

PSE: Analysesoftware für Logistiknetzwerke

IPD Böhm

Matthias Bracht



Institut für Programmstrukturen und Datenorganisation (IPD), Fakultät für Informatik, Lehrstuhl Prof. Böhm

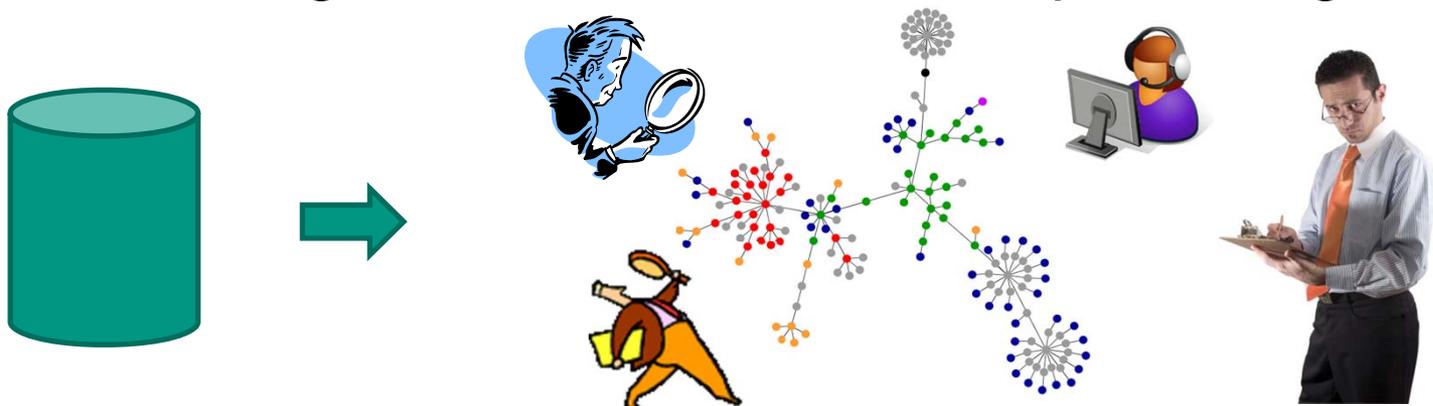


PSE: Analysesoftware für Logistiknetzwerke

Folie vom 18.10.
Auftaktveranstaltung



- Daten aus spannendem Forschungsgebiet vorhanden!
- Visualisierung aus verschiedenen Perspektiven gewünscht



Agenda für heute

- Organisatorisches
- Projekt *LogoTakt*
- Konzeptionelles: OLAP
- Aufgabenstellung
- Technisches: Datenbanken

Lehrstuhl für Systeme der Informationsverwaltung

■ Inhaber

- Prof. Dr.-Ing. Klemens Böhm



■ Vorlesungen

- Datenbanksysteme
- Data Warehousing und Mining
- Datenbankeinsatz
- Datenschutz und Privatheit
- Datenbankimplementierung und –Tuning
- Verteilte Datenhaltung
- Konzepte und Anwendungen von Workflowsystemen
- Datenbanken für räumlich-zeitliche Daten und für sich bewegende Objekte
- ...

Organisatorisches (1)

- Unsere Webseite:
<http://dbis.ipd.uni-karlsruhe.de/1628.php>
- Zusammenhang mit Diplomarbeit:
 - Ursula Kotzur [ursula.kotzur AT gmx.de]
- Ein Hiwi unterstützt euch:
 - Andreas Wixler [andreas.wixler AT student.kit.edu]
- Infrastrukturelles:
 - E-Mail-Verteiler bei der ATIS
 - SVN: Bitte einmal bis morgen einloggen mit eurem ATIS-Account und das „Profile“ ausfüllen: <https://svnserver.informatik.kit.edu/i40/login/>
Wir schalten euch dann für das PSE-Projekt frei.
 - Arbeiten im Poolraum möglich, falls gewünscht
 - Projektverwaltung mit Redmine: <http://redmine.ipd.uka.de>

Organisatorisches (2)

■ Kolloquium

- Termin: Donnerstags, 15:45 – 17:15 Uhr (SR 301, Geb. 50.34), im Wesentlichen zu Beginn/Ende jeder Phase
- Anderer Raum: 25.11., 9.12. – Alternativlösung wird spätestens auf der Webseite bekanntgegeben
- Weitere Termine nach Vereinbarung
- Anwesenheitsliste

■ Prüfung

- Anmeldung für PSE **und** TSE online – bitte bei der richtigen Veranstaltung anmelden (PSE Nr. 28)!
- Anmeldeschluss: 12.11.

■ Ernennung der Phasenverantwortlichen

- Jede Phase hat einen Hauptverantwortlichen
- Präsentation der Ergebnisse, Moderation der Diskussion
- Bitte ins Redmine-Wiki eintragen!

Vorläufiger Zeitplan

- KW 43 (28.10.): Erstes Gruppentreffen, Einführung

- KW 44-45 (01.11. – 12.11.): Pflichtenheft
 - 04.11.: Anforderungen, Interview, Redmine
 - 11.11.: Kolloquium zum Pflichtenheft

- KW 46-49 (15.11. – 10.12.): Entwurf
- KW 50-3 (13.12. – 21.01.): Implementierung
- KW 4-6 (24.01. – 11.02.): Qualitätssicherung
- KW 7-8 (14.02. – 25.02.): Klausurpause
- KW 9 (28.02. – 04.03.): Interne Abnahme
- KW 10 (07.03. – 11.03.): Abschlusspräsentation (evtl. erst 30.3.)

Agenda für heute

- Organisatorisches
- **Projekt *LogoTakt***
- Konzeptionelles: OLAP
- Aufgabenstellung
- Technisches: Datenbanken

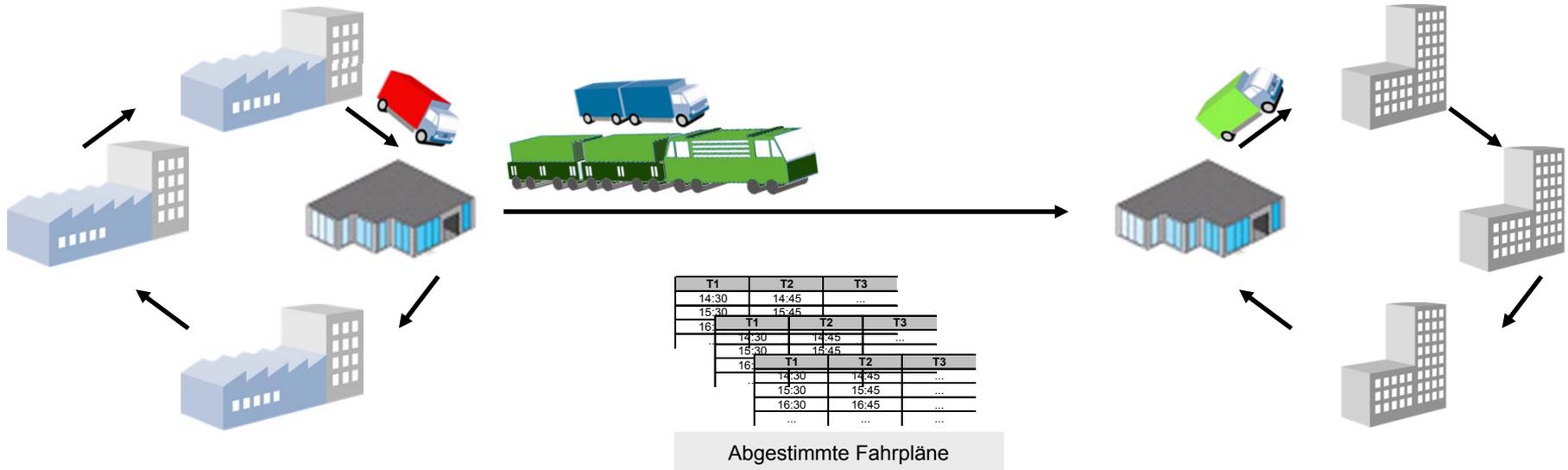
Projekt LogoTakt

- Ziel: Entwicklung von Technologien und Prozessen für robuste getaktete Logistiknetzwerke
- Vom BMWi gefördert



- Analyse der Robustheit abgelaufener Transportprozesse unter Einsatz von:
 - Graph Mining
 - Graphanfragesprachen

Kernmerkmale LogoTakt

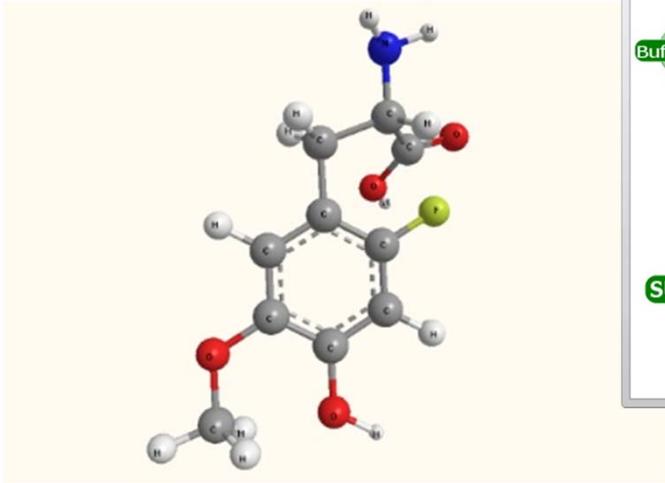


Taktung der Verkehre in Vor- Haupt- und Nachlauf

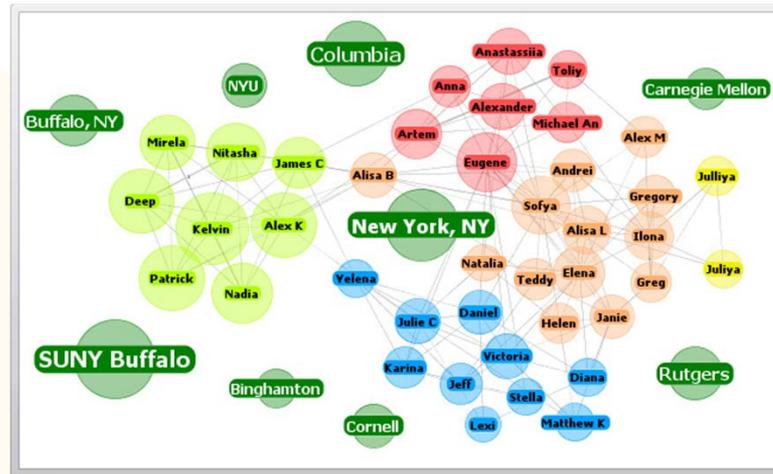
Hohe **Robustheit** durch vorkalkulierte Zeit- und Mengenpuffer

Intermodalität zur Verbesserung der Ökobilanz

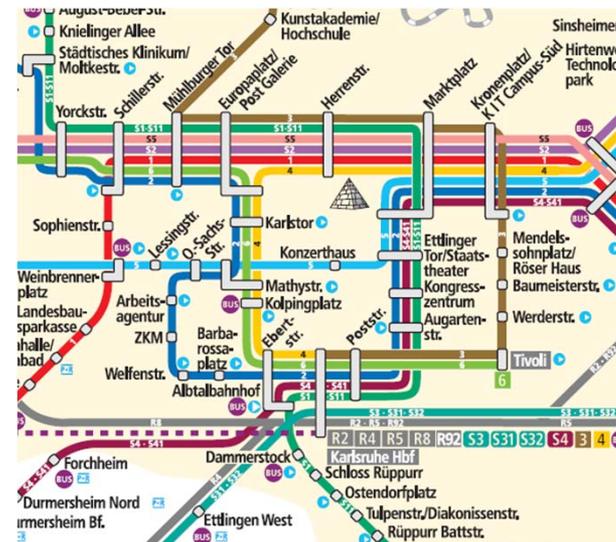
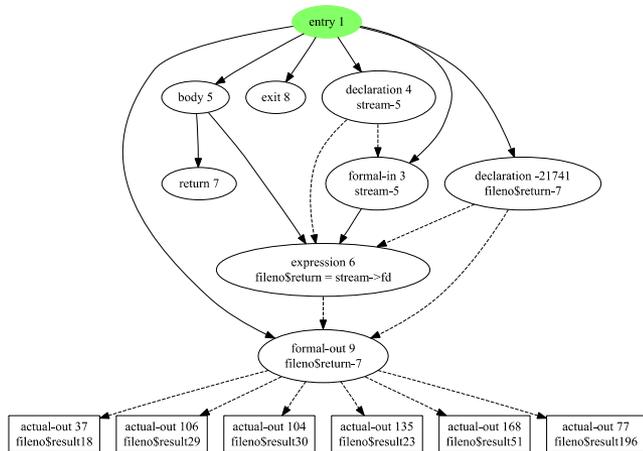
Unser Forschungsthema: Graphen



<http://www.fzd.de/db/Pic?pOid=26054>

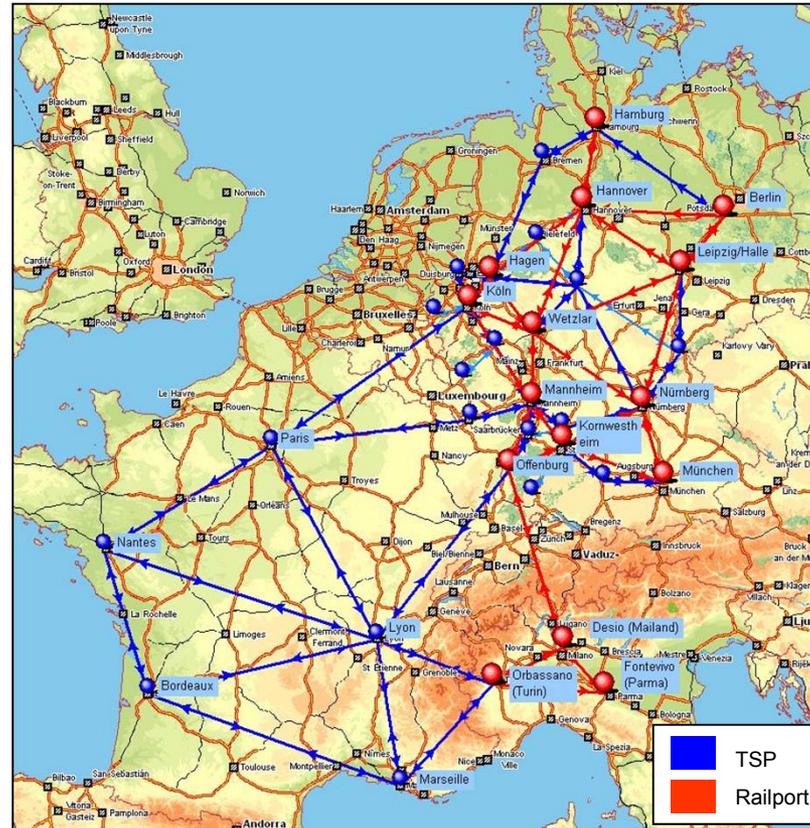


www.touchgraph.com



www.kvv.de

Netzwerkstruktur LogoTakt



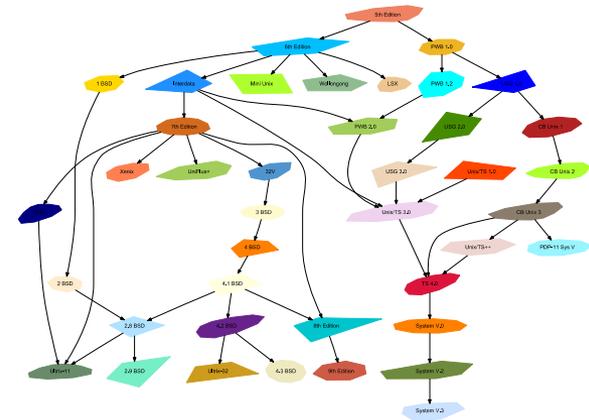
- Datengrundlage: „großer“ Basisgraph mit allen durchgeführten Fahrten.
- Ein Knoten pro Postanschrift.
- Eine Kante pro LKW-/Zugfahrt.

LogoTakt: Beispielhafte Informationsbedürfnisse

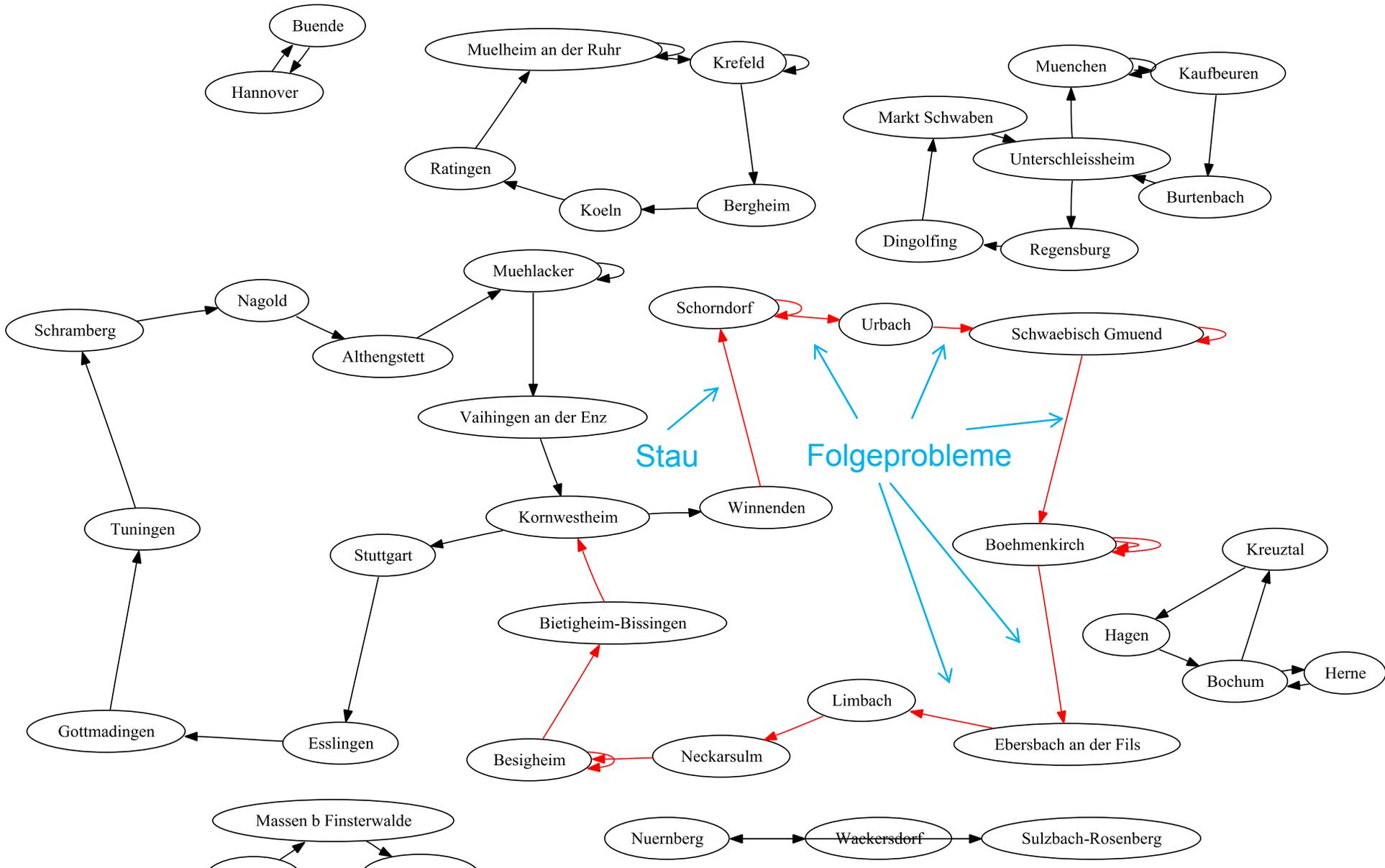
- Nur eigene LKW-Fahrten: **Schweitzer Spedition**
 - Nur Transporte mit eigenen Gütern:  **BOSCH**
 - Fahrplananalyse: Werden Anschlüsse verpasst? 
 - Umschlagpunktanalyse: Wie ausgelastet sind einzelne Railports? 
 - Volkswirtschaftliche Analyse: Welche Regionen tragen wieviel zum Gesamtaufkommen bei? 
 - Strukturelle Analyse: Welche Abläufe sind verspätungsanfällig? 
 - Gesamtanalyse: Wie zuverlässig ist das Netzwerk? 
-
- Bisher in LogoTakt: Manuelle Erzeugung von Graphen je nach Wunsch
 - Gewünscht: Einheitliches, flexibles Framework mit Visualisierung

Graph-Visualisierung

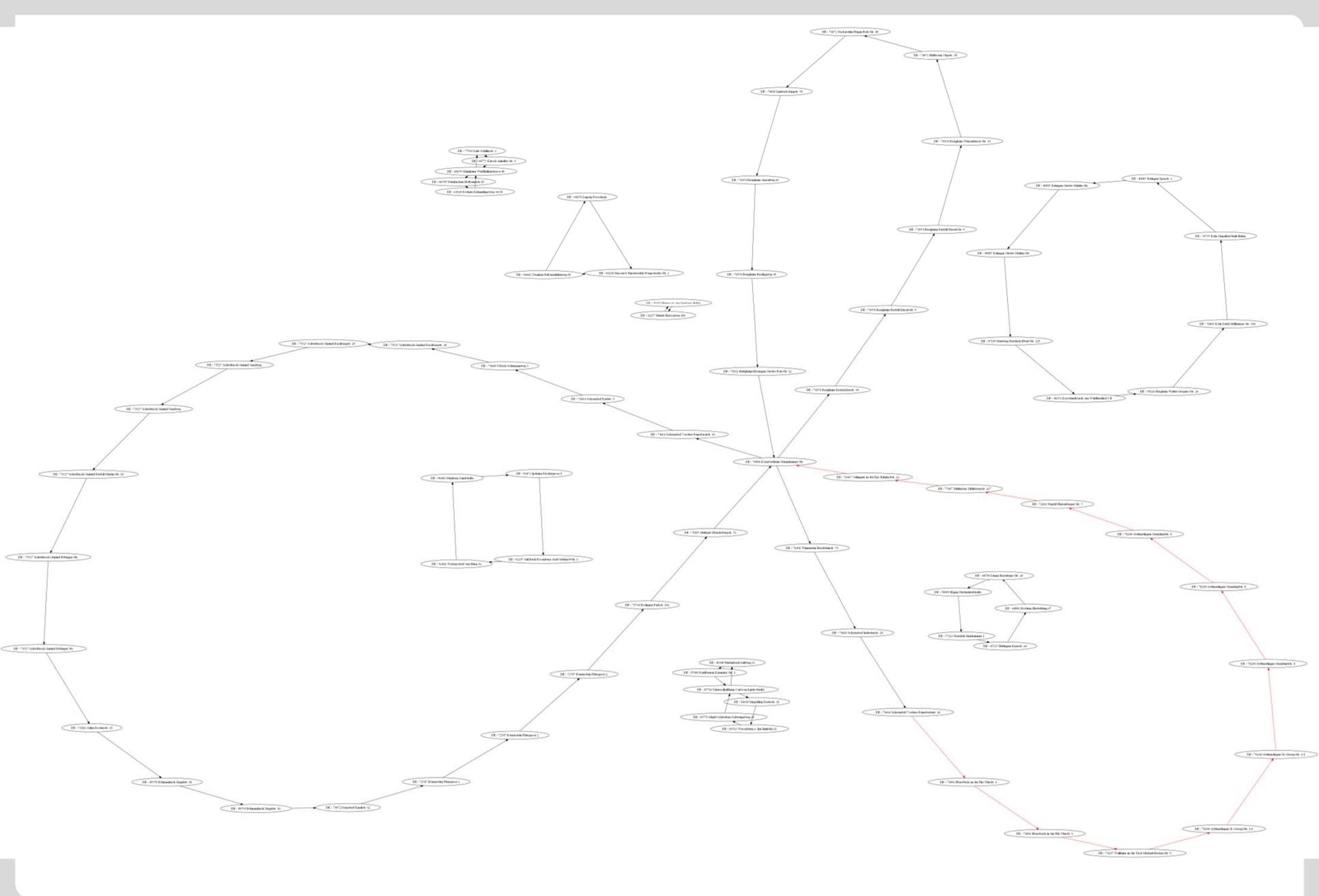
- Graphen als Menge von Knoten und Kanten sind ein Konstrukt der diskreten Mathematik.
- Graph-Visualisierung
 - ...ist oft sehr hilfreich!
 - ...verfolgt (ästhetische) Ziele, z.B.
 - Minimierung von Kreuzungen, Fläche
 - Maximierung des kleinsten Winkels
 - Vereinheitlichung von Kantenlängen
 - ...kann sehr aufwändig werden (NP-vollständig).
 - ...ist ein eigenes Forschungsgebiet.
 - ...kann von verschiedenen Tools erledigt werden.



Beispiel-Graph, 2010-04-06_Dienstag



Rückblick: Graphvisualisierung



LogoTakt: Beispielhafte Informationsbedürfnisse

- Nur eigene LKW-Fahrten: **Schweitzer Spedition**
- Nur Transporte mit eigenen Gütern:  **BOSCH**
- Fahrplananalyse: Werden Anschlüsse verpasst? 
- Umschlagpunktanalyse: Wie ausgelastet sind einzelne Railports? 
- Volkswirtschaftliche Analyse: Welche Regionen tragen wieviel zum Gesamtaufkommen bei? 
- Strukturelle Analyse: Welche Abläufe sind verspätungsanfällig? 
- Gesamtanalyse: Wie zuverlässig ist das Netzwerk? 

- Bisher in LogoTakt: Manuelle Erzeugung von Graphen je nach Wunsch
- Gewünscht: Einheitliches, flexibles Framework mit Visualisierung

Agenda

- Organisatorisches
- Projekt *LogoTakt*
- **Konzeptionelles: OLAP**
 - **Einführung Data Warehouses**
 - Multidimensionales Datenmodell
 - Graph OLAP
- Aufgabenstellung
- Technisches: Datenbanken

Eigenschaften eines Data Warehouse

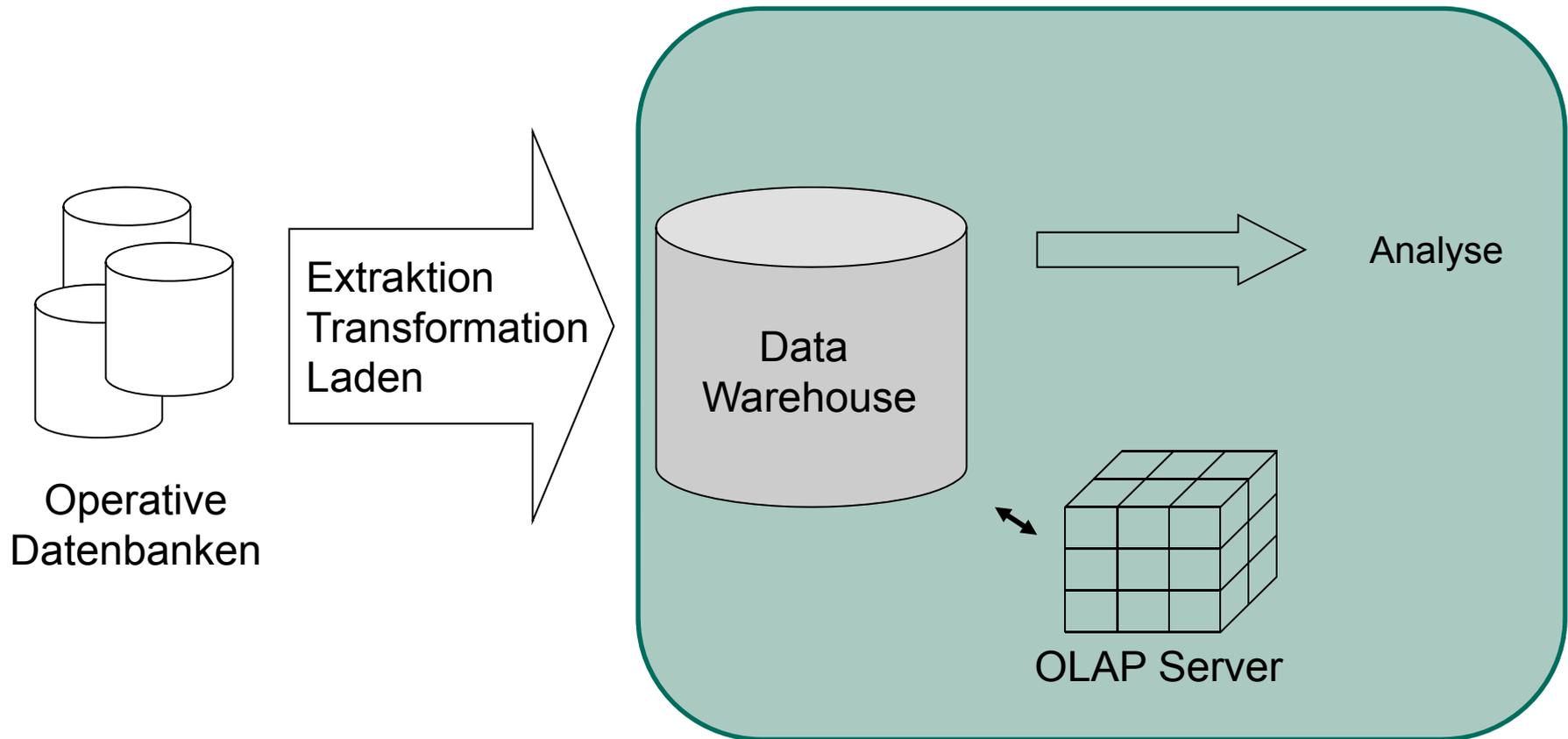
- Integrierte Sicht auf beliebige Daten
 - ...aus verschiedenen Datenbanken
 - ...Integration von Schemata und Daten aus Quellen

- Analyseaspekt
 - ...multidimensionales Datenmodell
 - ...Online Analytical Processing (OLAP)

- Stabile Datenbasis
 - Eingebachte Daten werden nicht mehr modifiziert
 - Neue Daten können aufgenommen werden

- Data-Warehouse-System
 - Komponenten zur Integration und Analyse + Data Warehouse

Vereinfachte Referenzarchitektur



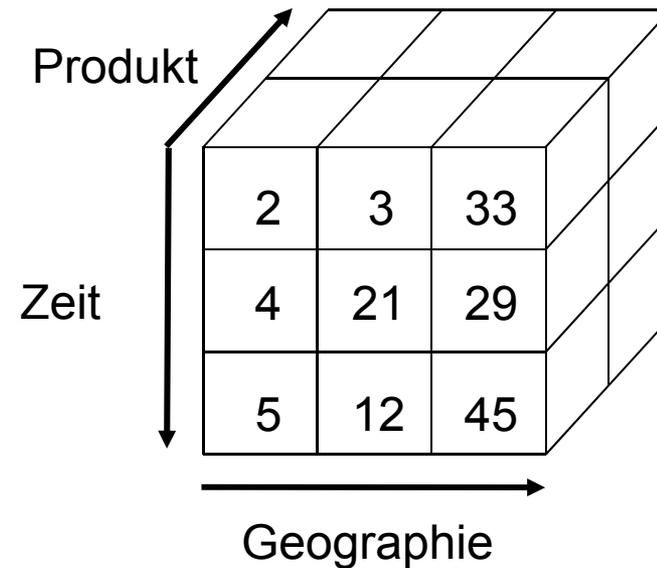
Anforderungen an Online Analytical Processing

- 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

- Organisatorisches
- Projekt *LogoTakt*
- **Konzeptionelles: OLAP**
 - Einführung Data Warehouses
 - **Multidimensionales Datenmodell**
 - Graph OLAP
- Aufgabenstellung
- Technisches: Datenbanken

Beispiel: Analyse von Verkäufen

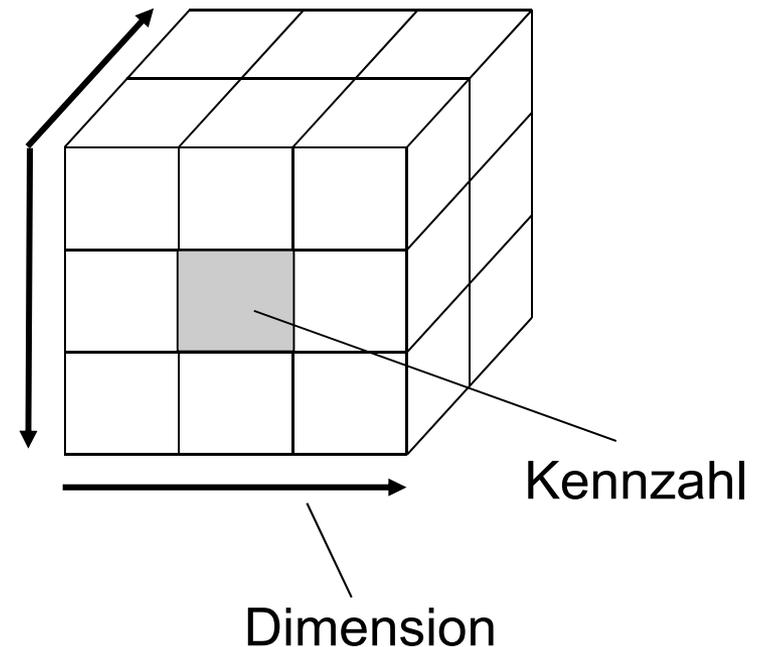


Multidimensionales Datenmodell - Begriffe

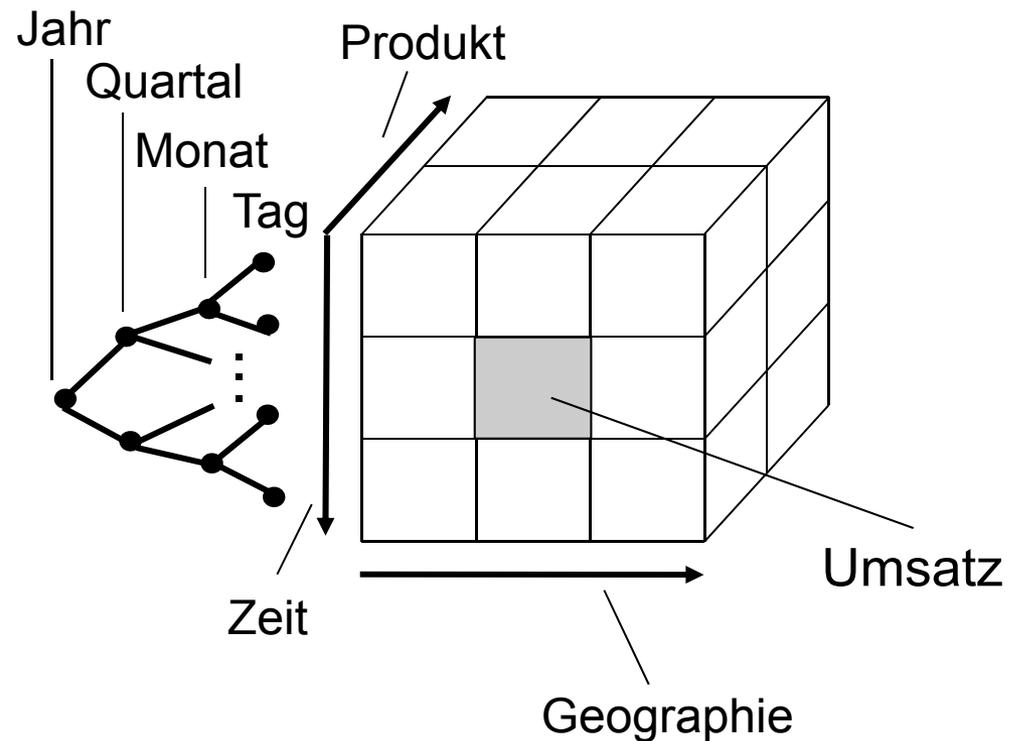
- Hilfsmittel zur Veranschaulichung von Daten
- verschiedene Aspekte auf gleiche Weise zugreifbar
- Einsatz bei OLAP-Anwendungen

- Kennzahlen
 - Elemente eines Würfels

- Dimensionen
 - Beschreiben Daten
 - Ermöglichen Zugriff auf Kennzahlen
 - Können zu Hierarchien gehören



Multidimensionales Datenmodell – Beispiel

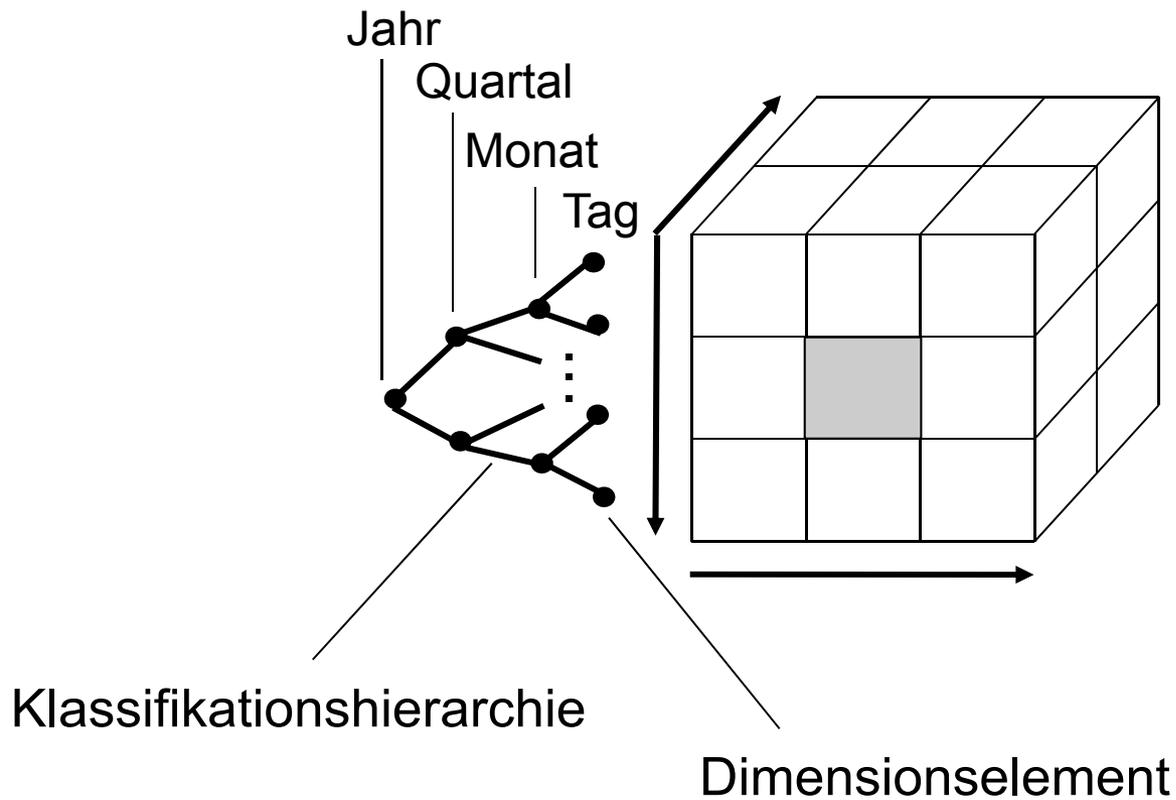


Dimensionen

- 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)

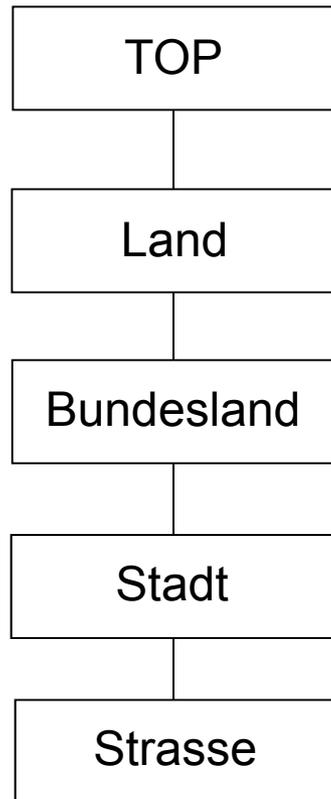
Dimensionen – Beispiel Zeit



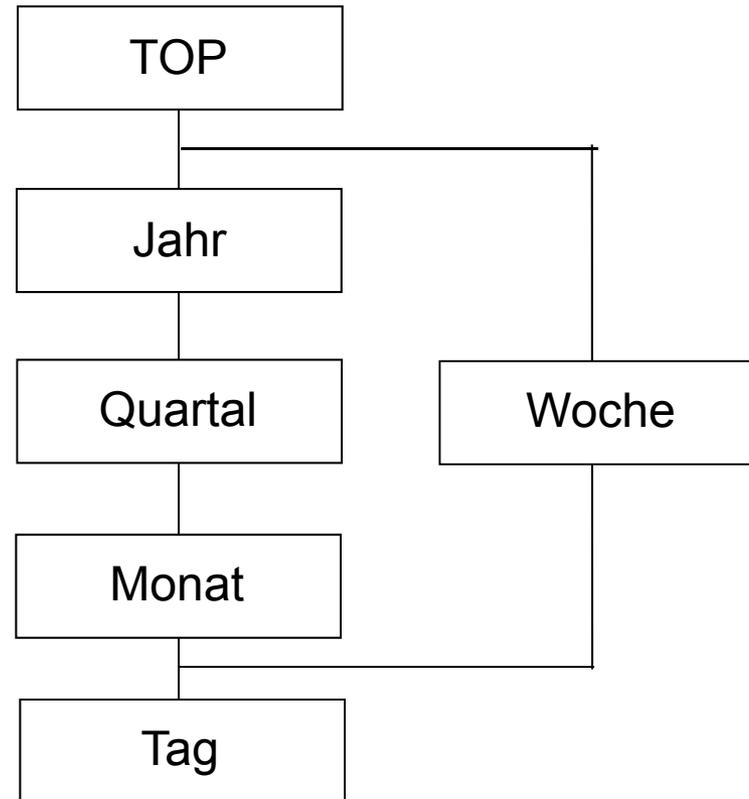
Arten von Klassifikationshierarchien

- 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



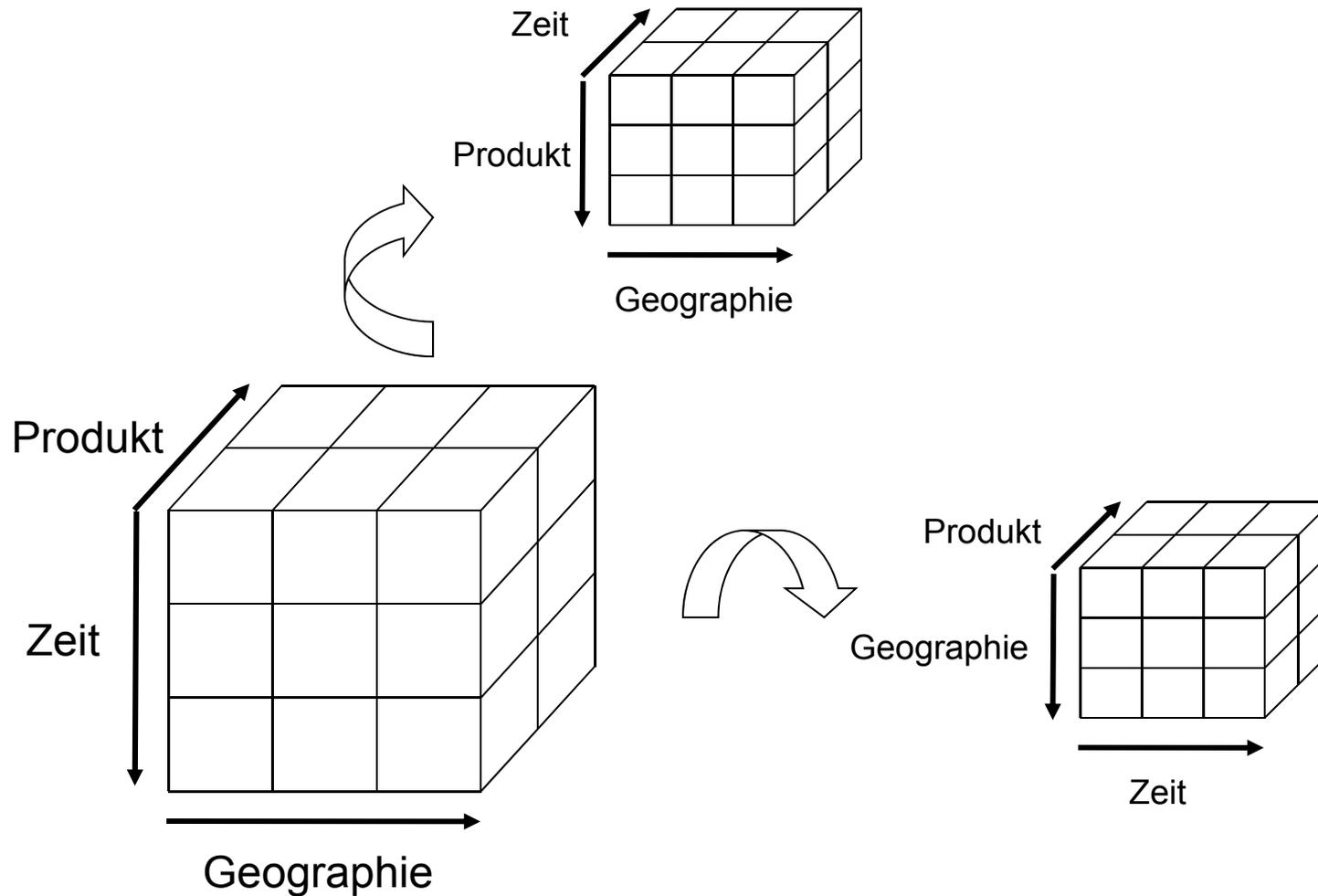
Parallele Hierarchie

Weitere Begriffe

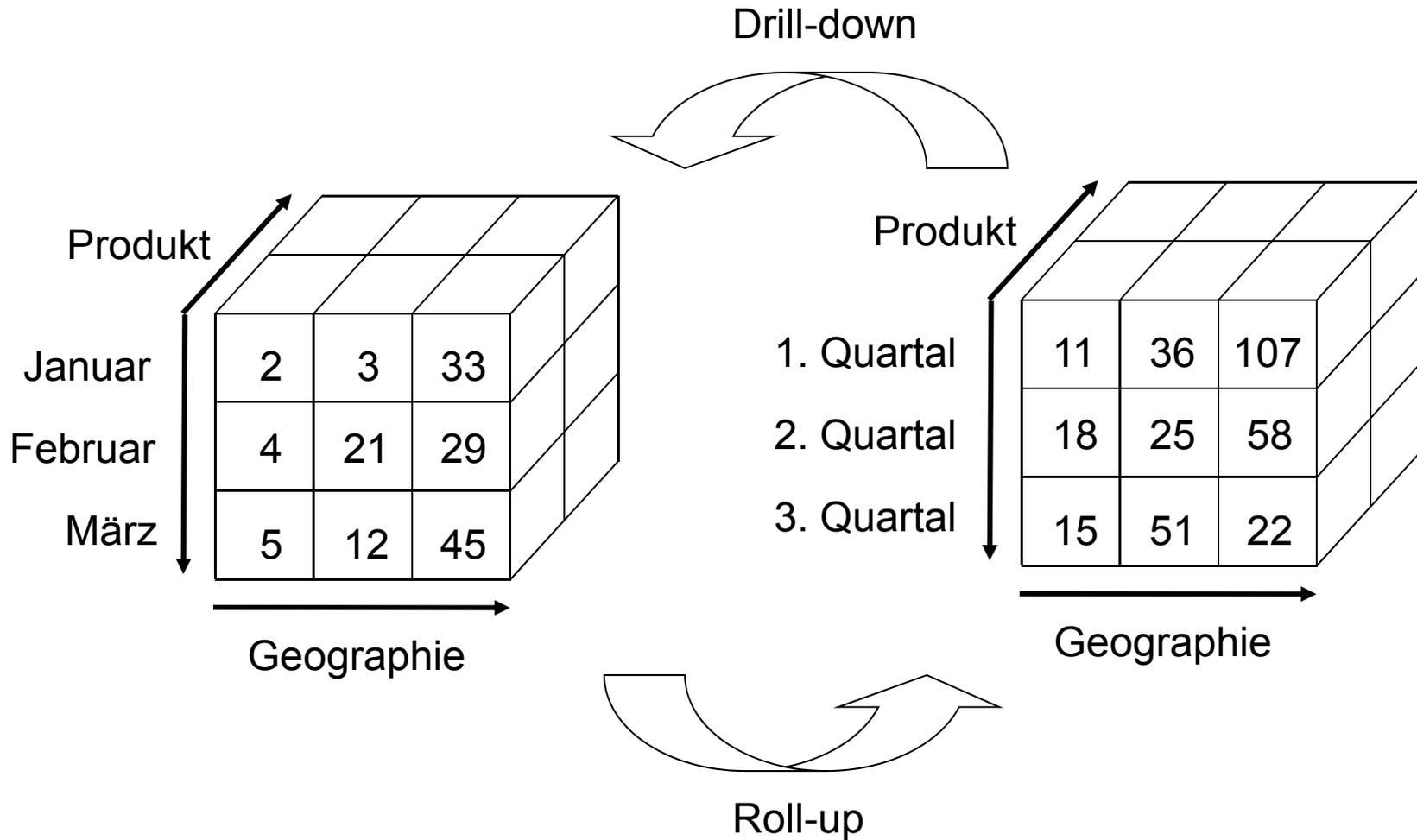
- Würfel
 - Kanten von Dimensionen aufgespannt
 - Kantenlänge entspricht Anzahl der Elemente in Dimension
 - Eine oder mehrere Kennzahlen pro Würfelzelle
 - Anzahl der Dimensionen heißt Dimensionalität

- Konsolidierungspfad
 - Pfade im Klassifikationsschema

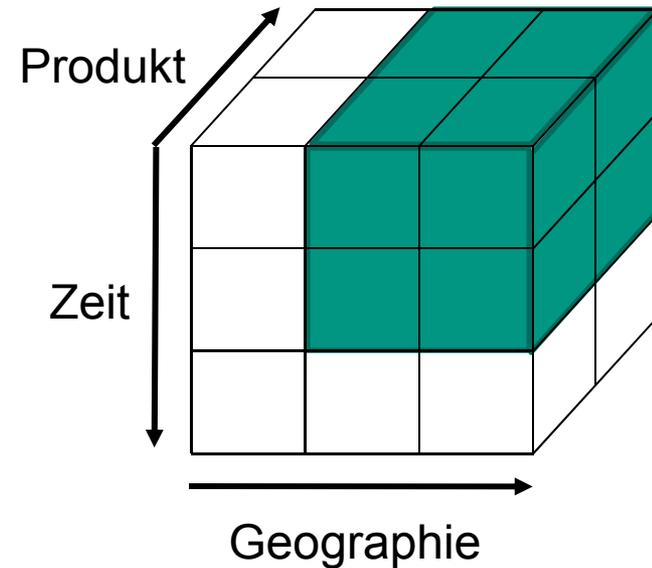
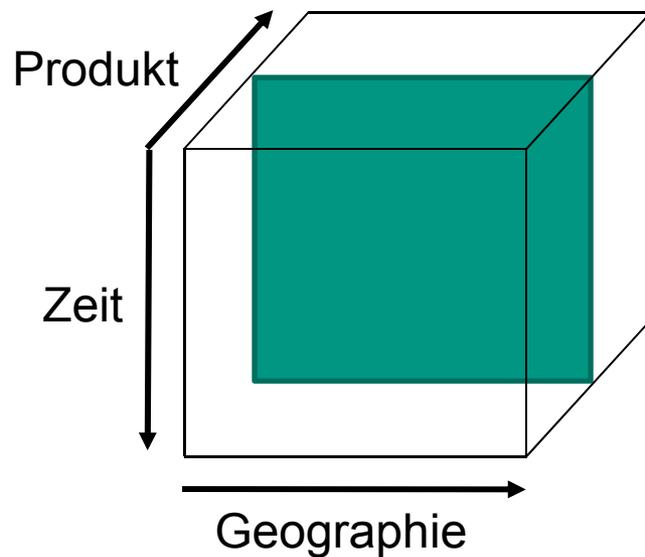
Multidim. Operatoren: Pivotieren



Multidim. Operatoren: Roll-up und Drill-down



Multidim. Operatoren: Slicing und Dicing



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

Zusammenfassung: multidim. Operatoren

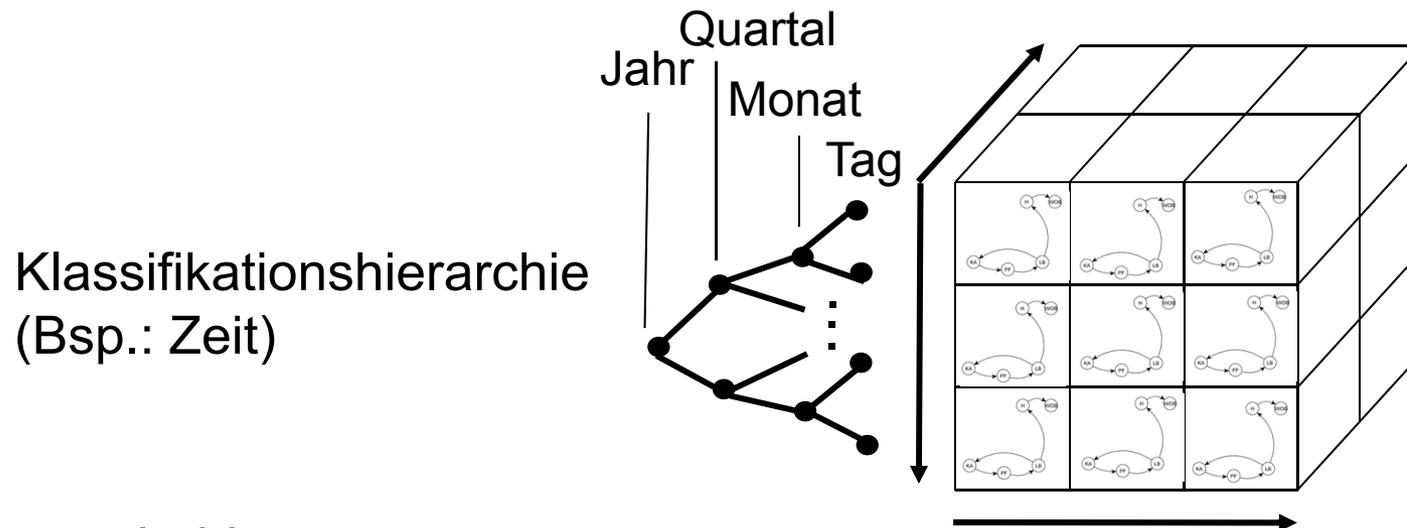
- Nutzen multidimensionale Struktur der Daten
- Erlauben einfache Modifikation von Anfragen
 - ...zu inhaltlich ähnlichen Anfragen
 - ...zu strukturell ähnlichen Anfragen
- Sind auf die Fragestellungen im Data Warehouse zugeschnitten

Agenda

- Organisatorisches
- Projekt *LogoTakt*
- **OLAP**
 - Einführung Data Warehouses
 - Multidimensionales Datenmodell
 - **Graph OLAP**
- Aufgabenstellung
- Technisches: Datenbanken

Jetzt: Graph OLAP

- Kennzahl „Verkäufe“ wird jetzt zu „Fahrten“.
- Eine Zelle eines Würfels entspricht einem *Graphen*.

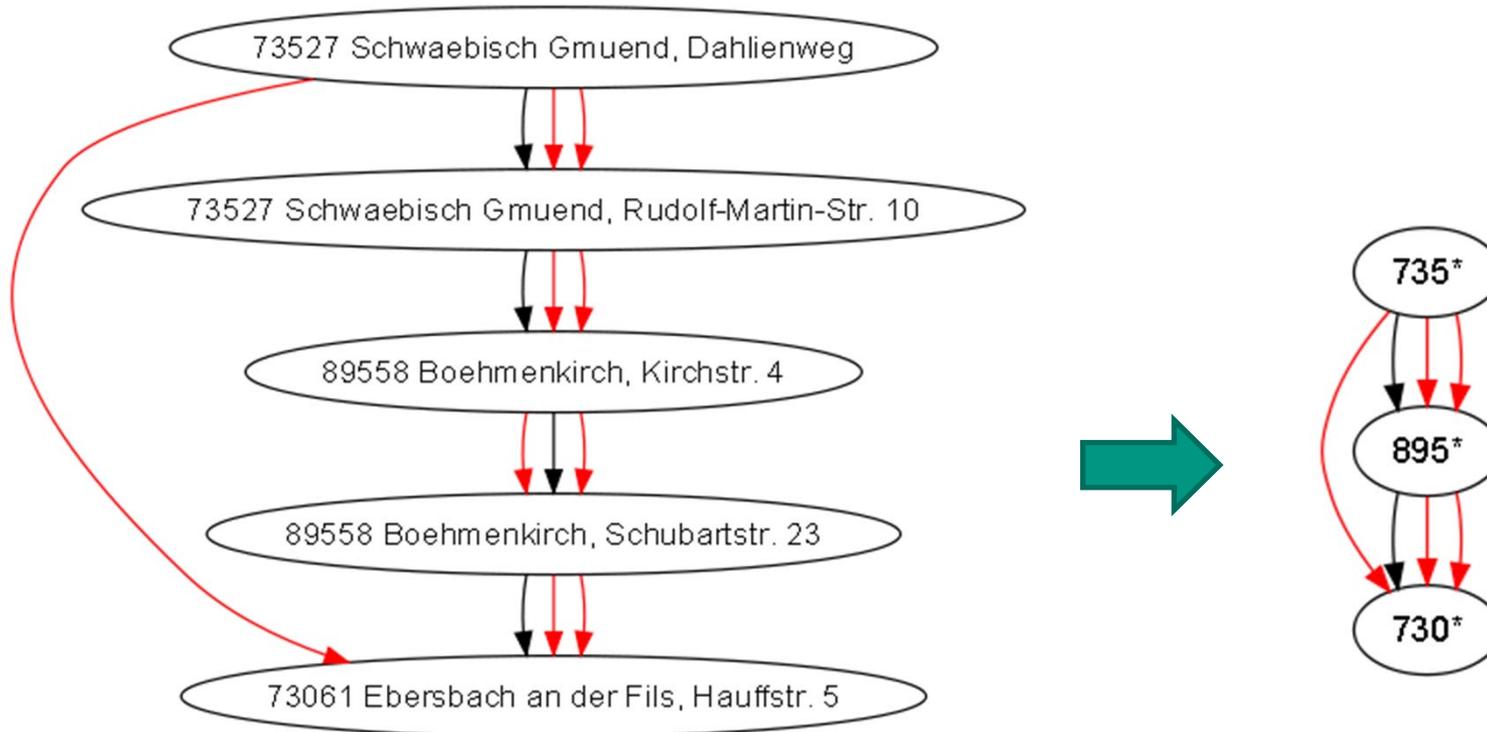


- Unterscheidung:
 - *Informationelle* Dimensionen lassen Topologie des Graphen unverändert (*I-OLAP*, Bsp.: Zeit).
 - *Topologische* Dimensionen fassen Knoten zusammen (*T-OLAP*, Bsp.: Geographie).

(Chen et al., *Graph OLAP: Towards Online Analytical Processing on Graphs*, ICDM'08)

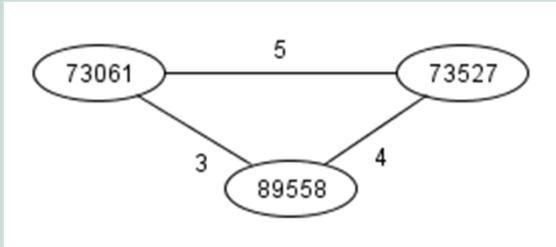
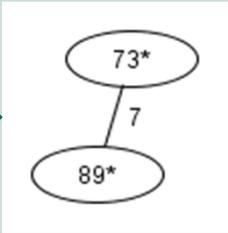
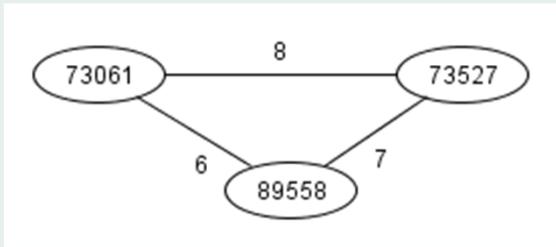
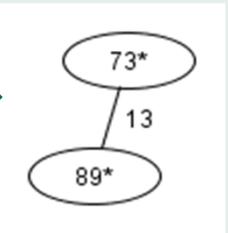
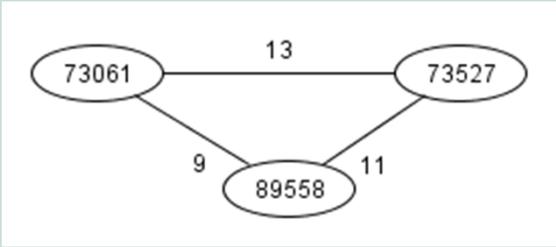
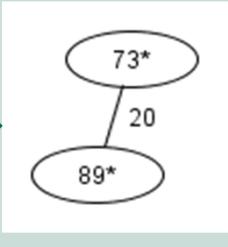
OLAP in der Logistik

- Mögliche Hierarchieebenen für topologische Dimension „Geographie“:
 Adresse → dreistelliger PLZ-Bereich



- Weiteres Beispiel für eine topologische Dimension:
 - Kunde (alle Standorte eines Kunden bilden einen neuen Knoten)

I-OLAP vs. T-OLAP

	PLZ		PLZ-Bereich	
Monat: Januar				
Monat: Februar				
Zeitraum: Frühjahr				
				

Agenda

- Organisatorisches
- Projekt *LogoTakt*
- OLAP
- **Aufgabenstellung**
- Technisches: Datenbanken

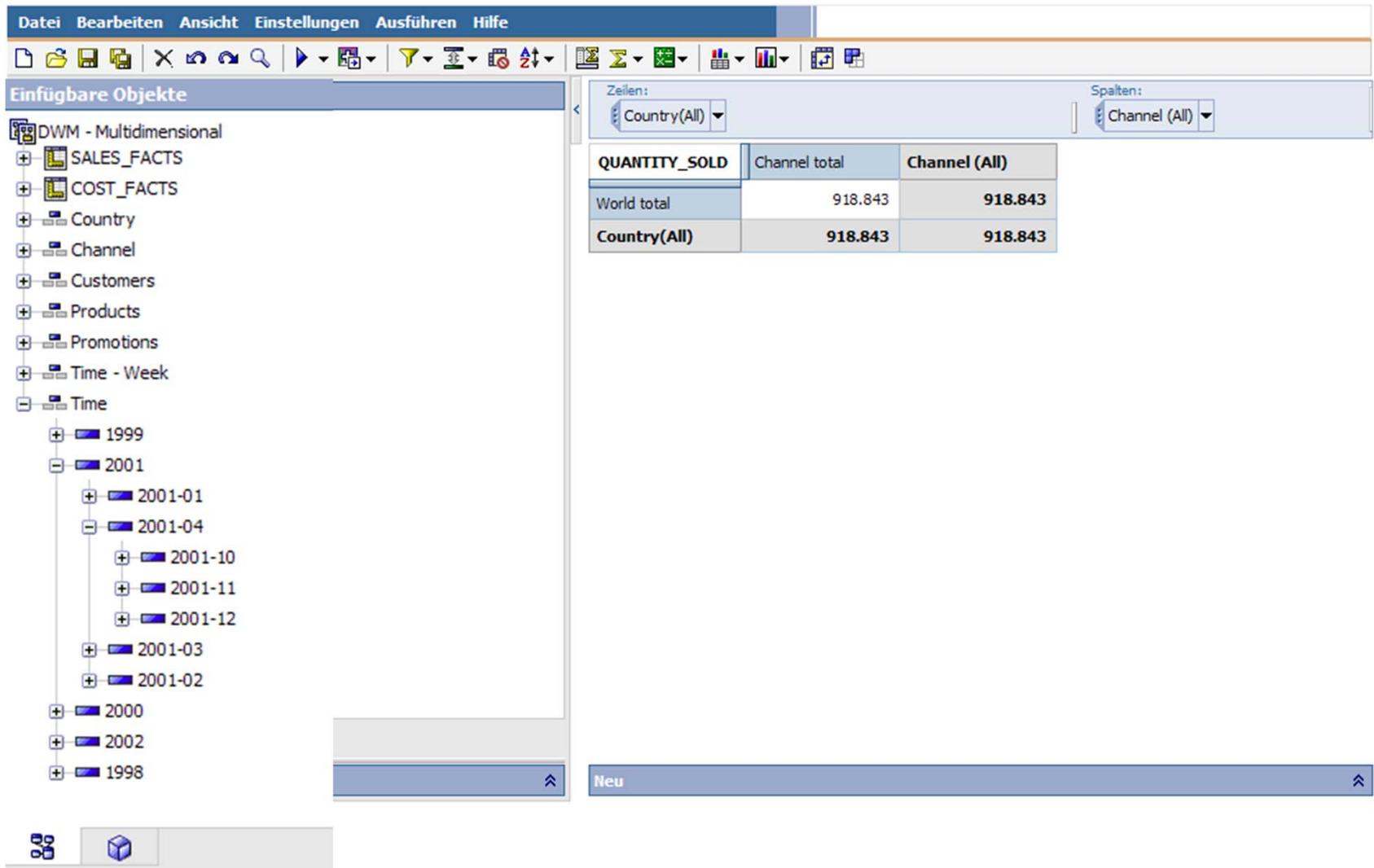
Szenario für die Aufgabe

- Benutzer wählt Perspektive aus.
 - Für I-Hierarchie (z.B. „ZeitMonate“, „ZeitWochen“, „ZeitWochentage“):
 - Hierarchieebene (z.B. „Monat“)
 - Datenelement (z.B. Monat „April“)
 - Für T-Dimension (z.B. „Geographie“):
 - Hierarchieebene (z.B. „Ein Knoten pro PLZ-Bereich“)
- Benutzer wählt Kantenattribut (z.B. „Gesamtgewicht“).
- Entsprechender Graph wird ausgelesen und angezeigt.
- (Dimensionen/Hierarchien/Hierarchieebenen sowie Kantenattribute sind vorgegeben. Siehe Dokumentation.)

- (Optionale!) Vision: „Zeig mir einen Graphen mit allen Fahrten an Werktagen im April 2010. Ein Knoten soll einen PLZ-Bereich umfassen. Kanten sollen mit dem Gesamtgewicht der transportierten Güter beschriftet sein.“



Zur Inspiration: Cognos



The screenshot displays the Cognos Business Intelligence interface. On the left, the 'Einfügbare Objekte' (Available Objects) pane shows a multidimensional cube structure under 'DWM - Multidimensional'. The cube includes dimensions: SALES_FACTS, COST_FACTS, Country, Channel, Customers, Products, Promotions, Time - Week, and Time. The Time dimension is expanded to show years from 1998 to 2002, with 2001 further broken down into months.

On the right, a pivot table is displayed. The 'Zeilen:' (Rows) dropdown is set to 'Country(All)' and the 'Spalten:' (Columns) dropdown is set to 'Channel (All)'. The table shows the following data:

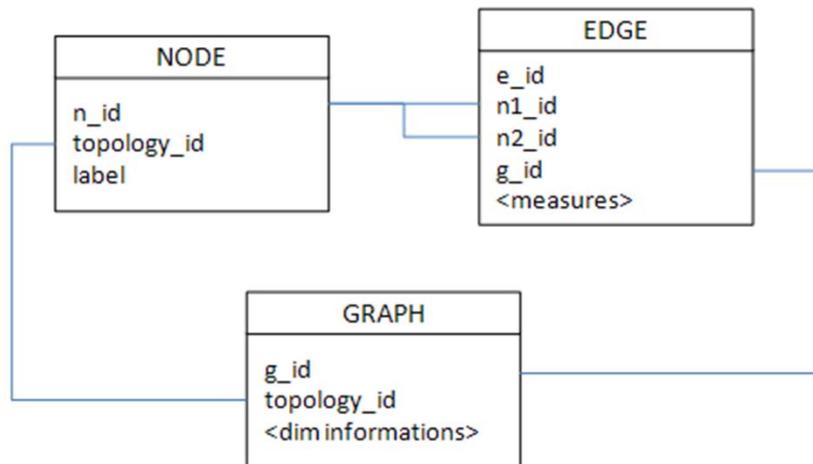
QUANTITY_SOLD	Channel total	Channel (All)
World total	918.843	918.843
Country(All)	918.843	918.843

The interface includes a menu bar (Datei, Bearbeiten, Ansicht, Einstellungen, Ausführen, Hilfe) and a toolbar with various icons for file operations, navigation, and analysis. At the bottom, there are icons for cube and report views, and a 'Neu' (New) button.

Daten

- Server: *i40db01.ipd.kit.edu*
- Datenbank: *logotakt*
- User: *psestud*
- Schema *golap*: Views namens *graph*, *node*, *edge*

- Erläuterungen der Attribute: Siehe Dokumentation.



Vielen Dank für die Aufmerksamkeit!

Fragen?

Anmerkungen?

Hinweise?

Vorschläge für erste Schritte:

- Graph OLAP **verstehen!**
- Prüfungsanmeldung (2x!)
- SVN-Anmeldung + „Profile“, Redmine-Wiki: Phasenverantwortliche
- MVC-Auffrischung
- Einblicke in die Datenbank, Skizzen eines Frontends
- Suche nach Bibliotheken für Visualisierung

Quellen:

- A. Bauer, H. Günzel: „Data Warehouse Systeme – Architektur, Entwicklung, Anwendung“, dpunkt.verlag, 2004.
- Chen et al., *Graph OLAP: Towards Online Analytical Processing on Graphs*, ICDM'08