

Einführung in Data Warehousing

Projektpraktikum im betrieblichen Umfeld 4.0 PR (WI/PPR) WS 2003 / 2004

Beate List

list@wit.tuwien.ac.at

Institut für Softwaretechnik & Interaktive Systeme
Wissenschaftlerinnenkolleg Internettechnologien
Technische Universität Wien

Data Warehouse - Agenda

- Einführung
- DWH Charakteristik
- DWH Architektur
- DWH Typen
- DWH Modellierung

Data Warehouse ?

„Single Source of Information“

Interessante Ergebnisse

Entscheidungsunterstützung

Unkomplizierter Zugriff

Riesige Datenmengen

Motivation: Data Warehouses

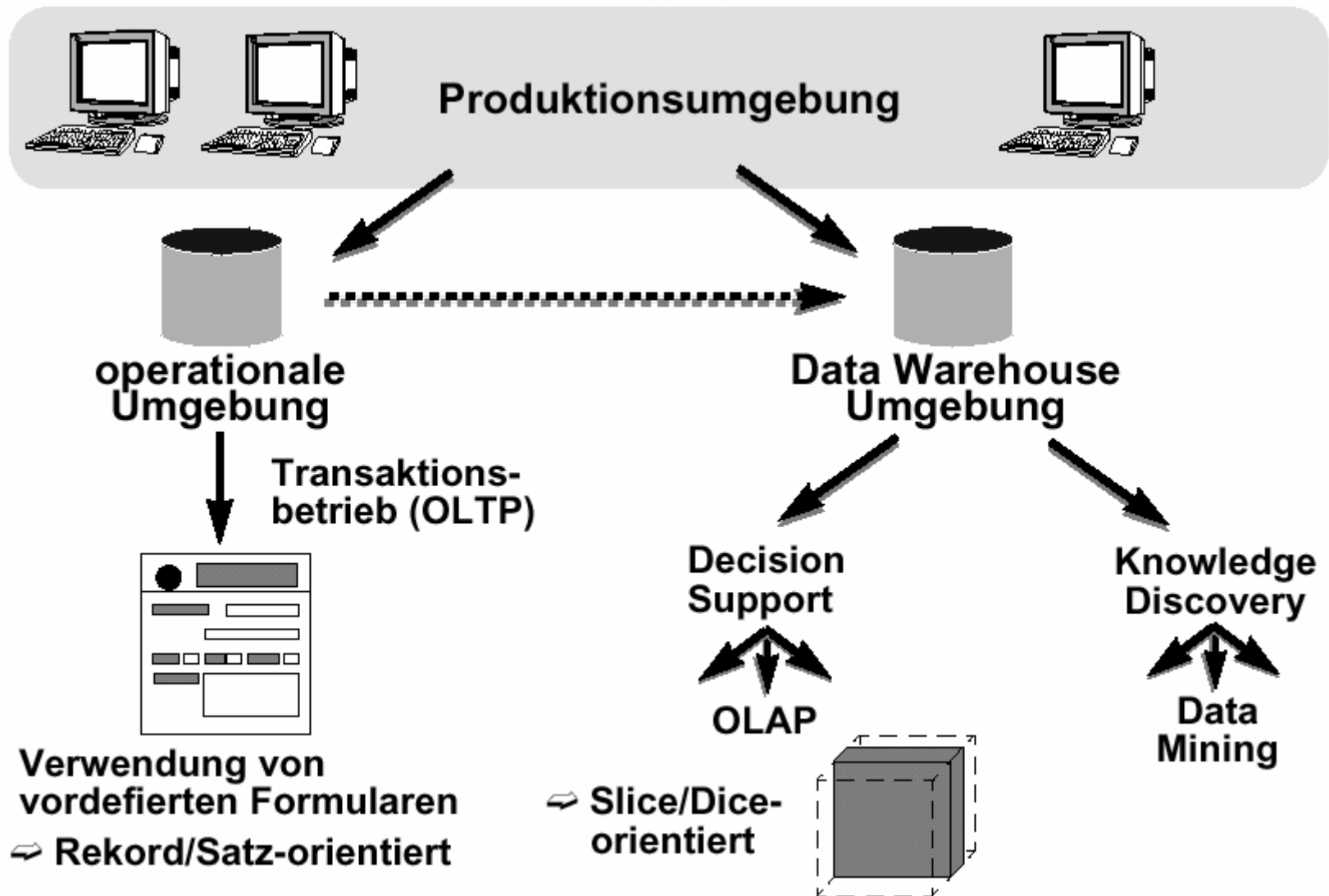
- Heterogene Daten integrieren und anreichern, um sie einer homogenen Analyse zuzuführen, die Informationen für (strategische) Entscheidungen liefert.

- Wettbewerb
 - Globale Märkte, schnelle Veränderungen
 - Information wird immer wichtiger für Unternehmen ⇔ Informationsüberflutung
 - Entscheidungsträger benötigen last-minute hoch-aggregierte Informationen
 - Kunden flexibler
 - Kürzere decision times basierend auf globaler (integrierter) Information
 - Mehr Wissen über den Kunden (CRM – Customer Relationship Management)

Motivation: Data Warehouses (2)

- Technologie
 - Effiziente Datenbanktechnologien verfügbar
 - Verteilte Architekturen
 - Hardware / Storage immer leistungsfähiger und billiger
 - User-Interfaces werden besser → Nutzung der IT-Infrastruktur durch mehr Personen
- Trends
 - OLAP (On-Line Analytical Processing)
 - Intranet
 - Internet / WWW
 - Data Mining
 - Knowledge Discovery in Databases (KDD)

Beispiel für Umgebung eines DWH

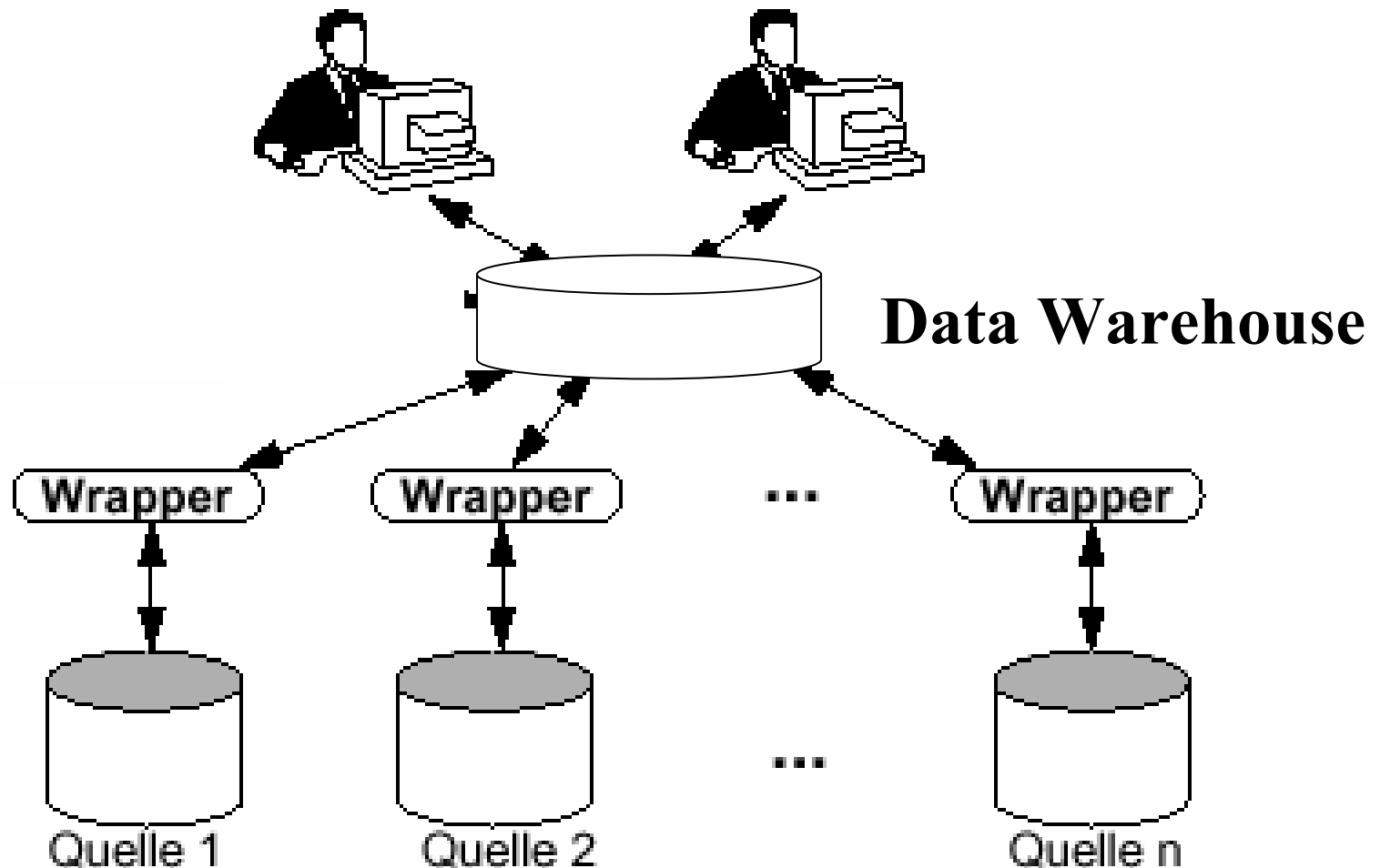


Definition nach Inmon

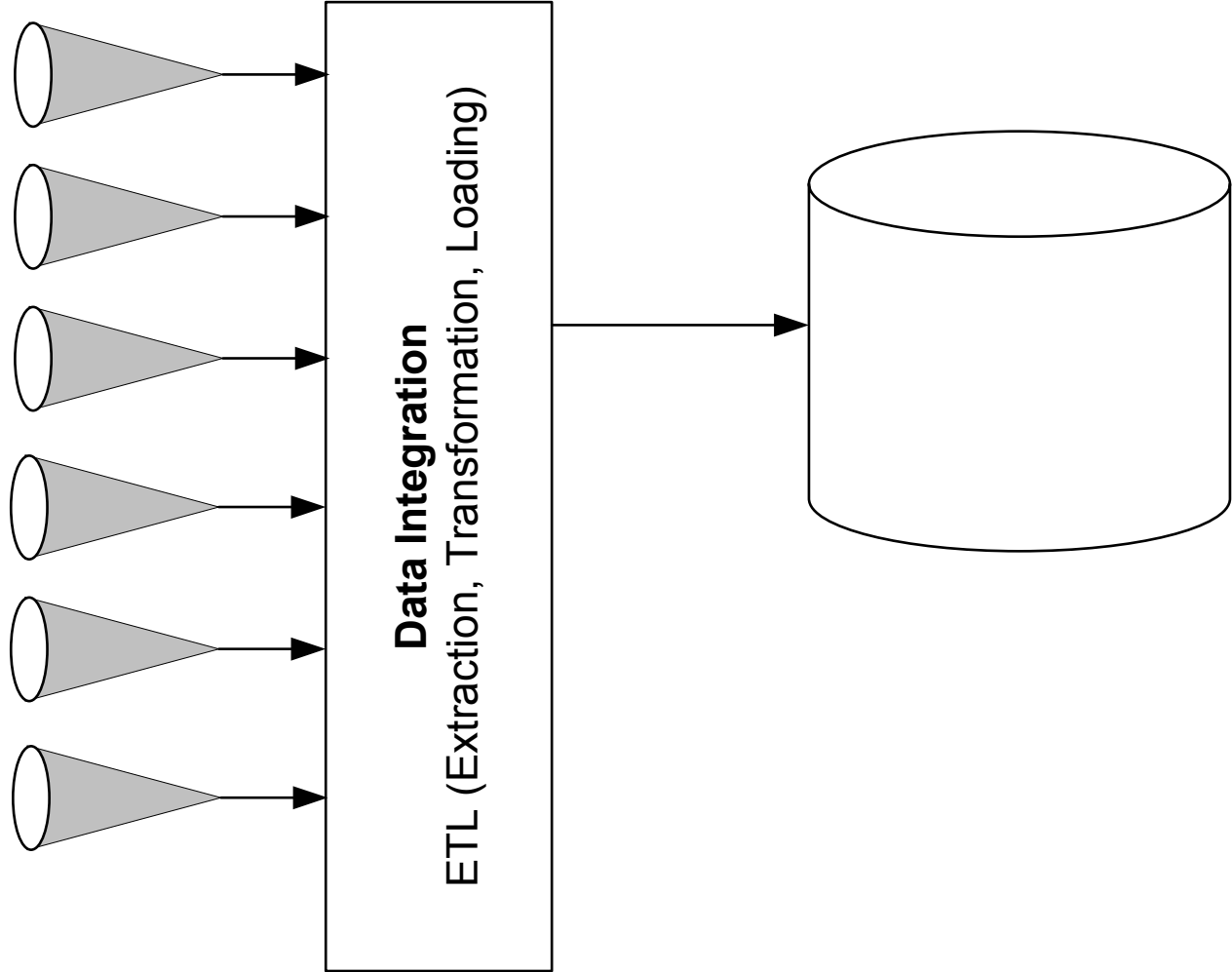
„A Data Warehouse is a subject-oriented, integrated, non-volatile, time-variant collection of data organized to support management needs“

- **themenorientiert**
alle Daten über ein Subjekt (z.B. Kunde) zentral gespeichert und nicht in einzelnen Applikationen “versteckt”
- **integriert**
Daten aus heterogenen Systemen und Formaten einheitlich in gemeinsamen Schema gespeichert und von Inkonsistenzen bereinigt
- **historisch**
Daten (Snapshots) über lange Zeit (5-10 Jahre) gespeichert, um Zeitanalysen vornehmen zu können
- **nicht flüchtig**
Das DWH wird periodisch aktualisiert; einmal eingebrachte Daten dürfen nicht verändert werden

Data Warehouse Architektur



Data Warehouse Architektur



Weitere Definitionen für DWH

A Data Warehouse is:

- a method for combining information from a number of different systems
- a central access point for an organization's data
- populated by data from otherwise incompatible information systems
- presented in a business relevant format
- **"read-only" information**, which means you can't affect the information sources
- enables reporting across many of these host applications

(Pandora, Swinburn University)

DWH Charakteristik

- Design für analytische Aufgaben (d.h. komplexe Abfragen)
 - Relativ kleine Anzahl an Benutzern
 - Lese-intensiver Zugriff
 - Periodische Updates (Hinzufügen von Daten)
 - Historisierung der Daten
 - Abfragen liefern umfangreiche Antwortmengen
 - Globale Sicht – Abteilungsübergreifend
- ... collection of data in support of management's decision making process.

Datenhaltung – Divergierende Ziele

OLTP (Online Transactional Processing)

- Technologie: Relationale DBMS
- Transaktionsorientiert
- Möglichst redundanzfrei, hochnormalisiert
- Fokus auf aktuelle Daten
- Abfragen liefern kleine Ergebnismengen
- Transaktionen: Integrität, Security, Concurrency, Locking
- Transaktionsdurchsatz
- Hoch-Verfügbarkeit (24x7)
- Workload auf gleichmäßig hohem Level

Data Warehouse / OLAP (Online Analytical Processing)

- RDBMS, MDBMS (multidimensionale DBMS)
- Teilweise hoch-redundant
- Vorberechnete Aggregationen
- Historische Daten / Analysen
- Analysen sehr komplex
- Gesamtintegration verschiedenster Systeme
- Themenorientiert
- Sehr hohe Datenvolumina
- Geringere Anforderungen an Verfügbarkeit
- Spitzenlasten, ad-hoc Abfragen

Datenhaltung – Divergierende Ziele (2)

OLTP (Online Transactional Processing)

- Hohe Zahl an parallelen Transaktionen
- Updates sehr häufig
- Zeitliche „Inkonsistenzen“: d.h. durch Daten-Updates kann es passieren, dass dieselbe Abfrage unterschiedliche Resultate zu unterschiedlichen Zeitpunkten liefert.
Bsp: Relation Part-Supplier
- Abfragen sind oftmals vorher bekannt und optimiert

Data Warehouse / OLAP (Online Analytical Processing)

- Antwortzeit ist nicht der primäre Fokus für Abfragen
- Keine Updates (traditionell) oder Deletes von DWH Daten
- Historisierung der Daten → im allgemeinen zeitlich konsistente Ergebnisse
- Neue Daten aus den Vorsystemen (DB, Legacy Systeme, sonstige Datenquellen, auch externe Datenquellen – z.B. Ortsdatenbank) werden zu vordefinierten Zeitpunkten (traditionell) integriert
- Ad-hoc Abfragen

Abgrenzung OLTP – DWH/OLAP

Operationale Systeme

- Schnelle Antwortzeit
- Applikationsorientiert
- Aktuelle Daten
- Daten in hoher Granularität (Detailierungsgrad)
- Häufige Updates
- Viele (kurze) Transaktionen zur Abwicklung des Tagesgeschäft
- Überwiegend konstante, hohe Last
- Oftmals isolierte Systeme

Informationssysteme (DWH/OLAP)

- Hohes Datenvolumen
- Themenorientiert (z.B. Kundensicht)
- Historische Daten
- Teilweise Vor-aggregierte Daten, konsolidierte Daten
- Keine Updates
- Komplexe Analysen für Entscheidungsunterstützung
- Spitzenlasten und geringe Last
- Integration von heterogenen Vorsystemen

Abgrenzung OLTP – DWH/OLAP

Charakteristika	OLTP	DWH/OLAP
Benutzertyp	Angestellter, IT-Professional	Manager, Decision Support
Benutzeranzahl	Hoch (Tausende), Concurrency	Wenige Benutzer
Antwortzeit	Sekunden	Sekunden – Minuten
Anwendung	Verwaltung, operatives Geschäft	Analyse und Entscheidungsunterstützung
Fokus	Dateneingabe	Informationsgewinn
Auswertungsmethode	Datensatz-orientiert	Multidimensional
Transaktionsart	Kurze Lese-/Schreibtransaktionen	Lange Lesetransaktionen
Funktion	Tägliche Operationen	Wissensgenerierung
Einheit	Einfache Transaktionen	Komplexe Anfragen
Ergebnismenge	Klein	Sehr groß
Anfragetyp	Strukturiert, vordefiniert	Ad-hoc

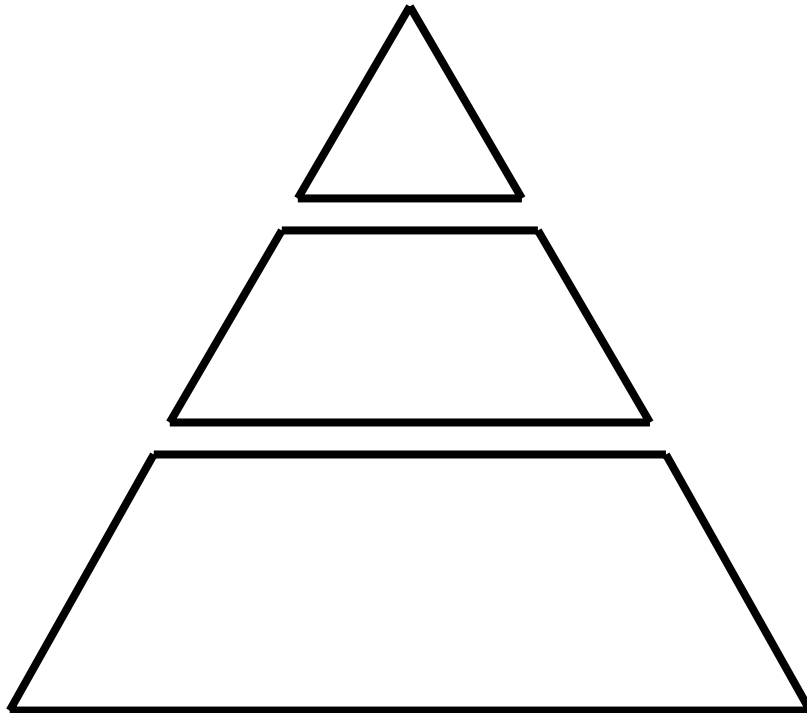
Abgrenzung OLTP – DWH/OLAP

Charakteristika	OLTP	DWH/OLAP
Datenverwaltungsziel	Transaktionale Konsistenzerhaltung	Zeitbasierte Versionierung
Daten	Aktuell, atomar, isoliert, dynamisch	Historisch, konsolidiert, stabil, aggregiert, integriert, multidimensional
Datenvolumen	MB – GB	GB – TB
Design	Anwendungsorientiert (ER basierend), hochnormalisiert	Subject-oriented (Star- /Snowflake), teilweise hochnormalisiert
Datensicht	Detailliert, komplex, relational	Einfach, summiert, multidimensional
Typische Operationen	Index/Hashzugriff auf Primärschlüssel	(Full-Table) Scans
Anpassungsfähigkeit	Begrenzt	Anwendungsabhängig

Gründe für ein separates DWH

- Konzeptionelle Differenzen im Datenbestand und dessen Nutzung
- Performance
 - OLTP optimiert für kurze Transaktionen und bekannte Lastprofile
 - komplexe OLAP-Anfragen degradieren die Performanz von Transaktionen des operationalen Betriebs
 - spezieller physischer und logischer Datenbankentwurf für multidimensionale Sichten und Anfragen notwendig
 - Transaktionseigenschaften (ACID) nicht wichtig
- Funktionalität
 - für OLAP-Anfragen werden historische Daten benötigt, die in OLTP-Systemen typischerweise nicht vorliegen
 - Konsolidierung (Integration, Bereinigung und Aggregation) von Daten aus heterogenen Datenquellen
- Sicherheit

DWH Schichtarchitektur



Presentation Layer

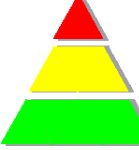
DWH / OLAP Layer

Extraction Layer



Laden des Data Warehouse

- Lesen von Daten aus beliebigen Quellen (ODBC, Excel, Internet, etc.)
 - Altsysteme oft undokumentiert und schwer zugänglich
- Entfernen von semantischen Inkonsistenzen (Integritätsprüfungen)
 - Außerhalb der Quellen mit überprüfbaren Qualitätssicherungsmethoden
- Effizienz des Ladevorgangs: “Kampf dem Ladefenster” (falls es eines gibt ...)
 - inkrementelles Laden; Erstellen/Aktualisieren von Aggregationen während des Ladeprozesses
 - online oder offline Laden? (Sperren von Datenpartitionen, blockieren von Anfragen)
 - paralleles Laden



Data Warehouse Datenhaltung

- Analysen benötigen oft Daten auf Rohdatenebene (feinste Granularität)
 - Verwaltung von sehr großen Datenvolumina (GB → TB)
- Datenorganisation innerhalb des Data Warehouses:
 - anwendungsorientierte Aufteilung nach Dimensionen (nach der Zeit, nach Produkten, etc.)
 - zeitbasierte Datenverwaltung ist unbedingt notwendig (Ladevorgänge sind häufig zeitorientiert)
- Verfügbarkeitsanforderung steigen (Trend 24x7 System)
 - DWH ist nicht mehr nur ein 'add-on' oder ein 'Spielzeug' für Entscheidungsträger
 - DWH wird dazu verwendet, Geld zu "machen" (break-even nach 3 Jahren; MetaGroup)



Analyse / Decision Support / Presentation

- “We are drowning in data, but starving for knowledge”
- Multidimensionalität als intuitiver Ansatz
 - effizienten Zugriff auf Datenwürfel bereitstellen
 - Unterstützung von verbreiteten Tabellenkalkulationen
- Typische Anfragen:
 - Wie haben sich dieses Jahr die Verkaufszahlen nach Verkaufskanäle und Produktgruppen entwickelt (im Vergleich zum letzten Jahr)?
 - Welche Geschäfte machen 80 Prozent des Umsatzes einer bestimmten Marke?
 - Wer sind die Hauptkonkurrenten eines Elektronikherstellers?
 - Distributionsüberschneidungen: Aufdecken von Korrelationen zwischen zwei Marken A und B

DWH Ansatz

- Information a priori integriert und im Data Warehouse gespeichert

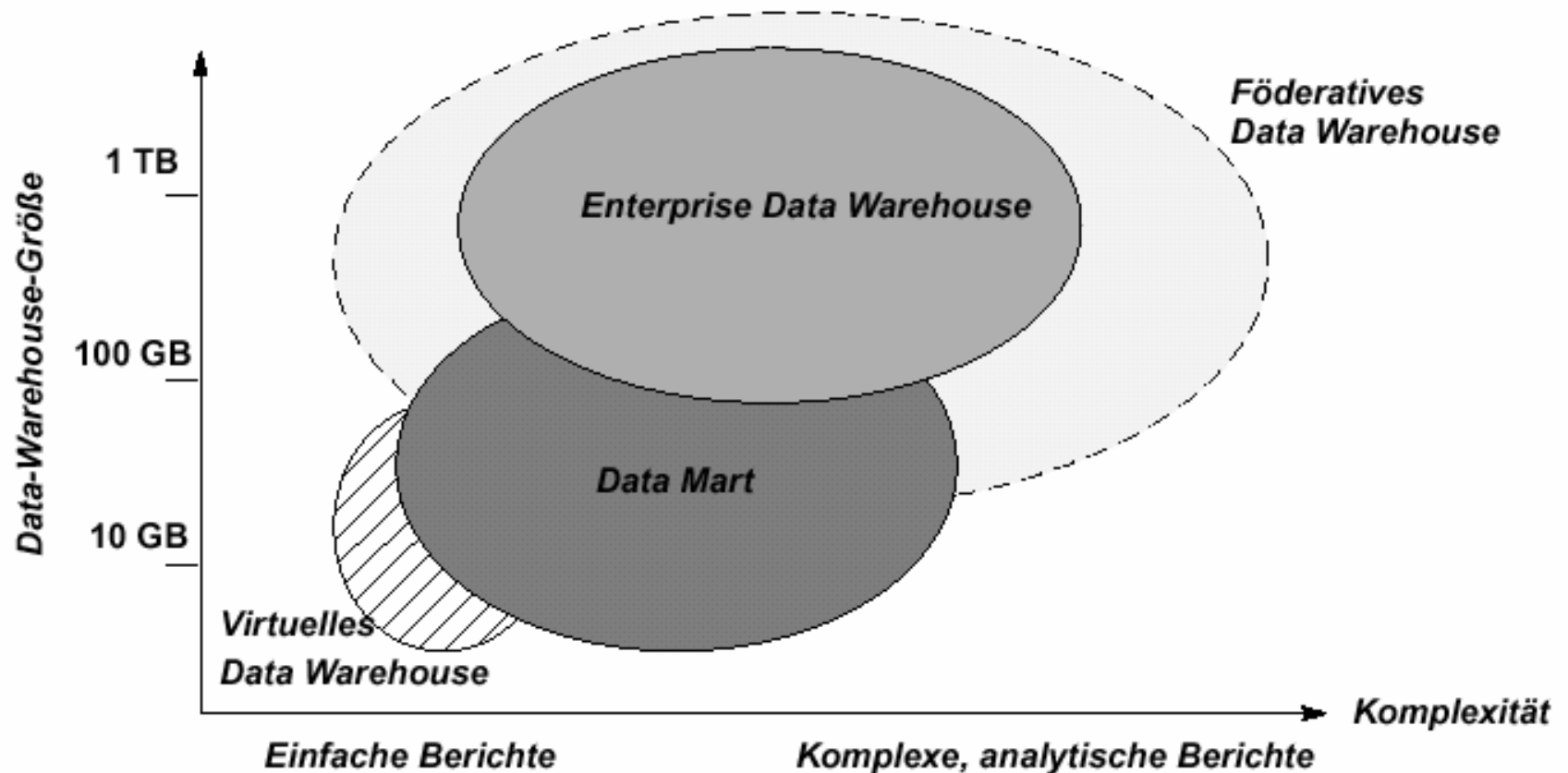
- Vorteile:
 - relativ kurze Anfragelaufzeiten
 - konkurriert nicht mit operationalem Betrieb
 - Externe Datenquellen leichter integrierbar

- Nachteil:
 - Redundanz und geringere Aktualität der Daten
→ neue Entwicklungen: z.B. (Near) Real-Time DWHs

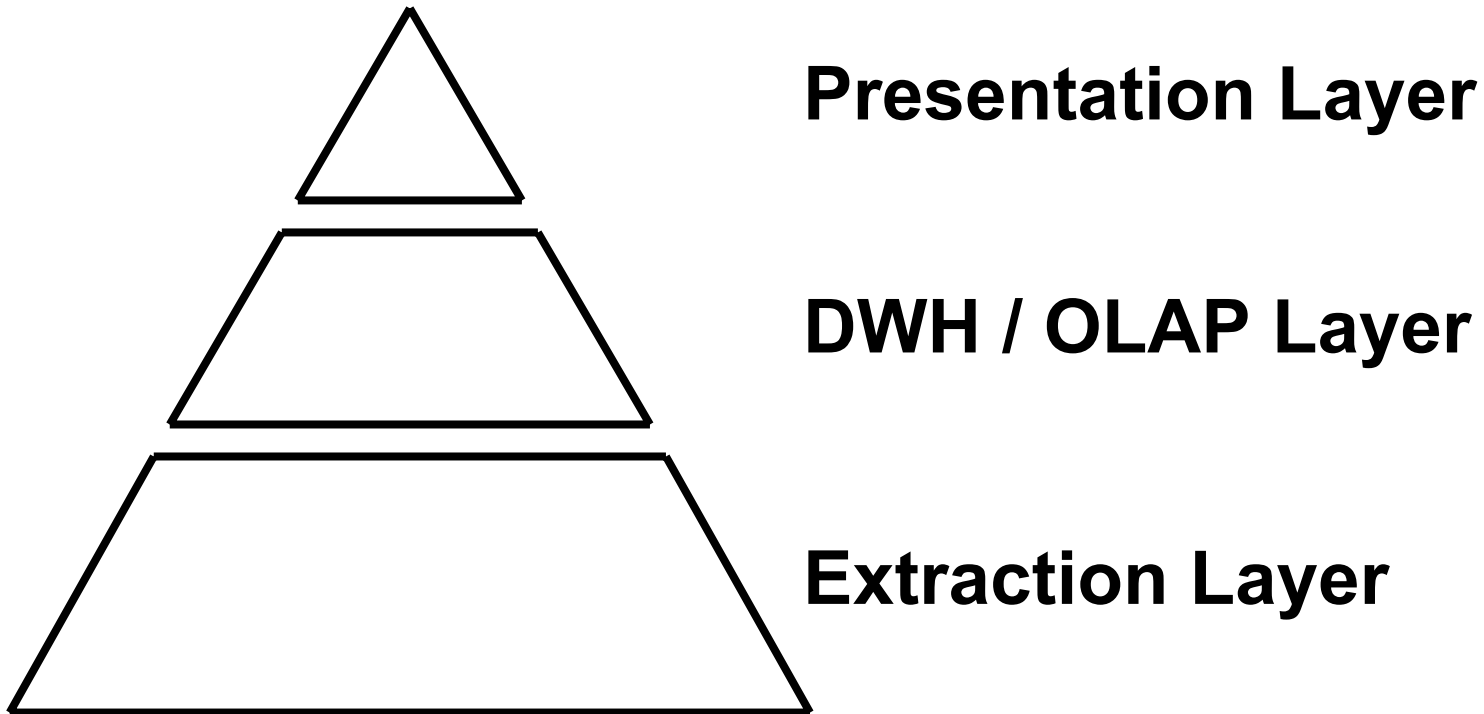
Verwendungszweck / Evolution von DWHs

- **Reporting**
Was ist passiert?
- **Analyse**
Warum ist etwas passiert?
- **Vorhersage / Prediction**
Was wird passieren?
- **Operationalisierung**
Was ereignet sich jetzt gerade?
- **Active DWH**
Was soll sich ereignen?
- **Closed Loop Decision Support**
Zielbasierter Regelansatz: Welche Effekte werden Entscheidungen haben?

DWH Typen - Evolution



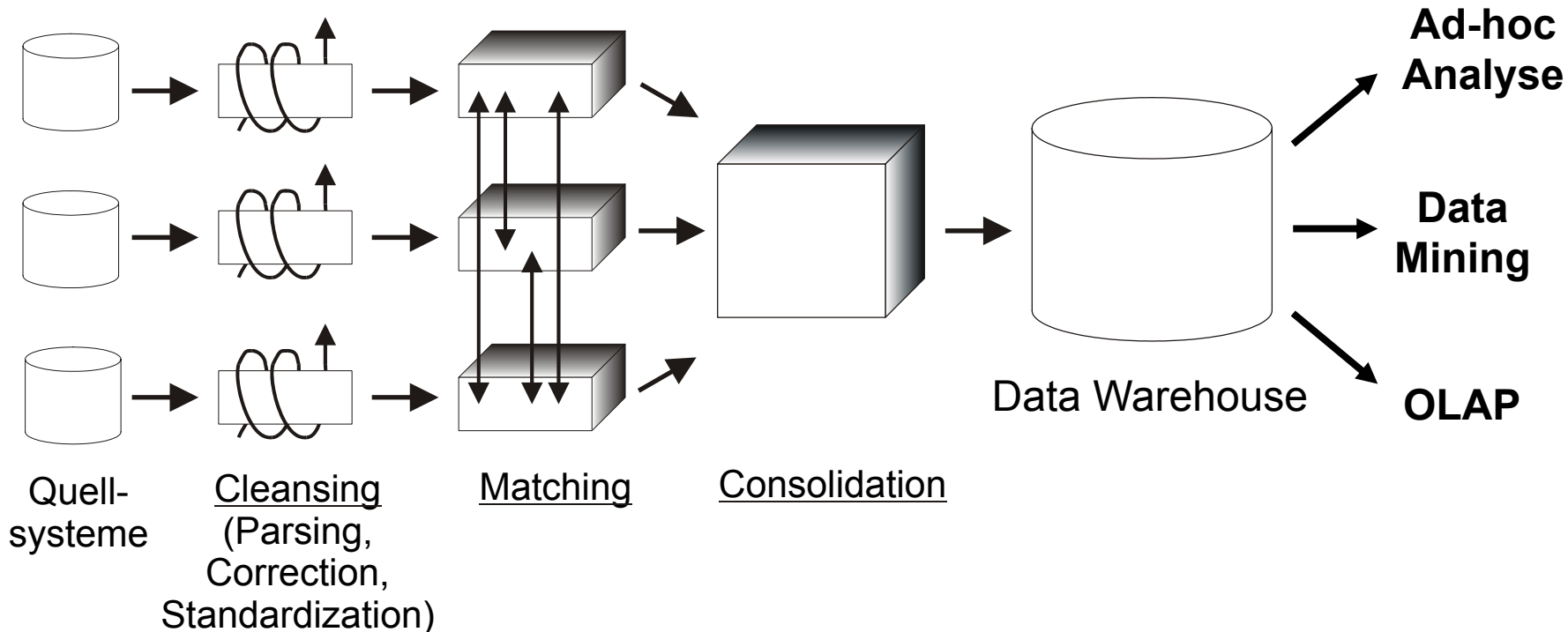
DWH Architektur im Detail



DWH: ETL / Datenqualität

ETL: Extract - Transform – Load

Datenqualität begrenzt die Verwend-/Verwertbarkeit der Daten



DWH: ETL / Datenqualität

- Methoden der Datenqualitätsanalyse
 - Sicht des Datenproduzenten: intuitiver (projektabhängiger) Ansatz und theoretischer Ansatz (Suche nach methodischen Fehlern)
 - Sicht des Datenkonsumenten: empirischer Ansatz des Endverbrauchers
- DQ-Indikatoren der Quellsysteme:
 - Daten sind präzise
 - Daten sind in einem bestimmten Datentyp gespeichert
 - Daten sind konsistent
 - sinnvolles Datenbankdesign
 - keine Redundanz
 - Daten folgen den Business Regeln
 - Domains werden eingehalten
 - Zeitrichtig (timeliness)
 - ...

DWH: ETL / Datenqualität

- DQ-Indikatoren (Fortsetzung):
 - gut verständlich
 - Daten sind integriert
 - Daten passen zum Geschäftsfall
 - Benutzer ist mit den Daten zufrieden
 - Daten sind komplett
 - keine doppelten Felder
 - keine Datenanomalien
- Daten der Quellsysteme sind normalerweise sehr schlecht
→ Erkennung ist notwendig
 - automatische Ausreißererkennung
 - Benutzer, die Anomalien erkennen oder erahnen
 - Datenalter
 - Programmabbrüche

DWH: ETL / Datenintegration

- Überlegung:
 - Laden von Daten aus unabhängigen Systemen,
 - inkompatiblen Datenformaten,
 - redundanten Daten
 - und aus unterschiedlichen Begriffswelten
- Ziel:
 - Integration heißt Datenmaterial und Verwendbarkeit verbessern
 - keine einheitliche Technologie, Schnittstellen, Datenformate
 - DWH braucht aber genau diese Vereinigung
(ca. 2/3 des Aufwandes beim Aufbau eines DWH)
- Vorgehensweise:
 - Schemaintegration
 - Laden der Daten

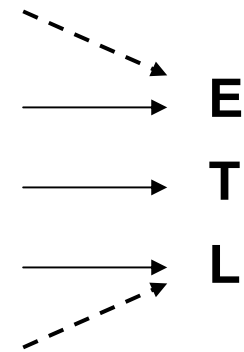
DWH: ETL / Schemaintegration

- Ziele:
 - Integration verschiedener lokaler Schemata (mit möglicherweise unterschiedlichen Datenmodellen) in ein gemeinsames globales Datenschema
 - Ausräumen aller Inkonsistenzen auf Schemaebene
- Hauptaufgaben:
 - Schema Translation:
 - Abbildung eines Schemas vom Quellschema in äquivalentes Zielschema
 - explizite Abbildung eines jeden Schemas
 - Entwicklung von Abbildungsregeln
 - Schemaanalyse
 - Analyse und Vergleich der Objekte der Schemata
 - Spezifikation von Beziehungen zwischen Objekten der einzelnen Schemata

DWH: ETL

- Im Verlauf der Ladephase müssen folgende Aufgaben bewältigt werden:

- Zugriff auf die operationalen Datenquellen
- Erkennen von Änderungen auf den Datenquellen
- Transformation der Daten
- Bereinigung der Daten
- Bilden von Aggregationen



- Probleme beim Ladevorgang:

- Zugriff auf riesige Datenvolumina bei kleinem Ladefenster
- sequentielles Laden dauert zu lange
- Parallelisierung (Füllen der Zielrelationen und Indexerstellung) und inkrementelles Laden notwendig

DWH: ETL - Datentransformation

- “Einfacher” Zugriff auf OLTP- und Altsysteme reicht nicht aus
- Benötigte Funktionalität der Transformation:
 - Änderung des Primärschlüssels
 - Primärschlüsselauflösung bei mehreren Eingabedateien
 - Umformatierung der Daten
 - unterschiedliche Datentypen
 - Bereitstellung von Default-Werten
 - Non-Standard-Formate der Eingabedateien
 - Homonym / Synonym – Problematik
 - Homonym: identischer Wert/Begriff für unterschiedliche Semantik
 - Synonym: unterschiedliche Werte/Begriffe mit derselben Bedeutung
 - Datenbereinigung

DWH: ETL – Data Cleansing

■ Ziele:

- Datenkonflikterkennung auf Instanzenebene
- Datenbereinigung auf Instanzenebene
- Erkennen von Anomalien
- Erkennen von fehlerhaften Daten und Ausreißern
- Semantische Probleme: Wortbedeutung; Maßeinheit; Genauigkeit, Format, Zeitpunkt

■ Ursachen für Dateninkonsistenzen:

- Fehler im Datenmaterial, die im OLTP-Betrieb nicht zum Tragen kommen
- Datenkonflikte zwischen Quellsystemen, die redundante Information enthalten (z.B. Kunde "Meier" hat in DB₁ eine andere Telefonnummer als in DB₂)
- Schemaevolution (z.B. Umstrukturierung des Produktkatalogs, Ausgliederung von Unternehmensteilen)

DWH: ETL – Data Cleansing

- **Dateninkonsistenzerkennung:**
 - Überprüfung semantischer Nebenbedingungen
 - Plausibilitätsprüfungen
 - Überprüfung von Toleranzwerten
 - Überprüfung von Summenwerten
 - Positionsvergleiche
 - nicht alle Datenkonflikte sind erkennbar!
- **Dateninkonsistenzbereinigung:**
 - Anwendungsspezifisches Wissen unbedingt erforderlich!
 - Einsetzen von Nullwerten, Defaultwerten
 - Ermittlung des korrekten Datenwerts
 - Interpolation von fehlenden Datenwerten

Fakten / Dimensionen

■ Fakten:

- Kennzahlen, Messgrößen
- Repräsentieren die primären Geschäftsbereiche
- Werden nicht geändert sobald sie im DWH sind
- In bestimmter Granularität gespeichert

■ Dimensionen:

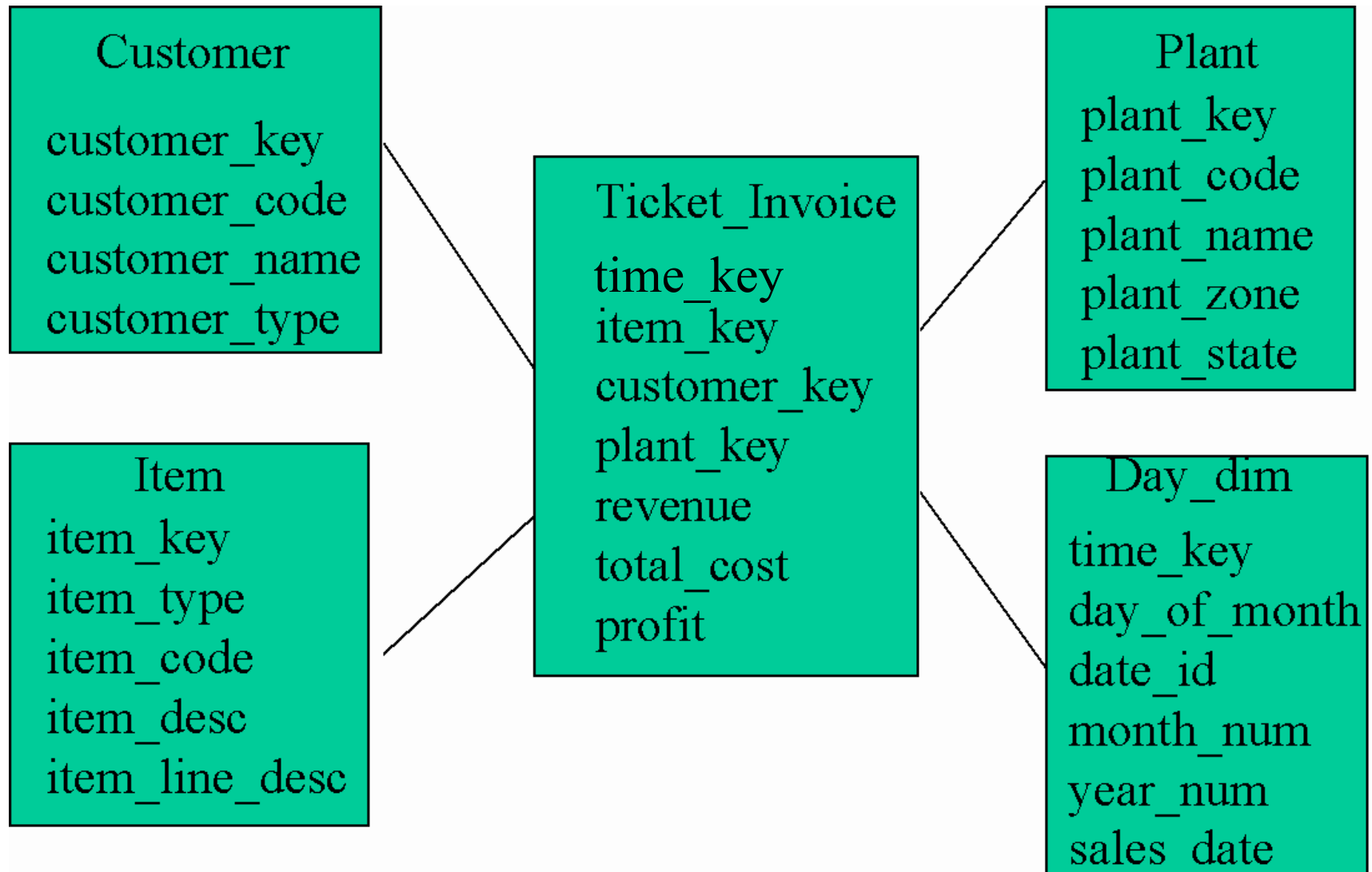
- Referenzieren auf Informationen, die zur Strukturierung / Charakterisierung der Fakten verwendet werden
- Definieren die Aggregations-Hierarchien
- Beispiele für Dimensionen: Zeit, Produktgruppen, Regionen

Star Schema

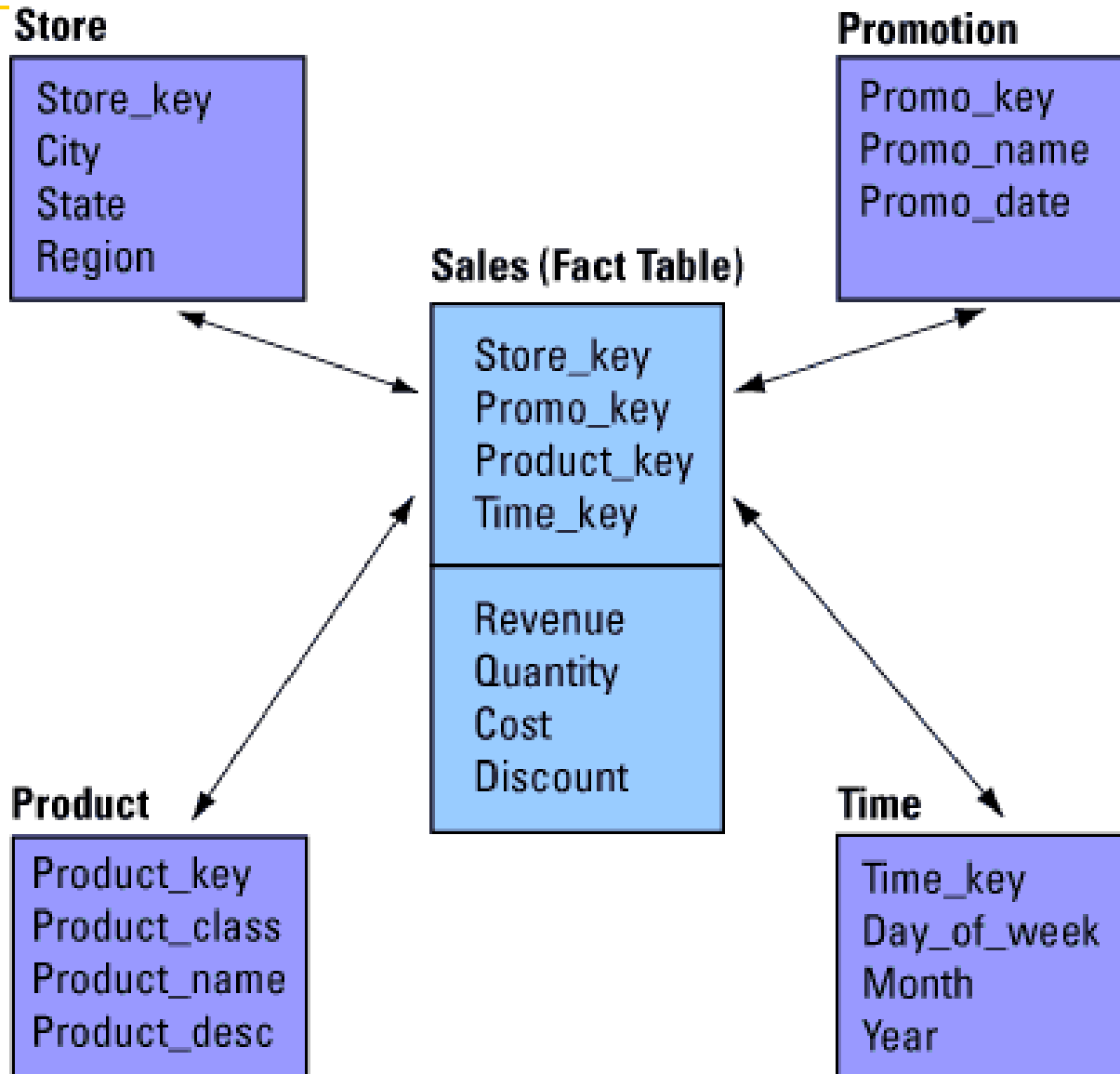
- Denormalisierte Dimensionstabellen
- Zentrale Fakt-Tabelle mit Verweisen in Dimensionstabellen
- Geringer Aufwand bei der Wartung
- Keine Joins über viele Tabellen notwendig
- Bildung von Aggregaten schwierig



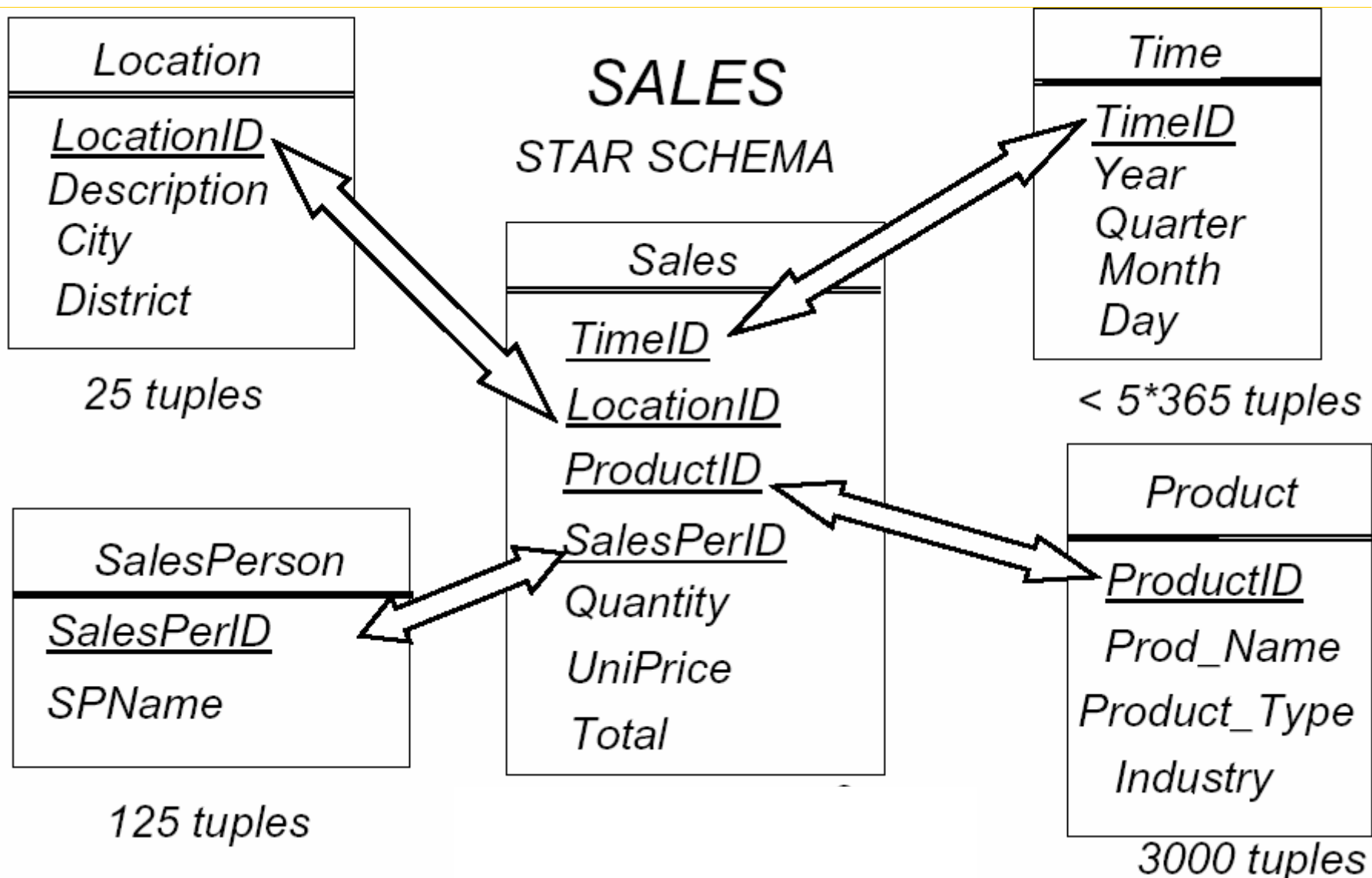
Star Schema - Beispiel



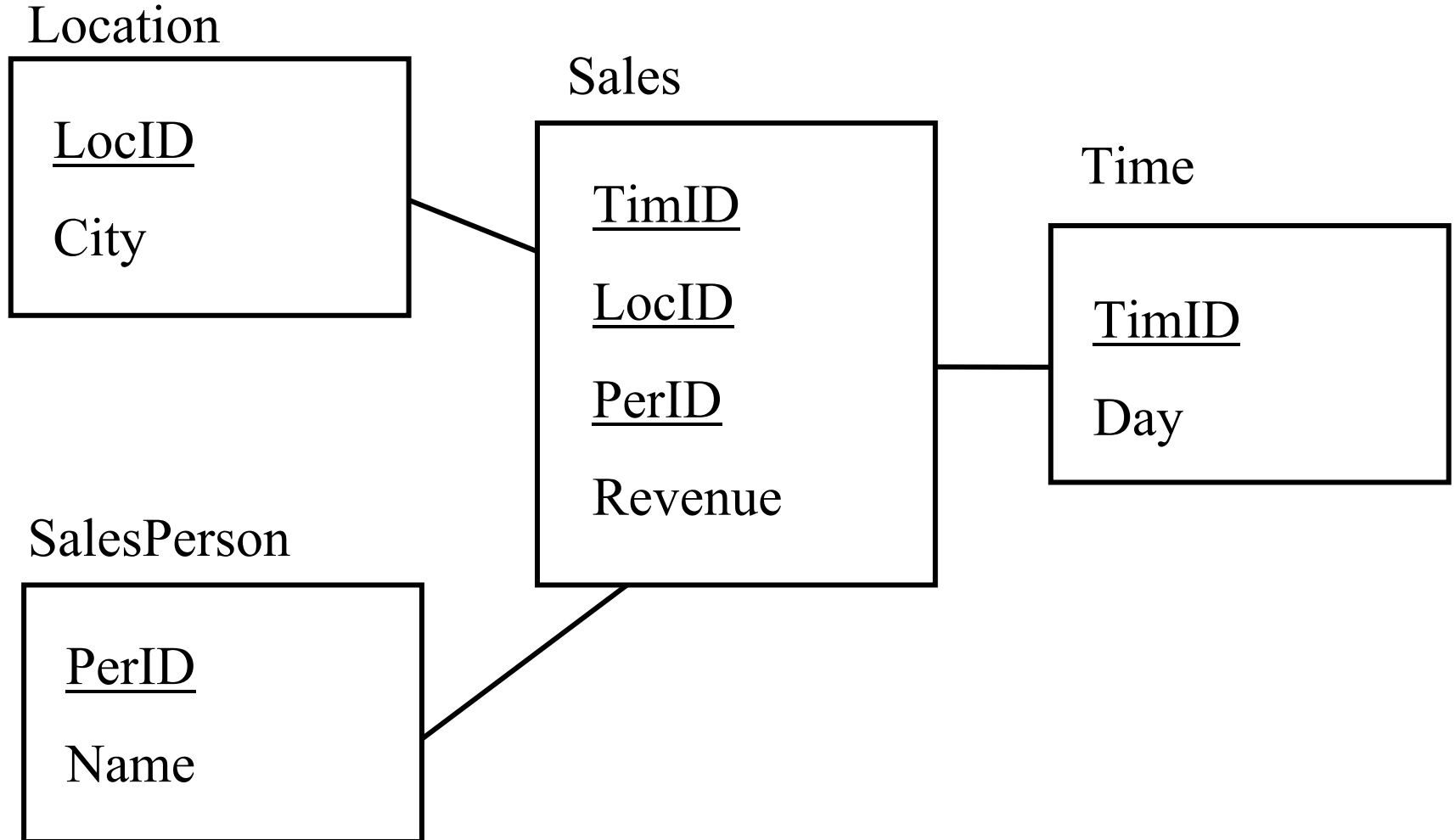
Star Schema - Beispiel



Star Schema - Beispiel



Star Schema – Beispiel „Sales“



Sales Star Schema: Tabellen und Records

Location

<i>LocID</i>	<i>City</i>
1	Well
2	Nels
3	Auck

SalesPerson

<i>PerID</i>	<i>Name</i>
1	John
2	Susan
3	James
4	Susan
5	Ann

Sales

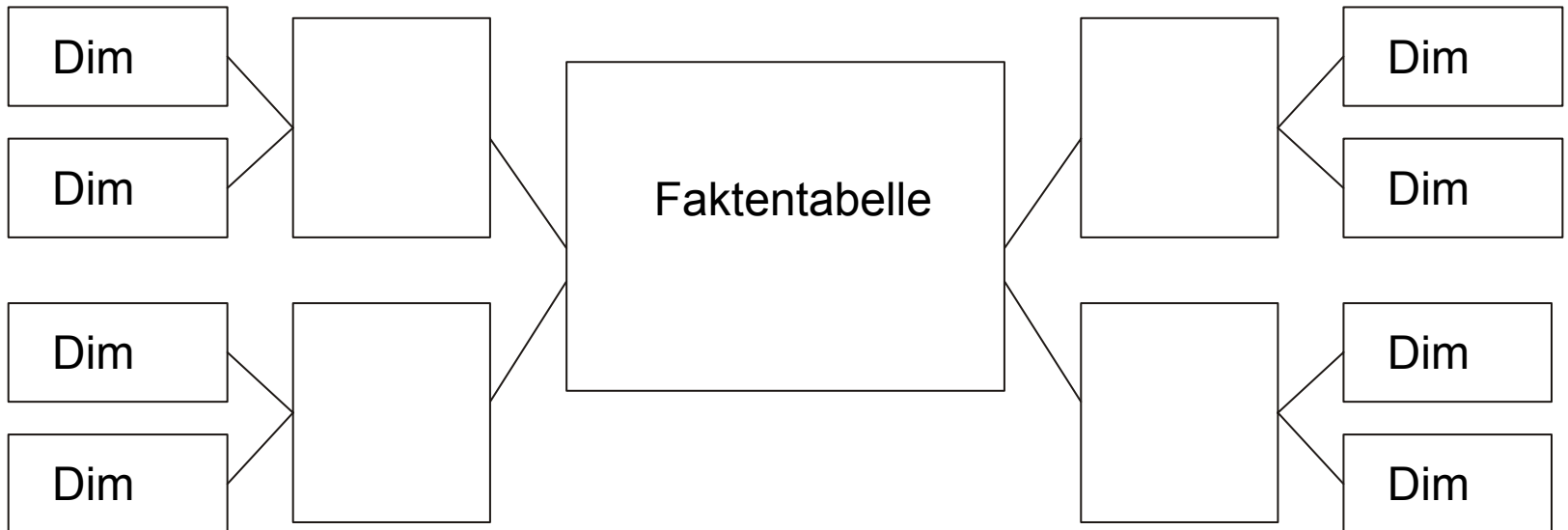
<i>LocID</i>	<i>PerID</i>	<i>TimID</i>	<i>Amnt</i>
1	1	1	230
1	1	2	300
1	1	8	310
1	2	7	50
2	3	1	550
2	3	5	100
3	4	6	880
3	5	1	60
3	5	2	60
3	5	4	140

Time

<i>TimID</i>	<i>Day</i>
1	Mon
2	Tue
3	Wed
4	Thu
5	Fri
6	Sat
7	San
8	Mon

Snowflake Schema

- Snowflake Schema basiert auf dem Star Schema
- Dimensionstabellen sind normalisiert (3. NF)
- Komplexere SQL Befehle, Hohe Anzahl an Join Operation
- Bildung von Aggregaten wird implizit, optimal unterstützt
- Erhöhte Anzahl an Tabellen



DWH / OLAP Schicht

Analysetoolübersicht

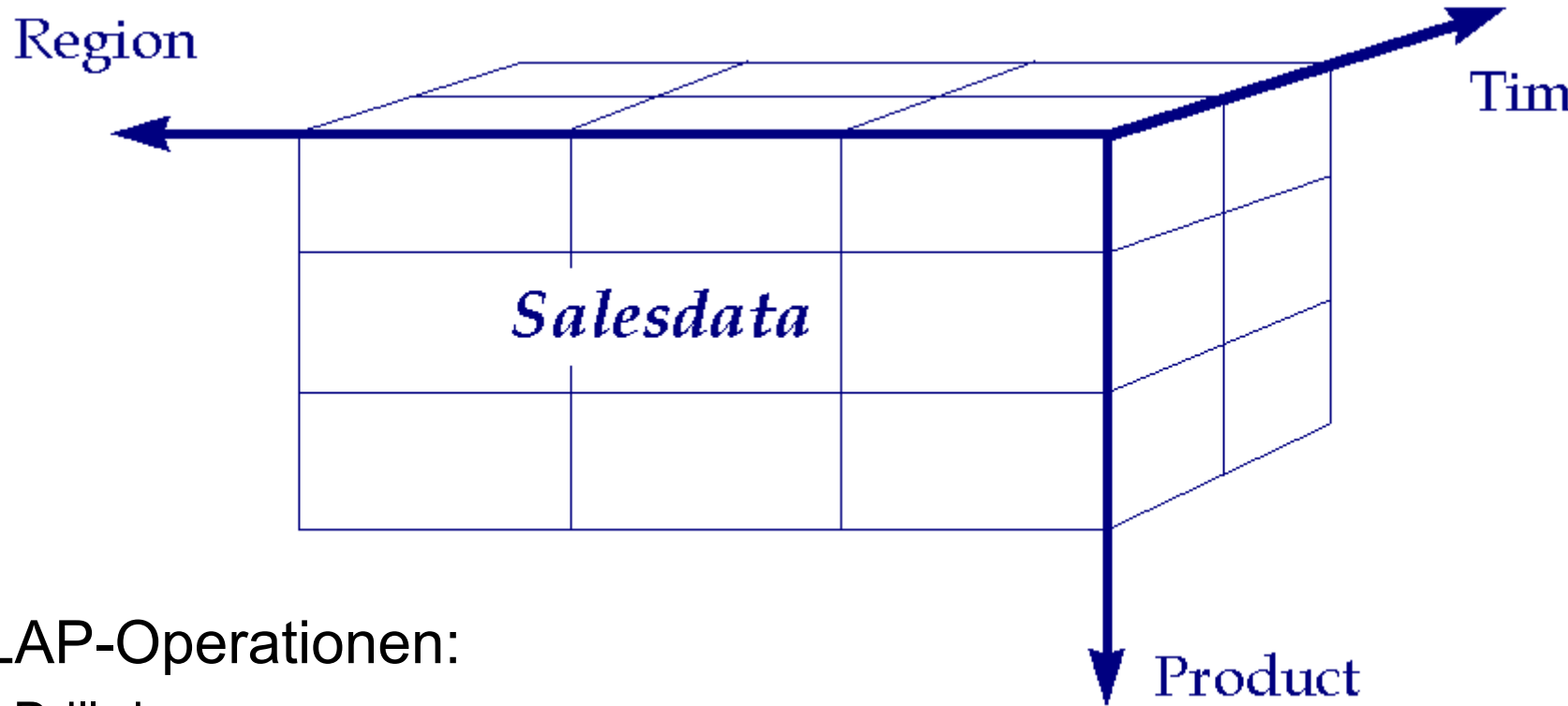
	Statische Abfragen	Dynamische Abfragen
Datenumfang gering	Anwendungen	OLTP Anfragen
Datenumfang komplex	Reporting	OLAP, Data Mining, Visualisierung

OLAP

- OLAP (on-line analytical processing) als “Anwendungsarchitektur” mit Data Warehouse als Datenquelle
- Business Probleme wie Marktanalysen oder Marktvorhersagen benötigen eine multidimensionale Sicht auf das Datenmaterial (Daten oftmals auch multidimensional modelliert)
- OLAP-Engines erlauben schnellen Zugriff auf Daten durch spezielle Operationen, wie roll-up, drill-down, slice, dice, etc.
- treibende Kräfte:
 - Funktionalität der Tools (Caching, Query Rewriting & Optimierung)
 - Antwortzeiten
 - Security

OLAP – multidimensionale Daten

■ OLAP-Cubes:



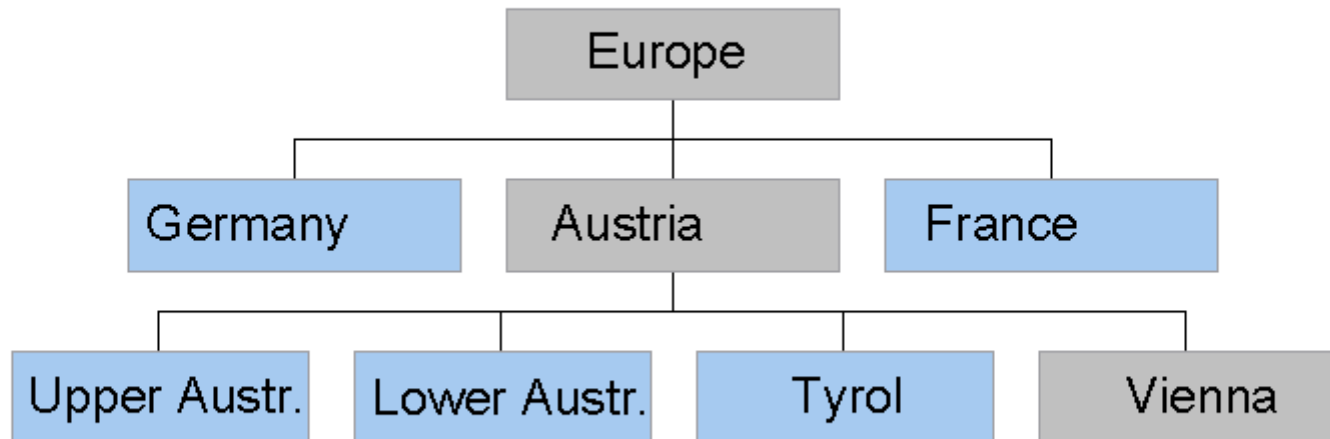
■ OLAP-Operationen:

- Drill-down
- Roll-up
- Slice
- Dice

OLAP Operationen

- Roll-Up / Drill-Down

Drill-Down



Roll-Up

OLAP Operationen

Cube Browser - Sales

	MeasuresLevel					
Product Family	Unit Sales	Store Cost	Store Sales	Sales Count	Store Sales Net	Profit
Products	266.773,00	225.627,23	85 565.238,13	86837	339.610,90	339.610,90
Drink	24.597,00	19.477,23	85 48.836,21	7978	29.358,98	29.358,98
Food	191.940,00	163.270,72	85 409.035,59	62445	245.764,87	245.764,87
Non-Consumable	50.236,00	42.879,28	85 107.366,33	16414	64.487,05	64.487,05

Double-click a member to drill up or down.

Close

Help

↓ Drill-down

↑ Roll-Up

Cube Browser - Sales

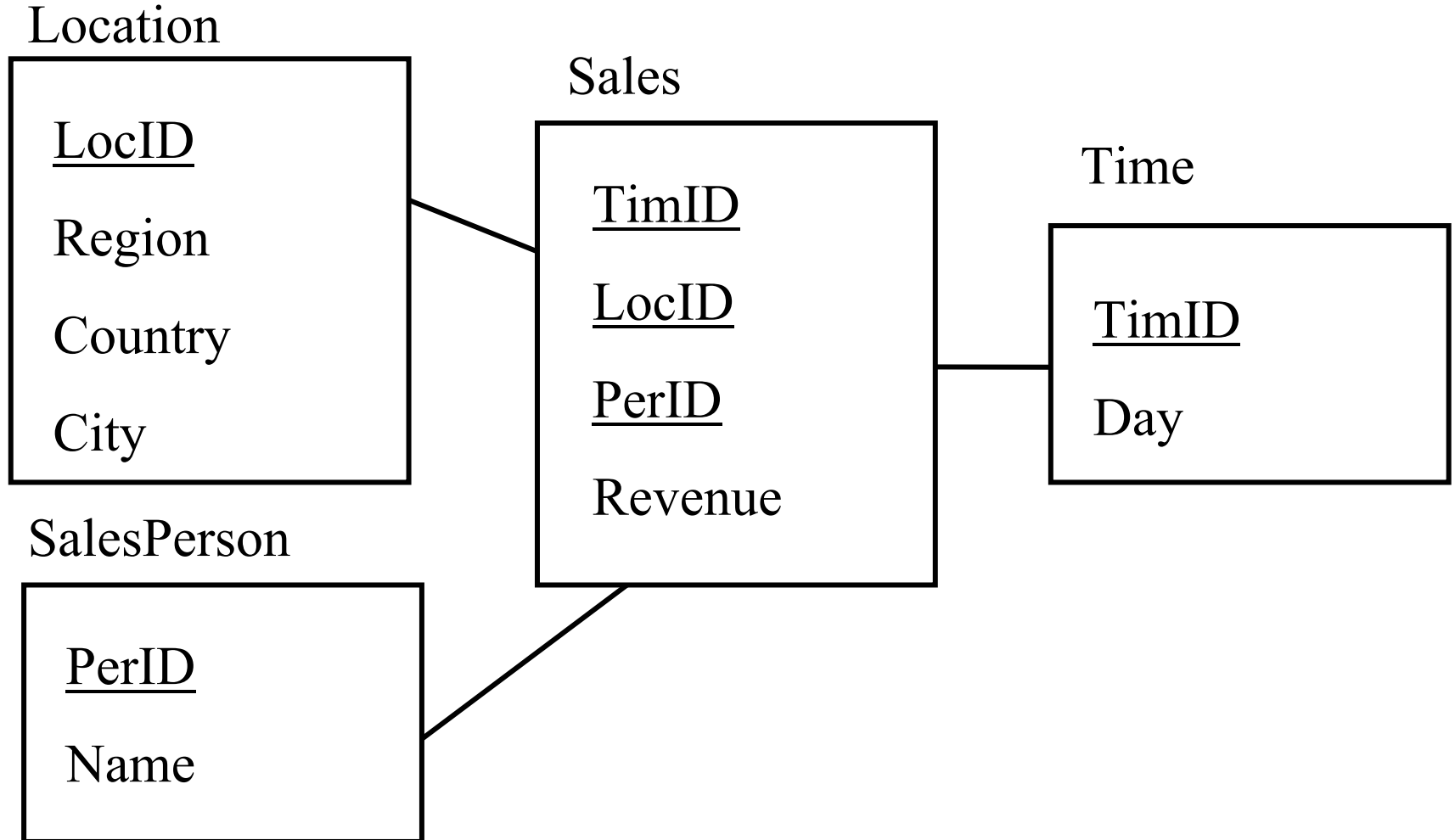
		MeasuresLevel				
Product Family	+ Product Department	Unit Sales	Store Cost	Store Sales	Sales Count	Store Sales Net
Products	All Products Total	266.773,00	225.627,23	85 565.238,13	86837	339.610,90
Drink	Drink Total	24.597,00	19.477,23	85 48.836,21	7978	29.358,98
Food	Food Total	191.940,00	163.270,72	85 409.035,59	62445	245.764,87
Non-Consumable	Non-Consumable Total	50.236,00	42.879,28	85 107.366,33	16414	64.487,05
	+ Carousel	841,00	595,97	85 1.500,11	272	904,11
	+ Checkout	1.779,00	1.525,04	85 3.767,71	569	2.242,65
	+ Health and Hygiene	16.284,00	12.972,99	85 32.571,86	5310	19.598,86
	+ Household	27.038,00	24.170,73	85 60.469,89	8862	36.299,11
	+ Periodicals	4.294,00	3.614,55	85 9.056,76	1401	5.442,21

Double-click a member to drill up or down.

Close

Help

Star Schema – Beispiel „Sales“



Star Schema – Beispiel „Sales“

Location Dim

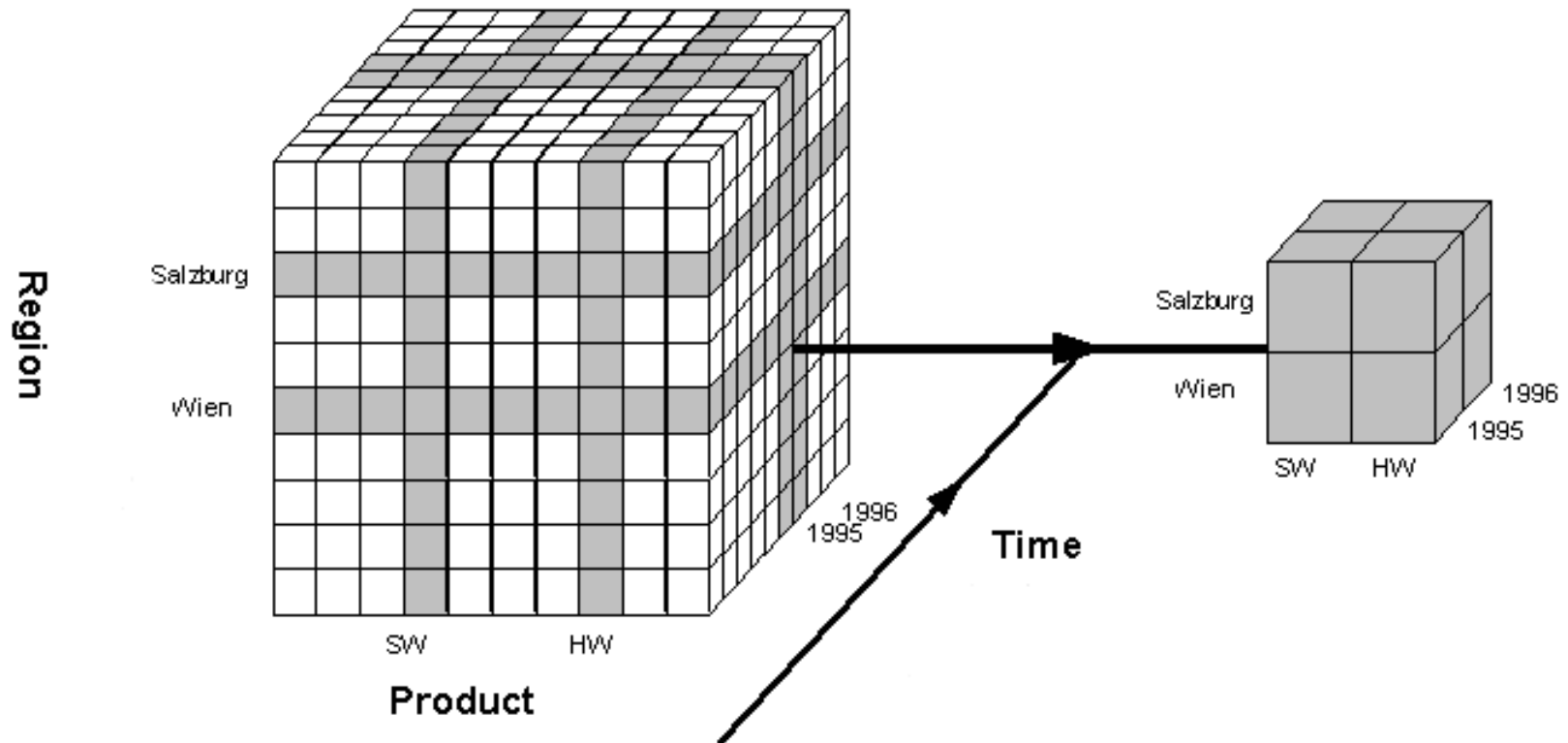
<u>LocID</u>
Region
Country
City

Location Fakt

<u>LocID</u>	Region	Country	City
1	Europa	Österreich	Wien
2	Europa	Österreich	Graz
3	Europa	Österreich	Linz
4	Europa	Schweiz	Zürich

OLAP Operationen

- Slice / Dice



OLAP Operationen

Cube Browser - Sales

Yearly ... \$150K + Time 1997 Store USA Gender M Marital ... All Marital Custom... All Custom

		MeasuresLevel				
Product Family	+ Product Department	Unit Sales	Store Cost	Store Sales	Sales Count	Store Sales Net
All Products	All Products Total	2.550,00	2.179,47	65 5.454,79	835	3.275,00
Drink	Drink Total	250,00	223,12	65 563,41	85	340,00
	+ Alcoholic Beverages	84,00	82,83	65 214,70	29	131,00
	+ Beverages	113,00	106,51	65 256,74	39	150,00
	+ Dairy	53,00	33,77	65 91,97	17	58,00
Food	Food Total	1.830,00	1.541,08	65 3.859,93	597	2.318,00
Non-Consumable	Non-Consumable Total	470,00	415,28	65 1.031,45	153	616,00

Double-click a member to drill up or down.

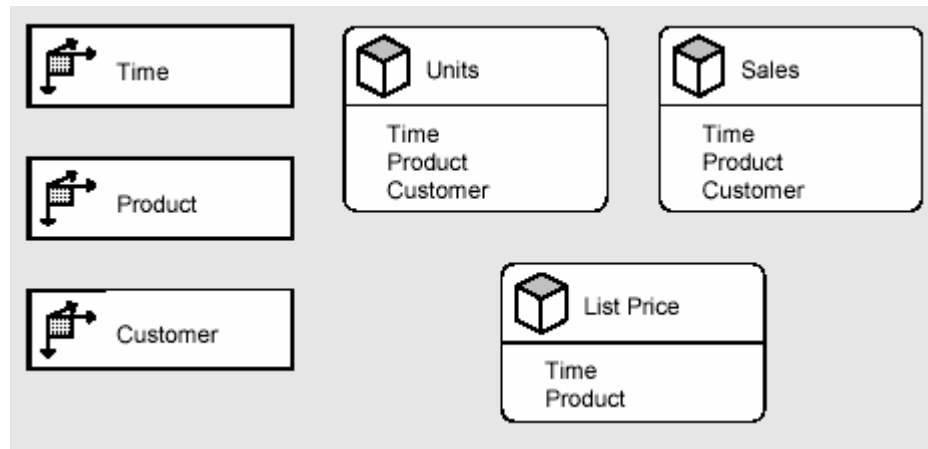
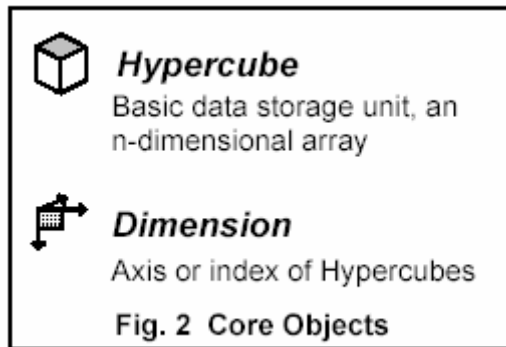
Close Help

OLAP Anforderungen

- Entwurf einer Datenbank, die „unbekannte“ Abfragen mit sehr guter Performance beantworten kann.
- OLAP Daten (multidimensional) aus dem DWH abgeleitet
- Daten spannen einen multidimensionalen Raum auf
- Eventuell unterschiedliche Speicherungsformen zwischen DWH und OLAP:
 - ROLAP (relationale Datenbank)
 - MOLAP (multidimensionale Datenbank)
 - HOLAP (hybrid OLAP – Kombination ROLAP, MOLAP)
 - DOLAP (Datenbank auf Desktop Computer / client side)

ADAPT – Application Design for Analytical Processing Technologies

Kernobjekte



ADAPT – Application Design for Analytical Processing Technologies

Hierarchien



Hierarchy

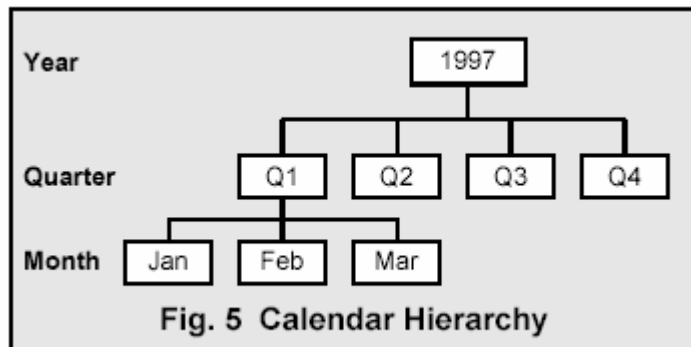
Set of parent/child member combinations that define aggregation



Level

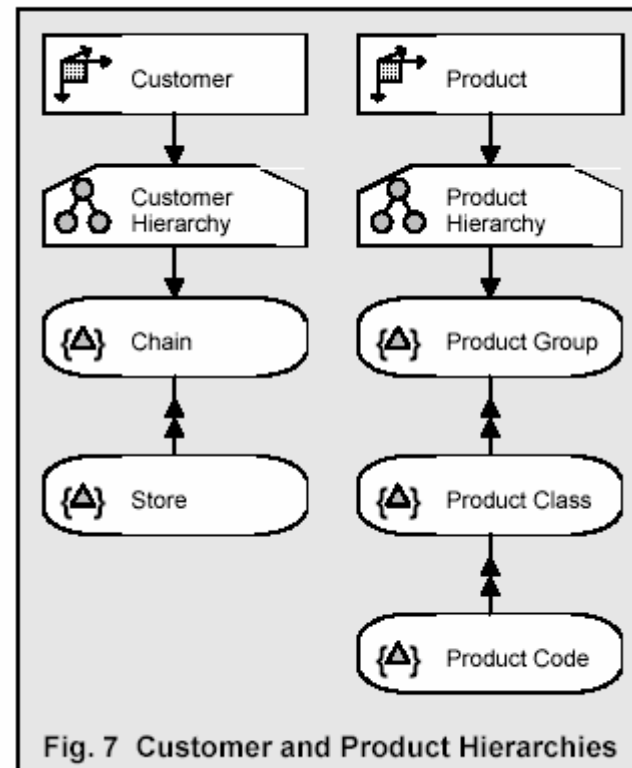
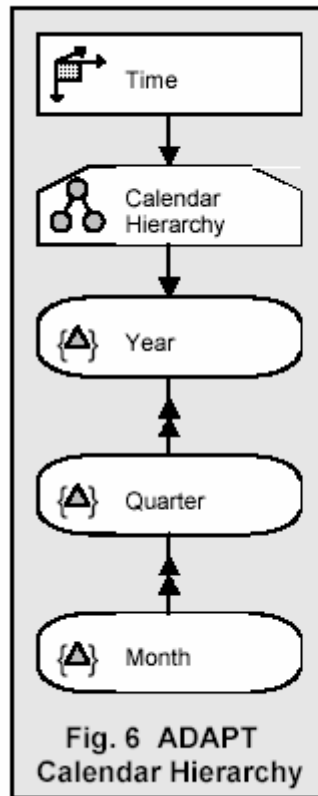
Collection of members used to define hierarchical precedence

Fig. 4 Dimensional Aggregation



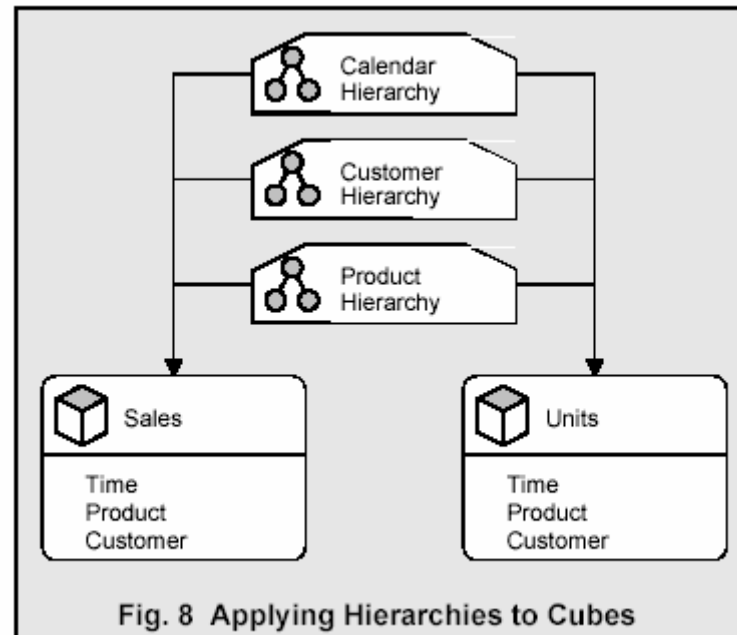
ADAPT – Application Design for Analytical Processing Technologies

Hierarchien



ADAPT – Application Design for Analytical Processing Technologies

Hierarchien und Würfel



Zusammenfassung

- Multidimensionale Analyse
 - DWH Dient der Entscheidungsunterstützung
 - ETL Prozess ist sehr aufwendig
 - Getrennt von operativen Systemen
 - DWH repräsentiert Große Datenmengen
- ➔ Ziel ist ein unternehmensweites Data Warehouse