

Einführung in OLAP und Business Analysis

Gunther Popp
dc soft GmbH

Überblick

- **Wozu Business Analysis mit OLAP?**
- **OLAP Grundlagen**
- **Endlich ... Technischer Background**
- **Microsoft SQL 7 & OLAP Services**

Klassische IT-Aufgaben

- Eine der klassischen IT-Aufgaben heute:
Daten sammeln und speichern
- Beispiele: Auftrags-, Kunden-,
Produktionsdaten
- Lebensnotwendig für das tägliche,
operative Geschäft
- Sammelbegriff: OLTP - Anwendungen

Business Analysis heute

- **Zusätzlich: Erstellung von Auswertungen und Reports**
- **Reports sind Grundlage zur Business Analysis**
- **Qualität der Reports kritisch für korrekte Entscheidungen im Unternehmen**
- **Reports maßgebend für die Zufriedenheit mit der IT-Abteilung**

Nachteile statischer Reports aus IT-Sicht...

- Erstellung guter Reports ist zeitaufwendig → Ressourcen werden gebunden
- Nicht alle gewünschten Zahlen können in Reports aufgenommen werden
- Resultat: Unzufriedenheit mit der IT-Abteilung. „Ihr liefert nicht die Infos, die ich wirklich brauche“

... und aus Anwender-Sicht



... und aus Anwender-Sicht

- **Wichtige Zahlen fehlen im Report**
- **Inkonsistente Datenquellen für die einzelnen Reports → Unterschiedliche Ergebnisse!**
- **Erstellung des Reports dauert zu lange**
- **Keine aktuellen Zahlen**
- **Um zu Ergebnissen zu gelangen, muss der Report erst in Excel eingetippt werden ...**

Business Analysis mit OLAP-Technologie

- Statische Reports werden durch OLAP-Server und passende Frontends ersetzt
- IT stellt zentralen Datenpool für Analysen zur Verfügung
- Anwender erstellen Auswertungen selbst



Was steckt hinter OLAP?

- **“On Line Analytical Processing”**
- **Flexible Analysen, ohne vorherige Kenntnis der Fragestellung**
- **Hauptvorteile**
 - Einfache Navigation durch komplexe Datenbestände
 - Effiziente Datenspeicherung komplexer Daten

FASMI-Test*):

- **Fast:** Antwortzeit kleiner 5 Sekunden
- **Analysis:** Einfache Datenanalyse ohne Programmierung
- **Shared:** Gleichzeitiger Zugriff mehrerer Nutzer
- **Multidimensional:** Mehrere, hierarchische Dimensionen
- **Information:** Einbindung aller notwendigen Daten

F

A

S

M

I

*) <http://www.olapreport.com>

Gunther Popp - Einführung in OLAP und Business Analysis

Folie 10 -

OLTP vs. OLAP (1)

- **Operationale Systeme (OLTP)**
 - Beispiel: Auftragsabwicklung, WaWi
 - Datenmodell optimiert für effiziente und sichere Datenhaltung (Normalisiert)
 - Schwerpunkt: Sichere & Schnelle Abarbeitung einzelner Transaktionen

OLTP vs. OLAP (2)

▪ Business Analysis (OLAP)

- Beispiel: Analyse von Produktverkäufen
- Datenmodell optimiert für schnelle Abfragen
- Schnelle Abfragen und beliebige Zusammenfassungen
- „Slice & Dice“: Veränderung des Blickwinkels zu jedem Zeitpunkt
- Schwerpunkt: Beantwortung flexibler, komplexer Fragestellungen

OLAP Grundlagen

- Dimensionen
- Hierarchien
- Meßwerte (Fakten)
- Aggregationen
- Würfel
- Begriffe aus dem Umfeld

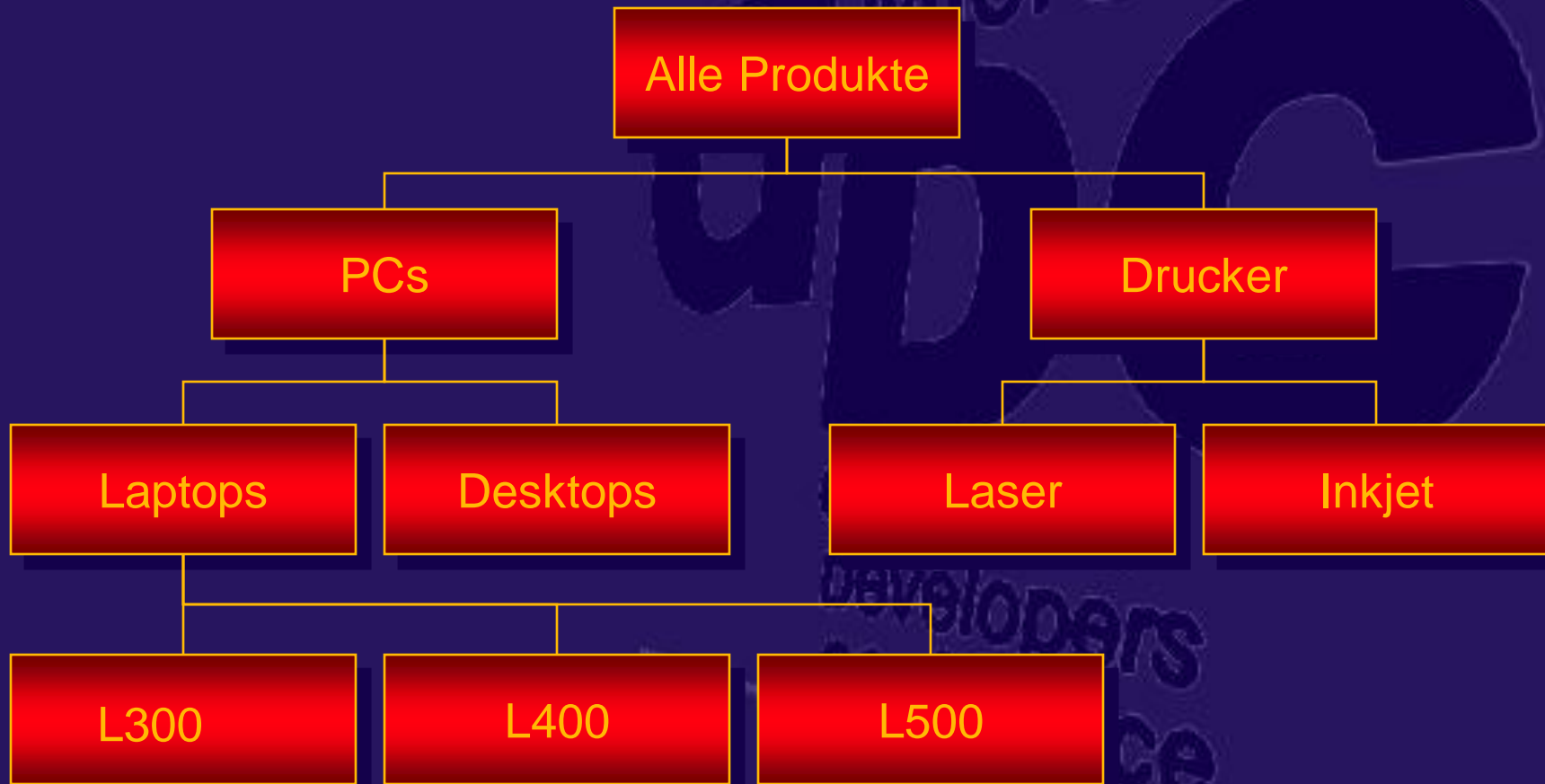
Dimensionen

- **Strukturierung von Daten nach Dimensionen ist das grundlegende Prinzip von OLAP**
- **Dimensionen im Sinne von OLAP sind die maßgeblichen Einflußgrößen, Prozesse, Faktoren oder Bestandteile eines Unternehmens**
- **Beispiel: Produkte, Regionen, Kunden**

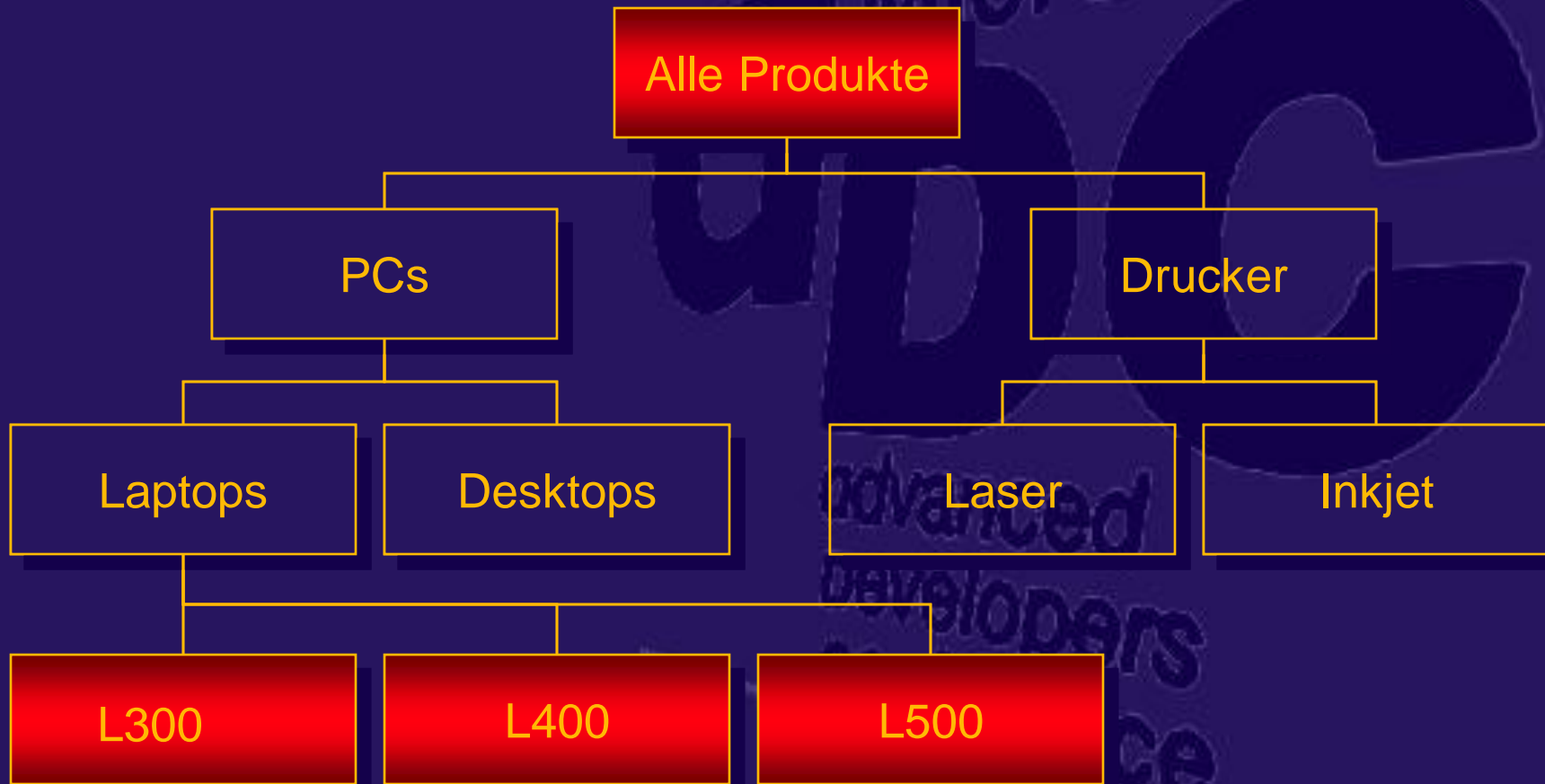
Hierarchien (1)

- Dimensionen sind hierarchisch strukturiert
- Hierarchien sind unverzichtbar, um die täglich anfallenden Datenmengen zu gruppieren und auf abstrakterem Niveau zu betrachten

Hierarchien (2)



Sekundäre Hierarchien



Sekundäre Hierarchien



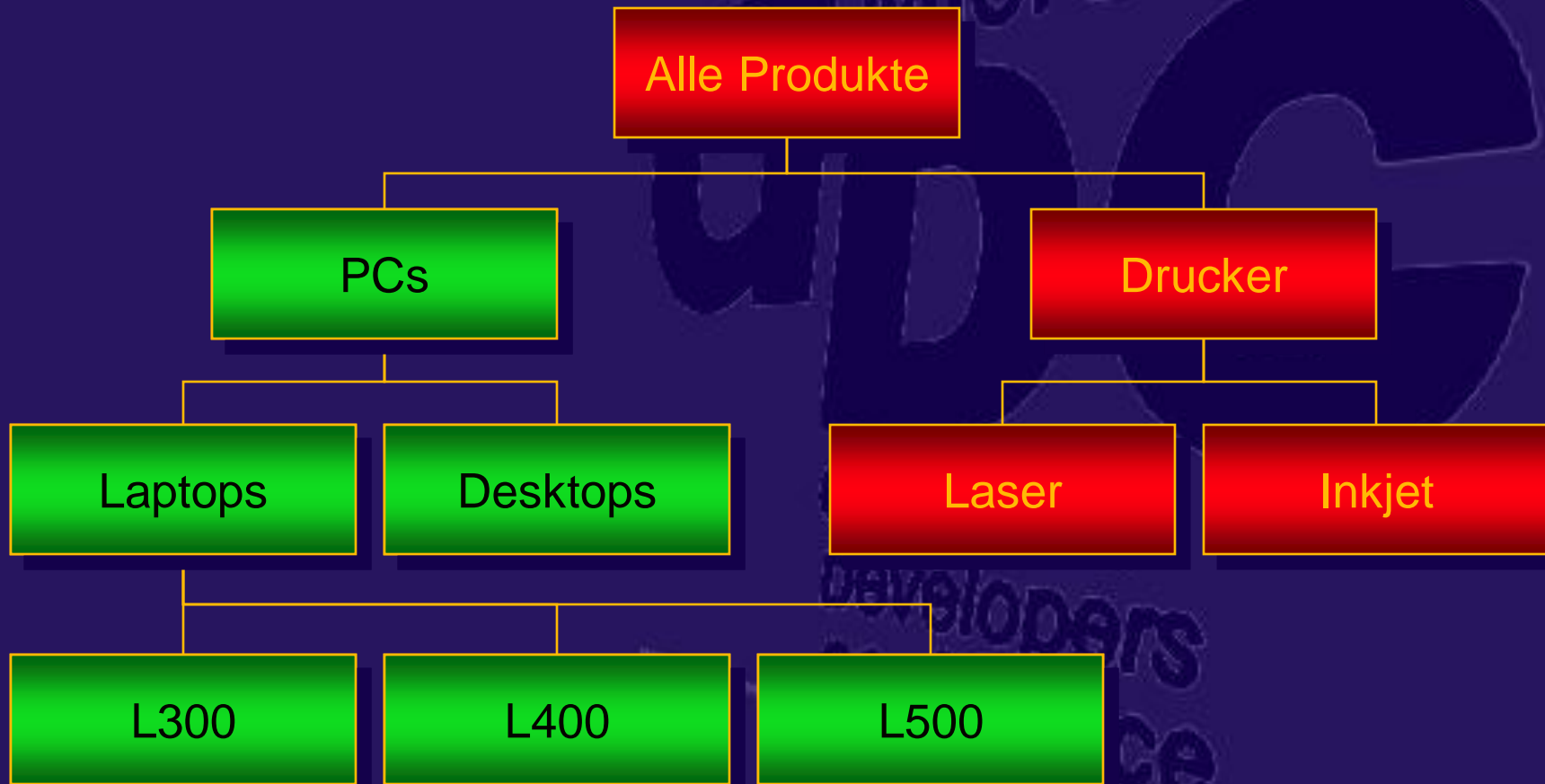
Meßwerte (Fakten)

- Dimensionshierarchien legen die Struktur einer OLAP-Datenbank fest, Meßwerte (engl. Measures) die zu speichernden Werte
- Beispiele: Umsatz, Stückzahlen oder Kosten
- Measures können auch berechnet werden
(z.B. Gewinn = Umsatz - Kosten)

Aggregationen (1)

- Vorbereitung von Aggregationen ermöglicht konstant kurze Antwortzeiten
- Dimensionshierarchien stellen die Basis für die Aggregationen dar
- Beispiel: Der Umsatz für „Laptops“ entspricht der Summe der Umsätze für die Modelle „L300“, „L400“, „L500“

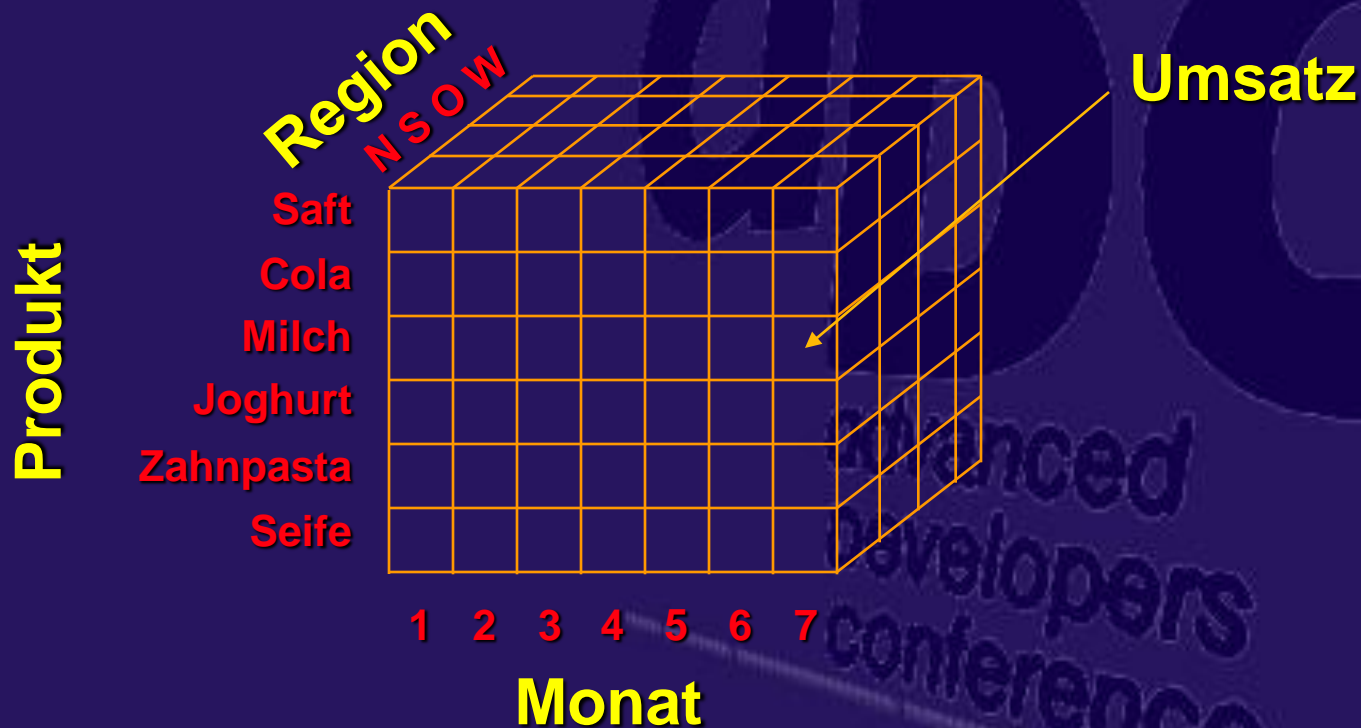
Aggregationen (2)



Würfel (Cubes)

- **Sammlung von Dimensionen und Measures**
- **Aufgebaut nach Hierarchien**
- **Kann aus 1, 2, 3, 25 oder mehr Dimensionen bestehen**
- **Aus Anwendersicht Bezeichnung für eine Art OLAP-Tabelle (z.B. Marketing-Cube, Produktions-Cube, etc.)**

Würfel (Cubes)



Begriffe aus dem OLAP Umfeld

- **Data Warehousing**

- Prozess zur Abbildung aller Daten im Unternehmen in einem zentralen, vereinheitlichten Datenpool

- **Data Marts**

- Fachspezifisches Data Warehouse (z.B. auf Abteilungsebene)

- **Data Mining**

- (Teilweise) automatisiertes Auffinden von Mustern, Ausnahmen und Trends in einem Data Warehouse bzw. Data Mart

Eine kleine Demo



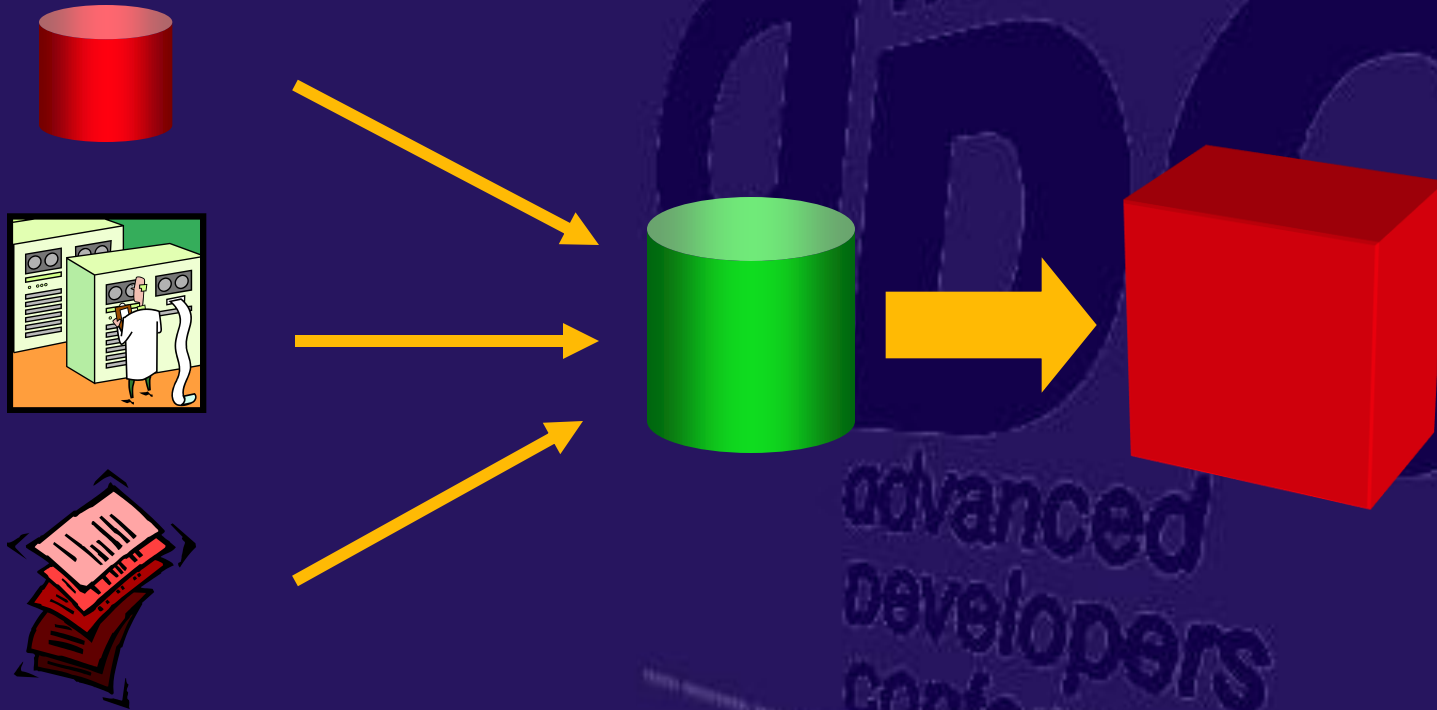
OLAP: Technische Realisierung

- **Operationale Daten werden kopiert und denormalisiert (→ ETL-Prozess)**
- **Hohe Redundanzen und vorberechnete Aggregationen ermöglichen schnelle Abfragen, unabhängig von der Datenmenge**
- **OLAP Server ist für multidimensionale Daten und Analyse-Anwendungen optimiert**

Technische Plattform: Microsoft SQL Server 7.0

- **Leistungsfähiges Backend, auch für sehr große Datenmengen**
- **Optimiert für denormalisierte Datenmodelle (Star-Schema)**
- **Enthält Tools für den ETL-Prozeß (Data Transformation Services)**
- **Enthält OLAP Server (Microsoft OLAP Services)**

Überblick



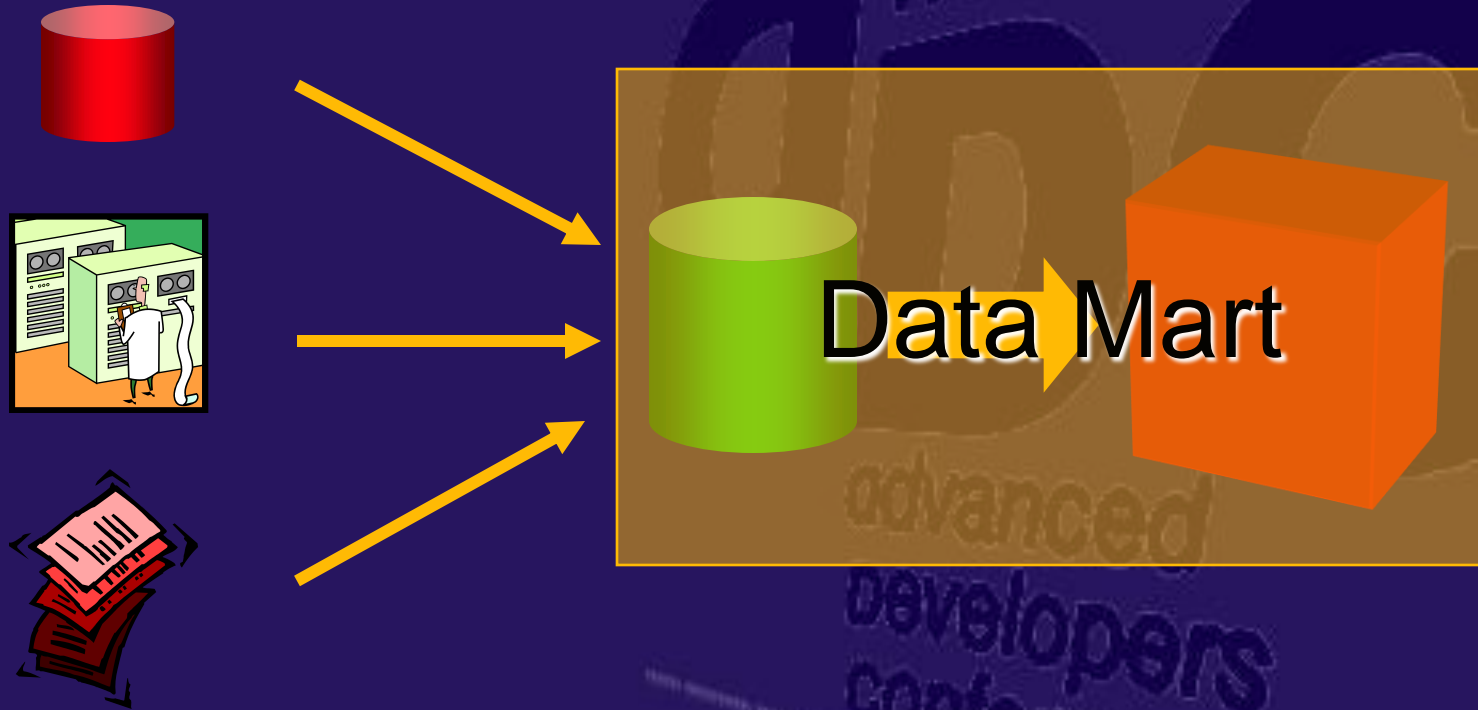
Operationale Systeme

ETL (DTS)

Star Schema (MS SQL 7.0)

OLAP Cube (MS OLAP Services)

Überblick



Operationale Systeme ETL (DTS) Star Schema (MS SQL 7.0) OLAP Cube (MS OLAP Services)

ETL-Prozess: Extraktion

- **Microsoft Data Transformation Services (DTS) können Daten aus heterogenen Quellen lesen**
- **Zugriff über OLE DB**
- **Für nahezu jede Datenquelle existiert eine OLE DB Anbindung**

ETL-Prozess: Transformation

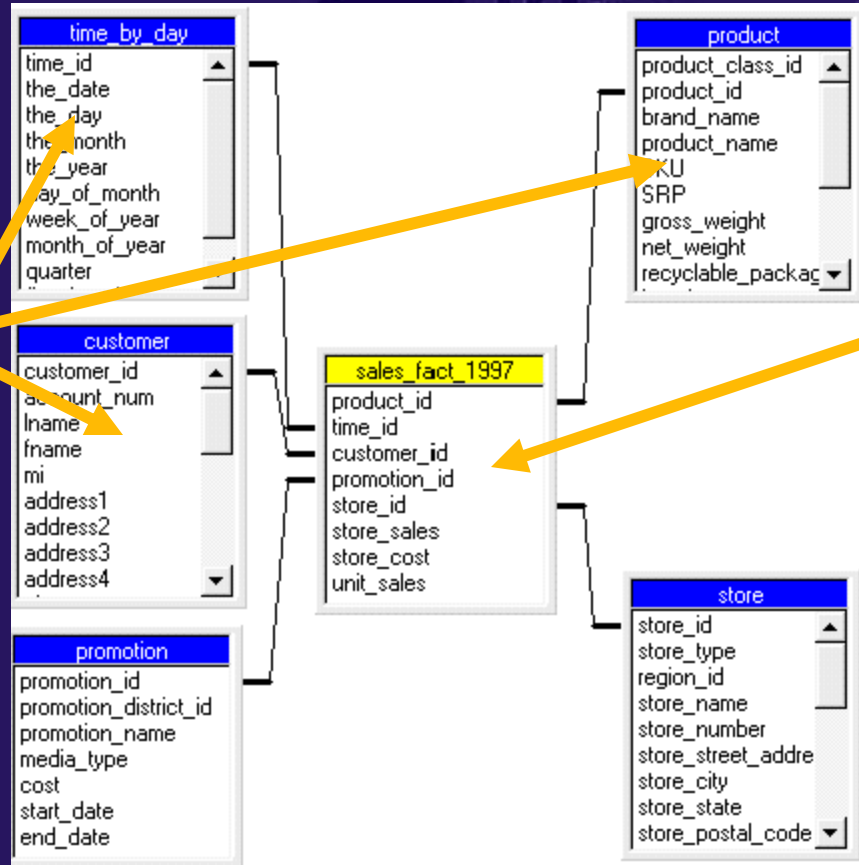
- Konvertierung von Datentypen
- Vereinheitlichung der Daten
- Data Cleansing
- Datenanreicherung
- Denormalisierung (→ Star Schema)
- Archivierung (z.B. Erstellung künstlicher Primary Keys)

ETL-Prozess: Laden

- **Operationale Daten wurden durch Extraktion und Transformation vorbereitet**
- **Laden des Data Marts und der OLAP Cubes muß jetzt schnell gehen!!**
- **BULK INSERT der vorbereiteten Daten ins Star-Schema**

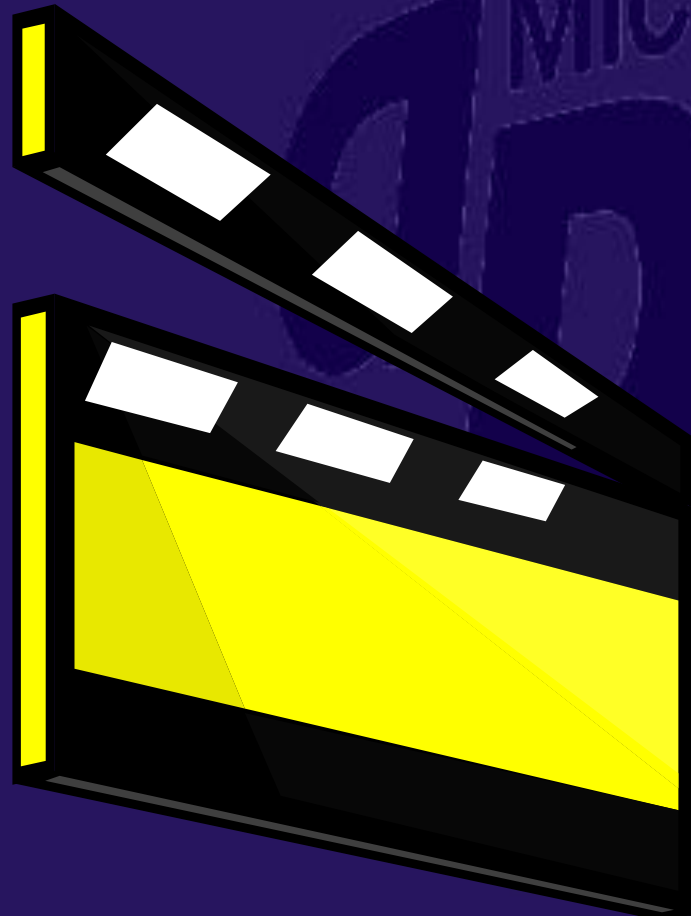
Ergebnis: Star-Schema

Dimensions-
Tabellen



Fakten-
Tabelle
(Detail-
daten)

Eine kleine Demo ...



Erstellung der OLAP-Cubes

- Aus einem Star-Schema werden in der Regel mehrere OLAP-Cubes „gefüllt“
- Ein Cube sollte max. 8-9 Dimensionen plus Meßwerte enthalten (Übersichtlichkeit!)
- Je nach Speicheroption werden die Daten der Fakten-Tabelle u.U. nochmals in den Cube kopiert!

OLAP Speicheroptionen

- **Multidimensionales OLAP (MOLAP)**
 - benutzt multidimensionalen Datenspeicher, um Detaildaten und Aggregationen zu speichern
- **Relationales OLAP (ROLAP)**
 - benutzt relationalen Datenspeicher, um Detaildaten und Aggregationen zu speichern
- **Hybrides OLAP (HOLAP)**
 - Detaildaten werden relational, Aggregationen multidimensional gespeichert (MOLAP+ROLAP)

Gemeinsamkeiten aller Methoden

- OLAP-Daten sind generell Kopien der operationalen Daten und werden in einem separaten System gespeichert und verwaltet
 - Performance
 - Verschiedene, heterogene Datenquellen
 - Data Cleansing
 - Vereinheitlichung der Daten
 - Zeitliche Koordination
 - Archivierung

Database Explosion (1)

- **Extremes Wachstum von OLAP-Datenbanken mit zunehmender Anzahl an Dimensionen**
- **Problem: OLAP Systeme erreichen schnelle Antwortzeiten nur, wenn Aggregationen vorberechnet werden**
- **Für optimale Performance generell alle möglichen Aggregationen vorberechnen?**

Database Explosion (2)

Umsatz		Region				
		Nord	Süd	West	Ost	Summe
Produkte	Home PC	10.000	8.000	12.000	7.000	37.000
	Laptop	2.000	5.000		2.000	9.000
	Server			20.000		20.000
	PDA		3.000		1.000	4.000
	Summe	12.000	16.000	32.000	10.000	70.000

- 2 Dimensionen, je 4 Elemente (16 mögliche Kombinationen), nur 1 Hierarchiestufe
- 10 Detailfelder sind mit Umsatzzahlen gefüllt (leere Datenfelder werden nicht gespeichert)
- Mit den Aggregationen enthält die OLAP-Datenbank aber insgesamt 19 Datenfelder (Wachstumsfaktor = 1,9!!)

Database Explosion (3)

- Je mehr Dimensionen und Hierarchiestufen, desto größer fällt der Wachstumsfaktor aus!
- Extremfall: OLAP Cube wächst im Vergleich zu den Detailsätzen um den Faktor 3.000

Database Explosion (4)

- **Gegenmittel: Einstellbarer Aggregationsgrad**
- **Meist reichen 20% - 30% Aggregationsgrad für gute Performance**
- **Microsoft OLAP Server bietet hierfür den Aggregation Wizard**

Building Killer Applications!