

# Data Warehousing

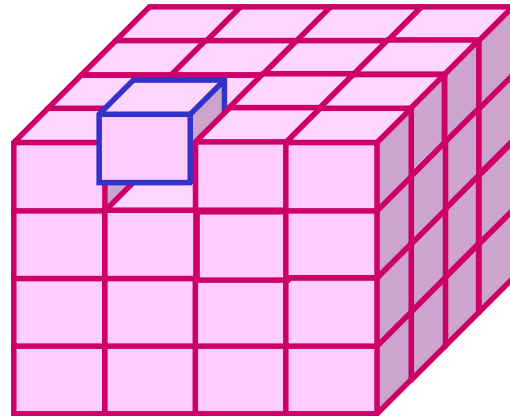
## Kapitel 0: Organisatorisches

Dr. Andreas Thor

Wintersemester 2009/10

Universität Leipzig

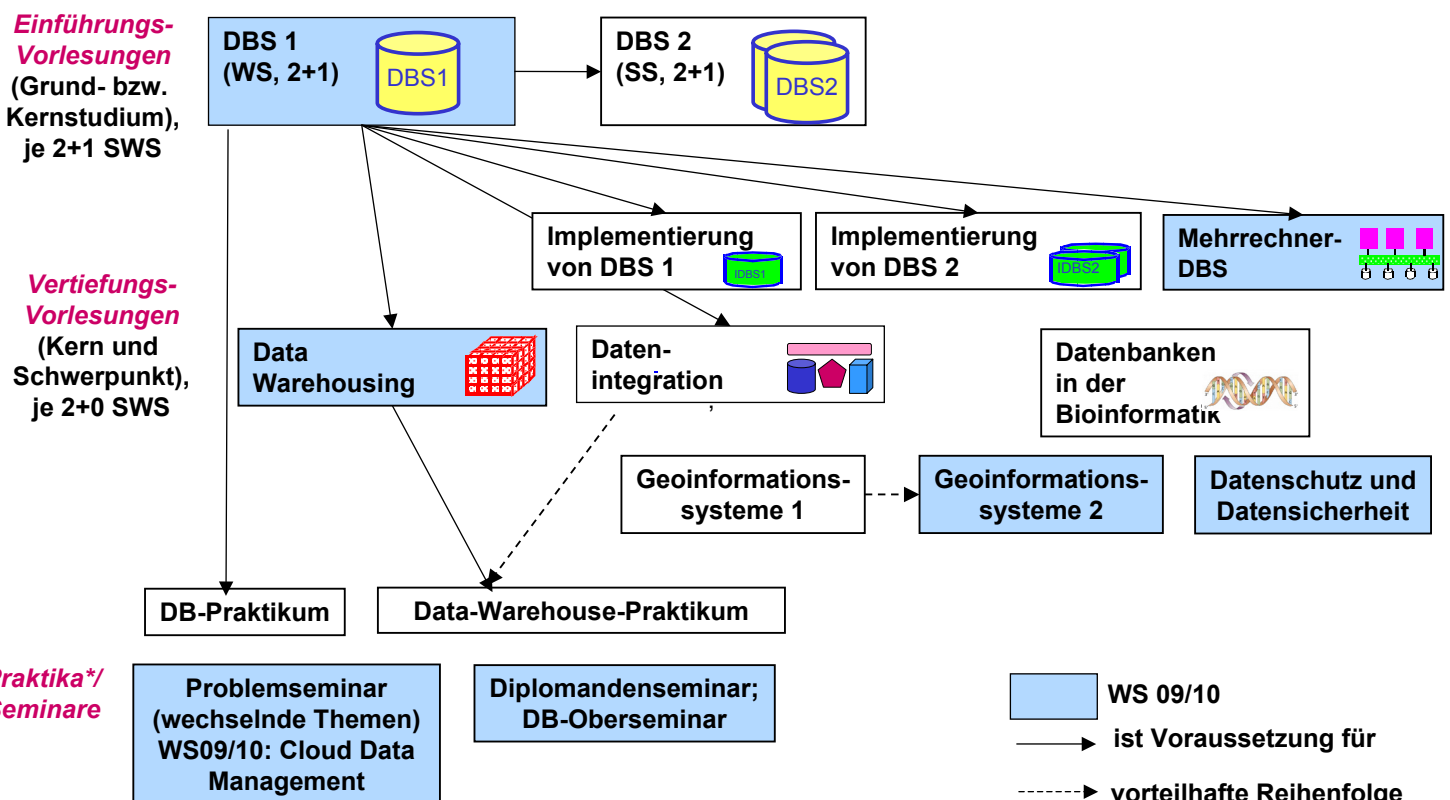
Institut für Informatik



<http://dbs.uni-leipzig.de>



## Lehrveranstaltungen zu „Datenbanken“ (WS09/10)



\* Detaillierter Praktikumsschein wird ausgestellt



# Organisatorisches

- Termin: dienstags, 9:15 Uhr, Felix-Klein-Hörsaal (Johannisgasse 26)
  - Beginn: 13. Oktober 2009
  - **Letzte Vorlesung** am 8. Dezember 2009
- Zusatztermine: montags, 11:15 Uhr, Raum 1-22 (Johannisgasse 26)
  - „Alle Montage im November“: 2., 9., 17., 24., 31. November 2009
- Klausurtermin:
  - **15. Dezember 2009**, 9:15 Uhr, Felix-Klein-Hörsaal (Johannisgasse 26)
- Webseite mit allen wichtigen Infos:
  - <http://dbs.uni-leipzig.de/stud/ws0910/dwh>



## Anrechnung / Leistungsbewertung

- Vorlesung kann angerechnet werden
  - im Diplom-, Master- oder Bachelor-Studiengang Informatik (Kern oder Vertiefung)
  - In weiteren Studiengängen mit Informatikanteilen
- Diplomstudiengang Informatik (ab 5. Semester)
  - Modulklausur (60 Min.) am **15. Dezember 2009** für Kerngebiet Praktische Informatik (2 SWS) oder
  - als Teil der mündlichen Schwerpunktprüfung
- Bachelor/Master Informatik (alte PO)
  - APL oder PL (3 credits) über Klausur
- Master Informatik (neue PO)
  - Teil der Module „Moderne Datenbanktechnologien“ und „Anwendungsbezogene Datenbankkonzepte“
  - Klausur (60 Minuten)
- geprüft werden konzeptionelles Wissen + Anwendungsfälle



# Modulzuordnung für neues Master-Studium (PO2006)

- Abt. Datenbanken bietet folgende Module im **Master-Studium** an
  - **10-202-2215: Moderne Datenbanktechnologien** - Kleines Modul (Kern, 5 LP)
  - **10-202-2216: Moderne Datenbanktechnologien** - Großes Modul (Vertiefung, 10 LP)
  - **10-202-2213: Anwendungsbezogene Datenbankkonzepte** - Kleines Modul (Kern, 5 LP)
  - **10-202-2214: Anwendungsbezogene Datenbankkonzepte** - Großes Modul (Vertiefung, 10 LP)
  - **10-202-2011: Masterseminar Informatik** (5 LP)
- Wahlmöglichkeiten **WS09/10** für „Moderne Datenbanktechnologie“ und „Anw. Datenbankkonzepte“
  - Mehrrechner-Datenbanksysteme
  - Data Warehousing
  - Datenschutz und Datensicherheit
  - Geo-Informationssysteme 2 (GIS2)
  - Problemseminar Cloud Data Management.
- **Musterstudienplan mit Schwerpunkt in Datenbanken/Informationssysteme**
  - 1. Semester: Moderne Datenbanktechnologien – Kleines (oder großes) Modul
  - 2. Semester: Anwendungsbezogene Datenbankkonzepte – Großes (oder kleines) Modul
  - 3. Semester: Moderne Datenbanktechnologien – Großes (oder kleines) Modul
  - 4. Semester: **Masterarbeit** / Anwendungsbezogene Datenbankkonzepte – Kleines (oder großes) Modul



## Lernziele der Vorlesung

- Überblicks- und Detailkenntnisse zum Aufbau, Betrieb, Funktionsweise und Einsatz von Data Warehouses
- Kenntnis von Architekturalternativen
- Modellierung von Data Warehouses
- Überblick zu Analyseverfahren (OLAP, Data Mining)
- Probleme der Datenintegration / Datenbereinigung
- Implementierungstechniken zur Sicherstellung einer hohen Performanz
- Sachkundige Beurteilung von kommerziell verfügbaren Data-Warehouse-Lösungen



# Data-Warehouse-Praktikum

- Data Warehouse ist eine Architektur zur Datenintegration
  - große wirtschaftliche Bedeutung
  - Unterstützung durch kommerzielle Datenbanksysteme
- Praktische Realisierung typischer Elemente eines DWH-Projektes
  - Datenimport, Datenextraktion
  - Data Cleaning (Objektkonsolidierung, Datennormalisierung, Ableitung neuer Attribute)
  - Konzeption eines Warehouse-Schemas mit entsprechender Datenbefüllung
  - Analysen mittels MDX-Anfragen
- Technisches
  - Ca. 200 MB Daten (CSV, XML)
  - Software: SQL Server 2008 und Visual Studio
  - Gruppenarbeit (2 Studenten) mit 3 Testaten
- Organisatorisches
  - Sommersemester 2010 (geplant)



## Vorläufige Vorlesungsübersicht

1. Einführung
2. Architektur von Data Warehouse-Systemen
3. Modellierung / Operationen
  - MOLAP, ROLAP, hybride Ansätze
  - SQL-Erweiterungen, Operatoren
4. Metadaten und ETL (Extraktion, Transformation, Laden)
  - Metadaten-Integration, Schema Matching
  - Datenbereinigung (Data Cleaning)
5. Performance-Techniken
  - Indexstrukturen
  - Materialisierte Sichten
  - Nutzung paralleler Datenbanken
6. Data Mining-Verfahren
7. Data-Warehouse-Anwendung (Website Recommendations)



# Literatur

- Bauer/Günzel (Hrsg.):  
*Data Warehouse Systeme*.  
2. Auflage, dpunkt 2004
- Lehner, W.: *Datenbanktechnologie für Data-Warehouse-Systeme, Konzepte und Methoden*, dpunkt, 2003
- Jarke / Lenzerini / Vassiliou / Vassiliadis:  
*Fundamentals of Data Warehouses*.  
2. Auflage, Springer 2003
  
- Eigene Forschungsarbeiten: <http://dbs.uni-leipzig.de>
  - Schema Matching: COMA, COMA++
  - Data Cleaning (Object Matching): MOMA, STEM, FEVER
  - Website Recommendations: AWESOME



# Online-Übungen

LOTS (Leipzig Online Test System), <http://lots.uni-leipzig.de>

- Kennung: dwh#0910

The screenshot shows the homepage of the Leipzig Online Test System (LOTS). The header features the text 'Leipzig Online Test System' and 'UNIVERSITÄT LEIPZIG' with the 'Institut für Informatik Abteilung DBS' logo. A navigation bar includes 'Home', 'Registrierung', and 'Impressum'. The main content area is divided into three sections: 'Login' with a form for 'Username:' and 'Passwort:' and a 'login' button; 'Gast Login' with a text block explaining guest access and a 'als Gast einloggen' button; and 'System Info' showing '# Benutzer: 2' and a clock icon with the date '12.10.2005 15:27:05'. A 'News' section at the bottom contains the text 'Bitte loggen Sie sich ein, um personalisierte News lesen zu können.'




# Lehrstuhl Datenbanksysteme

- Seit 1994 am Institut für Informatik
- Ca. 12-15 Mitarbeiter (ab 2010: > 20)
- Umfangreiches Lehrangebot
  - Vorlesungen, Praktika, Seminare
  - Online-Übungssystem LOTS
  - Eigene Lehrbücher: MRDBS, IDBS, ...




## Forschung



### Abteilung Datenbanken Leipzig

am Institut für Informatik



Startseite

### Inhalte

- ▶ Mitarbeiter
- ▼ **Forschung**
  - [Publikationen](#)
  - ▶ Projekte
  - Jahresberichte
  - Kooperationen
  - ▶ Promotionen
  - ▶ [Colloquia](#)
  - [Conferences](#)
- ▶ Studium
- ▶ Service

### Forschung

Publications (2009, 2008, 2007, 2006, 2005, 2004, 2003, ...)

WDI Lab: Innovation Lab on Web Data Integration

Mashup-based data integration

Schema and Ontology Matching (COMA++)

Object Matching / Entity Resolution (FEVER/STEM, MOMA)

Evolution of ontologies and mappings (Schema Evolution bibliography [↗](#))

Citation Analysis

See [labs.dbs.uni-leipzig.de](http://labs.dbs.uni-leipzig.de) [↗](#) for current prototypes and implementations



## Database Group Leipzig — Labs

### [Online Citation Service \(OCS\)](#)

Get the DBLP publications with Google Scholar citation counts. (experimental)

### [Leipzig Online Test System \(LOTS\)](#)

Multiple Choice Tests, SQL query and XQuery training

### [Teiresias: Neugriechisch — Deutsch](#)

Online Wörterbuch

### [Publication Categorizer](#)

Integrate and categorize publications specific to some topic, e.g. [Schema Evolution](#) and [Data Cleaning](#)

### [COMA++ Prototype](#)

Schema and Ontology Matching

<http://labs.dbs.uni-leipzig.de>



## WDI-Lab: Innovationslabor Webdatenintegration

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung



Wir suchen zum 1.1.2010 (Bewerbungen bis 15.10.2009):

- 1 Mitarbeiter/in für BWL als Projektunterstützung im Innovationslabor (WDI-Lab)
- 9 Mitarbeiter/innen für Forschung und Entwicklung im Innovationslabor (WDI-Lab)

## Arbeitsgruppen

- AG1: Workflowbasiertes Integrationsframework (Mashups)
- AG2: Schema- und Ontologie-Matching
- AG3: Objekt-Matching und Dublettenbehandlung



# Wir suchen studentische Mitarbeiter!

- Was wir uns wünschen:
  - Studium der Informatik oder Wirtschaftsinformatik
  - Kenntnisse in Datenbanken, Web-Technologie, Software-Technologie
  - Engagement und Interesse an wissenschaftlicher Arbeit
  
- Was wir bieten:
  - SHK-Verträge über 10h/Woche
  - Mitarbeit im BMBF-geförderten WDI-Lab zur Webdatenintegration
  - Kooperative Arbeitsatmosphäre
  - Unterstützung bei der Suche nach passendem Bachelor-, Master- bzw. Diplomarbeitsthema
  
- Ansprechpartner: Prof. Rahm oder Frau Massmann



## Abschlussarbeiten

- Themen für Diplom-, Bachelor- und Master-Arbeiten aus aktuellen Forschungsbereichen des Lehrstuhls, z.B.
  - Webdatenintegration
  - Bioinformatik
  - Zitierungsanalyseu.a.
  
- Kooperationen mit regionalen Unternehmen
  
- Anfragen per Formular oder persönlich

