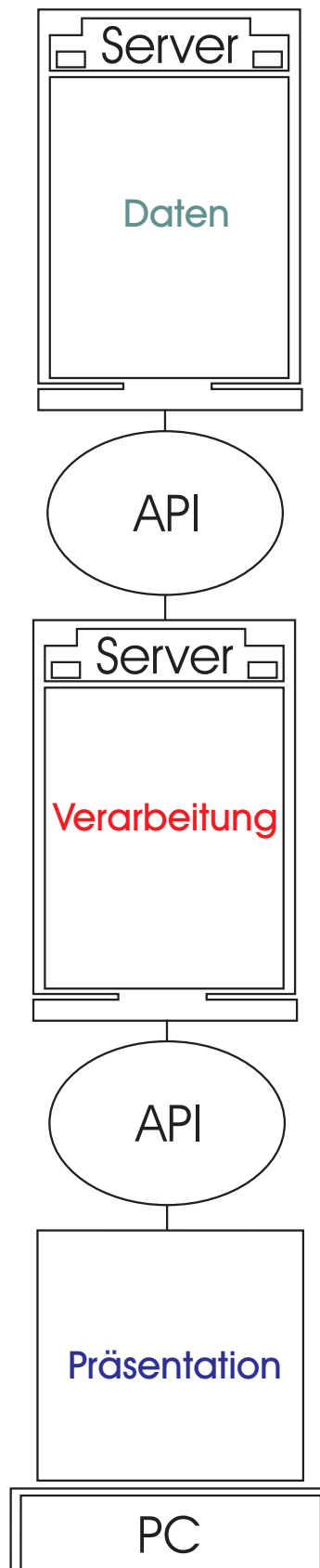
API := Application Programming Interface

5.74 Zweistufige Data Warehouse-Architektur

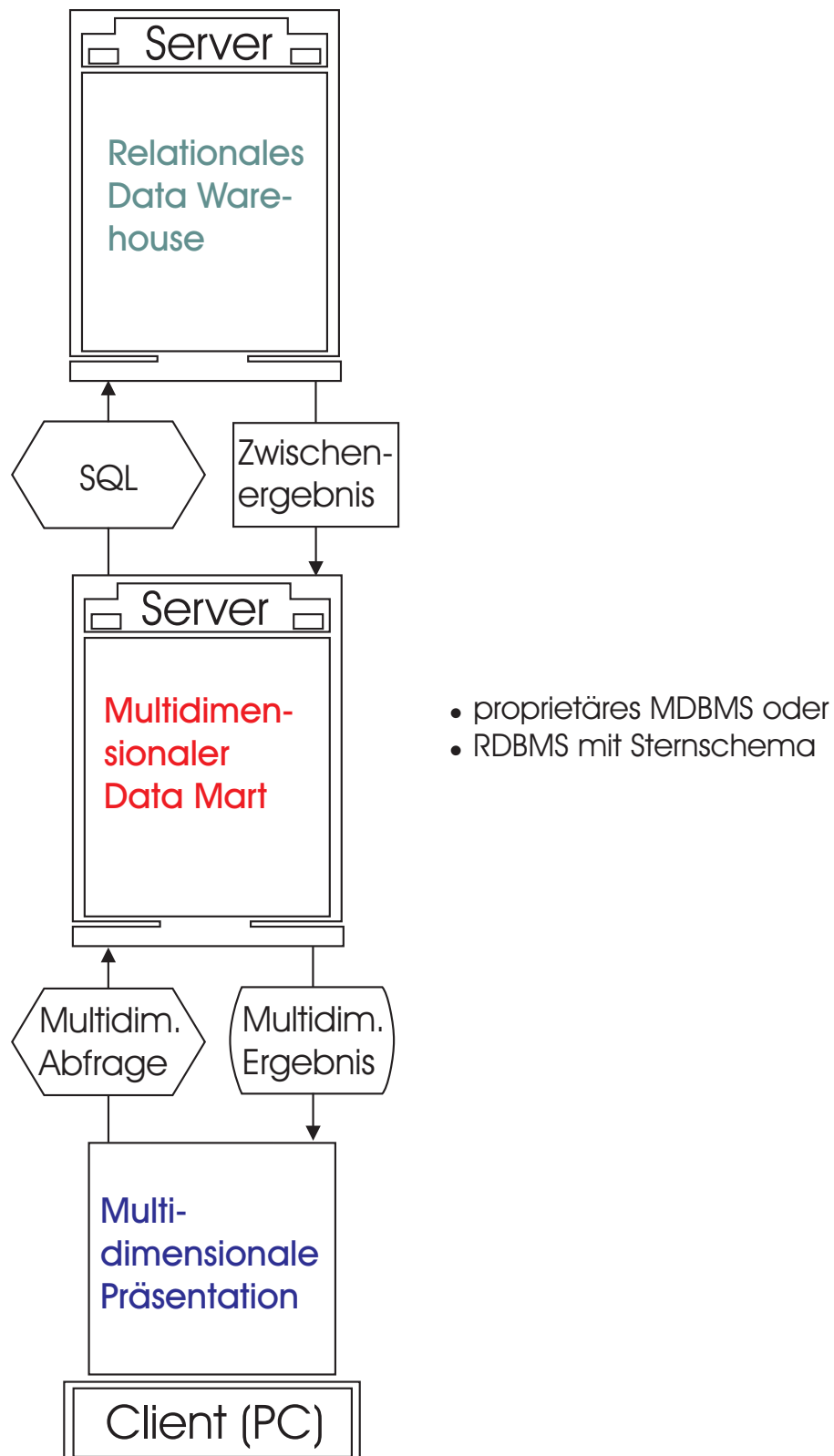


Bsp. BrioQuery von Brio Technology

Dreistufige Client/Server-Architektur



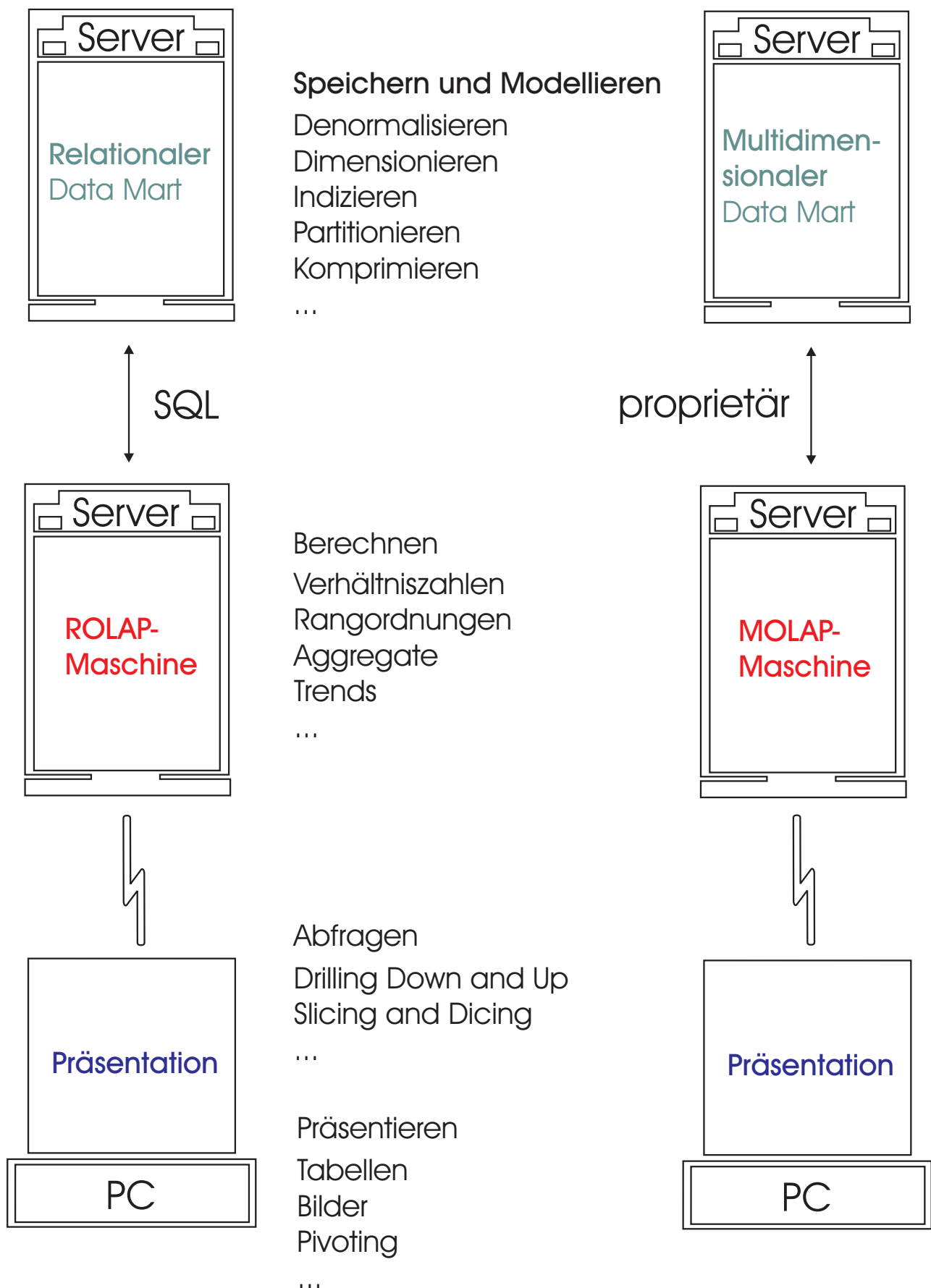
5.75 Dreistufige Data Warehouse-Architektur



vgl. dreistufige OLAP-Architekturen

226

Dreistufige ROLAP- und MOLAP-Architektur



Clientzugriff auf Data Warehouses

⇒ OLAP-Werkzeug

Endbenutzerwerkzeuge wie *Cognos* PowerPlay

⇒ Berichtsgenerator

⇒ Tabellenkalkulationspaket und OLAP-Add Ins

Microsoft Excel

Microsoft Excel Add In BusinessQuery von *Business Objects*

⇒ Zweidimensionales Abfragewerkzeuge

QBE, SQL

⇒ endbenutzernahes Client-Datenbanksystem

Microsoft Access

⇒ Groupware

Lotus Notes

⇒ Eigenentwicklung mit 4GL-Umgebungen

Microsoft Visual Basic, *PowerSoft* PowerBuilder

⇒ Web Browser und OLAP Add Ins →

ActiveX-Komponenten mit OLAP-Funktionalität

⇒ Data Mining-Werkzeug

Cognos Scenario

⇒ Text Mining-Werkzeug

Oracle ConText

⇒ Statistiksoftware-Frontend

SAS, SPSS

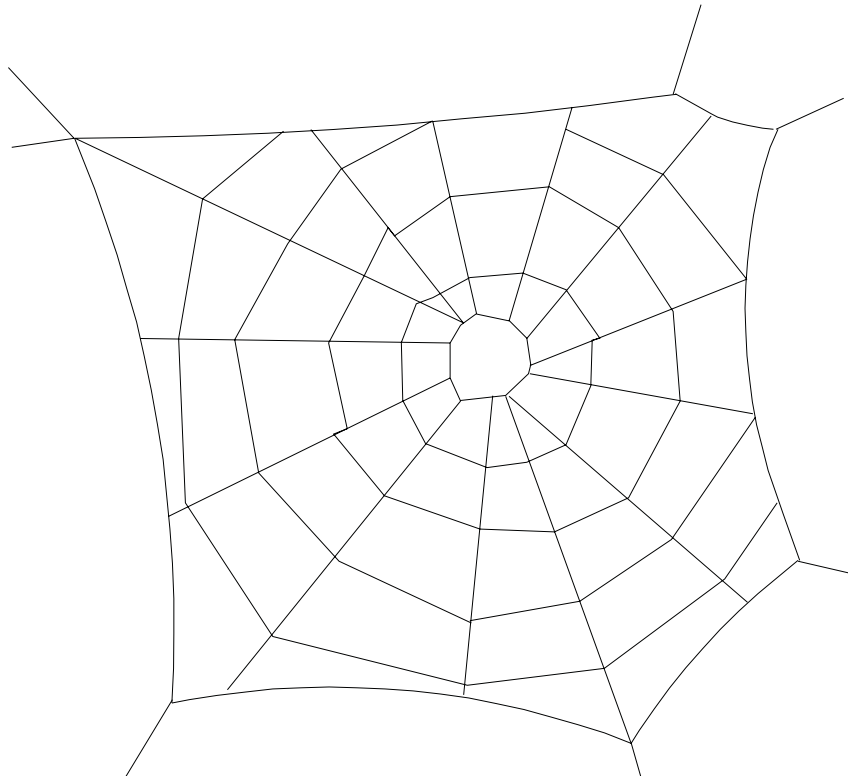
Endbenutzerzugriff auf Data Warehouses

	Lokales C/S-Netz	Intranet/Internet
<i>Plattformunabhängig</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>GUI weit verbreitet</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>GUI-Funktionalität hoch</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<i>Werkzeugbeispiele</i>	komfortables OLAP-Werkzeug	Internet Browser (ev. mit OLAP- Erweiterungen)



Webzugriff

Endbenutzerzugriff über Intranet und Internet



Der ideale Zugriff ist ...

- ✓ plattformunabhängig
- ✓ benutzerfreundlich
- ✓ kostengünstig



Internet und Intranet

Intranet für Data Warehouses

✓ Benutzerfreundlichkeit

- bedienerfreundlicher hypermedialer Web Browser

✓ Plattformunabhängigkeit

- einheitliche Frontend-Software
(v.a. für MS Windows, Mac und Unix)
- einheitliche Webserver-Software

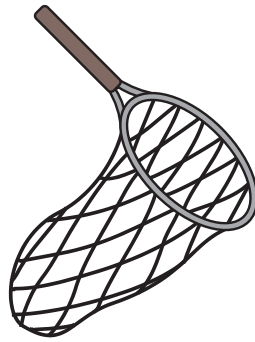
✓ Produktivität

- bekannte Technologie
- portable Entwicklungswerkzeuge
- kostengünstige Entwicklungswerkzeuge

✓ Sicherheit

- Zugriff aus dem öffentlichen Internet nicht möglich
- Lesezugriff auf Data Warehouses genügend sicher

Exkurs - Das Internet verbindet Rechnernetze



Internet :=

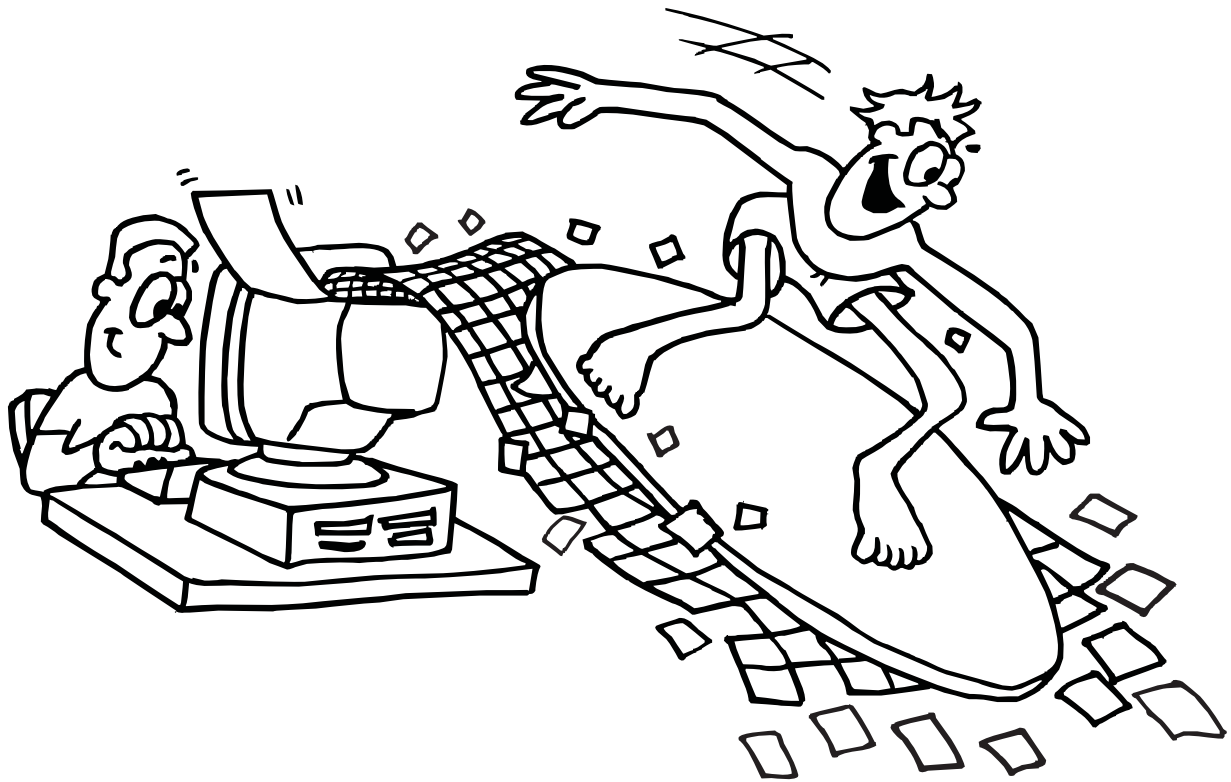
- ✓ weltweites ...
- ✓ öffentliches Netz ...
- ✓ aus Tausenden von Netzen und Millionen von Teilnehmern, ...
- ✓ die fehlertolerant kommunizieren und ...
- ✓ Dienste wie elektronische Post und Webpublikation teilen

Geschichte

- **60er Jahre**
Sichere Kommunikation zwischen Militärs
- **70er und 80er Jahre**
Schnelle Kommunikation zwischen Hochschulangehörigen
- **90er Jahre**
Sichere und schnelle Kommunikation für jedermann

Eine kommerzielle Anwendung

Merrill Lynch Financial Services kommunizieren mit Zehntausenden von Finanzberatern und Kunden über das Internet und sparen so Millionen von Kilo Papier pro Jahr.



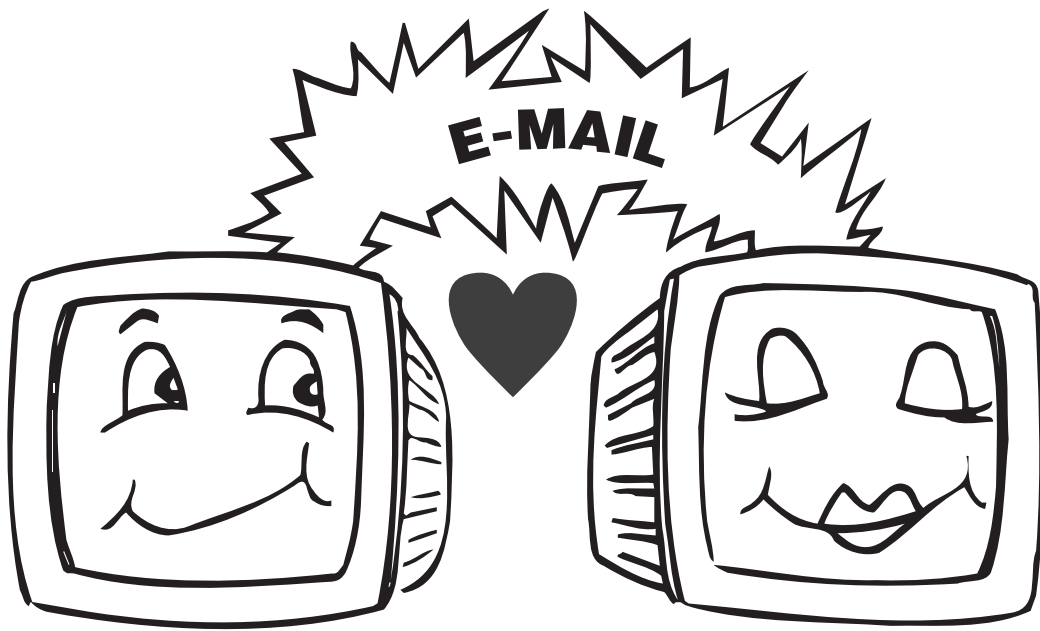
World Wide Web (WWW) :=

- ✓ benutzerfreundliche Oberfläche, ...
- ✓ auf der Endbenutzer hypermedial durch
- ✓ die wichtigsten Internetdienste navigieren

Geschichte

- **1989:** *Konzept* am CERN Genf (Berners-Lee und Cailliau)
- **1993:** *Browser* am nationalen US-Hochleistungsrechenzentrum (Andreessen als Millionär in Jeans und Turnschuhen)

Exkurs - Wichtige Dienste des Internet



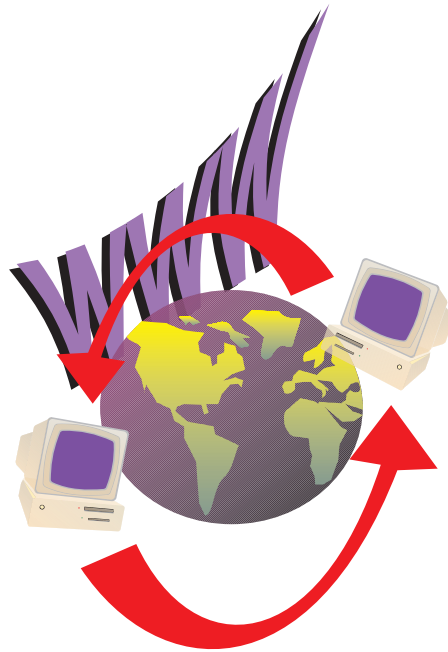
- Elektronische Post (E-Mail)
- Publikation
Bsp. *WWW* (World Wide Web) →
- Diskussion
Bsp. *News Groups*
- Datentransfer
Bsp. *FTP* (File Transfer Protocol) → Data Warehouse
- Bestellung und Bezahlung
Bsp. *Pay-Per-View Publishing*
- ...

Exkurs - Wer regiert das Internet?



Internet Activities Board (IAB)

- Das Internet **gehört** nicht einzelnen Unternehmungen oder Gruppen
- Internetgremien können nicht vorschreiben und sanktionieren sondern nur **empfehlen**
- **Hersteller** dominieren die Internetgremien
- Der **Ausbau** des Internet ist unkoordiniert



Das WWW ist ...

☺ plattformunabhängig

MS Windows, OS/2, MacOS, Unix, ...

☺ benutzerfreundlich

☺ kostengünstig

aber (noch) ...

☹ oft langsam

☹ manchmal unsicher

☹ nicht genügend funktional

Seiten laden und anzeigen

- durch Anklicken von Hyperlinks
- durch Eingabe von URLs

Navigieren

- innerhalb einer Webseite
- über mehrere Webseiten

Seiten durchsuchen

- Volltextsuche auf der laufenden Seite
- Suche über mehrere Webseiten (Suchmaschinen)

Seiten oder Ausschnitte speichern

- Speicherungsformat wählen
(HTML, ASCII/ANSI oder ein Textverarbeitungsformat)
- Gewählte Texte und Grafiken speichern

Browser anpassen, z.B.

- Startseite wählen
- Favoriten speichern

Exkurs - Arbeitsplatz auf dem WWW einrichten

Hardware- und Softwareanforderungen

- PC
- Modem oder LAN-Anschluss
- Anschluss bei einem Internet Service Provider
- Browser

Laufende Kosten

- Grundgebühren je nach Funktionalität
- Benützungsgebühren

Probleme

- Übersicht auf dem WWW (“getting lost in the hyperspace”)
- Überlastung zu Spitzenzeiten
- Urheberrecht, Vertraulichkeit, Sittlichkeit

Exkurs - Mit *analogen* Modem auf das Internet



Modem (Modulator und Demodulator) :=

Gerät, das zwischen *analogen* Telefon- und *digitalen* Computersignalen konvertiert

Modem

- **internes** - (als Karte am Bus)
- **externes** - (über einen Stecker an der seriellen Schnittstelle) ¹

Datenübertragung

- Verbindung **aufnehmen**
- Geschwindigkeit **aushandeln**
- Verbindung **abbrechen**

Anforderungen

- **hohe Geschwindigkeit** (Bsp. 28'800 bps)
- **Faxfähigkeit**
- **Hayes-Kompatibilität**

¹ Vor- und Nachteile interner und externer Modems?

Exkurs - *Digitales* ISDN statt analoges Modem

ISDN (Integrated Services Digital Network) :=

- ✓ öffentliches Netz, das ...
- ✓ verschiedene Telecom-Dienste integriert

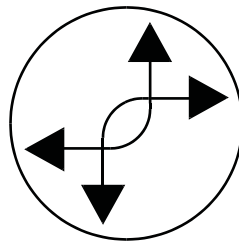
Integrierte Dienste

- ✓ Telefon
- ✓ Übertragung von Computerdaten (Text, Grafik, Video, Audio)
- ✓ Videokonferenz
- ✓ Bildschirmtext
- ✓ Videotex

Exkurs - Internetknoten fehlertolerant verbinden



① Sender K_1 **verpackt**



② Internet **transportiert**



③ Empfänger K_n **entpackt**

- ① K_1 **adressiert** Datenpakete an Empfänger K_n
- ② **Router** verbindet K_1 und K_n über die Knotenrechner $K_2 \dots K_{n-1}$.
(Bei einem Ausfall ändert das Transportprotokoll TCP / IP → den Weg automatisch)
- ③ K_n **empfängt** die Datenpakete



TCP verpackt

(Transmission Control Protocol)

- ① Nachricht in Datenpakete



IP transportiert

(Internet Protocol)

- ② Verbindung wählen
- ③ Geschwindigkeit wählen



TCP entpackt

- ④ Nachricht aus Datenpaketen

Exkurs - Internetadressen

Jeder Internetrechner benötigt eine eindeutige Adresse

(URL, Uniform Resource Locator), zum Beispiel ...

`http: // www.wwz.unibas.ch / wi/eus / home.html`

- ❶ Hypertext Transfer Protocol zwischen Client und Webserver
- ❷ **Domäne** (WWW-Angebot des WWZ der Universität Basel, CH¹)
- ❸ ein **Verzeichnis** (wi/eus) auf dem Webserver
- ❹ Webseite HOME des Typs HTML (Hypertext Markup Language)

¹ Oft steht statt CH die Domäne COM. Was bedeutet .COM ?

Exkurs - Webadressen und -inhalte versenden

Sender einer Webseite

Anwendungsschicht (HTTP-Protokoll)



LAN- oder Modemschnittstelle



Transportschicht (TCP-Protokoll)



Internetschicht (IP-Protokoll)



LAN- oder Modemschnittstelle



Anwendungsschicht (HTTP-Protokoll)

Empfänger der Webseite

Exkurs - Webseiten erstellen

- ① Zielgruppen bestimmen
- ② Inhalt spezifizieren
 - Vorhandene Seiten übernehmen
 - Neue Seiten gestalten
- ③ Betreuer der Web Site bestimmen
- ④ Aufbau der Web Site entwerfen
 - Home Page (Einstiegsseite) und übrige Seiten
 - Lesestrassen
- ⑤ Werkzeuge auswählen
 - Browser → (Bsp. [MS Internet Explorer](#))
 - Konventioneller Editor (Bsp. [MS Office](#))
 - Dedizierter Editor (Bsp. [MS FrontPage](#) →)
- ⑥ Vorlagen erstellen
 - Bildvorlagen (v.a. Symbole und Logos)
 - Seitenvorlagen
- ⑦ Inhalt und Form implementieren und testen
- ⑧ Bei Suchmaschinen anmelden
- ⑨ Feedback ermöglichen

Exkurs - Geschäft auf dem WWW gründen

1. Merkfähige Adresse finden

- Bsp. McDonald

2. Verbindungsart wählen

- Eigene Server-Hardware und -software
Bsp. MS NT Server
- Leitung zum WWW-Server eines Anbieters (engl. Provider)

3. Aktive und passive Dokumente erstellen

- Text-, Bild- und Tonfolgen in HTML
Bsp. MS FrontPage

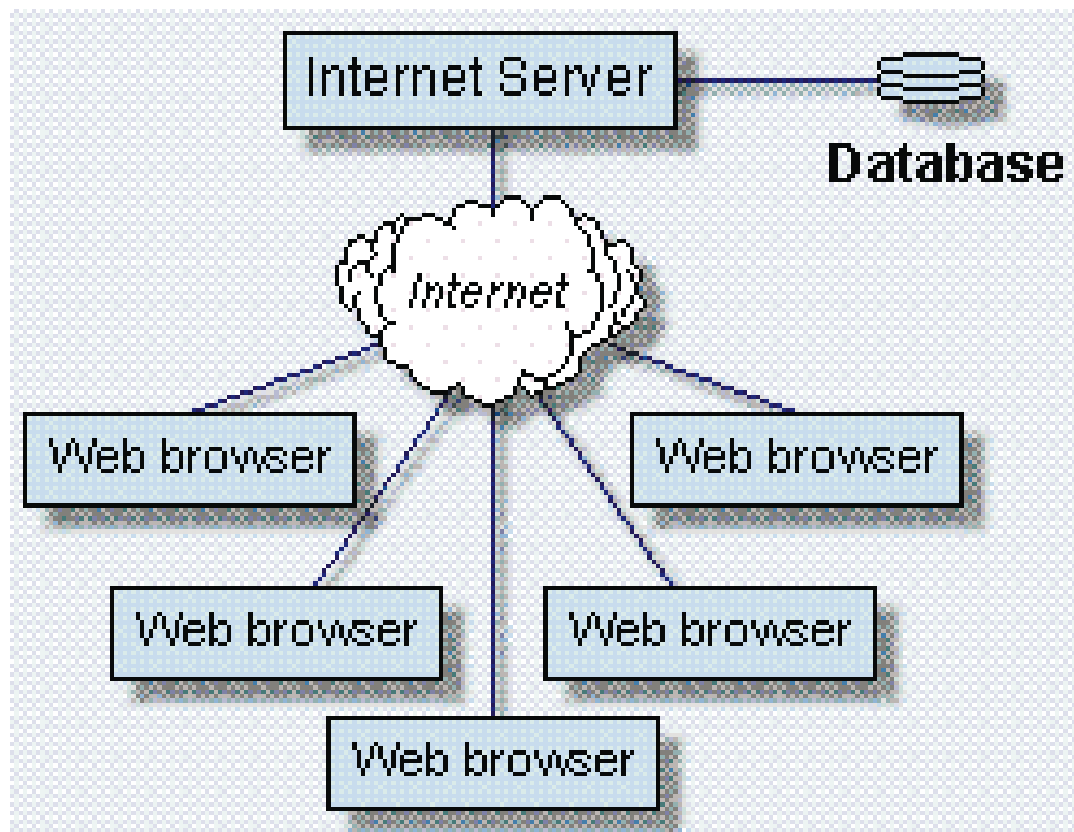
4. Zahlungsverkehr regeln

- Kreditkarte
- e-cash

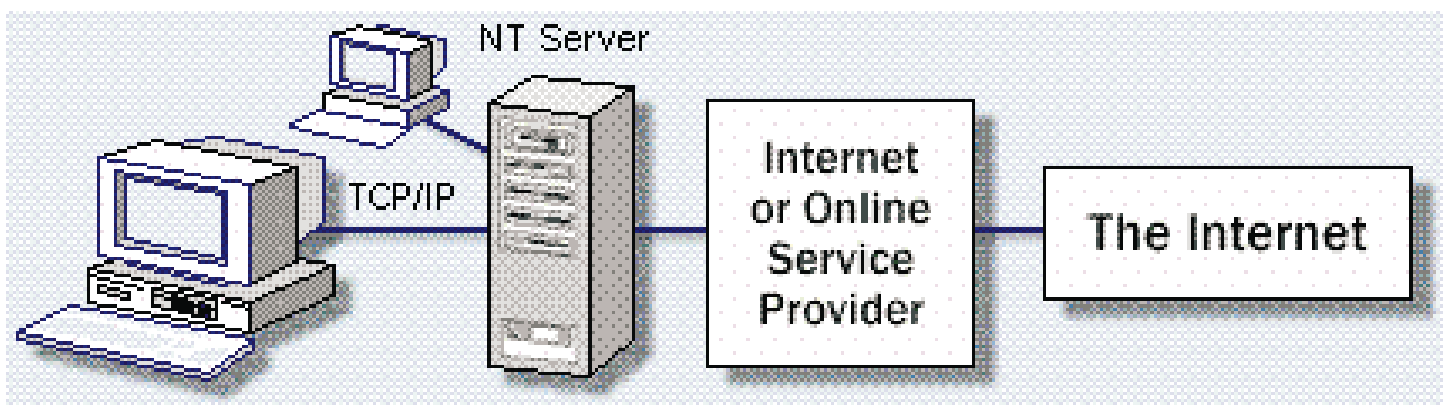
5. Werben

- Home Page bei Suchmaschinen anmelden
Bsp. Alta Vista
- ...

Intranet - Ein Internet für die Unternehmung



Internet



Intranet

Internet

- PC
- Grossrechner-Terminals
- LANs
- Zugang zum Internet

Exkurs - Betriebliche Webanwendungen

Published Information	
<i>Static Information:</i>	
Corporate policies and procedures	Budgeting process
Customer and product information	Bulletin boards
Human resources info. and job listings	Engineering design and manufacturing
Instructional materials	Group communications through e-mail
Research materials and library system access	Group scheduling including mobile workers
Maps and directions	Human resources processes
Org. structures and employee data	Marketing plan
<i>Information supporting ad-hoc analysis:</i>	
Corporate financial information	Expense report process (administrative workflow)
Interactive Business Processes	
Corporate templates	Electronic commerce (e.g. company store) and order tracking
Real time inventory and sales data	Live links between bus. partner systems
Scheduling data	Customer support services
Research data	Help desk services
	Manufacturing inventory system
	Selecting health benefits

Exkurs - HTML als 1. Web-Beschreibungssprache

Hypertext Markup Language (HTML) := geschlossene deklarative Sprache, die vor allem die *Erscheinung* (das Format) und weniger den Inhalt (die Daten) von Webseiten beschreibt

```
<HTML>
  <HEAD>
    <TITLE>Formatierungsbeispiel</TITLE>
  </HEAD>
  <BODY>
    ...
    <P ALIGN="LEFT">
      <FONT COLOR="000000" SIZE="3">...</FONT>
    </P>
    ...
  </BODY>
</HTML>
```


Exkurs - XML - Neuere Beschreibungssprache

EXtensible **M**arkup **L**anguage (XML) := erweiterungsfähige deklarative Sprache, die nicht nur die Form sondern auch den *Inhalt* von Webseiten beschreiben kann

Bsp. Domänenspezifische Beschreibungen

<BESTELLUNG>

<VON>

<PERSON>

<GESCHLECHTSNAME>Müller</GESCHLECHTSNAME>

<VORNAME>Hans</VORNAME>

</PERSON>

</VON>

<PRODUKT>

<GRUPPE>Buch</GRUPPE>

<NAME>Einführung in XML</NAME>

...

</PRODUKT>

...

</BESTELLUNG>

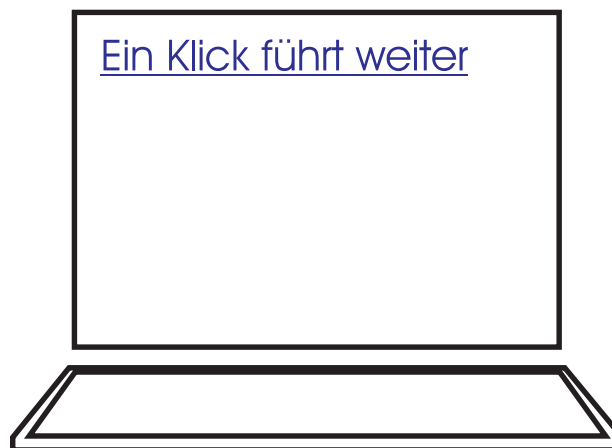
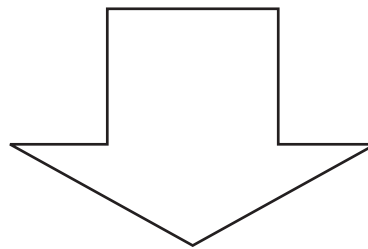
5.77 Spezifikation/Verarbeitung von Webseiten



Dedizierter WYSIWYG-Editor
(z.B. MS FrontPage) formatiert
Text und Bilder (*.gif, *.jpg)
mit *verborgenen* Codes

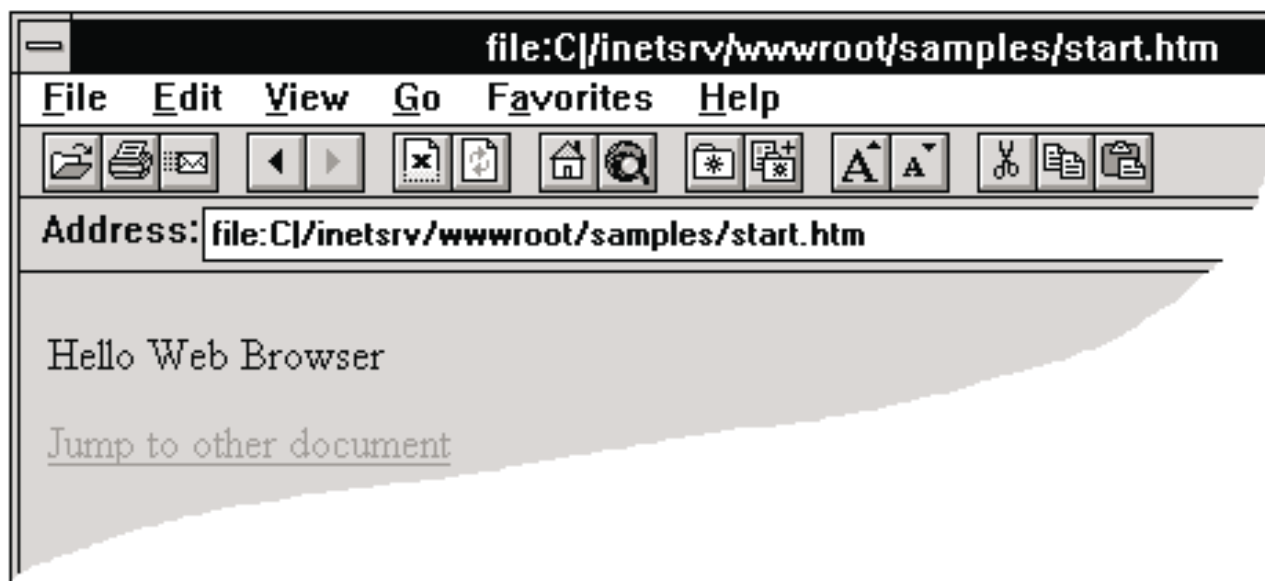


Allgemeiner Texteditor
(z.B. MS Write) formatiert
Text und Bilder (*.gif, *.jpg)
mit *sichtbaren* Codes



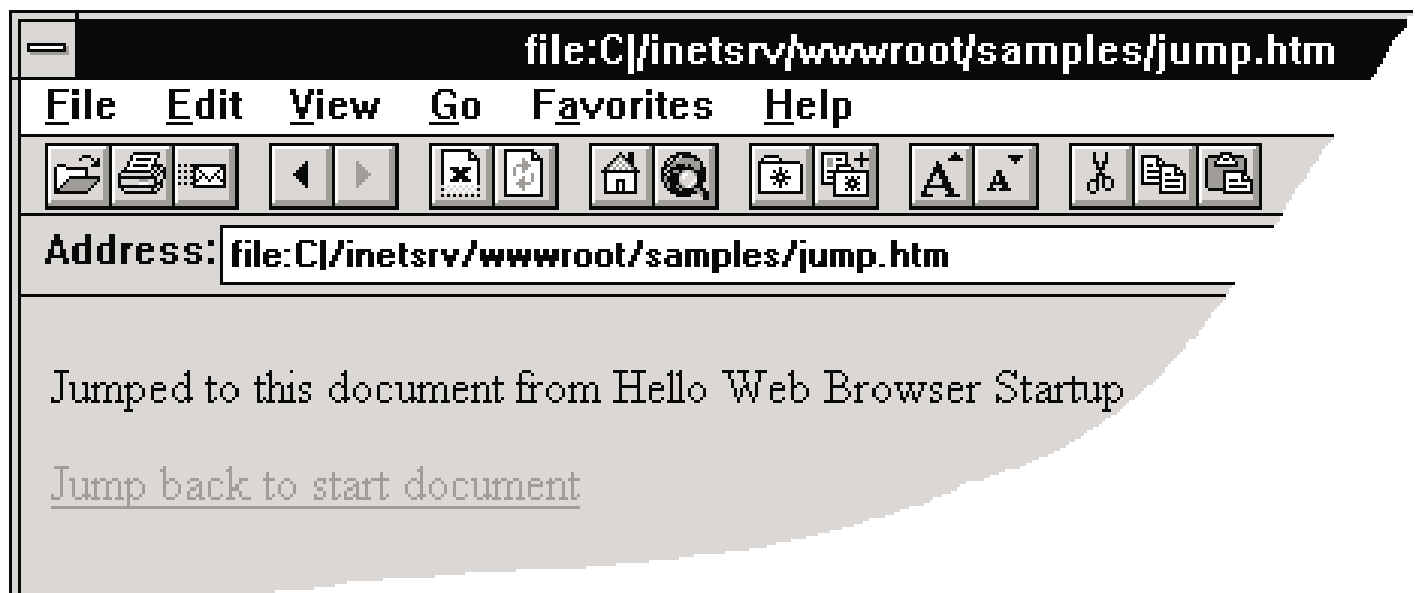
Browser (z.B. MS Internet Explorer oder
Netscape Navigator / Communicator)
analysiert den mit HTML markierten
Text der empfangenen Webseiten mit
einem Parser und interpretiert die
eingebetteten Formatcodes (<...>)

Exkurs - Hyperlinks in HTML darstellen



```
<p>Hello Web Browser Startup</p>  
<A HREF="jump.htm">Jump to other document</A>
```

START.HTM document with HTML source code.



```
<p>Jumped to this document from Hello Web Browser Startup</p>  
<A HREF="start.htm">Jump back to start document</A>
```

JUMP.HTM document with HTML source code

Datenzugriff über das Web - Überblick

① HTML

Browser interpretiert in den Text eingebettete ▶HTML-Befehle

⇒ Netz stark beansprucht

⇒ Interaktivität stark eingeschränkt

② HTML und Ergänzungen

Clientseitige Ergänzungen

a) Browser interpretiert zusätzlich eingebettete Skriptbefehle

Bsp. ▶JavaScript, ▶VBScript

b) Browser ruft Bibliotheksfunktionen auf

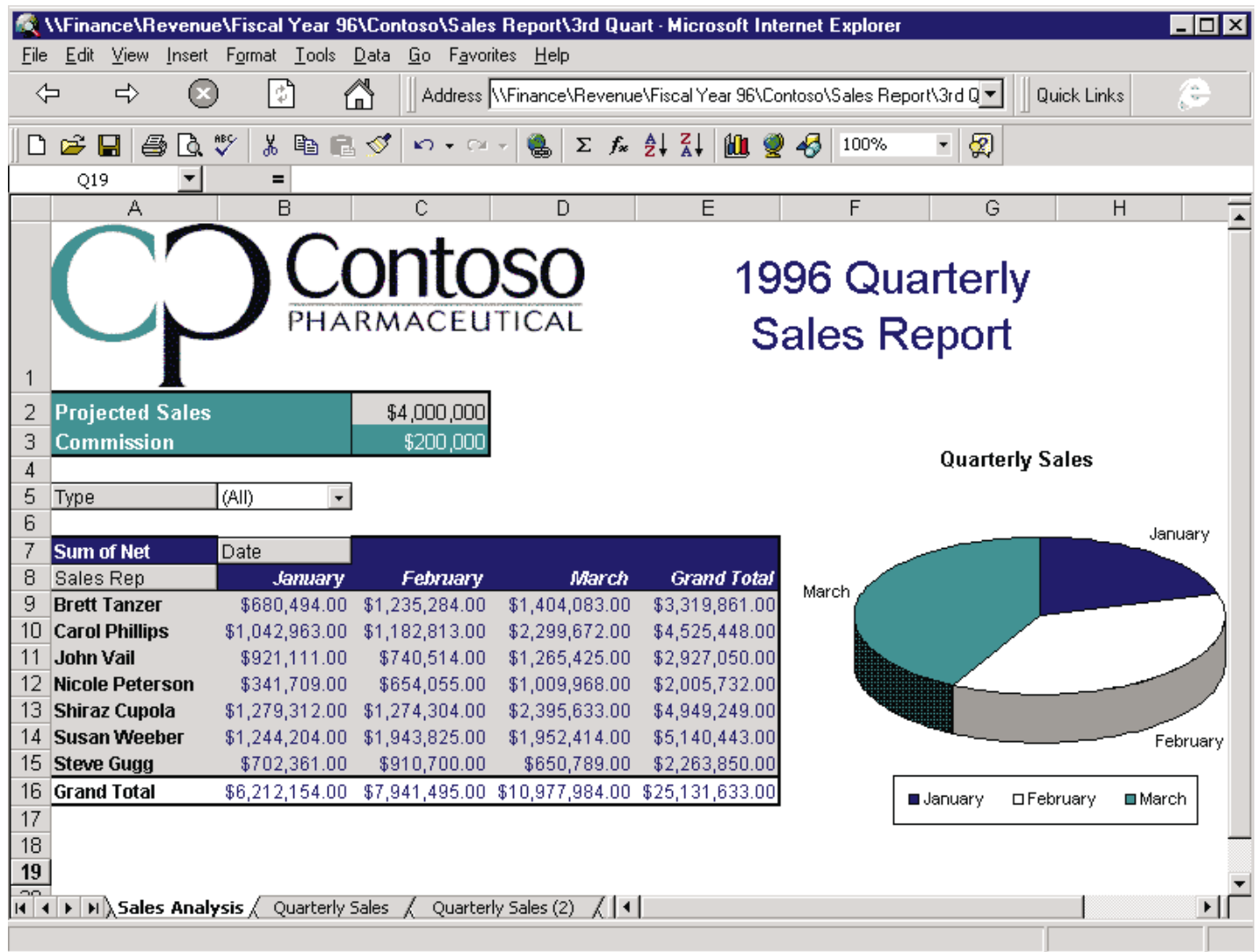
Bsp. ▶ActiveX, Applets

Serverseitige Ergänzungen

Datenbankserver interpretiert Befehle einer ...

- *portablen* Schnittstelle des Webserver
▶CGI (langsam)
- *prioprietären* Schnittstelle des Webserver
▶ISAPI, ▶NSAPI, ... (schneller)

Interaktive statt passive Webseiten



Arbeitsmappe aus MS Excel

Standard-Steuerelemente von HTML

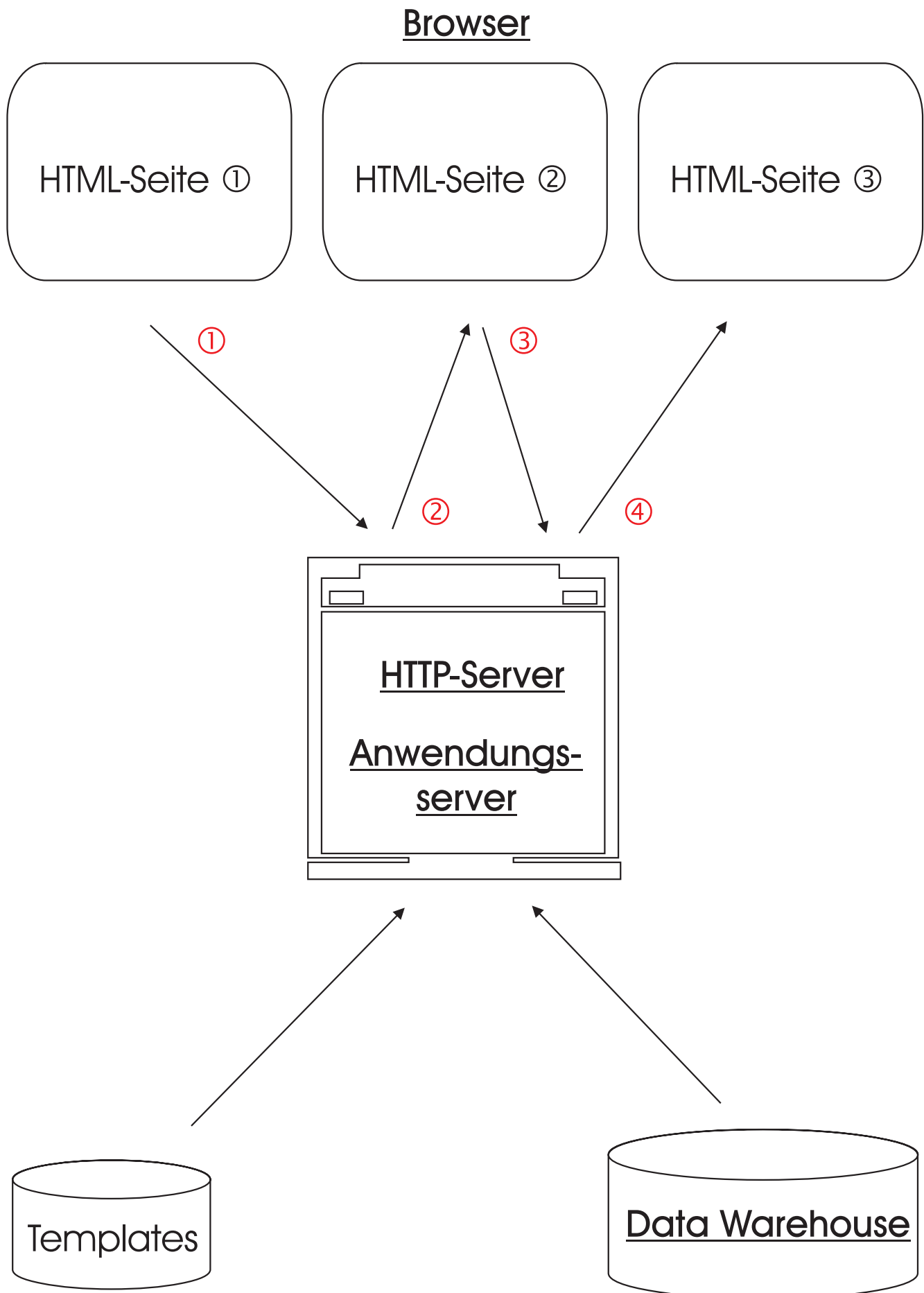


Grenzen von HTML

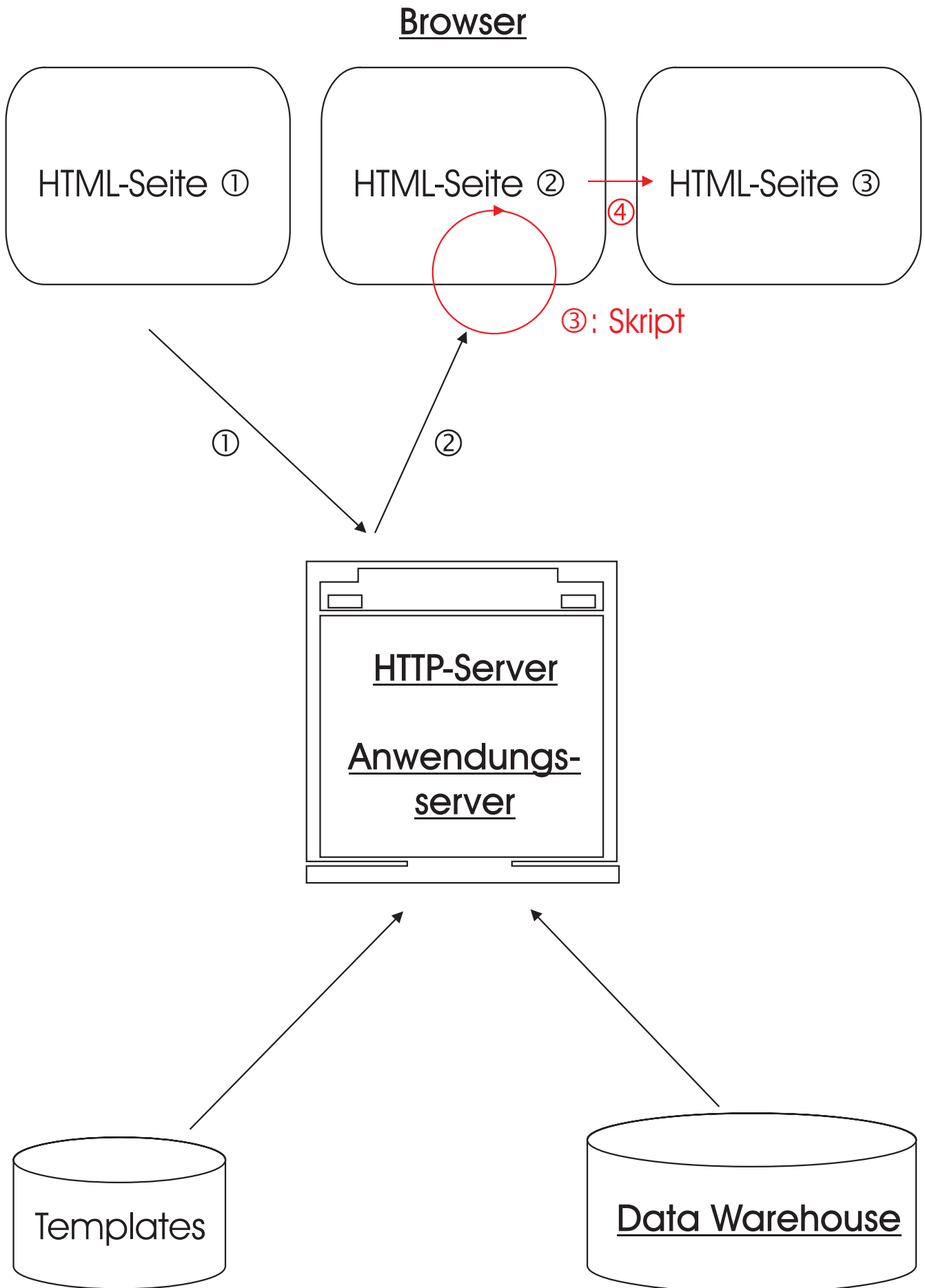


Ergänzungen von HTML machen
Intranet-Warehousing erst möglich

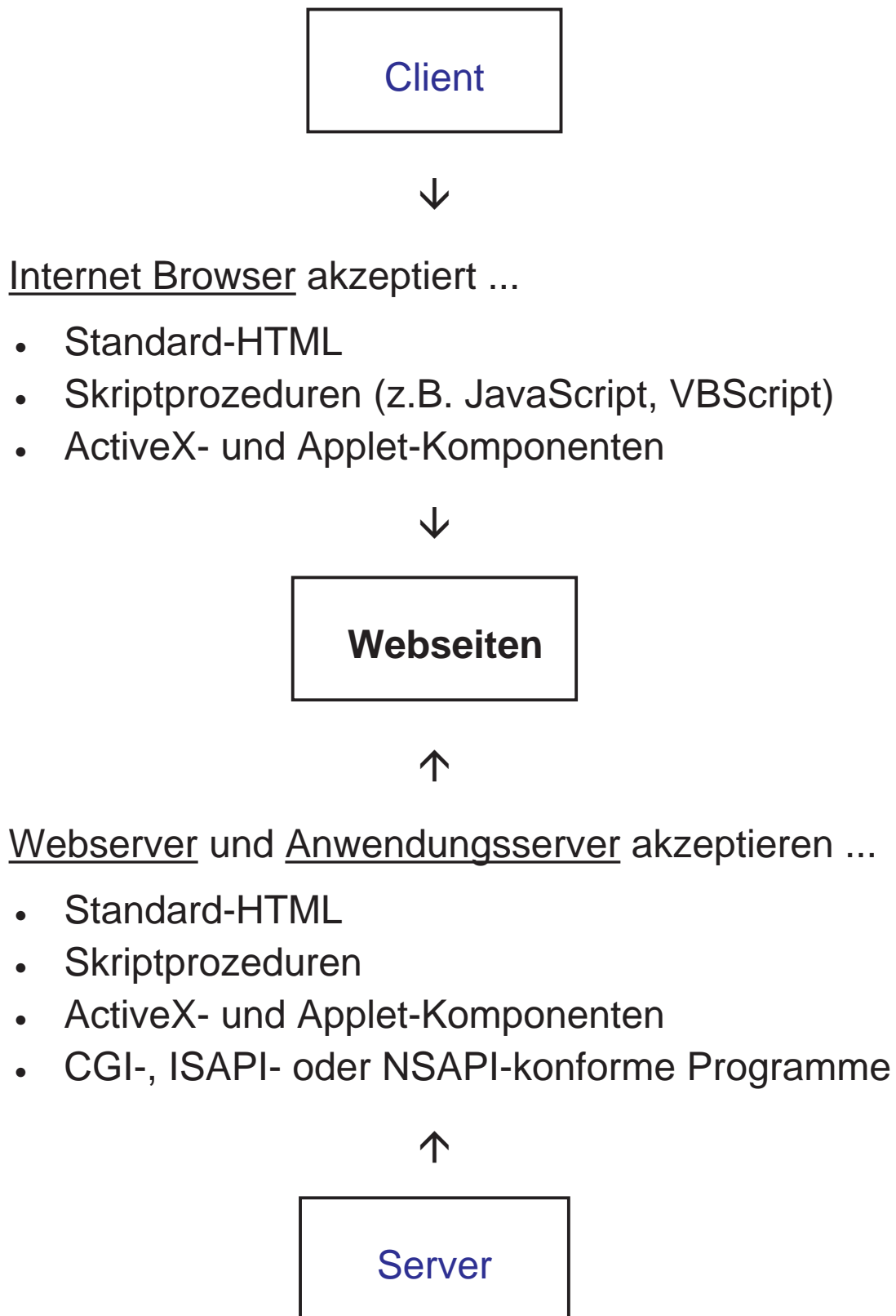
5.78 Interaktivität mit Standard-HTML



5.79 Interaktivität mit HTML und Skriptsprachen



Client- und serverseitige HTML-Ergänzungen



Steuerelemente in HTML integrieren

Steuerelemente (controls)

mit vordefinierten Ereignissen

Bsp. *Edit Box*, *Schaltfläche*

Werkzeuge

zur Integration von Web-Steuerelementen

① *Standard HTML* zur Verarbeitung *einfacher* Ereignisse

⇒ HTML-Befehle einfügen

② *Skriptsprachen* zur Verarbeitung *komplexer* Ereignisse

⇒ Skriptbefehle an Client oder Server in Webseite einbetten

Bsp. JavaScript, VBScript (Teilmenge von Visual Basic)

③ *Sprachen* zur *anspruchsvollen* Programmierung

⇒ Programm als Modul zur Verfügung stellen

Bsp. Java-Programm als Applet zur Verfügung stellen

Bsp. DLL¹ als ActiveX-Komponente zur Verfügung stellen

¹ Dynamic Link Library

VBScript in eine HTML-Seite einbetten

```
<SCRIPT LANGUAGE="VBScript">  
  'Initialisierung  
  ...  
  'Prozedur mit Steuerelement MsgBox  
Sub Hallo_Klick()  
  MsgBox "Hallo!"  
End Sub  
  ...  
</SCRIPT>
```

ActiveX aus einer HTML-Seite aufrufen

Aufruf einer ›ActiveX-DLL

<OBJECT

ID="SteuerelementXY"

CLASSID="clsid:123-1234-123-123-"

CODEBASE="http://MyServ/MyDLL.OCX">

</OBJECT>

Ausschnitt aus der ActiveX-DLL *MyDLL.OCX*

```
Sub antworte(Abfrage As String, Antwort as
String)
    'Abfrage interpretieren und verarbeiten
    ...
    Antwort =
        "Content-Type: text/html
        <HTML><BODY>
            'Antwort zeigen
            ' (d.h. folgendes Template instanzieren)
            ...
        </BODY></HTML>"
End Sub
```

Serverprogramme ergänzen HTML

Webserver-Programme ...

ergänzen HTML, z.B. bei der

- Interpretation von Datenbankabfragen
- Indizierung von Webseiten

steuern den Webserver über Schnittstellen, z.B. ...

- **CGI** (Common Gateway Interface für Serverprogramme)
 - ☺ portabel
 - ☹ weniger effizient (Aufspannen eines neuen Prozesses)
- **ISAPI** (Internet Server Application Programming Interface)
 - ☹ proprietär (Microsoft)
 - ☺ effizienter

sind geschrieben in ...

- einer Skriptsprache wie ›JavaScript oder ›VBScript
- einer konventionellen Programmiersprache wie C

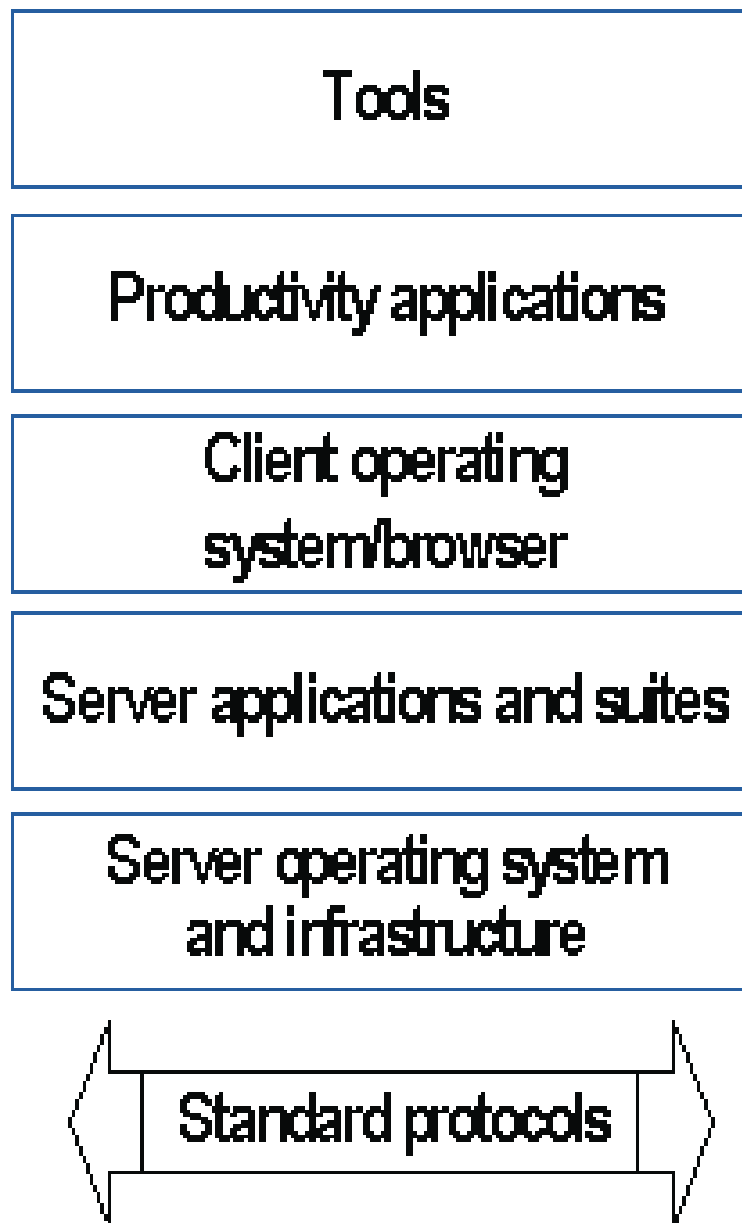
Verbesserungen des HTML-Konzepts

	<i>HTML ohne Scripts</i>	<i>Neue Konzepte</i>
<i>Sprachumfang</i>	vordefiniert	erweiterbar
<i>Syntaxkontrolle</i>	schwach	streng
<i>Strukturierbarkeit</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>Hyperlink-Adressen</i>	lokal und absolut	auch zentral und symbolisch
<i>Sonderzeichen</i>	schlecht unterstützt	gut unterstützt
<i>Trennung von Format und Inhalt</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>Formatierung</i>	aufwendig und fix	Stylesheets
<i>Suche</i>	ungenau/aufwendig, weil Inhalt schlecht beschreibbar	direkter, weil Inhalt gut beschreibbar
<i>Datenaustausch</i>	Tags formatbezogen	Tags inhaltsbezogen (Datenfelder)
<i>Änderbarkeit von Seiten durch den Client</i>	Format und Inhalt statisch → Netzverkehr gross	Format und Inhalt dynamisch → Netzverkehr kleiner
<i>Mehrfachverwendung (z.B. CD-ROM, Druck)</i>	Reformatierung aufwendig	Stammdokument mit Stylesheets
<i>Objektorientierung</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Beispiele von Neuentwicklungen

XML (eXtensible Markup Language),
DHTML (Dynamic HTML), DOM (Document Object Model)

Aufgabe Softwareklassen auf dem Web

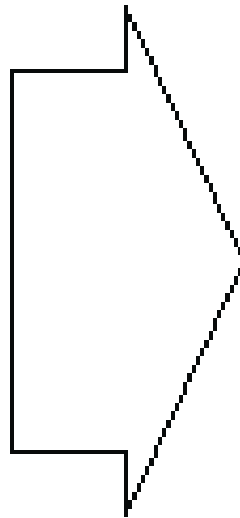


Ordnen Sie die Software der richtigen Protokollebene zu

- a) MS Windows NT Server
- b) TCP/IP
- c) MS Exchange Server
- d) MS Internet Explorer
- e) MS SQL Server
- f) MS Office
- g) ActiveX - Steuerelemente

Content Publishing

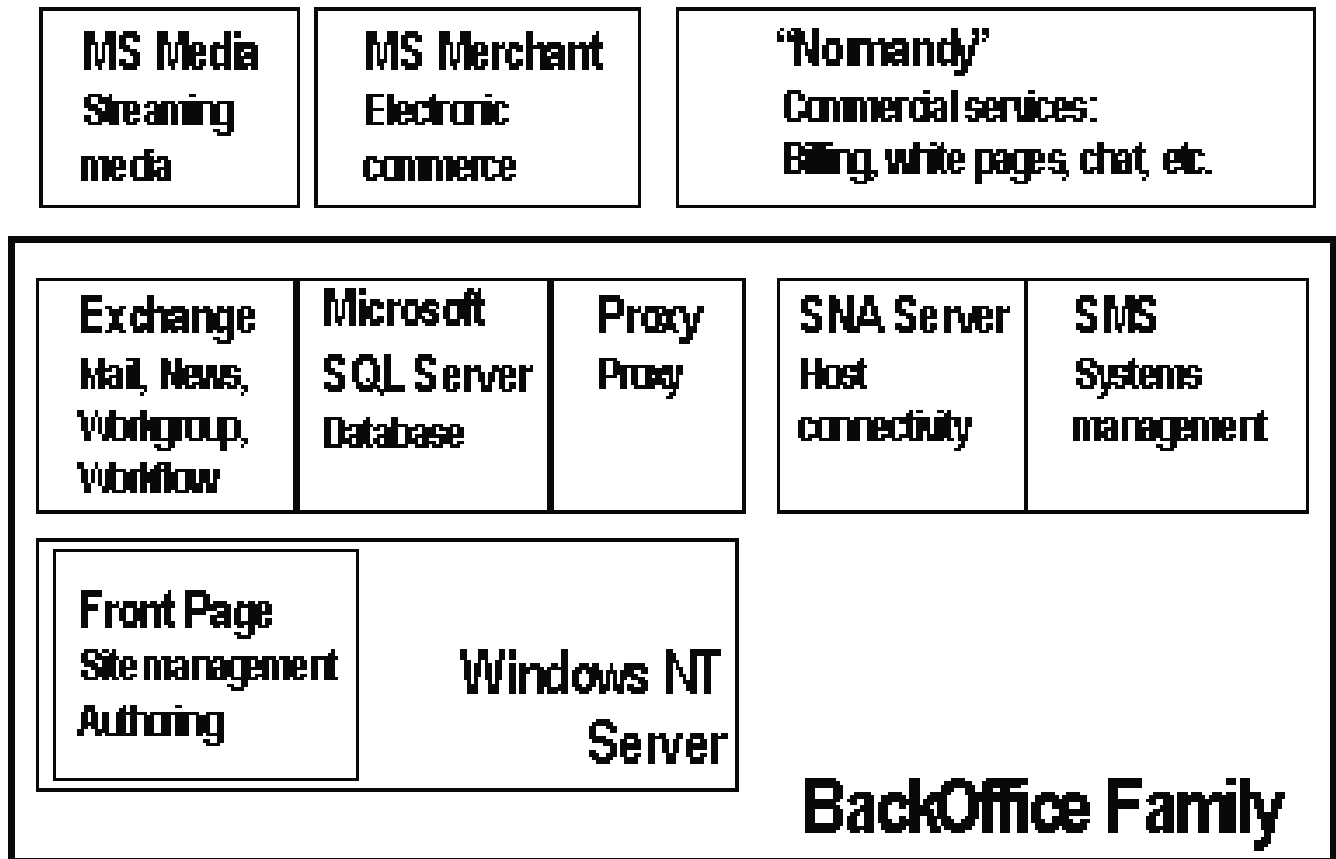
- HR handbook
- Real estate advertising
- Events calendar
- Hard goods catalog
- Advertising-sponsored content
- Flight schedules
- Mutual fund advertising and prospectus



Business Applications

- HR benefits management
- Appointment scheduling, mortgage applications
- Online ticketing
- Order capture and fulfillment
- Fee-based content (micropayments)
- Reservations and ticketing
- Portfolio management

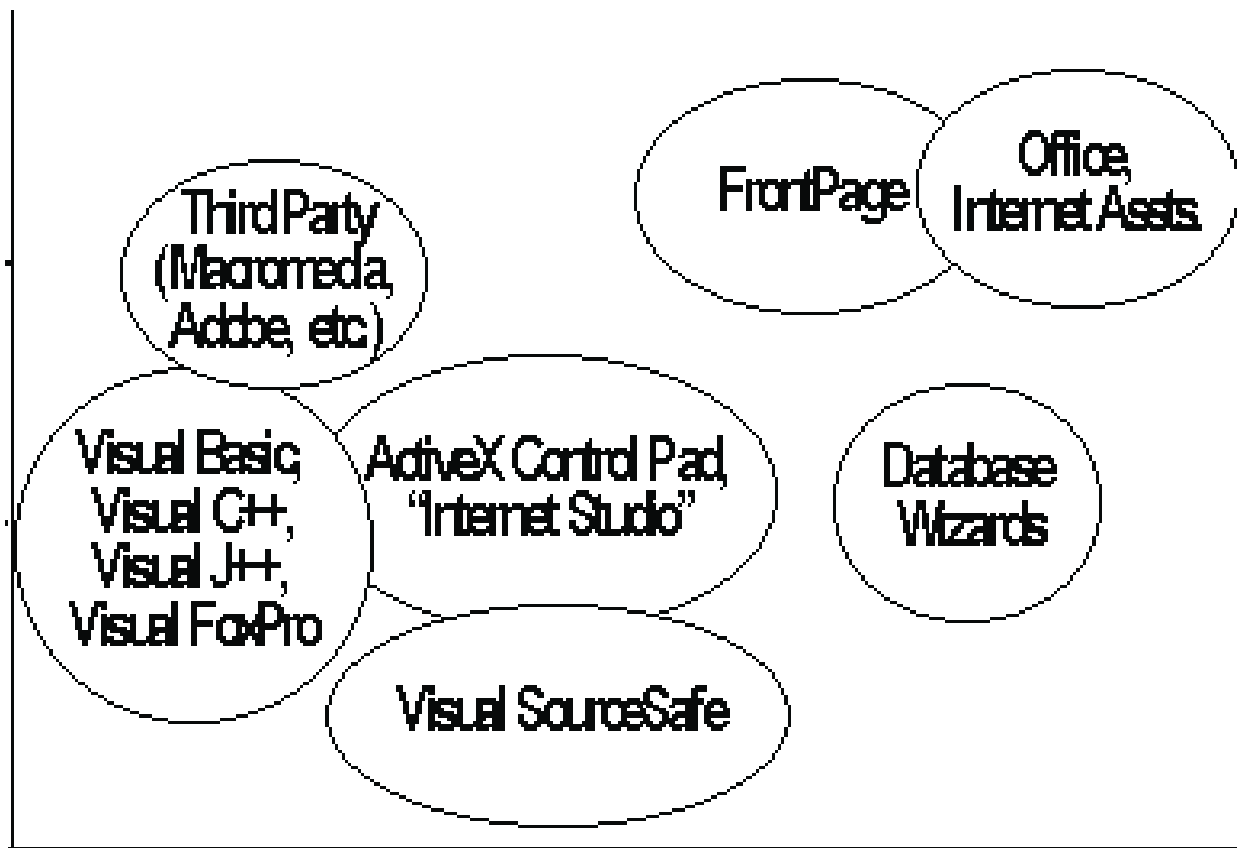
Inter- und Intranet - *MS BackOffice*



Ordnen Sie zu

- a) Sicheren Internetzugang gewährleisten
- b) Elektronische Post verwalten
- c) Webseiten erstellen verwalten
- d) Elektronischen Handel unterstützen
- e) Multimediale Internet-Anwendungen unterstützen
- f) Professionelle Internet-Verleger unterstützen
- g) IBM-Welt einbinden
- h) Netzbetriebssystem
- i) Server-Datenbanken verwalten
- k) Netzressourcen verwalten

Entwicklungswerkzeuge



Ordnen Sie zu

- a) Endbenutzerwerkzeuge
- b) Programmierwerkzeuge
- c) Datenbankwerkzeuge
- d) Sicherheitsmanagement
- e) Rich Content ?
- f) ›ActiveX Control ?

Entwicklungstechnologien von Microsoft

ActiveX :=

- ✓ Binärcode-Modul (▷DLL), das ...
- ✓ mit anderen Anwendungen “zusammengesteckt” wird und
- ✓ klein genug für die Übertragung auf dem Internet ist

Im Gegensatz zu einem ▷Applet sind ActiveX-Komponenten Binärcode-Module und für die Microsoft-Betriebssysteme optimiert

Aufgabenbeispiele

- Verhalten des Internet Browsers ergänzen
- Neues Steuerelement auf einer Webseite einsetzen
- Internet Browser aus einer anderen Anwendung manipulieren

ISAPI (Internet Server Application Programming Interface) :=

- ✓ ▷proprietäre
- ✓ ▷API-Schnittstelle zwischen einem
- ✓ Microsoft-Webserver und einer
- ✓ Applikation, die auf einen Datenbankserver oder
- ✓ eine andere ▷Serverapplikation zugreift

Aufgabenbeispiel

- Server-Datenbank auf dem WWW publizieren

5.81 Mehrstufiger Zugriff auf Web DW

① Browser (Client)
liest ein Formular oder
eine frei formulierte Abfrage

Antwort in HTML →



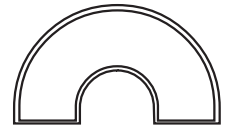
② Web Server (HTTP Server)
erkennt im HTML-Strom die Abfrage

Abfrage in HTML



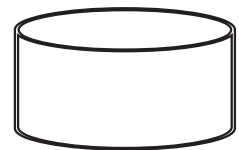
③ Anwendungsserver
übersetzt die HTML- in eine SQL-Abfrage

Abfrage in SQL →



④ Data Warehouse Server
führt die Datenbankabfrage aus

Ergebnis in einer Variable



③ Anwendungsserver erstellt mit dem
Abfrageergebnis ein HTML-Dokument

Ergebnis in HTML



② Web Server leitet das HTML-Abfrageergebnis weiter



① Browser interpretiert das HTML-
Dokument und zeigt es an →



Ein Template in erweitertem HTML

Fakten

Dimensionen

```
<HTML>
  <TITLE>Monatsverkäufe</TITLE>
  <BODY>
    <!--SQL
      SELECT PRODUKT.Name,
      SUM ( VERKAEUFE.$Umsatzart )
      FROM VERKAEUFE, PRODUKTE, FILIALEN, PERIODEN
      WHERE PERIODEN.Monat = VERKAEUFE.Monat
      GROUP BY PRODUKT.Name
      ORDER BY PRODUKT.Name
    -->
    <!--Formatierung des SQL-Ergebnisses in einer Tabelle>
    ...
  </BODY>
</HTML>
```

- ① Der Webserver extrahiert die SQL-Abfrage aus der Webseite
- ② Der Anwendungsserver interpretiert die SQL-Abfrage

Die Variable \$Umsatzart wird vom Anwendungsserver mit einem im Abfrageformular gewählten Wert - zum Beispiel *Geldumsatz* oder *Mengenumsatz* - instanziiert

Die Komplexität einer Abfrage kann von einer einfachen Menüabfrage bis zu jener von QBE reichen

5.81 HTML-Template aufrufen und instanzieren

① Der Benutzer füllt ein Abfrageformular aus. Der **Browser** sendet die Benutzereinträge 1...N als Teil der folgenden URL an den Web Server: *http://<Web Site>/<Templatename><Var1, Wert>...<VarN, Wert>*. Der Anwendungsserver wird später mit den Paaren <Var, Wert> ein SQL-Template instanzieren.



② Der **Web Server** leitet die *URL* an den Anwendungsserver weiter

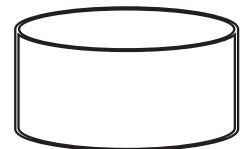


③ Der **Anwendungsserver** ...

- a) isoliert die Paare <Var, Wert>
- b) sucht in der Metadatenbank das passende SQL-Template
- c) instanziiert dessen Variablen mit den entsprechenden Werten
- d) sendet die SQL-Abfrage an den Data Warehouse Server



④ Der **Data Warehouse-Server** führt die SQL-Abfrage aus und sendet das *Ergebnis* dem Anwendungsserver



③ Der **Anwendungsserver** instanziiert das Template mit dem SQL-Ergebnis und sendet das *HTML-Ergebnis* an den Web Server



② Der **Web Server** leitet das *HTML-Ergebnis* an den Browser des Benutzers weiter



① Der **Browser** zeigt das *HTML-Ergebnis* meist in Form einer Tabelle an



Internet-Adressen

Produktunabhängige Information

<http://www.dwinfocenter.org/>

(The Data Warehousing Information Center)

White Papers, Literaturverweise, Software

<http://www.dw-institute.com> (Data Warehousing Institute)

Material aus der Sicht einer Beratungsunternehmung

<http://www.itmanagement.earthweb.com/>

Beiträge der Zeitschrift *Data at o* , insbesondere über Software

<http://altaplana.com/olap/> (Alta Plana Online Analytical Processing)

Überblick, insbesondere OLAP

Produktübersichten

<http://www.olapreport.com> (The OLAP Report)

umfangreiche Marktübersicht (Verweise auf Anbieter finden Sie auf den betreffenden Folien)

Produktunabhängige News Groups

<http://www.winf.ruhr-uni-bochum.de/olap/>

Verweise auf White Papers, Forschungsinstitute, Unternehmungen und Produkte zu Warehousing, OLAP und Data Mining

<news:comp.databases.olap>

 [Web Quiz](#)

Folienverzeichnis (Ein Klick führt zur gewünschten Folie)

Data Warehousing und Data Mining	2
Einordnung	3
Unterrichtsmaterial	4
Daten- und modellgetriebene Methoden	5
Grundlagen	6
Data Warehouses	7
Data Warehouses und Data Mining	8
Exkurs - Operative und strategische Planung	9
Produktionsdatenbank	10
Mängel von Produktionsdatenbanken	11
Eine mögliche Definition	12
5.1 Operative vs. analytische Datenbanken	13
Adressaten	14
Gründe für zwei Buzzwords ...	15
5.3 Eine einfache Architektur	16
5.2 HANDEL - Fallbeispiel	17
Data Warehouses sind speicheraufwendig	18
Entwicklung	19
5.4 HANDEL - Eine Entwicklungsskizze	20
Data Warehouses verschiedener Grösse	21
Data Mart	22
Ziele modularisierter Data Marts	23
Enterprise Data Warehouse	24

5.6 Zentralisierte EDW-Architektur	25
5.7 Hierarchische EDW-Architektur (HEDW)	26
5.10 Koordinierte EDW-Architektur (KEDW)	27
Architekturen	28
Data Mart-Koordination - Voraussetzungen	29
5.9 Data Mart-Koordination - Abfragebeispiel	30
Anforderungen an Attribute	31
Aufgabe <i>Anforderungen an Attribute</i>	32
5.8 Data Mart vs. Enterprise Data Warehouse	33
5.5 Data Mart-Entwicklung	34
EDW und Data Mart - Ein Kontinuum	35
Beurteilung von Data Marts	36
5.12 Betriebsnähe durch Indikator u. Dimensionen	37
Würfel - Begriffe	38
Einen Indikator nach seinen Dimensionen messen	39
HANDEL - Fakten nach Dimensionen auswerten	40
Indikator versus Dimensionen	41
Aufgabe <i>Fakt oder Dimension (A 5.1)</i>	42
Mehrdimensionale Daten speichern	43
Relationale versus mehrdimensionale DB	44
Anforderungen an RDBMS	45
Grundlagen	50
5.15 Zugriff auf operative und analytische Daten	51
Datenbankabfragen	52

Abfragesprachen	53
5.16 Prozedurale und deklarative Abfragen	54
5.17 Query by Example (QBE)	55
SQL - Grobübersicht	56
5.18 SQL - Teilsprachen	57
SQL - Vorteile	58
Mängel von SQL	59
5.19 Abfragearten	60
Endbenutzerzugriff mit OLAP - Definition	61
5.20 OLAP versus OLTP	62
Anwendungen	63
Definitionselemente	64
📌 ANLAGEBERATUNG - Fallbeispiel	65
📌 Einfache Darstellung von Ergebnissen	66
Importierte und abgeleitete Daten	67
Zusammenfassungen und Vergleiche	68
Zusammenfassung - Ein Beispiel	69
Zusammenfassungen können verfälschen	70
Browsing	71
5.21 📌 Filtern beschränkt die angezeigten Werte	72
Drilling Up and Down	73
5.21 📌 Drilling Down - Ausgangslage	74
5.22 📌 Drilling Down - Ergebnis	75
📌 Drilling navigiert in Kategorienhierarchien	76

Dimensionen mit Kategorien beschreiben	77
Drilling Down kann verfälschen	78
5.23 † Slicing und Dicing analysiert / synthetisiert	79
† Pivoting tauscht Zeilen und Spalten	80
5.24 † Eine Ebene ist eine Zusatzdimension	81
5.25 † Die Visualisierung hebt hervor	82
OLAP-Arten - Definitionen	83
OLAP-Arten - Vergleich	84
5.28 OLAP-Architekturen	85
5.29 OLAP-Architekturen und Produkte	86
DOLAP-Frontend Cognos <i>PowerPlay</i>	87
5.30 Entwicklung eines DOLAP-Würfels	88
🔗 ANLAGEBERATUNG mit <i>Cognos PowerPlay</i> (5.2)	89
Funktionalität von <i>PowerPlay</i>	93
① Ausgewählte Kriterien für OLAP-Clients	94
② Ausgewählte Kriterien für OLAP-Server	96
OLAP-Produkte und Hersteller	98
5.31 Methode im Vergleich	99
Modellierung relationaler Data Warehouses	100
Modellierung von Informationssystemen	101
Operative und analytische Datenmodelle	102
Unwünschbare und wünschbare Redundanz	103
Data Warehouses führen zu Redundanz	104
Ziele der Datenmodellierung	105

5.32 Ein operatives Unternehmungsdatenmodell	106
5.34 Ein operatives Bereichsmodell <i>Fakturierung</i>	107
Data Warehouse-Modellierung	108
Transformationen des operativen Modells	109
5.35 Operatives Bereichs- → partielles DW-Modell	110
5.36 Vom partiellen DW-Modell zum EDW-Modell	111
5.37 Logische Modellierung im Vergleich	112
5.38 Sternschemata für relationalen Data Mart	113
HANDEL - Betriebliche Anforderungen sammeln	114
5.40/5.41 Anforderungsdiagramm	115
5.42 Ein Sternschema für HANDEL	116
5.43 Mehrere Faktzeilen pro Dimensionszeile	117
Sternschema-Tabellen für HANDEL	118
5.44 Sternschema allgemein	119
5.45 Sternschemata filtern Abfragen ...	120
LIEFERFRIST - Sternschema (A 5.3)	121
VERKAUF - EDW-Modell und Sternschema (A 5.4)	122
Relationales OLAP erfordert Sternschemata	124
OLAP-Abfragen auf einem Sternschema	125
Eine SQL-Abfrage auf einem Sternschema	126
5.48 Sternschemata durch Denormalisierung	127
Stern- versus normalisiertes Schema I	128
Stern- versus normalisierte Schemata II	129
Vom normalisierten zum Sternschema	130

<u>Erweiterte Sternschemata</u>	131
<u>Beurteilung erweiterter Sternschemata</u>	132
<u>5.49 Erweitertes Sternschema - Normalisierung</u>	133
<u>Erweitertes Sternschema - m:n-Beziehung (A 5.6)</u>	134
<u>KONTO - Auflösung einer m:n-Beziehung (A 5.7)</u>	135
<u>Metadaten</u>	136
<u>5.52 Einige Klassen von Metadaten</u>	137
<u>Vorkommen von Metadaten</u>	138
<u>Data Dictionary</u>	139
<u>Data Dictionary-System</u>	140
<u>5.53 Data Dictionary-System</u>	141
<u>Anforderungen an DD-Systeme</u>	142
<u>Beispiele von DD-Systemen</u>	143
<u>ROLAP-Frontend <i>if...Synchrony</i></u>	144
<u>Architektur von <i>Synchrony</i></u>	145
<u>Wichtige Metadaten von <i>Synchrony</i></u>	146
<u>Funktionalität von <i>Synchrony</i></u>	147
<u>🖱 EINZELHANDEL mit <i>if...Synchrony</i> (A 5.8)</u>	148
<u>Entwicklung und Betrieb</u>	152
<u>Von Produktionsdatenbanken zum EDW</u>	153
<u>Data Warehouse-Entwicklungszyklus</u>	154
<u>Phasenunabhängige Tätigkeiten</u>	155
<u>5.54 Data Warehouse-Ziele</u>	156
<u>Endbenutzerbefragung - Ablauf</u>	157

<u>Endbenutzerbefragung - Inhalte</u>	158
<u>Spezifikationsdokument</u>	159
<u>Entwicklungsprobleme</u>	160
<u>Rollen der DW-Entwicklung und -Verwaltung</u>	161
<u>Rollen der DW-Entwicklung (A 5.8)</u>	162
<u>Software für Data Warehouses</u>	163
<u>Eine integrierte Softwarelösung</u>	164
<u>Implementation und Betrieb</u>	165
<u>5.58 Operative Daten laden</u>	166
<u>ZEITSCHRIFTEN - 1. Fallbeispiel</u>	167
<u>2. Daten auswählen</u>	168
<u>3. Daten transformieren</u>	169
<u>4. Daten transformieren und analysieren</u>	170
<u>Funktionalität von Ladeoperationen</u>	171
<u>Beispiel Validitätsprüfungen</u>	173
<u>Ladesoftware</u>	174
<u>Einige Ladeprodukte</u>	175
<u>Physische Modellierung</u>	176
<u>Performanceprobleme</u>	177
<u>Performanceproblem Aggregation</u>	178
<u>Performanceproblem Sternschema</u>	179
<u>5.59 Performanceproblem Mehrdimensionalität I</u>	180
<u>Performanceproblem Mehrdimensionalität II</u>	181
<u>Data Warehouse-Performance</u>	182

<u>Datenmodell-Performance verbessern</u>	183
<u>Optimierungsmassnahme Denormalisierung</u>	184
<u>5.61 Normalisierung verlangsamt Abfragen</u>	185
<u>5.62 Denormalisierung beschleunigt</u>	186
<u>Optimierungsmassnahme Partitionierung</u>	187
<u>Partitionierungsbeispiele</u>	188
<u>5.63 Spaltenpartitionierung</u>	189
<u>Optimierungsmassnahme Indexierung</u>	190
<u>5.64 Optimierungsmassnahme <i>Bitmuster</i></u>	191
<u>Komprimierte Bitmusterindizes</u>	192
<u>Wann Bitmuster-Indizes?</u>	193
<u>Kosten und Nutzen der Bitmuster-Indexierung</u>	194
<u>Performancedeterminanten Hardware</u>	195
<u>Hardwareperformance verbessern</u>	196
<u>Komponentenperformance verbessern</u>	197
<u>Optimierungsmassnahme Parallelisierung</u>	198
<u>Data Warehouse-Parallelisierung</u>	199
<u>5.65 Beispiel eines Data Warehouse-MPS</u>	200
<u>Arten von Mehrprozessorsystemen</u>	201
<u>5.66 Symmetrisches MPS (SMP)</u>	202
<u>Eine <i>mittlere</i> SMP-Konfiguration (1998)</u>	203
<u>5.67 Clustering</u>	204
<u>5.68 Massiv paralleles MPS</u>	205
<u>5.70 NUMA (Non Uniform Memory Access)</u>	206

<u>5.69 Symmetrische vs. massiv parallele MPS</u>	<u>207</u>
<u>Parallelisierungsverantwortliche</u>	<u>208</u>
<u>Parallelisierungsvarianten</u>	<u>209</u>
<u>Datenparallelisierung</u>	<u>210</u>
<u>Partitionierung für die Datenparallelisierung</u>	<u>211</u>
<u>Partitionierungskriterien</u>	<u>212</u>
<u>Trade Offs einiger Optimierungsmassnahmen</u>	<u>213</u>
<u>Netzzugriff</u>	<u>214</u>
<u>5.71 Vom Host zum Client/Server-System</u>	<u>215</u>
<u>Rechnernetze</u>	<u>216</u>
<u>Rechnernetze teilen Ressourcen</u>	<u>217</u>
<u>Rechnernetze für Data Warehouses</u>	<u>218</u>
<u>Client/Server-Systeme</u>	<u>219</u>
<u>5.72 Alternative Modelle der Aufgabenteilung</u>	<u>220</u>
<u>5.73 Client/Server-Architekturen</u>	<u>221</u>
<u>Zweistufige Client/Server-Architekturen</u>	<u>222</u>
<u>5.74 Zweistufige Data Warehouse-Architektur</u>	<u>223</u>
<u>Dreistufige Client/Server-Architektur</u>	<u>224</u>
<u>5.75 Dreistufige Data Warehouse-Architektur</u>	<u>225</u>
<u>Dreistufige ROLAP- und MOLAP-Architektur</u>	<u>226</u>
<u>Clientzugriff auf Data Warehouses</u>	<u>227</u>
<u>Endbenutzerzugriff auf Data Warehouses</u>	<u>228</u>
<u>Endbenutzerzugriff über Intranet und Internet</u>	<u>229</u>
<u>Intranet für Data Warehouses</u>	<u>230</u>

<u>Exkurs - Das Internet verbindet Rechnernetze</u>	<u>231</u>
<u>Exkurs - WWW - Populärster Internet-Aspekt</u>	<u>232</u>
<u>Exkurs - Wichtige Dienste des Internet</u>	<u>233</u>
<u>Exkurs - Wer regiert das Internet?</u>	<u>234</u>
<u>Exkurs - Beurteilung des WWW</u>	<u>235</u>
<u>Exkurs - Endbenutzerwerkzeug <i>Browser</i></u>	<u>236</u>
<u>Exkurs - Arbeitsplatz auf dem WWW einrichten</u>	<u>237</u>
<u>Exkurs - Mit <i>analogen</i> Modem auf das Internet</u>	<u>238</u>
<u>Exkurs - <i>Digitales</i> ISDN statt analoges Modem</u>	<u>239</u>
<u>Exkurs - Internetknoten fehlertolerant verbinden</u>	<u>240</u>
<u>Exkurs - Der Transporteur TCP / IP</u>	<u>241</u>
<u>Exkurs - Internetadressen</u>	<u>242</u>
<u>Exkurs - Webadressen und -inhalte versenden</u>	<u>243</u>
<u>Exkurs - Webseiten erstellen</u>	<u>244</u>
<u>Exkurs - Geschäft auf dem WWW gründen</u>	<u>245</u>
<u>Intranet - Ein Internet für die Unternehmung</u>	<u>246</u>
<u>Exkurs - Betriebliche Webanwendungen</u>	<u>247</u>
<u>Exkurs - HTML als 1. Web-Beschreibungssprache</u>	<u>248</u>
<u>Exkurs - XML - Neuere Beschreibungssprache</u>	<u>249</u>
<u>5.77 Spezifikation/Verarbeitung von Webseiten</u>	<u>250</u>
<u>Exkurs - Hyperlinks in HTML darstellen</u>	<u>251</u>
<u>Datenzugriff über das Web - Überblick</u>	<u>252</u>
<u>Interaktive statt passive Webseiten</u>	<u>253</u>
<u>5.78 Interaktivität mit Standard-HTML</u>	<u>254</u>

<u>5.79 Interaktivität mit HTML und Skriptsprachen</u>	<u>255</u>
<u>Client- und serverseitige HTML-Ergänzungen</u>	<u>256</u>
<u>Steuerelemente in HTML integrieren</u>	<u>257</u>
<u>VBScript in eine HTML-Seite einbetten</u>	<u>258</u>
<u>ActiveX aus einer HTML-Seite aufrufen</u>	<u>259</u>
<u>Serverprogramme ergänzen HTML</u>	<u>260</u>
<u>Verbesserungen des HTML-Konzepts</u>	<u>261</u>
<u>Aufgabe Softwareklassen auf dem Web</u>	<u>262</u>
<u>Internettrends</u>	<u>263</u>
<u>Inter- und Intranet - <i>MS BackOffice</i></u>	<u>264</u>
<u>Entwicklungswerkzeuge</u>	<u>265</u>
<u>Entwicklungstechnologien von Microsoft</u>	<u>266</u>
<u>5.81 Mehrstufiger Zugriff auf Web DW</u>	<u>267</u>
<u>Ein Template in erweitertem HTML</u>	<u>268</u>
<u>5.81 HTML-Template aufrufen und instanzieren</u>	<u>269</u>
<u>Internet-Adressen</u>	<u>270</u>