

Data Warehousing

Kapitel 0: Organisatorisches

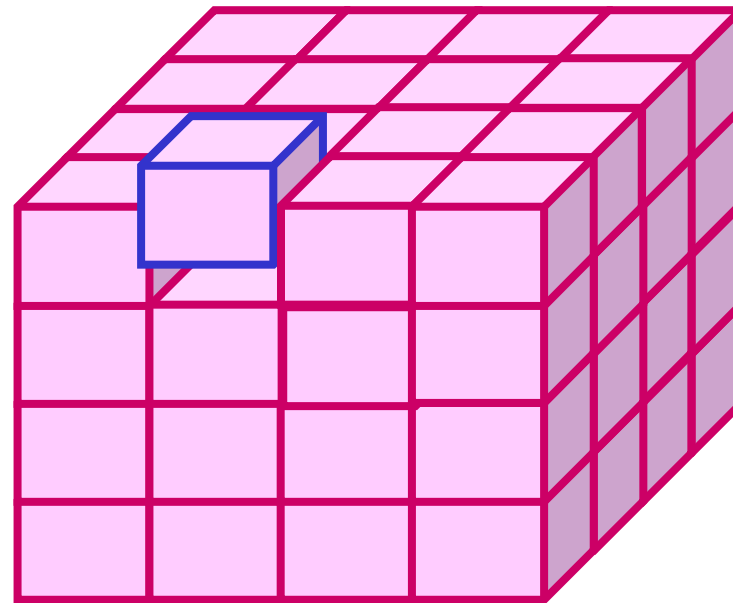
Michael Hartung

Sommersemester 2011

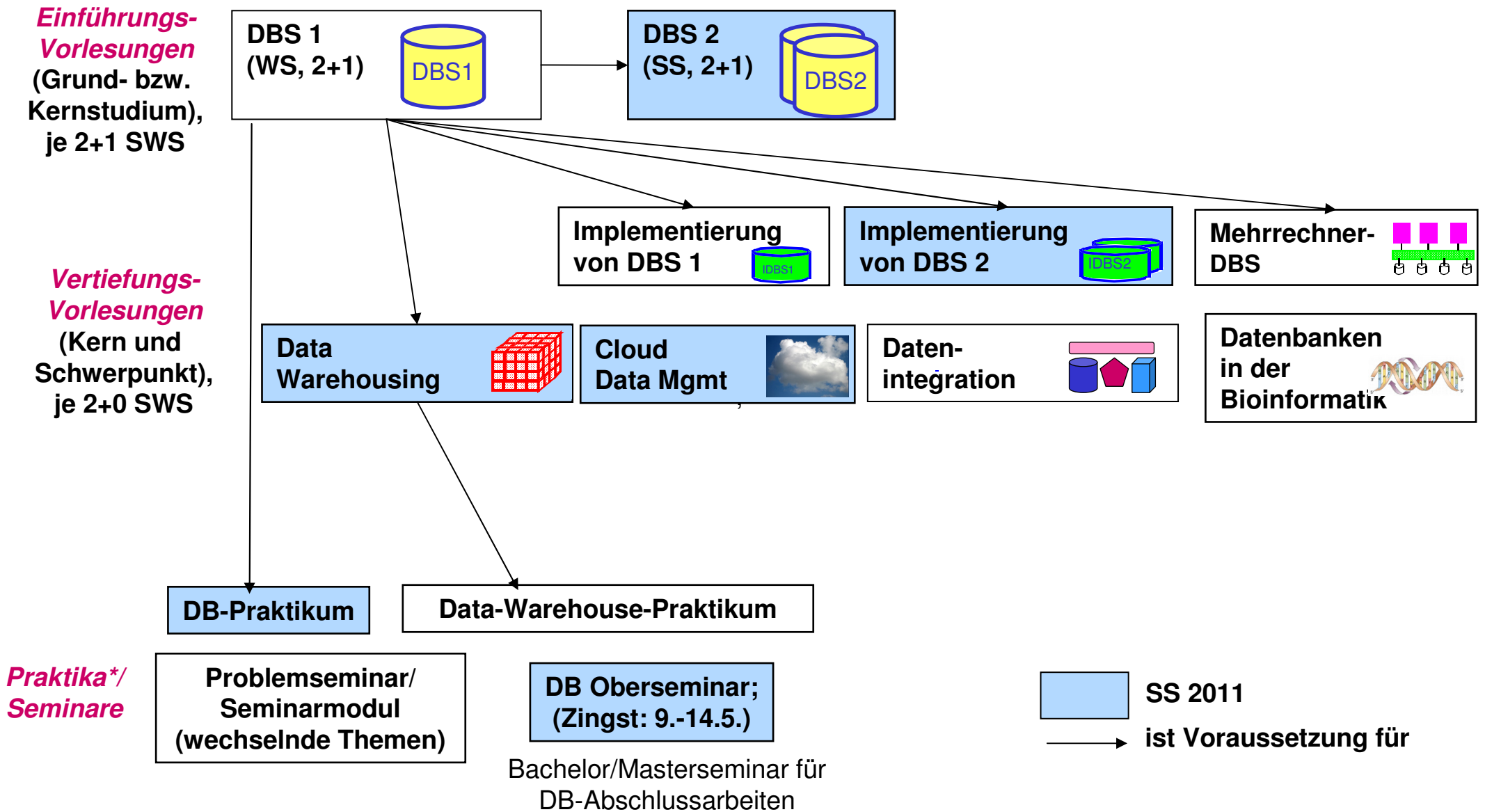
Universität Leipzig

Institut für Informatik

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



Lehrveranstaltungen zu „Datenbanken“ (SS2011)



* Detaillierter Praktikumsschein wird ausgestellt



Organisatorisches

- Termin: montags, 11:15 Uhr, SG 1-10 (Seminargebäude)
 - **Beginn:** 4. April 2011
 - **Letzte Vorlesung** am 11. Juli 2011
- Prüfung:
 - Klausur 60min am Ende des Semesters (konzeptionelles Wissen + Anwendungsfälle)
- Webseite mit allen wichtigen Infos:
 - <http://dbs.uni-leipzig.de/stud/2011ss/dwh>

Contents
▶ Staff
▶ Research
▼ Study
▶ Klausurtermine
▼ SS 2011
◻ DBS2
◻ DWH
◻ IDBS2
◻ dbprak
▶ Previous terms
▶ Module der Abteilung
▶ Study material
▶ E-Learning
▶ Theses
◻ Erasmus
▶ Service

Recent publications
• AMC – A Framework for Modelling and Comparing Matching Systems as Matching Processes
• ATOM: Automatic Target-driven Ontology Merging
• Multi-page Content

Data Warehousing SS 2011 (DWH)

- Dozent: M. Hartung
- Ort: SG 1-10 (Seminargebäude)
- Zeit: montags, 11:15 Uhr
- Beginn: 04.04.2011

Teilnehmerkreis:

Studenten der Informatik mit Abschluss Diplom/Bachelor/Master (Kern oder Vertiefung) und alle Interessenten

Übersicht:

Data Warehouses werden in zahlreichen Unternehmen zur Konsolidierung und Integration unterschiedlicher Daten (Entscheidungsunterstützung, Decision Support) eingesetzt. Die Vorlesung bietet eine Einführung in Architektur, ETL und sich an Informatiker, Wirtschaftsinformatiker und alle, die sich einen fundierten Einstieg in das wichtige Thema "DWH" wünschen.

Gliederung / Skript

Kapitel	Thema
0	Organisation/Inhalt
1	Einführung: Data Warehouses
2	Architektur von Data Warehouse-Systemen
3	Mehrdimensionale Modellierung und Operationen
4	ETL: Schemaintegration und Data Cleaning
5	Performance-Techniken: Indexstrukturen, Fragmentierung, materialisierte Sichten



DBS-Module

■ Master-Studium

- 10-202-2215 – Moderne Datenbanktechnologien (Kleines Modul)
- 10-202-2216 – Moderne Datenbanktechnologien (Großes Modul)
- 10-202-2213 – Anwendungsbezogene Datenbankkonzepte (Kleines Modul)
- 10-202-2214 – Anwendungsbezogene Datenbankkonzepte (Großes Modul)
- 10-202-2011 – Masterseminar Informatik / Seminarmodul

■ Bachelor-Studium

- 10-201-2211 – Datenbanksysteme 1
- 10-201-2212 – Datenbanksysteme 2
- 10-201-2210 – Datenbankpraktikum
- 10-201-2224 – *Realisierung von Informationssystemen*
- 10-201-2010 – Bachelorseminar Informatik / Seminarmodul



Mapping Module – Lehrveranstaltungen SS11

- Master-Module “Anwendungsbezogene Datenbankkonzepte” (zwei bzw. drei Veranstaltungen für kleines bzw. großes Modul):
 - Implementierung von DBS 2
 - Data Warehousing
 - Cloud Data Management
 - ggf. DBS2
- Bachelor-Modul “Realisierung von Informationssystemen”
 - Implementierung von DBS 2
 - Data Warehousing
- Bachelorseminar / Masterseminar
 - Vortrag über laufende Bachelor/Masterarbeit



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 Testdaten
- Organisatorisches
 - Wintersemester 2011/12



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>

The screenshot shows the homepage of the Leipzig Online Test System (LOTS). The header features the system name and the University of Leipzig logo. A navigation menu includes 'Home', 'Registrierung', and 'Impressum'. The main content area is divided into three sections: 'Login', 'Gast Login', and 'System Info'. The 'Login' section contains input fields for 'Username:' and 'Passwort:' with a 'login' button. The 'Gast Login' section provides information about guest accounts and a 'als Gast einloggen' button. The 'System Info' section displays the number of users and the current date and time.

Leipzig Online Test System **UNIVERSITÄT LEIPZIG**
Institut für Informatik
Abteilung DBS

Home **Registrierung** **Impressum**

Login

Username:

Passwort:

login

Gast Login

Sie können sich als Gast einloggen, um LOTS ohne vorherige Anmeldung zu testen.
Der Gast Account ist auf 30 min Benutzung und im Funktionsumfang beschränkt.
Bei weiterem Interesse sollten Sie sich [registrieren](#).

Viel Spass!

Ihr LOTS Team

als Gast einloggen

System Info

Benutzer: 2

 12.10.2005
15:27:05

News

Bitte loggen Sie sich ein, um personalisierte News lesen zu können.



Lehrstuhl Datenbanksysteme

- Seit 1994 am Institut für Informatik
- > 20 Mitarbeiter
- Umfangreiches Lehrangebot
 - Vorlesungen, Praktika, Seminare
 - Online-Übungssystem LOTS
 - Eigene Lehrbücher: MRDBS, IDBS, ...



Forschung



Abteilung Datenbanken Leipzig

am Institut für Informatik



Startseite

Inhalte

- ▶ Mitarbeiter
- ▼ Forschung
 - Prototypes
 - Publikationen
 - ▶ Projekte
 - Jahresberichte
 - Kooperationen
 - ▶ Promotionen
 - Colloquia
 - ▶ Conferences
- ▶ Studium
- ▶ Service

Forschung

Anzeigen

translation

Bearbeiten

Gliederung

Versionen

[Publications \(2011, 2010, 2009, 2008, 2007, 2006, 2005, 2004, 2003, ...\)](#)

[WDI Lab: Innovation Lab on Web Data Integration](#) 

[Mashup-based data integration](#)

[Schema and Ontology Matching \(COMA++\)](#)

[Object Matching / Entity Resolution \(FEVER, MOMA\)](#)

[Large Scale Entity and Ontology Matching](#)

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



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/WHK-Verträge
 - Mitarbeit im BMBF-geförderten WDI-Lab zur Webdatenintegration oder in Rahmen von DFG-Projekten
 - Kooperative Arbeitsatmosphäre
 - Unterstützung bei der Suche nach passendem Bachelor-, Master- bzw. Diplomarbeitsthema
- Ansprechpartner: Prof. Rahm oder Frau Röllig



Abschlussarbeiten

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

