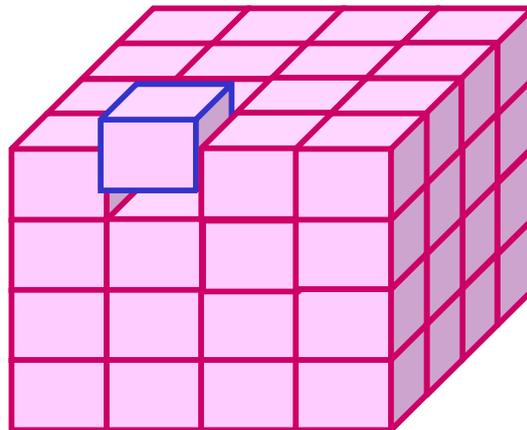


Data Warehousing

Sommersemester 2016

Prof. Dr. E. Rahm

Universität Leipzig
Institut für Informatik



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



DBS-Module

■ Master-Studium Informatik

- 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 / Seminar modul

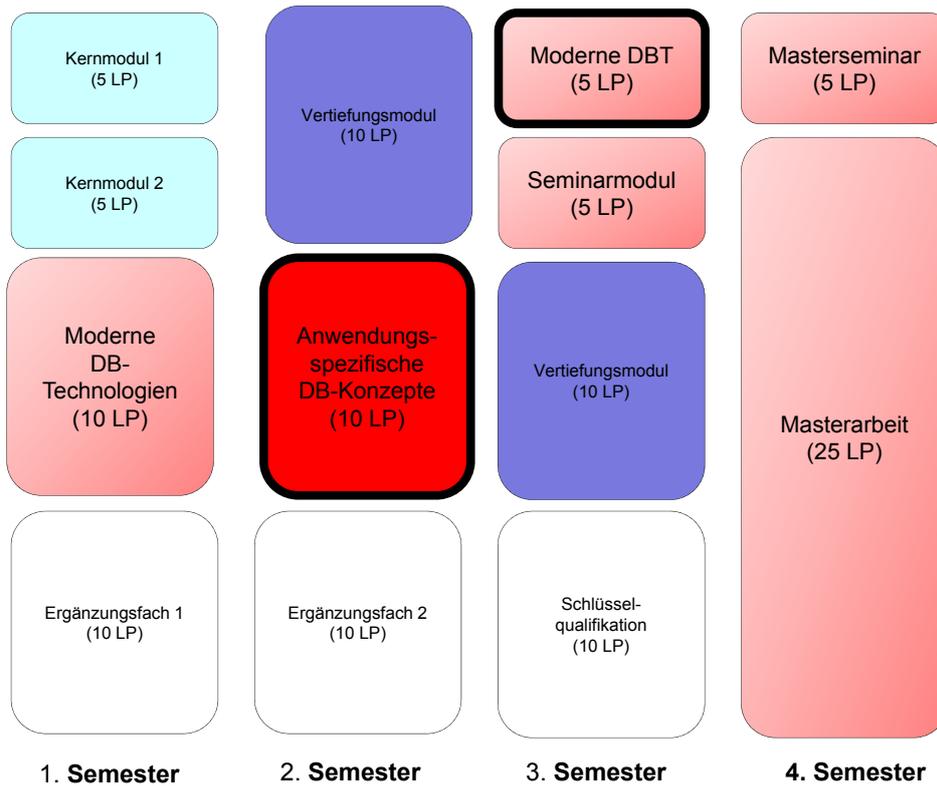
■ Bachelor-Studium Informatik

- 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 / Seminar modul

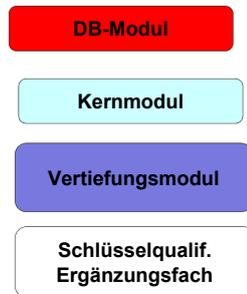
■ Anrechenbarkeit für Bachelor/Master Wirtschaftsinformatik und Master Bioinformatik



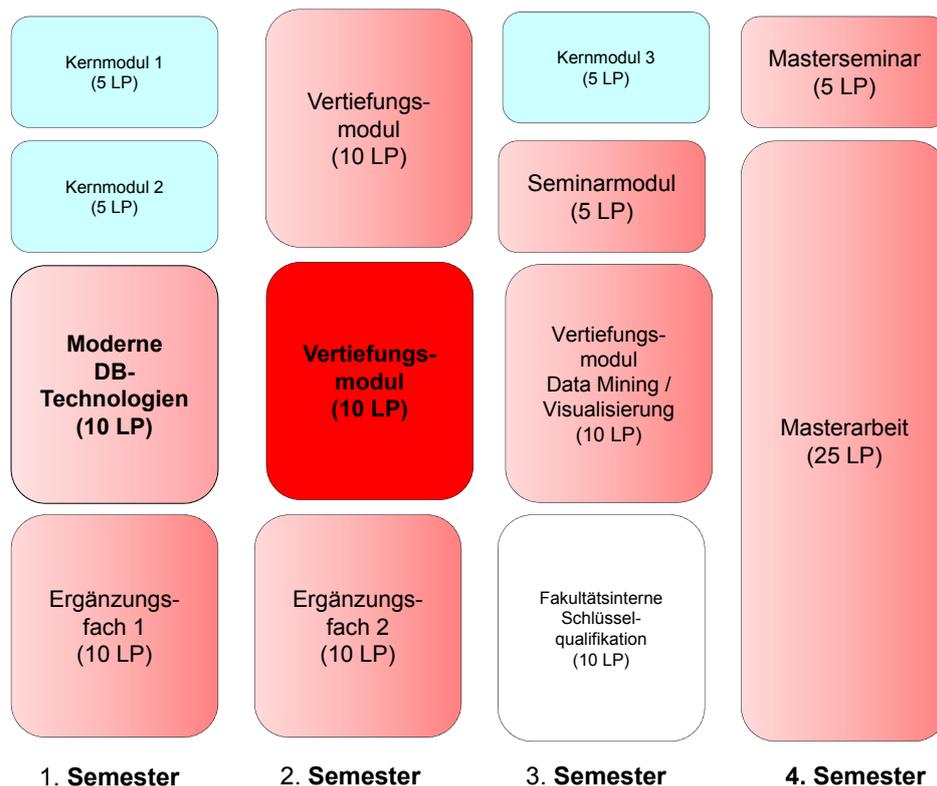
Informatik Masterstudium DBS-Profil



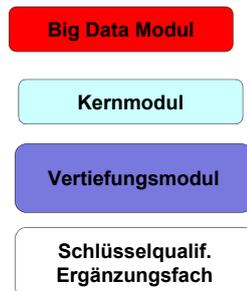
Legende:



Masterstudium Informatik: Schwerpunkt Big Data (NEU)



Legende:



Big Data Schwerpunkt: wählbare Module

■ Datenmanagement

- Moderne Datenbanktechnologien (Pflichtmodul, 10 LP)
- **Anwendungsbezogene Datenbankkonzepte (10 LP)**

Die beiden Module können auch in einer Kernfachvariante (5 LP) belegt werden und beinhalten verschiedene Vorlesungen, Praktika und Seminare. Besonders relevant sind die Vorlesungen *Cloud Datenmanagement*, *NoSQL-Datenbanken*, *Mehrrechner-DBS*, *Data Warehousing* und *Datenintegration* sowie das *Big Data Praktikum* und *Data Warehouse Praktikum*.

■ Data Mining / Visualisierung (je 10 LP), Wahlpflicht

- Neuroinspirierte Informationsverarbeitung
- Statistisches Lernen (inkl. R-Kurs)
- Visualisierung (inkl. Praktikum)

■ Web / Information Retrieval

- Fortgeschrittene Methoden des Information Retrieval
- Semantic Web
- Textdatenbanken (5 LP)
- Computational Advertising (5 LP)

■ Ergänzungsfächer

- Sequenzanalyse und Genomik
- Statistische Aspekte der Analyse molekularbiologischer und genetischer Daten
- Medizinische Bildaufnahme und Bildverarbeitung
- Graphen und biologische Netzwerke
- Alternativ: Module zu Digital Humanities



BACHELOR DBS-Profil

Algorithmen und Datenstrukturen 1 (5 LP)	Algorithmen und Datenstrukturen 2 (5 LP)	DBS 1 (5 LP)	DBS 2 (5 LP)	Realisierung von IS (5 LP)	Kernmodul 5 (5 LP)
Modellierung und Programmierung 1 (5 LP)	Modellierung und Programmierung 2 (5 LP)	Softwaretechnik (5 LP)	Softwarepraktikum (5 LP)	Seminarmodul (5 LP)	Bachelorseminar (5 LP)
Technische Informatik 1 (5 LP)	Technische Informatik 2 inkl. Hardwarepraktikum (5 LP)	Kommunikationssysteme (5 LP)	Kernmodul (5 LP)	Vertiefungsmodul (10 LP)	Bachelorarbeit (10 LP)
Logik (5 LP)	Java-Praktikum (5 LP)	Automaten und Sprachen (5 LP)	Berechenbarkeit (5 LP)		
Analysis 1 (10 LP)	Lineare Algebra 1 (10 LP)	Diskrete Strukturen (5 LP)	DB-Praktikum (5 LP)	Ergänzungsfach (10 LP)	Fakultätsübergreifende Schlüsselqualifikation (10 LP)
		Wahrscheinlichkeitstheorie (5 LP)	Ergänzungsfach (5LP)		
1. Semester	2. Semester	3. Semester	4. Semester	5. Semester	6. Semester

Legende:

Inf-Pflichtmodul	MI-Modul
DBS-Modul	Mathematikmodul
	Schlüsselqualif. Ergänzungsfach



DBS-Lehrveranstaltungen

Logo	Name	Typ	SWS	Sem.
	Datenbanksysteme 1	Einführung	2+1	WS
	Datenbanksysteme 2	Einführung	2+1	SS
	Implementierung von DBS 1	Vertiefung	2	WS
	Implementierung von DBS 2	Vertiefung	2	SS
	Mehrrechner-DBS	Vertiefung	2	WS
	Data Warehousing	Vertiefung	2	SS
	Datenintegration	Vertiefung	2	SS
	Bio Data Management	Vertiefung	2	WS
	Ontologie-Management	Vertiefung	2	WS
	Cloud Data Management	Vertiefung	2	SS
	NoSQL-DB	Vertiefung	2	WS

Name	Typ	Sem.
DB-Praktikum	Praktikum	SS
Data-Warehouse-Praktikum	Praktikum	WS
Big Data Praktikum	Praktikum	WS
Problemseminar	Seminar	WS
Bachelorseminar	Seminar	SS+WS
Masterseminar	Seminar	SS+WS

LV im SS2016



SS16, © Prof. Dr. E. Rahm

Mapping: Module – Lehrveranstaltungen SS16

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

SS16, © Prof. Dr. E. Rahm



Leistungsbewertung

- Prüfungsklausur von 60 Minuten
 - Juli 2016
 - überprüft konzeptionelles Wissen + Anwendungsfälle
- Klausurerfolg durch
 - Vorlesungsteilnahme und –nachbearbeitung
 - Online-Übungen
 - Literatur



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) / Instanz-Matching

5. Performance-Techniken

- Indexstrukturen
- Materialisierte Sichten
- Nutzung paralleler Datenbanken

6. Data Mining-Verfahren

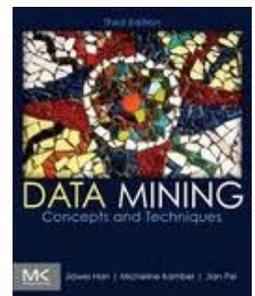
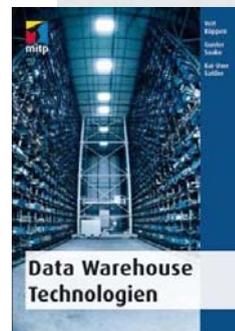
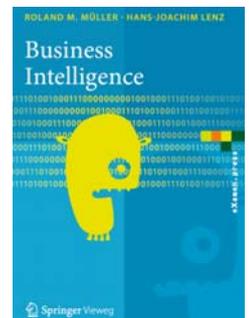
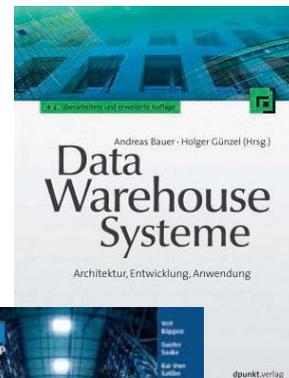
- Clusterung
- Assoziationsregeln



Literatur

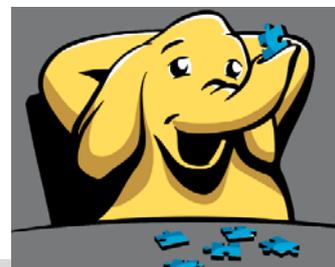
■ Buchauswahl

- Bauer/Günzel (Hrsg.): *Data Warehouse Systeme*. 4. Auflage, dpunkt 2013
- Köppen/Saake/Sattler: *Data Warehouse Technologien*. Mitp 2012
- Müller/Lenz: *Business Intelligence*. Springer 2013
- Han / Kamber / Pei: *Data Mining: Concepts and Techniques*. 3rd edition, Morgan Kaufmann 2011



■ Eigene Forschungsarbeiten: <http://dbs.uni-leipzig.de>

- Schema Matching (COMA++, GOMMA)
- Data Cleaning / Instanz-Matching (DEDOOP)



Online-Übungen

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



Leipzig Online-Test-System

UNIVERSITÄT LEIPZIG
Fakultät für Mathematik und Informatik
Institut für Informatik
Abteilung Datenbanken

Home | **Registrierung** | Impressum

Login

Username:

Password:

[Passwort vergessen?](#)

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 als Gast [registrieren](#).

Viel Spass!

Ihr LOTS Team

System Info

#Benutzer: 31

 05.02.2007
12:15:21

News

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



LOTS: Online-Übungen

Aufgabe 3 (Nested-Block-Join)

Für den Gleichverbund zwischen R und S (je 100.000 Sätze, Blockungsfaktor 100) soll ein Nested-Block-Join genutzt werden. Welche Kombinationen zwischen verfügbarer Hauptspeichergröße M und Anzahl erreichbarer Plattenzugriffe (ohne Schreiben des Resultats) treffen zu?

0 von 2

- M=1001; 1 Million Plattenzugriffe
- M=5001; 2000 Plattenzugriffe
- M=1001; 2000 Plattenzugriffe
- M=501; 3000 Plattenzugriffe



Aufgabe 4 (Hash-Join)

Markieren Sie die zutreffenden Aussagen.

0 von 2

- Durch die Nutzung von Bitvektoren lässt sich der Umfang der Hash-Tabellen kleiner halten
- Hash-Joins sollten nur eingesetzt werden, wenn die kleinere Eingabetabelle im Hauptspeicher gespeichert werden kann
- Hash-Joins können effektiv zur Realisierung von Mehr-Wege-Joins genutzt werden
- Hash-Join-Verfahren eignen sich zur Beantwortung von Equi-Join-Anfragen



Forschung



Database Group Leipzig

within the department of computer science

UNIVERSITÄT LEIPZIG

[help](#) | [register](#)



Home

Contents

- ▶ Staff
- ▼ Research
 - Publications
 - ▶ Projects
 - Prototypes
 - Annual reports
 - Cooperations
 - ▶ Graduations
 - Colloquia
 - ▶ Conferences
 - ▶ Study
 - ▶ Service

Research

[Publications \(2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005, ...\)](#)

[Big Data Center ScaDS Dresden/Leipzig](#)

[Entity Matching for Big Data \(Dedoop\)](#)

[Graph-based data analysis \(GRADOOP, BIIIG\)](#)

[Evolution of Semantic Annotations \(ELISA\)](#)

[Schema and Ontology Matching \(COMA++, GOMMA\), Ontology Merging \(ATOM\)](#)

[Semantic Enrichment of Ontology Mappings \(STROMA\)](#)

[Background Knowledge Exploitation \(SemRep\)](#)

[LOD Link Discovery](#)

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

[Web Data Integration \(WDI\) Lab](#)

[Bibliometric Analysis](#)



Recent publications

- [Semi-Automatic Identification of Counterfeit Offers in Online Shopping Platforms](#)

SS16, © Prof. Dr. E. Rahm

0-15



UNIVERSITÄT LEIPZIG



BIG DATA CENTER

- Life sciences
- Material and Engineering sciences
- Environmental / Geo sciences
- Digital Humanities
- Business Data



Big Data Life Cycle Management and Workflows

Data Quality / Data Integration	Knowledge Extraktion	Visual Analytics
------------------------------------	-------------------------	---------------------

Efficient Big Data Architectures



Abschlussarbeiten / SHK-Jobs

- Wir suchen Studenten mit guten DB-Kenntnissen für
 - Bachelor- oder Masterarbeiten
 - SHK/WHK-Tätigkeiten

- Mitarbeit an aktuellen Forschungsthemen
 - des Lehrstuhls bzw
 - des BigData-Zentrums ScaDS

