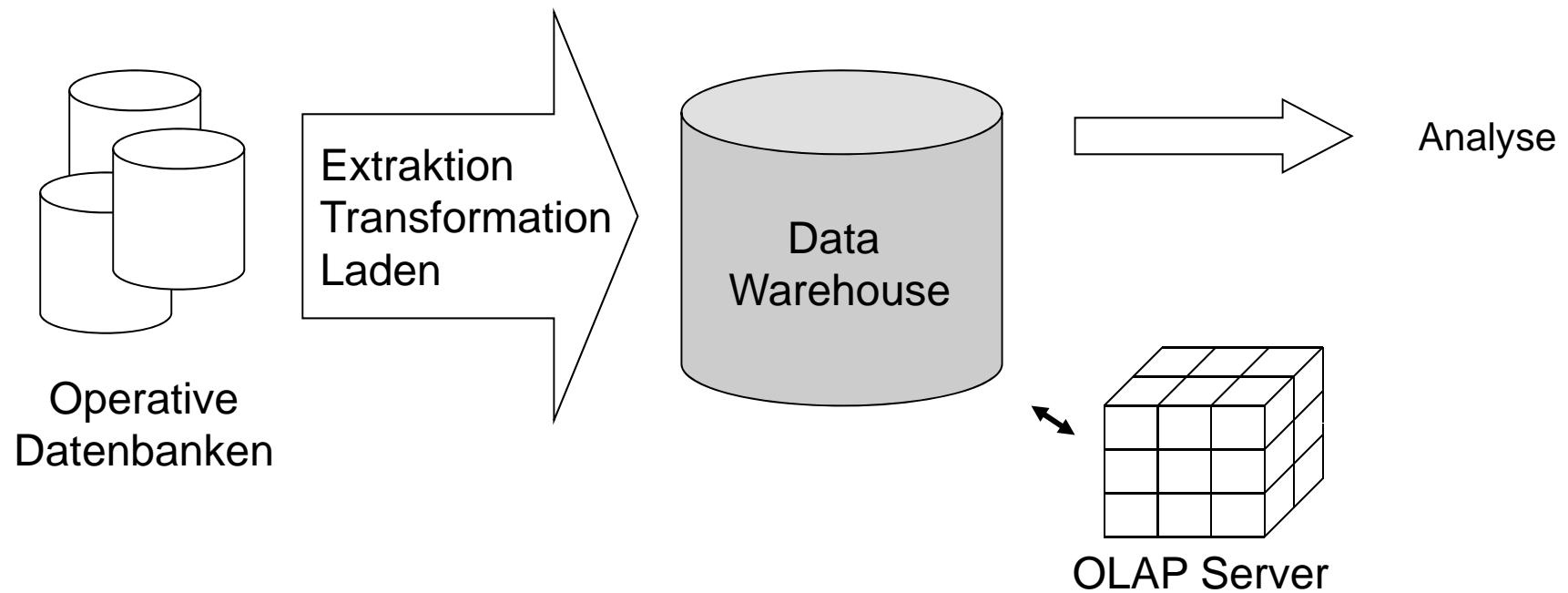


Data Warehousing (II): Multidimensionales Datenmodell, Cognos

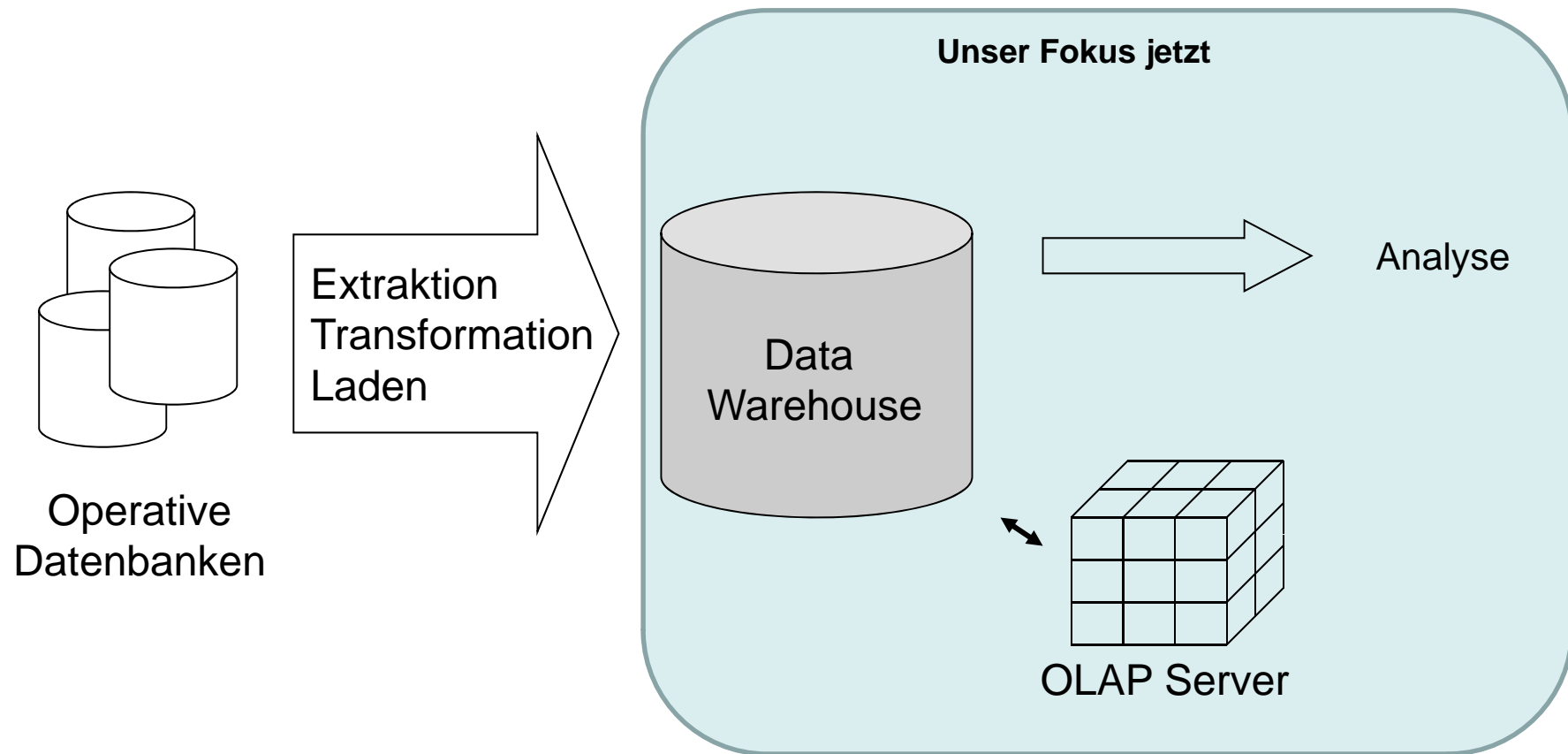
Praktikum: Data Warehousing und Mining

Vereinfachte Sicht auf die Referenzarchitektur

Folie aus der letzten Vorlesung



Fokus jetzt im Praktikum



Agenda

- **Online Analytical Processing (OLAP)**
 - Multidimensionales Datenmodell
 - Konzeptionelle Modellierung
 - Relationale Umsetzung des multidim. Modells
 - Multidimensionale Operatoren
- Cognos
- Hinweise zur Bearbeitung

Analysephase

- Unterschiedliche Ansätze:
 - Data Mining – **Erster Block**
 - Suche nach unbekanntem Mustern im Datenbestand
 - „Welche Eigenschaften zeichnen Kunden aus, die wieder bei uns einkaufen werden?“
 - Online Transactional Processing (OLTP) – **Vorheriger Block**
 - Zugriff auf vorhandenen Datenbestand
 - Nutzung von Datenmanipulationssprachen (z.B. SQL)
 - „Wieviele Einheiten von Artikel X wurden vom Kunden Y im Zeitraum Z gekauft?“
 - Online Analytical Processing (OLAP) – **Jetzt**
 - Anpassung des Datenbestands an die Analyse
 - Dynamischer, flexibler, interaktiver Zugriff auf eine Vielzahl von Werten
 - „In welchem Gebiet macht Warengruppe X den größten Umsatz?“

Data Access

- Tools:
 - Anfragesprachen (z.B. SQL) – Vorheriger Block
 - Lesen von Daten
 - Arithmetische Operationen auf Daten
 - Zunächst keine Präsentationsmöglichkeit
 - Reporting Tools (z.B. Cognos) – Jetzt
 - Lesen der Daten
 - Anreicherung der Daten durch arithmetische Operationen
 - Präsentation der Daten in Berichten
 - Unterstützung von Ampelfunktionalität

Anforderungen an Online Analytical Processing

- Geschwindigkeit
 - Anfragen sollten in 5 Sekunden beantwortet sein
- Analysemöglichkeit
 - Ermöglichung anwenderfreundlicher und intuitiver Analyse
- Sicherheit
 - Sicherer Mehrbenutzerbetrieb
 - Stabile Sicherungsmechanismen
- Multidimensionalität
 - Multidimensionale Sicht auf die Daten
- Kapazität
 - Hohe Skalierbarkeit der verwalteten Daten

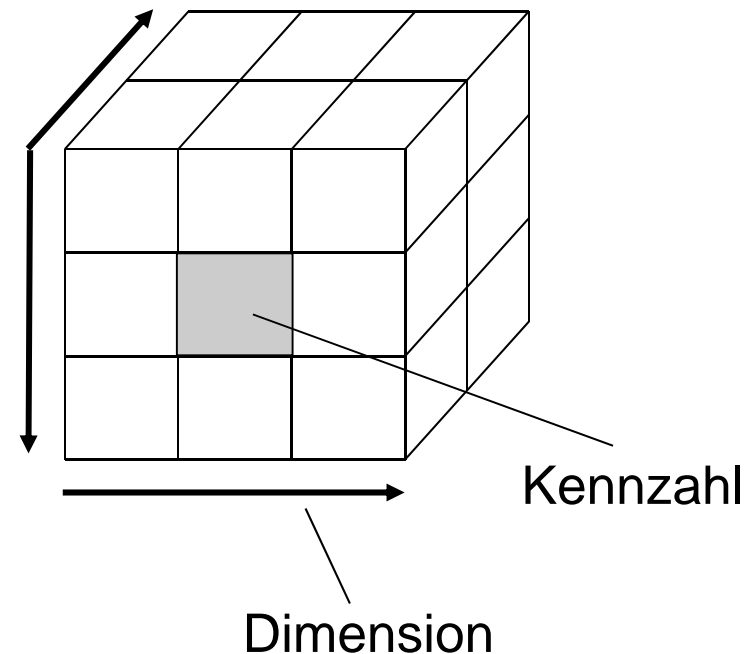
→ FASMI: Fast Analysis of Shared Multidimensional Information (Pendse/Creeth, 1995)

Agenda

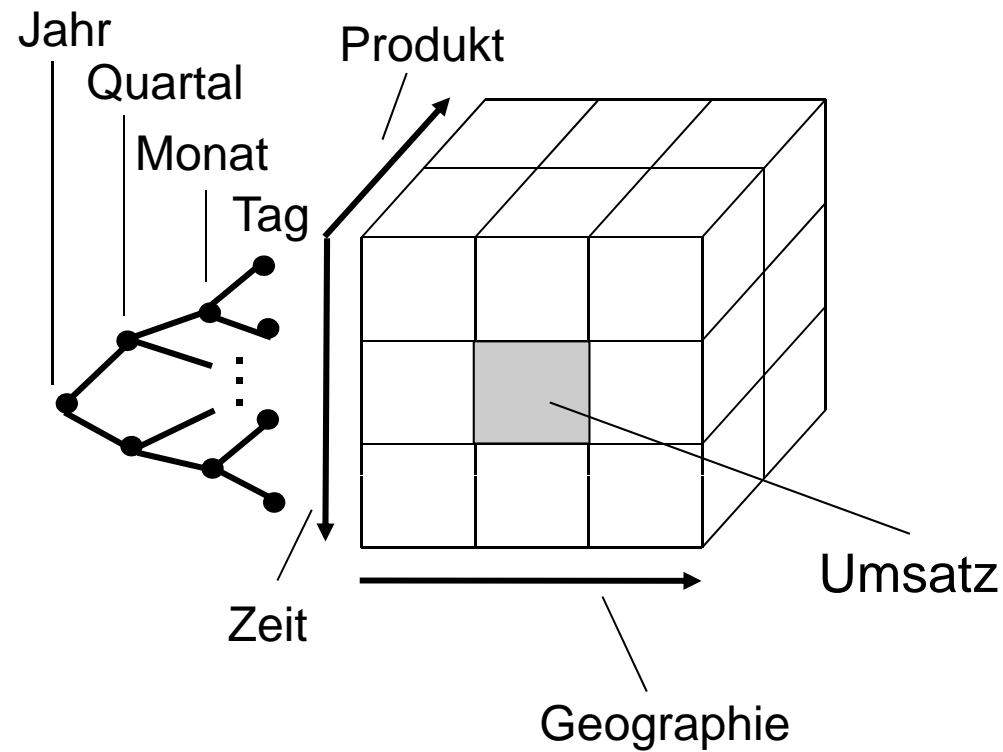
- Online Analytical Processing (OLAP)
 - **Multidimensionales Datenmodell**
 - Konzeptionelle Modellierung
 - Relationale Umsetzung des multidim. Modells
 - Multidimensionale Operatoren
- Cognos
- Hinweise zur Bearbeitung

Multidimensionales Datenmodell - Begriffe

- Hilfsmittel zur Veranschaulichung von Daten
- verschiedene Aspekte auf gleiche Weise zugreifbar
- Einsatz bei OLAP-Anwendungen
- Kennzahlen
 - Elemente eines Würfels
- Dimensionen
 - Beschreiben Daten
 - Ermöglichen Zugriff auf Kennzahlen
 - Können zu Hierarchien gehören



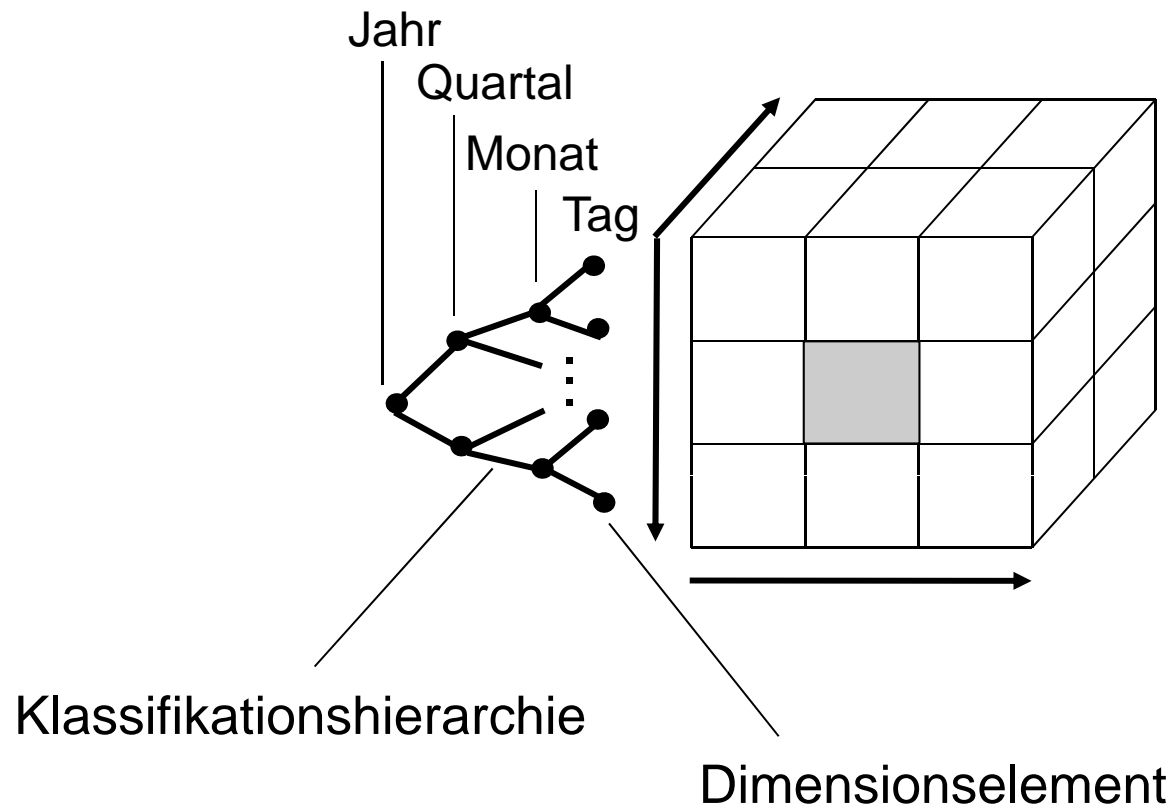
Multidimensionales Datenmodell – Beispiel



Dimensionen

- Einordnung
 - Bewertung der Analysedaten durch Kenngrößen (z.B. Umsatz, Kosten)
 - Untersuchung der Kenngrößen aus verschiedenen Perspektiven (z.B. Stadt, Bundesland, Zeitachse)
 - Betrachtungsperspektive heißt Dimension
- Eigenschaften
 - Mindestens 2 Dimensionselemente
 - Dimensionselemente bilden Blätter eines Baums (sog. Klassifikationshierarchie)

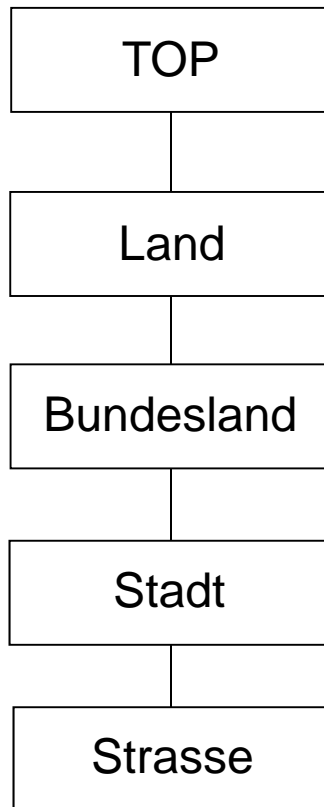
Dimensionen – Beispiel Zeit



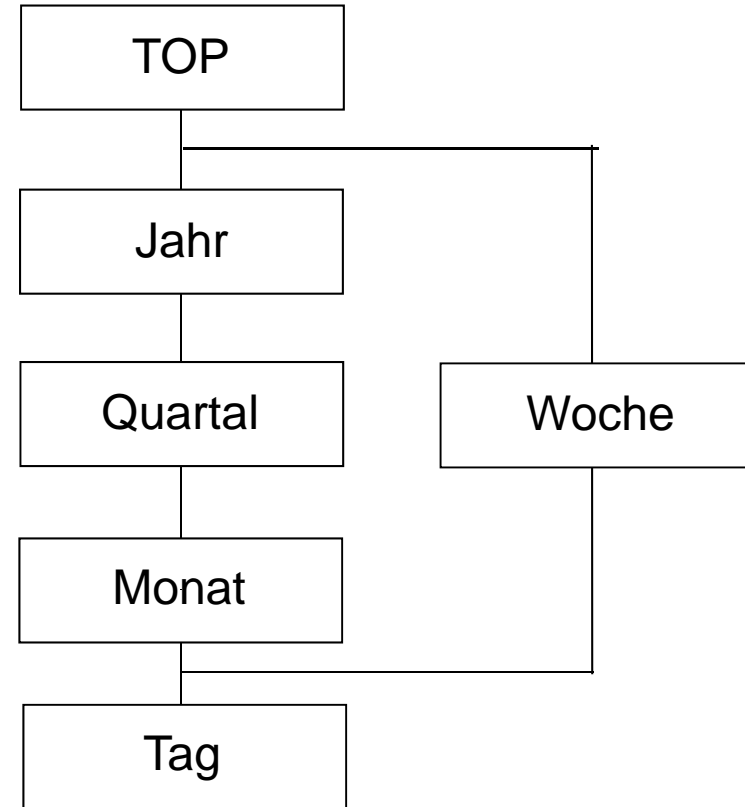
Arten von Klassifikationshierarchien

- Einfache Hierarchien
 - Höhere Hierarchieebenen enthalten die aggregierten Werte der jeweils niedrigeren Ebenen
 - Oberster Knoten: Gesamtknoten
 - Verdichtung aller Werte einer Dimension
- Parallele Hierarchien
 - Entstehen bei unterschiedlicher Art der Gruppierung
 - Parallele Äste ohne Beziehung
 - Betrachtung eines Teilaspekts der Hierarchie pro Ast

Klassifikationshierarchie - Beispiele



Einfache Hierarchie



Parallele Hierarchie

Weitere Begriffe

- Würfel
 - Kanten von Dimensionen aufgespannt
 - Kantenlänge entspricht Anzahl der Elemente in Dimension
 - Eine oder mehrere Kennzahlen pro Würfelzelle
 - Anzahl der Dimensionen heißt Dimensionalität
- Konsolidierungspfad
 - Pfade im Klassifikationsschema

Agenda

- Online Analytical Processing (OLAP)
 - Multidimensionales Datenmodell
 - **Konzeptionelle Modellierung**
 - Relationale Umsetzung des multidim. Modells
 - Multidimensionale Operatoren
- Cognos
- Hinweise zur Bearbeitung

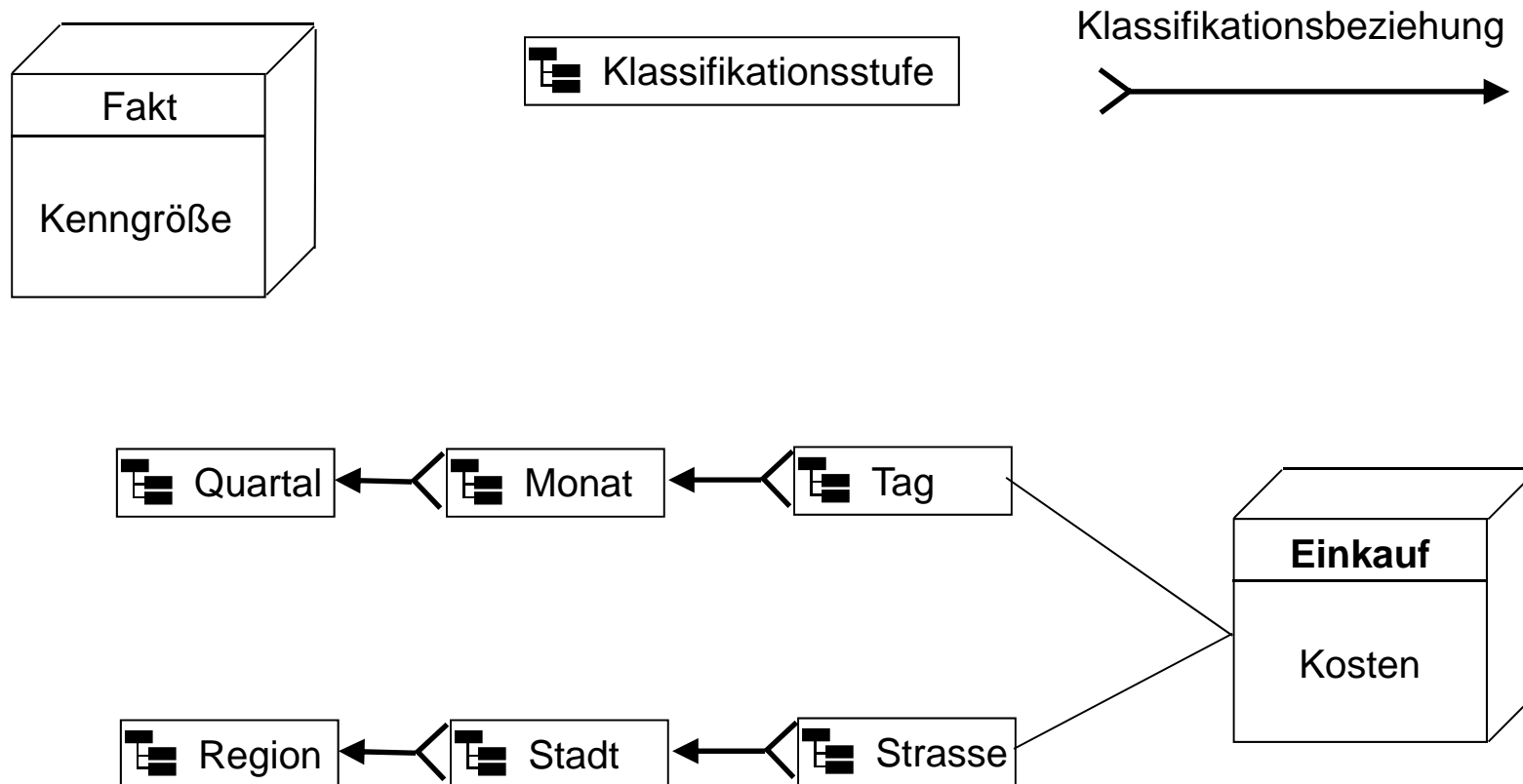
Konzeptionelle Modellierung

- Einsatz von Entity/Relationship-Modell oder UML
 - Probleme:
 - Modellierung der Konsolidierungspfade nicht möglich
 - Entitäten besitzen keine Semantik
 - Wir wollen aber: Höheren Automatisierungsgrad durch Verzicht auf universelle Anwendbarkeit
 - Unterscheidung zwischen Klassifikationsstufen, beschreibenden Attributen und Kennzahlen nicht möglich
- Daher eigene Modellierungsmodelle
 - Multidimensionales Entity/Relationship-Modell (ME/R)
 - Multidimensionale Unified Modeling Language (mUML)
 - Ansatz von Totok (objektorientiert, notiert in UML)
 - ...
- Hier: ME/R

ME/R-Modell

- Weiterentwicklung des E/R-Modells
 - Anforderungen
 - Spezialisierung:
Alle eingeführten Elemente sind Spezialfälle von E/R-Konstrukten
 - Minimale Erweiterung:
Leicht erlernbar für erfahrene E/R-Modellierer
 - Darstellung der multidimensionalen Semantik:
Klassifikationsschema, Würfelstruktur muss abbildbar sein
 - Eingeführte Konstrukte
 - Entitätenmenge „Dimension Level“ (Klassifikationsstufe)
 - n-äre Faktenbeziehung
 - Binäre Klassifikationsbeziehungsmenge

Visualisierung der ME/R - Konstrukte



Agenda

- Online Analytical Processing (OLAP)
 - Multidimensionales Datenmodell
 - Konzeptionelle Modellierung
 - **Relationale Umsetzung des multidim. Modells**
 - Multidimensionale Operatoren
- Cognos
- Hinweise zur Bearbeitung

Relationale Umsetzung des multidim. Modells

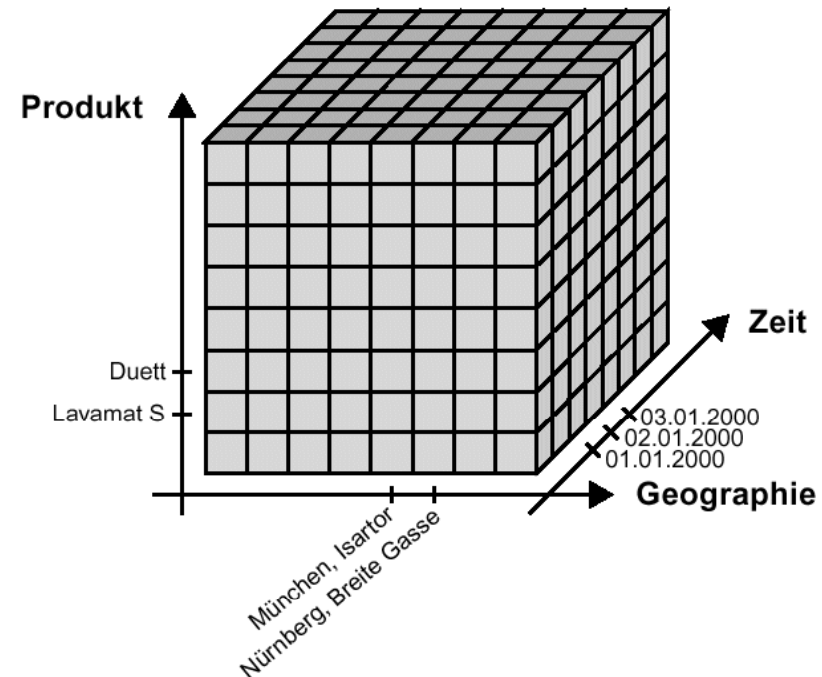
- Anforderungen
 - Beibehaltung der Semantik
 - z.B. Hierarchien
 - Effiziente Umsetzung von Anfragen
 - Effiziente Verarbeitung von Anfragen
 - Einfache Wartung
 - z.B. beim Nachladen von Daten

→ Relationales OLAP (ROLAP)

Relationale Umsetzung: Faktentabelle

- Beispiel zeigt einen typischen Datensatz
- Kennzahlen, Dimensionen \Rightarrow Spalten
- Zellen \Rightarrow Tupel
- Zusätzlich existieren Hierarchien
 - Z.B.: Artikel – Produktgruppe – Produktkategorie
- Wie kann der Datacube mit Hierarchien in einem DBMS gespeichert werden?

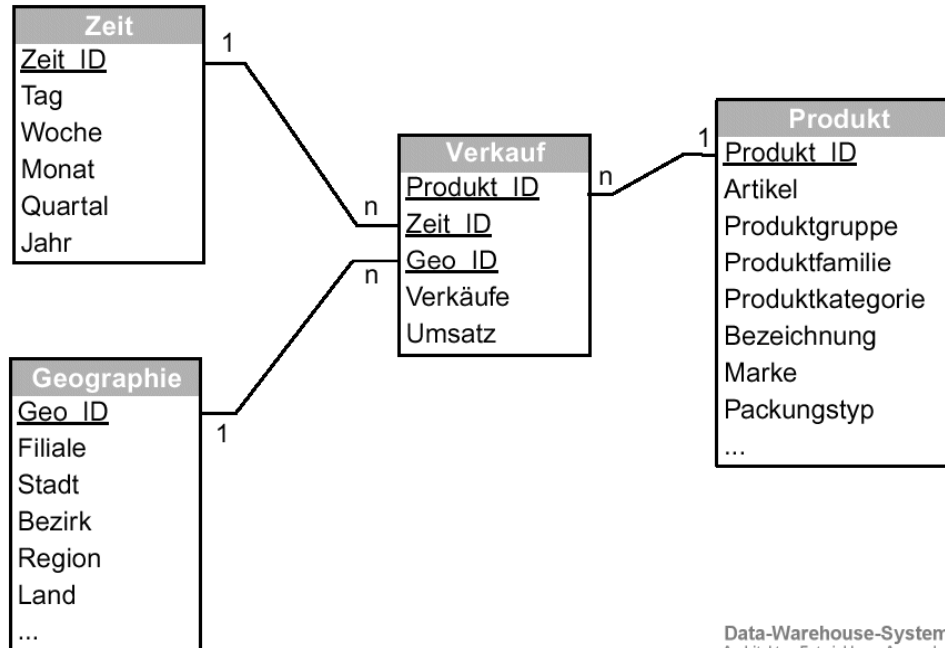
Data-Warehouse-Systeme; Architektur, Entwicklung, Anwendung; ISBN 3-932588-76-2



Artikel	Filiale	Tag	Verkäufe
Duett	Nürnberg, Breite Gasse	03.01.2000	7
Duett	München, Isartor	03.01.2000	3
Lavamat S	München, Isartor	03.01.2000	2

Faktentabelle ohne Hierarchien

Relationale Umsetzung: Star-Schema



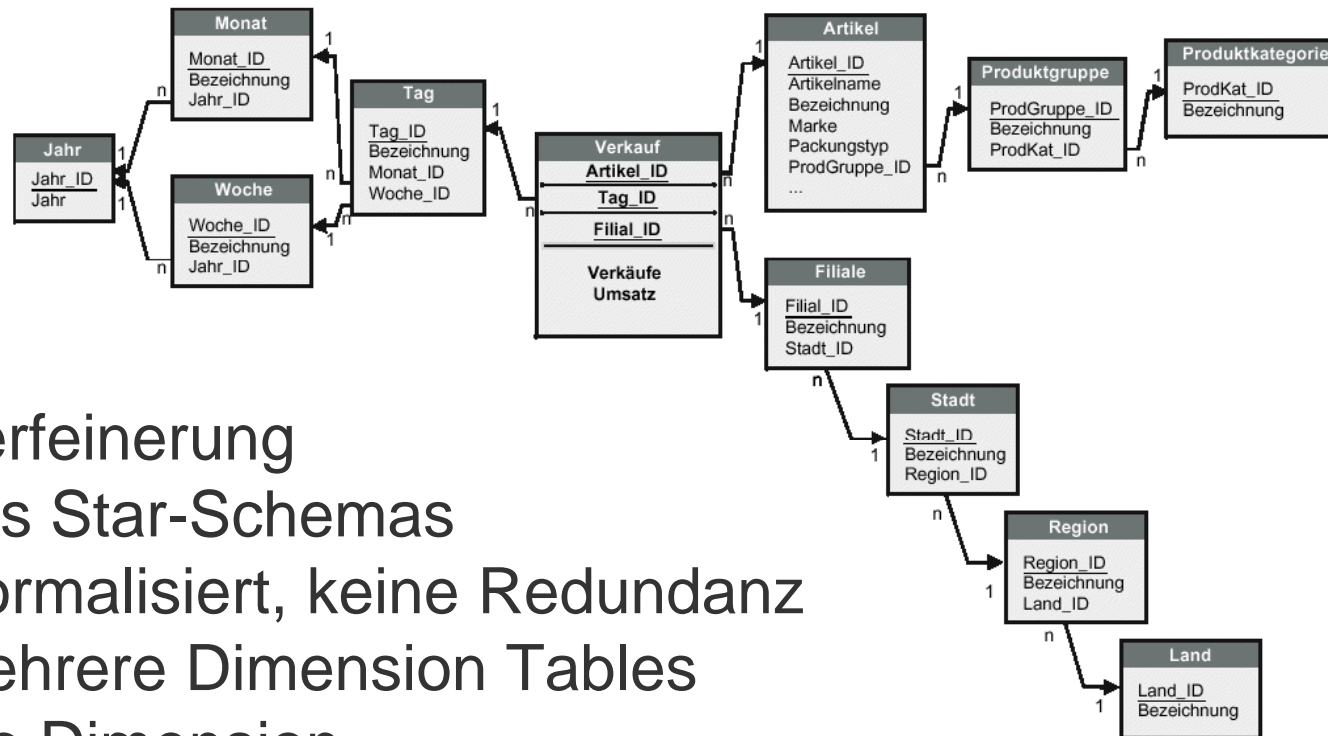
- Eine Relation pro Dimension
- Nicht normalisiert
 - Redundanz
 - Gefahr von Anomalien
- Nur wenige Joins notwendig
- Nachteile werden in Warehouses oft in Kauf genommen

Data-Warehouse-Systeme
Architektur, Entwicklung, Anwendung
ISBN 3-932588-76-2

<u>Produkt ID</u>	Artikel	...	Produktgruppe	Produktfamilie	Produktkategorie	...
1235	Lavamat S	...	Waschmaschinen	Waschgeräte	weiße Ware	...
1236	Duett	...	Waschmaschinen	Waschgeräte	weiße Ware	...
1237	Novotronic	...	Trockner	Waschgeräte	weiße Ware	...
1238	Vento 500	...	Trockner	Waschgeräte	weiße Ware	...
...

Data-Warehouse-Systeme; Architektur, Entwicklung, Anwendung; ISBN 3-932588-76-2

Relationale Umsetzung: Snowflake-Schema



- Verfeinerung des Star-Schemas
- Normalisiert, keine Redundanz
- Mehrere Dimension Tables pro Dimension
 - Relation pro Ebene einer Hierarchie
- Viele Joins:
 - 11 Tabellen bei Gruppierung nach Kategorie, Land und Jahr

Relationale Umsetzung: Semantikverluste

- Verluste in...
 - Faktentabelle
 - Unterscheidung von Dimensionen und Kenngrößen nicht ersichtlich
 - Dimensionstabelle
 - Unterscheidung zwischen beschreibendem Attribut und Attribut der Klassifikationsebene nicht möglich
 - Aufbau der Dimensionen geht verloren
- Lösung:
 - Erweiterung des Systemkatalogs in relationalen DBMS
 - Multidimensionales OLAP (MOLAP)
 - Hybrides OLAP (HOLAP)

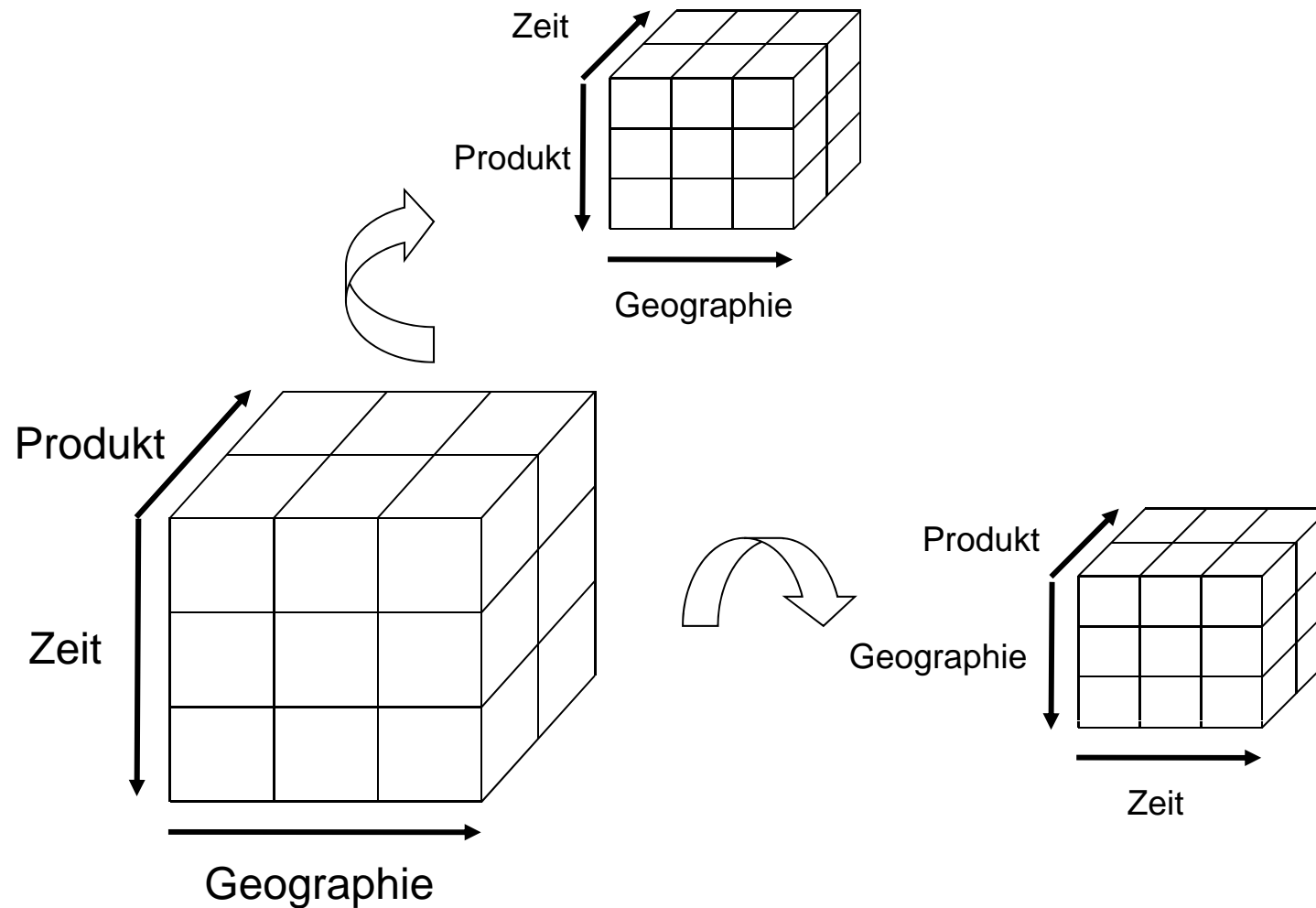
Überblick: OLAP-Techniken

- Relationales OLAP (ROLAP):
 - Relationale Datenbank
 - Skalierbarkeit gegeben, aber oft langsam (viele Joins)
- Multidimensionales OLAP (MOLAP):
 - Multidimensionale Datenbank
 - Persistente Speicherung aggregierter Kennzahlen
 - Oft schneller, aber „curse of dimensionality“
- Hybrides OLAP (HOLAP):
 - Mischform, Tradeoff: Geschwindigkeit vs. Flexibilität
- Desktop OLAP (DOLAP):
 - Lokale multidimensionale Analysen z.B. für Mobile User
- Memory-Based OLAP:
 - Datenhaltung komplett im RAM (Beispiel: QlikView)

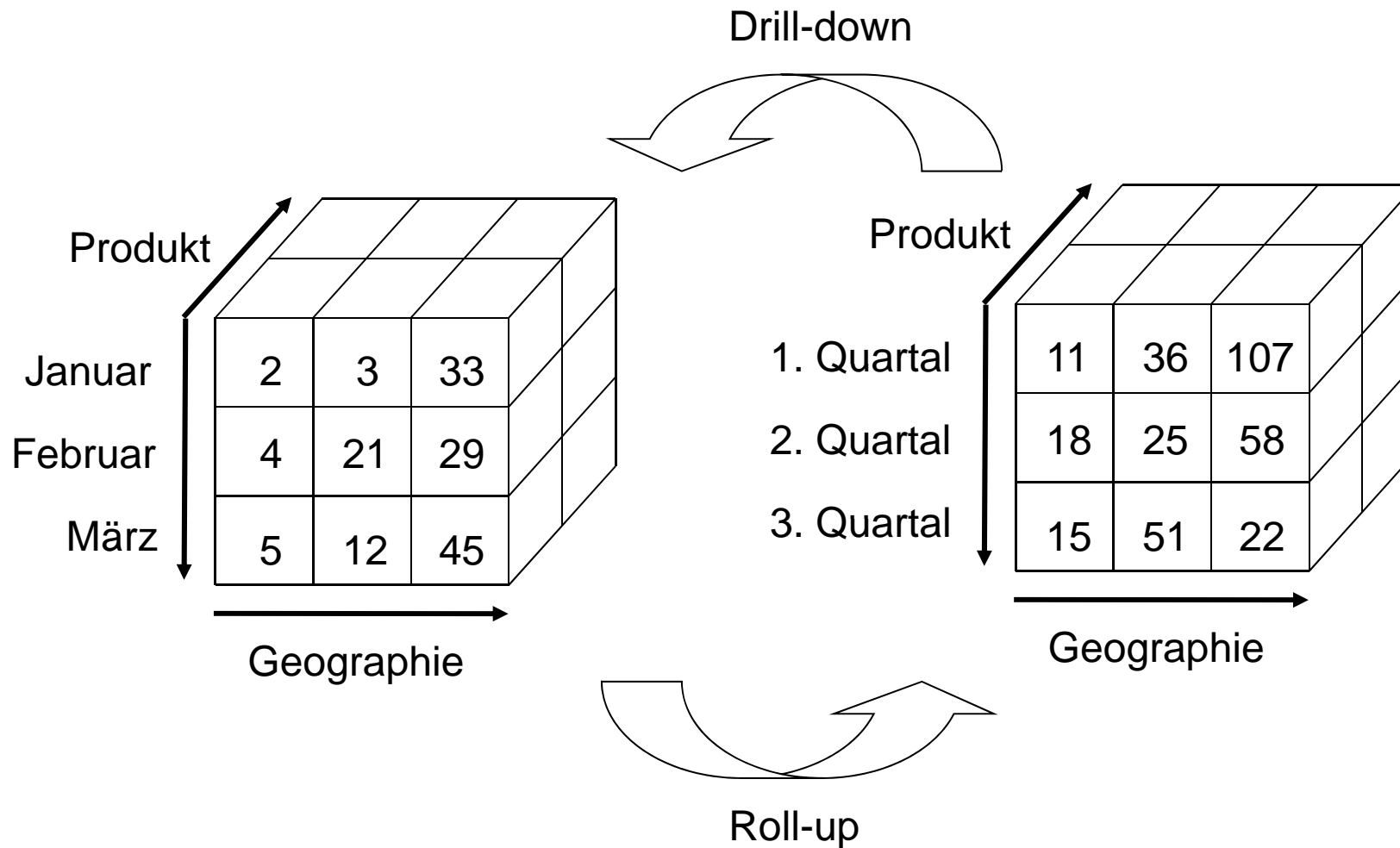
Agenda

- Online Analytical Processing (OLAP)
 - Multidimensionales Datenmodell
 - Konzeptionelle Modellierung
 - Relationale Umsetzung des multidim. Modells
 - **Multidimensionale Operatoren**
- Cognos
- Hinweise zur Bearbeitung

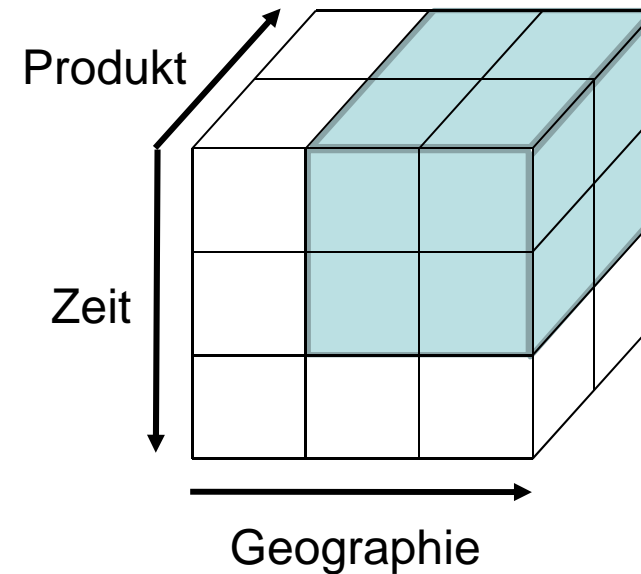
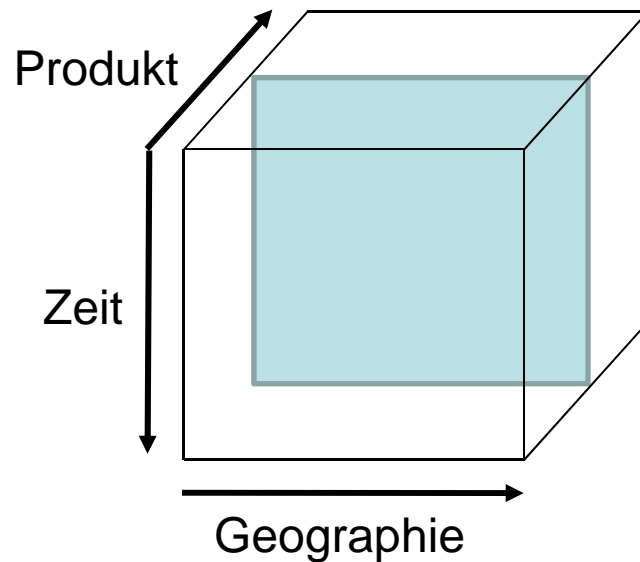
Pivotieren



Roll-up und Drill-down



Slicing und Dicing



- Slicing: Einschränkung entlang einer Dimension
- Dicing: Einschränkung entlang zweier oder mehr Dimensionen

Zusammenfassung: multidim. Operatoren

- Nutzen multidimensionale Struktur der Daten
- Erlauben einfache Modifikation von Anfragen
 - ...zu inhaltlich ähnlichen Anfragen
 - ...zu strukturell ähnlichen Anfragen
- Sind auf die Fragestellungen im Data Warehouse zugeschnitten

Rückblick: MDX (I)

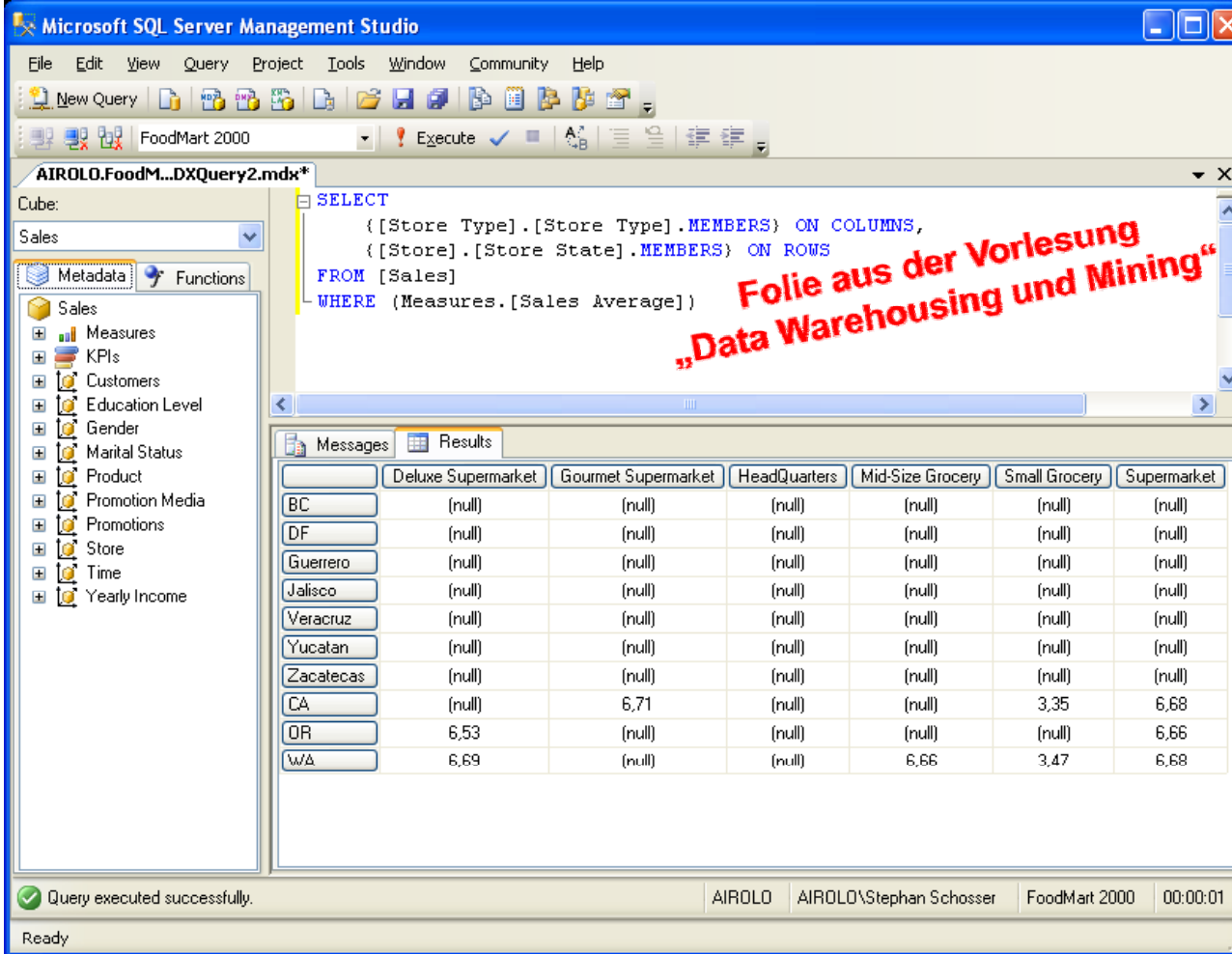
Kapitel 8 - Gliederung

*Folie aus der Vorlesung
„Data Warehousing und Mining“*

Operationen auf dem Cube: Multidimensional Expressions (MDX)

- Was ist MDX?
- Das mehrdimensionale Datenmodell
 - Microsoft Terminologie/ Konzepte
- MDX – Basiskonstrukte
- MDX – Weiterführende Konstrukte
 - Named Sets und Calculated Members
 - Hierarchisches Navigieren
 - Zeitserien
 - Tuples und CROSSJOIN
 - Filtern und Sortieren
 - Top und Bottom Analysen
 - Numerische und bedingte Abfragen
- Fazit

Rückblick: MDX (II)



The screenshot shows the Microsoft SQL Server Management Studio interface. The query editor contains the following MDX query:

```
SELECT
  ([Store Type].[Store Type].MEMBERS) ON COLUMNS,
  ([Store].[Store State].MEMBERS) ON ROWS
FROM [Sales]
WHERE (Measures.[Sales Average])
```

The Results pane displays a table with the following data:

	Deluxe Supermarket	Gourmet Supermarket	HeadQuarters	Mid-Size Grocery	Small Grocery	Supermarket
BC	(null)	(null)	(null)	(null)	(null)	(null)
DF	(null)	(null)	(null)	(null)	(null)	(null)
Guerrero	(null)	(null)	(null)	(null)	(null)	(null)
Jalisco	(null)	(null)	(null)	(null)	(null)	(null)
Veracruz	(null)	(null)	(null)	(null)	(null)	(null)
Yucatan	(null)	(null)	(null)	(null)	(null)	(null)
Zacatecas	(null)	(null)	(null)	(null)	(null)	(null)
CA	(null)	6,71	(null)	(null)	3,35	6,68
QR	6,53	(null)	(null)	(null)	(null)	6,66
\w\Δ	6,69	(null)	(null)	6,66	3,47	6,68

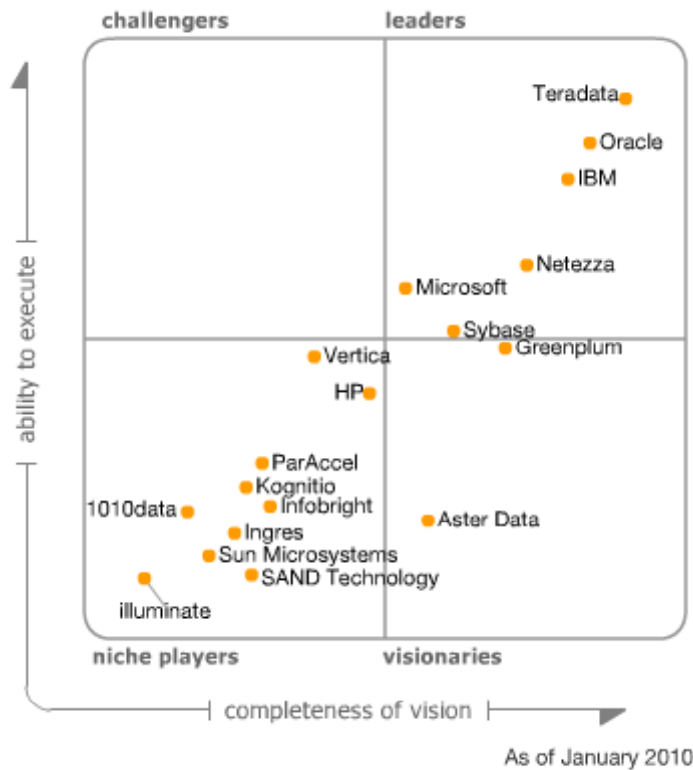
Messages: Query executed successfully.

Ready

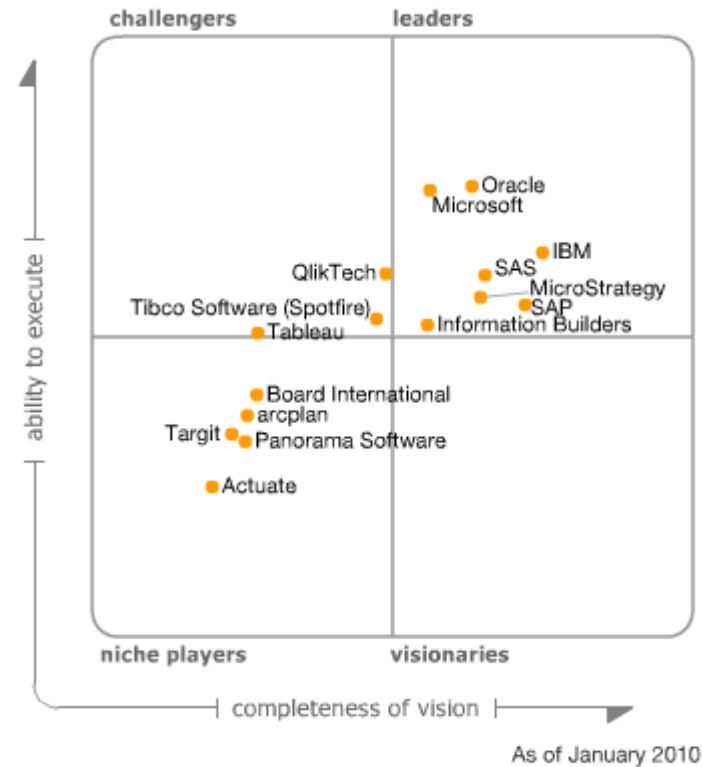
Agenda

- Online Analytical Processing (OLAP)
 - Multidimensionales Datenmodell
 - Konzeptionelle Modellierung
 - Relationale Umsetzung des multidim. Modells
 - Multidimensionale Operatoren
- Cognos
 - Cognos Report Studio
 - Cognos Analysis Studio
 - Internals
- Hinweise zur Bearbeitung

Der Business-Intelligence-Markt 2010



Magic Quadrant for Data Warehouse Database Management Systems, Gartner 2010



Magic Quadrant for Business Intelligence Platforms, Gartner 2010

Cognos

- Übernahme durch IBM 2008 („Cognos, an IBM Company“)
- Cognos BI ermöglicht Erstellen von
 - Ad-hoc-Anfragen (Query Studio)
 - Berichten (Report Studio)
 - Multidimensionalen Anfragen (Analysis Studio)
- Backend
 - Data Cube
 - Relationale Daten
- Anmeldung
 - Internet-Explorer
 - URL: <http://i40virt01.ipd.kit.edu/cognos8>
 - Benutzername: <Benutzername>
 - Passwort: <Passwort>

Cognos - Startbildschirm



The screenshot shows the Cognos 8 start screen. The top navigation bar is dark blue with the text "COGNOS® 8" on the left, "DBKurs30" and "Abmelden" on the right. The main content area is divided into two columns. The left column has a header "Eigene Inhalte" and contains two items: "Eigene Startseite" with a house icon and "Cognos-Inhalt" with a folder icon. Below this is a header "Administration" and one item: "Cognos-Inhalte verwalten" with a wrench and screwdriver icon. At the bottom left of the main area is a checkbox labeled "Diese Seite in Zukunft anzeigen" which is checked. The right column has a header "Eigene Aktionen" and contains five items: "Eigene Metriken verwalten" with a traffic light icon, "Eigene Daten abfragen" with a pen icon, "Eigenes Unternehmen analysieren" with a 3D cube icon, "Professionelle Berichte erstellen" with a set square icon, and "Eigene Ereignisse verwalten" with a circular arrow icon. A "Einführungstour" button is located at the bottom right of the main area. The footer is dark blue and contains copyright information on the left and the Cognos logo with the tagline "THE NEXT LEVEL OF PERFORMANCE™" on the right.

COGNOS® 8

DBKurs30 Abmelden

Eigene Inhalte

- Eigene Startseite
- Cognos-Inhalt

Administration

- Cognos-Inhalte verwalten

Diese Seite in Zukunft anzeigen

Eigene Aktionen

- Eigene Metriken verwalten
- Eigene Daten abfragen
- Eigenes Unternehmen analysieren
- Professionelle Berichte erstellen
- Eigene Ereignisse verwalten

Einführungstour

Copyright (C) 2007 Cognos Incorporated. Alle Rechte vorbehalten.
Cognos und das Cognos Logo sind Marken von Cognos Incorporated.

COGNOS
THE NEXT LEVEL OF PERFORMANCE™

Cognos-Komponenten

- Query Studio (für einfache Anfragen)
 - „In Echtzeit“
 - Visuell
- Report Studio (für komplexe Berichte)
 - Visuell
 - Ergebniserzeugung auf Befehl
 - Bedingte Variablen
 - Dynamische Berichte
 - Layout individueller anpassbar
- Analysis Studio (für multidimensionale Sicht)
 - „In Echtzeit“
 - Visuell

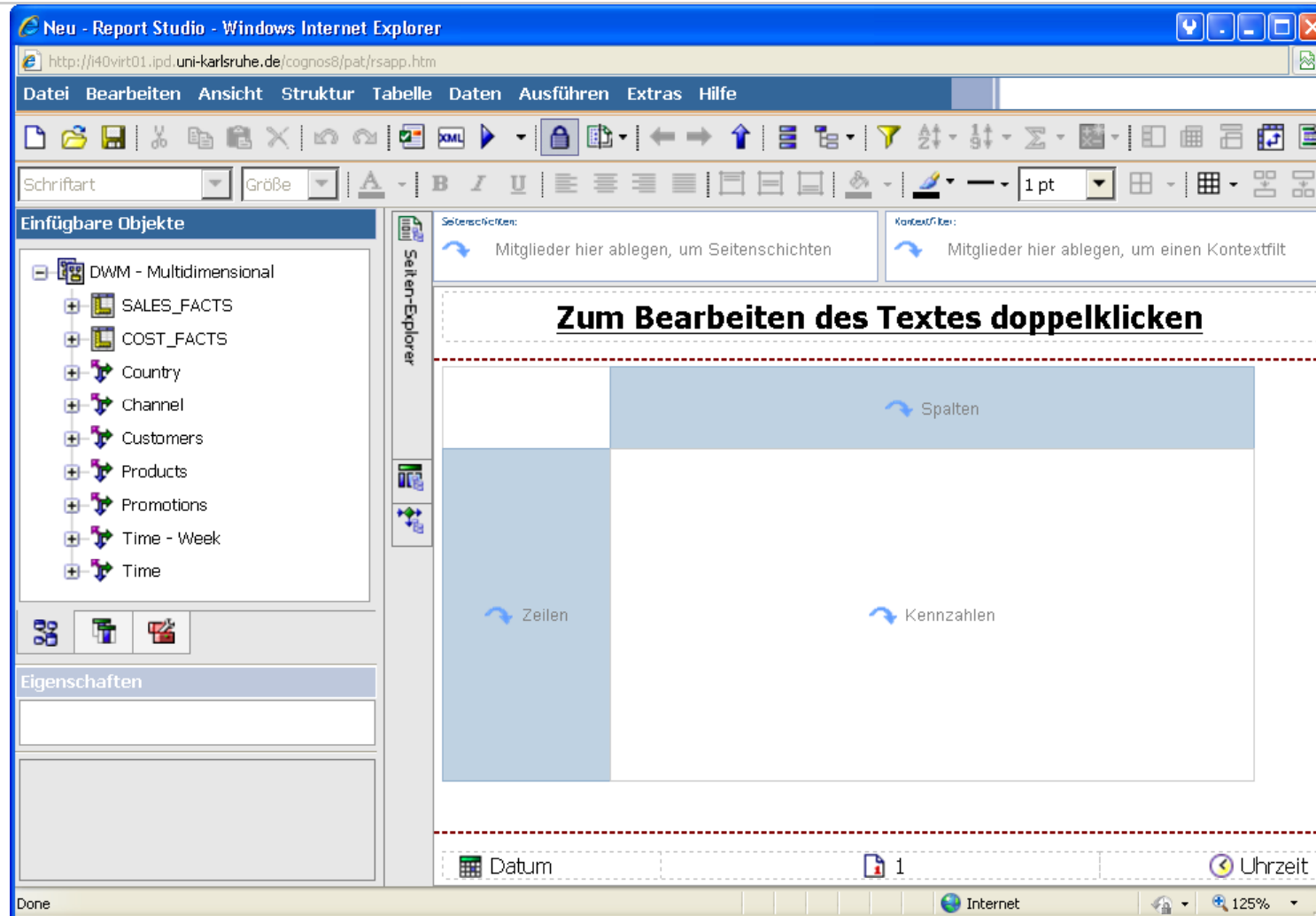
Agenda

- Online Analytical Processing (OLAP)
 - Multidimensionales Datenmodell
 - Konzeptionelle Modellierung
 - Relationale Umsetzung des multidim. Modells
 - Multidimensionale Operatoren
- Cognos
 - **Cognos Report Studio**
 - Cognos Analysis Studio
 - Internals
- Hinweise zur Bearbeitung

Report Studio

- Erstellen komplexer Berichte
- Vielfältige Ausgabe- und Darstellungsoptionen
- Parameterabfrage bei Anfrageaufruf

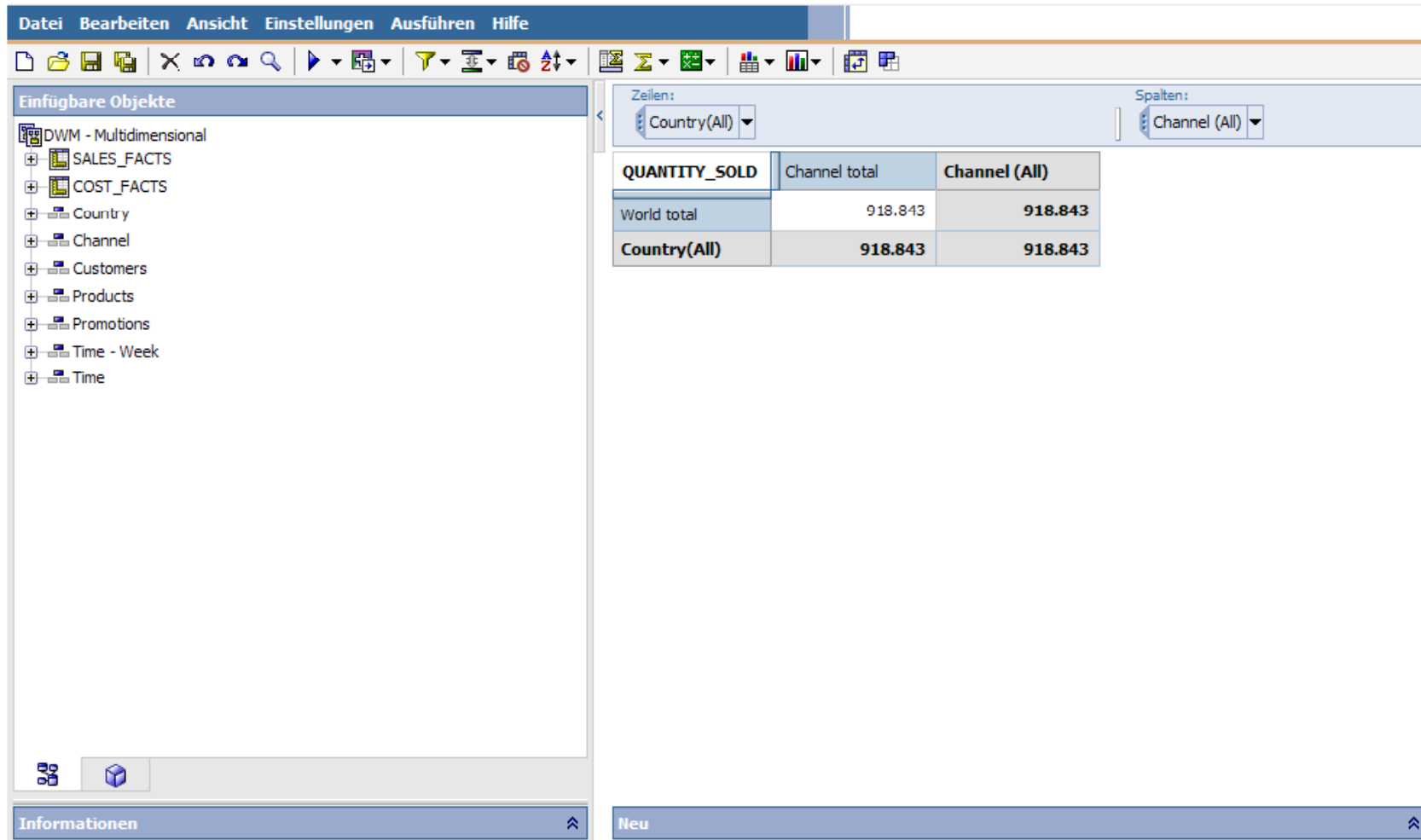
Report Studio - Aufbau



Agenda

- Online Analytical Processing (OLAP)
 - Multidimensionales Datenmodell
 - Konzeptionelle Modellierung
 - Relationale Umsetzung des multidim. Modells
 - Multidimensionale Operatoren
- Cognos
 - Cognos Report Studio
 - **Cognos Analysis Studio**
 - Internals
- Hinweise zur Bearbeitung

Analysis Studio – Aufbau

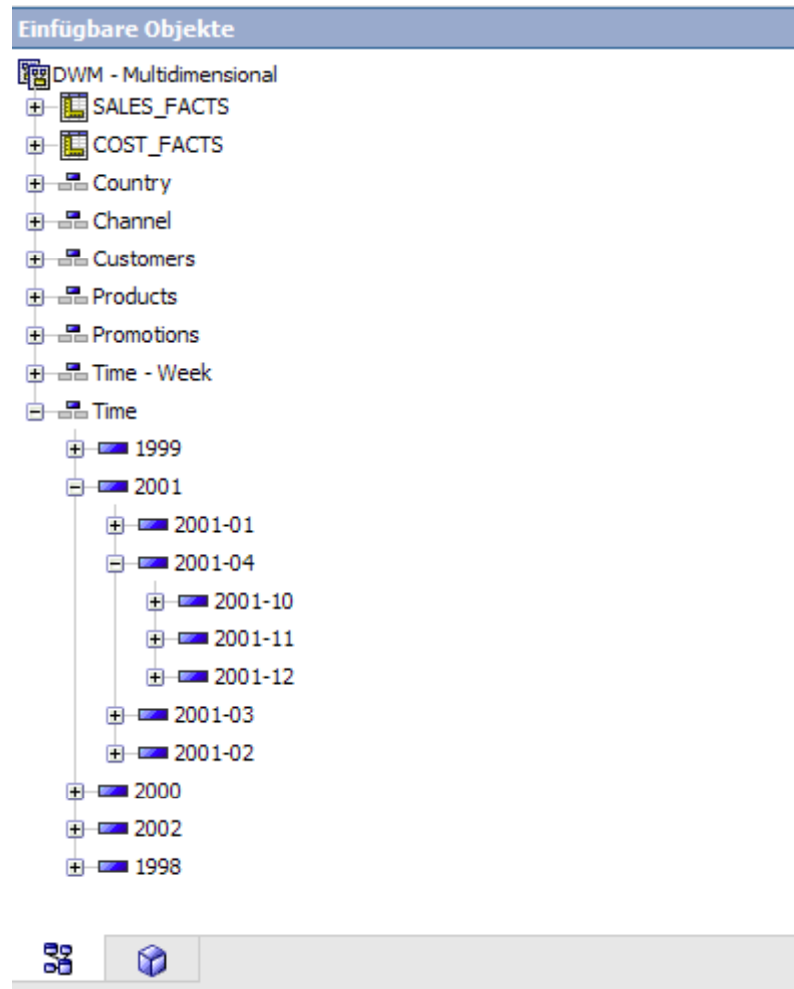


The screenshot shows the Analysis Studio interface with a menu bar (Datei, Bearbeiten, Ansicht, Einstellungen, Ausführen, Hilfe) and a toolbar. On the left, the 'Einfügbare Objekte' (Available Objects) pane lists a multidimensional model 'DWM - Multidimensional' with dimensions: SALES_FACTS, COST_FACTS, Country, Channel, Customers, Products, Promotions, Time - Week, and Time. The main workspace displays a pivot table with 'Country(All)' selected in the 'Zeilen:' (Rows) dropdown and 'Channel (All)' in the 'Spalten:' (Columns) dropdown. The table shows the following data:

QUANTITY_SOLD	Channel total	Channel (All)
World total	918.843	918.843
Country(All)	918.843	918.843

At the bottom, there are 'Informationen' and 'Neu' tabs.

Sicht auf die Daten



- Blättern in der Hierarchie möglich
- Ähnlich MDX Tools in SQL Server
- Unterscheidung zwischen Fakten und Dimensionen

Multidimensionale Operatoren

QUANTITY_SOLD	Channel total	Channel (A
Asia	85.455	
Middle East	7	
Americas	549.452	5
Oceania		
Europe		
World		

Als benutzerdefinierten Satz speichern...
Ausblenden
Beibehalten
Ausschließen
Sortieren
Berechnen
Spalten unterdrücken
Drilldown
Drillup
Gehe zu
Attribute anzeigen
Eigenschaften (World total)

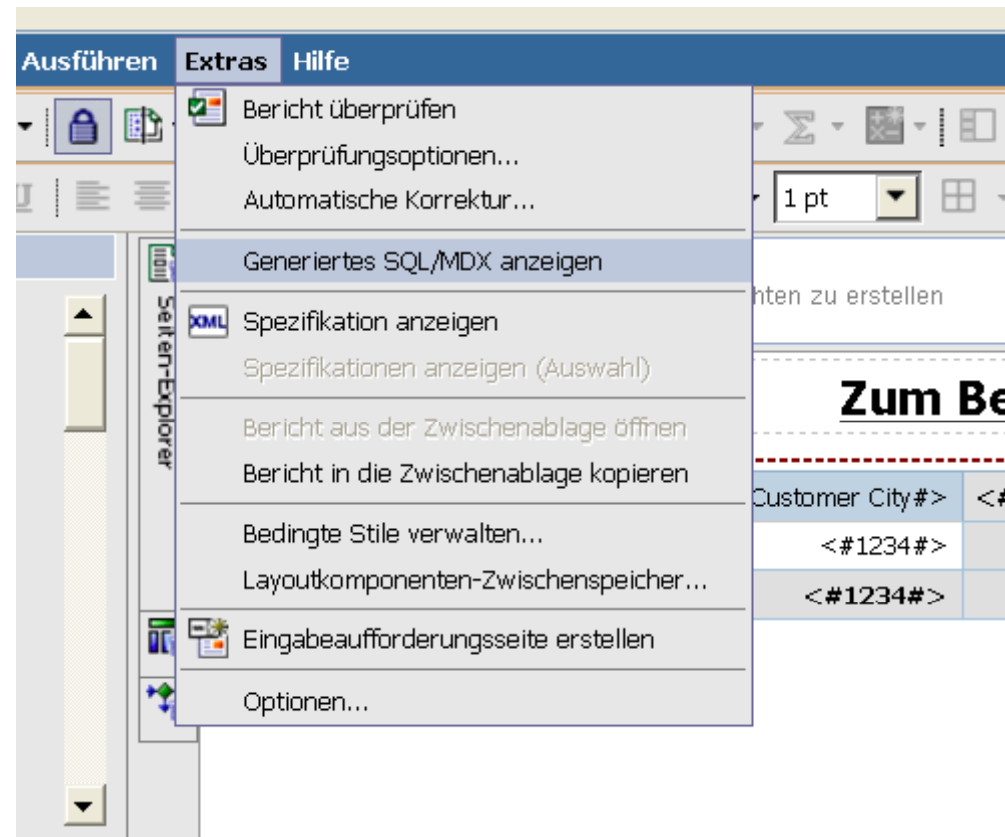
- Rechtsklick auf Spalten- bzw. Zeilennamen

- Kontextmenü erlaubt
 - Drilldown und Drillup
 - Slicing und Dicing (über „Ausschließen“)
- Pivotieren wie im Report Studio
 - Entsprechendes Icon im Menü anklicken

Agenda

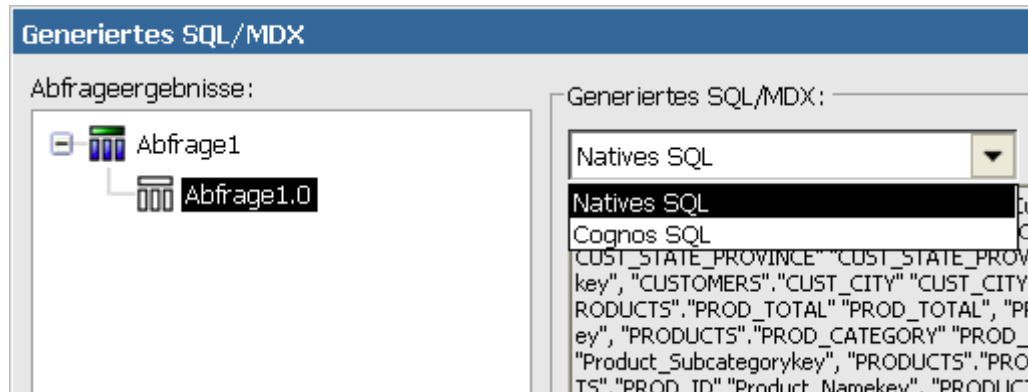
- Online Analytical Processing (OLAP)
 - Multidimensionales Datenmodell
 - Konzeptionelle Modellierung
 - Relationale Umsetzung des multidim. Modells
 - Multidimensionale Operatoren
- Cognos
 - Cognos Report Studio
 - Cognos Analysis Studio
 - **Internals**
- Hinweise zur Bearbeitung

Cognos: Anfragesprachen



...MDX auch von Cognos unterstützt!

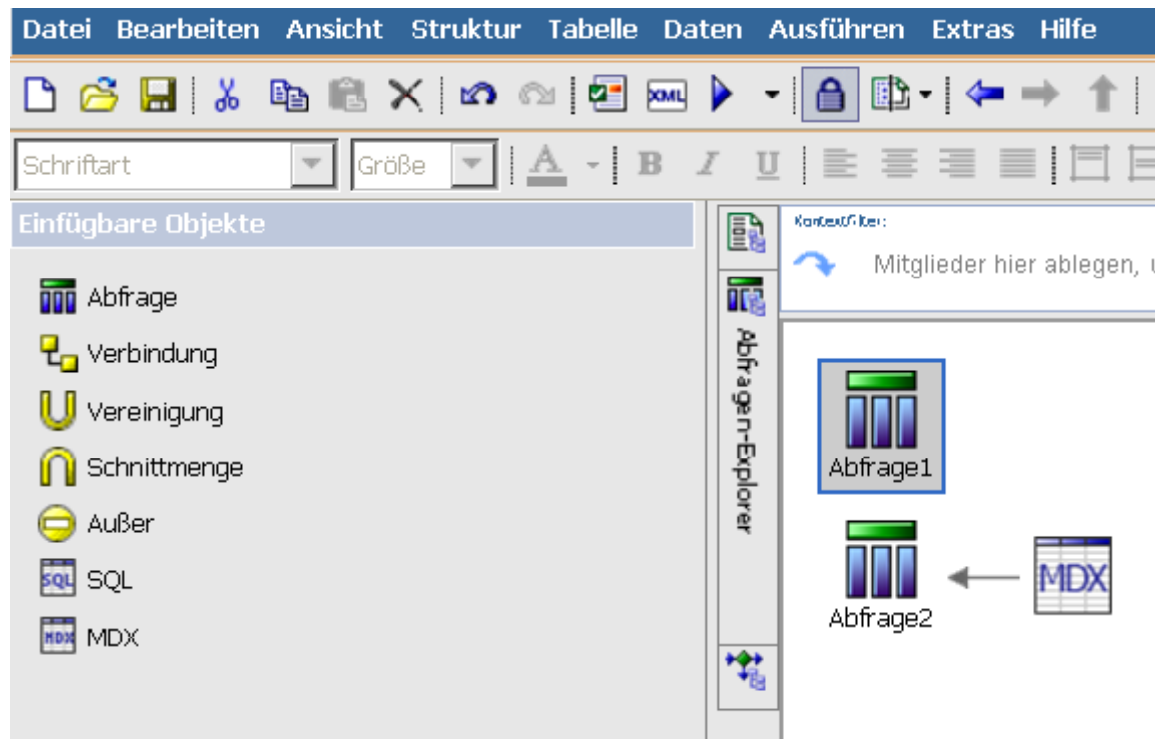
Natives SQL vs. Cognos SQL



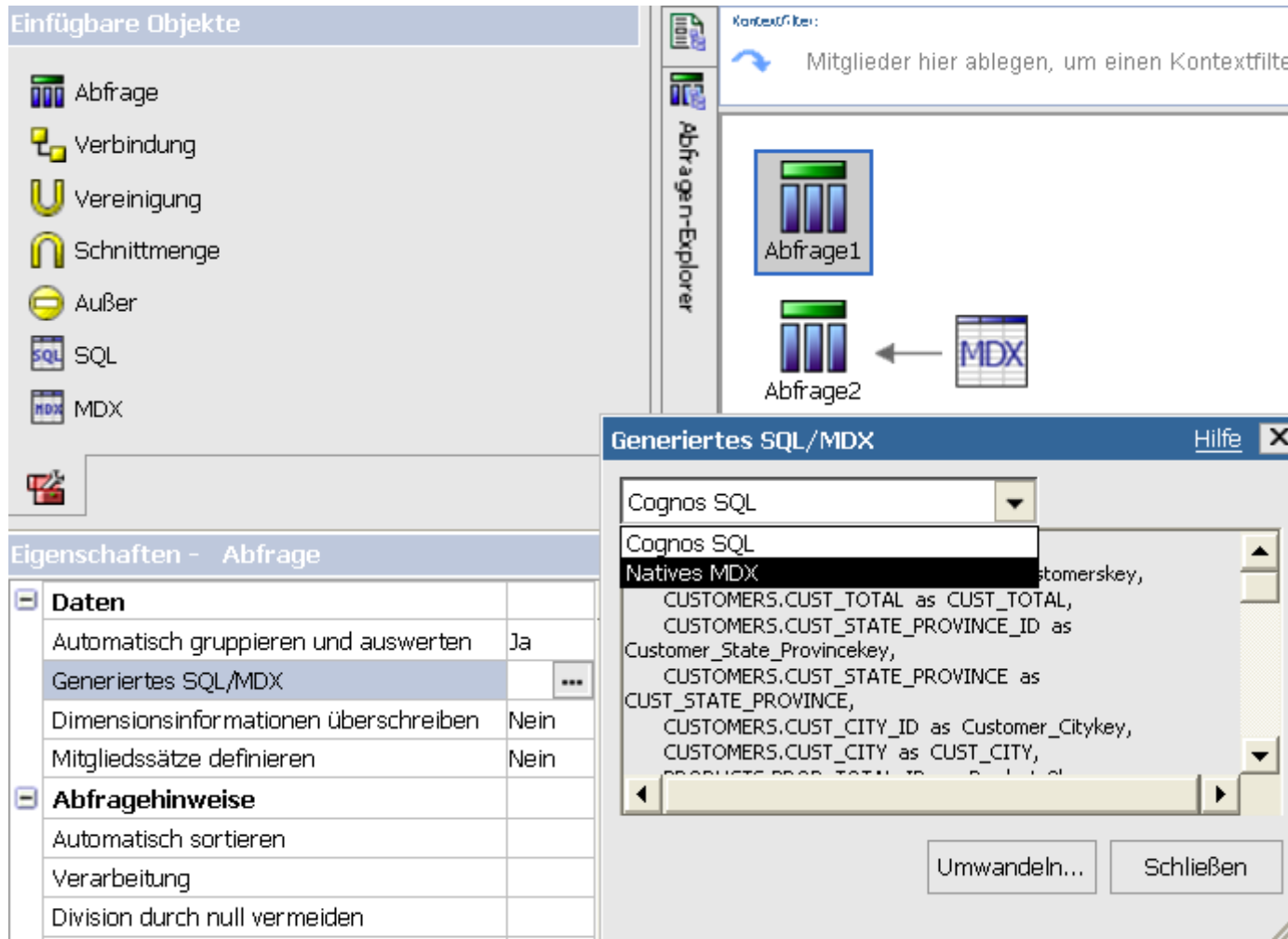
- Natives SQL
 - Für einzelne Anfragen an Datenbank
- Cognos SQL
 - Generische Cognos-Erweiterung von SQL
 - Für mehrere Datenquellen
 - Natives SQL ist Untermenge

MDX-Unterstützung

- In Abfrageexplorer: SQL oder MDX möglich



Natives MDX vs. Cognos SQL



Einfügbare Objekte

- Abfrage
- Verbindung
- Vereinigung
- Schnittmenge
- Außer
- SQL
- MDX

Eigenschaften - Abfrage

Daten	
Automatisch gruppieren und auswerten	Ja
Generiertes SQL/MDX	...
Dimensionsinformationen überschreiben	Nein
Mitgliedssätze definieren	Nein
Abfragehinweise	
Automatisch sortieren	
Verarbeitung	
Division durch null vermeiden	

Generiertes SQL/MDX

Cognos SQL

Cognos SQL

Natives MDX

CUSTOMERS.CUST_TOTAL as CUST_TOTAL,
CUSTOMERS.CUST_STATE_PROVINCE_ID as
Customer_State_Provincekey,
CUSTOMERS.CUST_STATE_PROVINCE as
CUST_STATE_PROVINCE,
CUSTOMERS.CUST_CITY_ID as Customer_Citykey,
CUSTOMERS.CUST_CITY as CUST_CITY,
PRODUCTS.PROD_TOTAL as CUST_TOTAL

Umwandeln... Schließen

Erstellung von Berichten: Vorgehensweisen

- Intuitives Vorgehen:
 - „Zusammenklicken“
 - im Hintergrund Formulierung in Cognos SQL
 - Generierung von nativem SQL/MDX für DB-Anfrage
 - Optimierung der DB-Anfragen durch Benutzer möglich
- Alternativ:
 - Einbindung existierender SQL-/MDX-Anfragen über Abfrageexplorer
 - Erst danach Berichterstellung

Agenda

- Online Analytical Processing (OLAP)
 - Multidimensionales Datenmodell
 - Konzeptionelle Modellierung
 - Relationale Umsetzung des multidim. Modells
 - Multidimensionale Operatoren
- Cognos
 - Cognos Report Studio
 - Cognos Analysis Studio
 - Internals
- **Hinweise zur Bearbeitung**

Hinweise zur Bearbeitung

- Erreichen der Tools
 - Internet Explorer
 - URL: <http://i40virt01.ipd.kit.edu/cognos8>
- Report Studio bevorzugt nutzen
 - Performance!
- Daten des Data Cubes aus
 - Package: DWM Multidimensional
- Ergebnis in „Eigenem Ordner“ speichern

Quellenangaben

- A. Bauer, H. Günzel: „Data Warehouse Systeme – Architektur, Entwicklung, Anwendung“, dpunkt.verlag, 2004.
- J. Han, M. Kamber: „Data Mining – Concepts and Techniques“, Morgan Kaufmann, 2006.
- Cognos Report Studio Professional Authoring User Guide, 2007.