

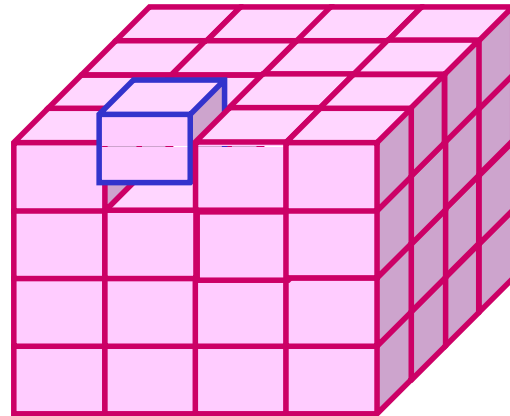
# Data Warehousing

Sommersemester 2008

Prof. Dr. E. Rahm

Universität Leipzig  
Institut für Informatik

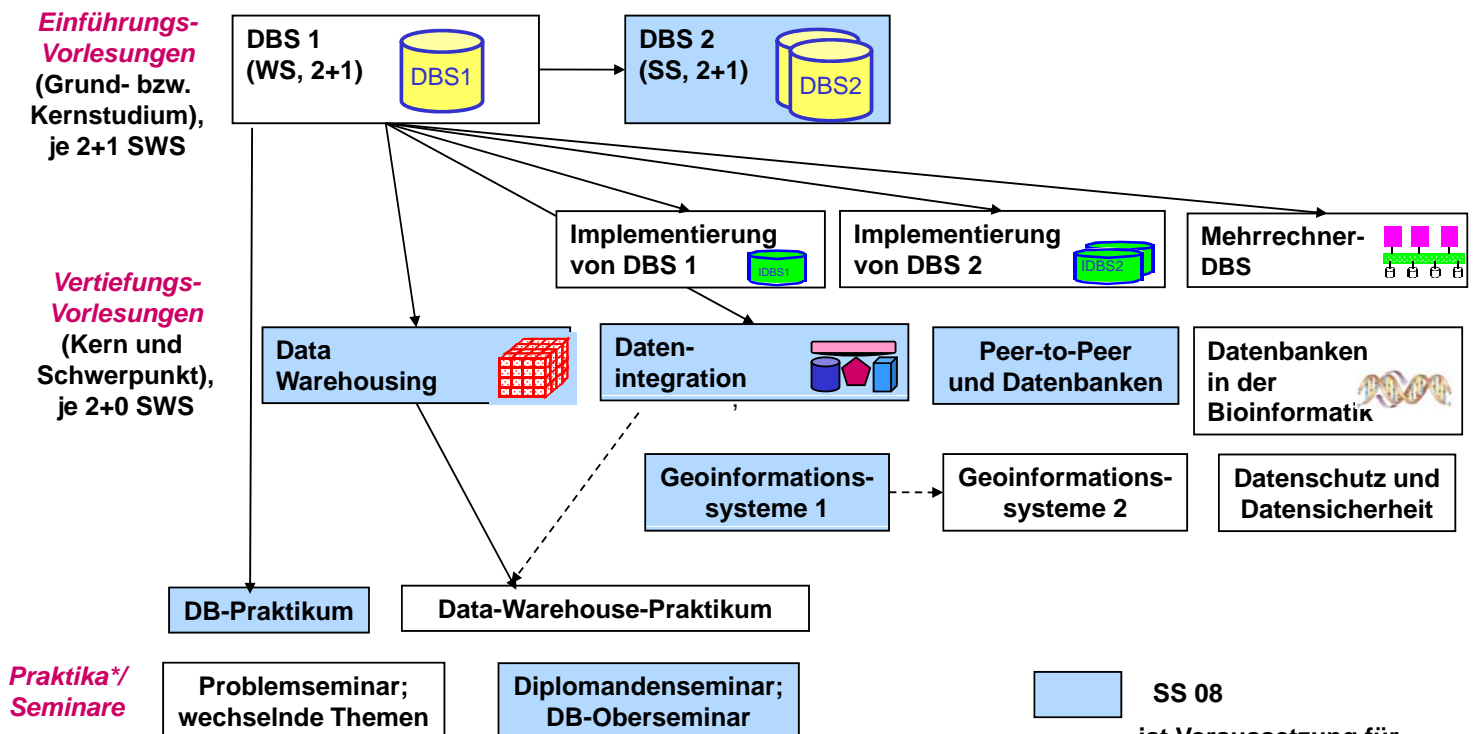
<http://dbs.uni-leipzig.de>



## Lehrveranstaltungen zu „Datenbanken“ (SS08)

**Einführungs-  
Vorlesungen**  
(Grund- bzw.  
Kernstudium),  
je 2+1 SWS

**Vertiefungs-  
Vorlesungen**  
(Kern und  
Schwerpunkt),  
je 2+0 SWS



\* Detaillierter Praktikumschein wird ausgestellt



# Anrechnung / Leistungsbewertung

- Vorlesung kann angerechnet werden
  - im Diplom-, Master- oder Bachelor-Studiengang Informatik K(ern oder Vertiefung)
  - In weiteren Studiengängen mit Informatikanteilen
- Diplomstudiengang Informatik (ab 5. Semester)
  - Modulklausur (60 Min.) im Juli-Sep. 2008 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“
  - 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)
- Wahlmöglichkeiten **SS08** für „Anw. Datenbanktechnologien“ aus
  - Data Warehousing
  - Datenintegration
  - Geoinformationssysteme 1
  - P2P und Datenbanken
  - ggf. Datenbanksysteme 2

Daneben kann bei Belegung der Vorlesung „Data Warehousing“ noch das im WS08/09 angebotene Data-Warehouse-Praktikum eingebracht werden.
- **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



# 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 2005 und Visual Studio
  - Gruppenarbeit (2 Studenten) mit 3 Testaten
- Organisatorisches
  - Wintersemester 2008/09 (evtl. Beginn bereits ab August 2008 möglich)



## 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



# 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: MOMA
  - Website Recommendations: AWESOME



# Online-Übungen

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

The screenshot shows the homepage of the Leipzig Online Test System. At the top, there is a blue header with the text 'Leipzig Online Test System' and 'UNIVERSITÄT LEIPZIG'. Below the header, there is a navigation bar with links for 'Home', 'Registrierung', and 'Impressum'. The main content area is divided into several sections: 'Login' with input fields for 'Username:' and 'Passwort:' and a 'login' button; 'Gast Login' with a text block explaining guest login and a 'als Gast einloggen' button; 'System Info' showing '# Benutzer: 2' and a clock icon; and 'News' with a text block encouraging users to log in for personalized news.



## Freie SHK-Stelle

- DFG-Projekt Dynamische Fusion verteilter Webdaten
- Voraussetzungen
  - Diplom, Bachelor- oder Master-Studium der Informatik
  - Gute Datenbankkenntnisse
  - Gute Kenntnisse bei der Erstellung von Web-Applikationen
  - Verantwortungsbewusstsein, Zuverlässigkeit
- Aufgaben
  - Implementationen zur Anbindung neuer Web-Quellen
  - Evaluation von Query- und Match-Verfahren zur Datenintegration
- Kontakt: Andreas Thor ([thor@informatik.uni-leipzig.de](mailto:thor@informatik.uni-leipzig.de))

