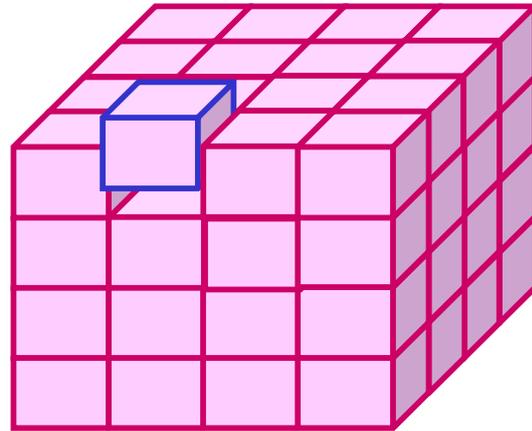


# Data Warehousing

Sommersemester 2014

**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

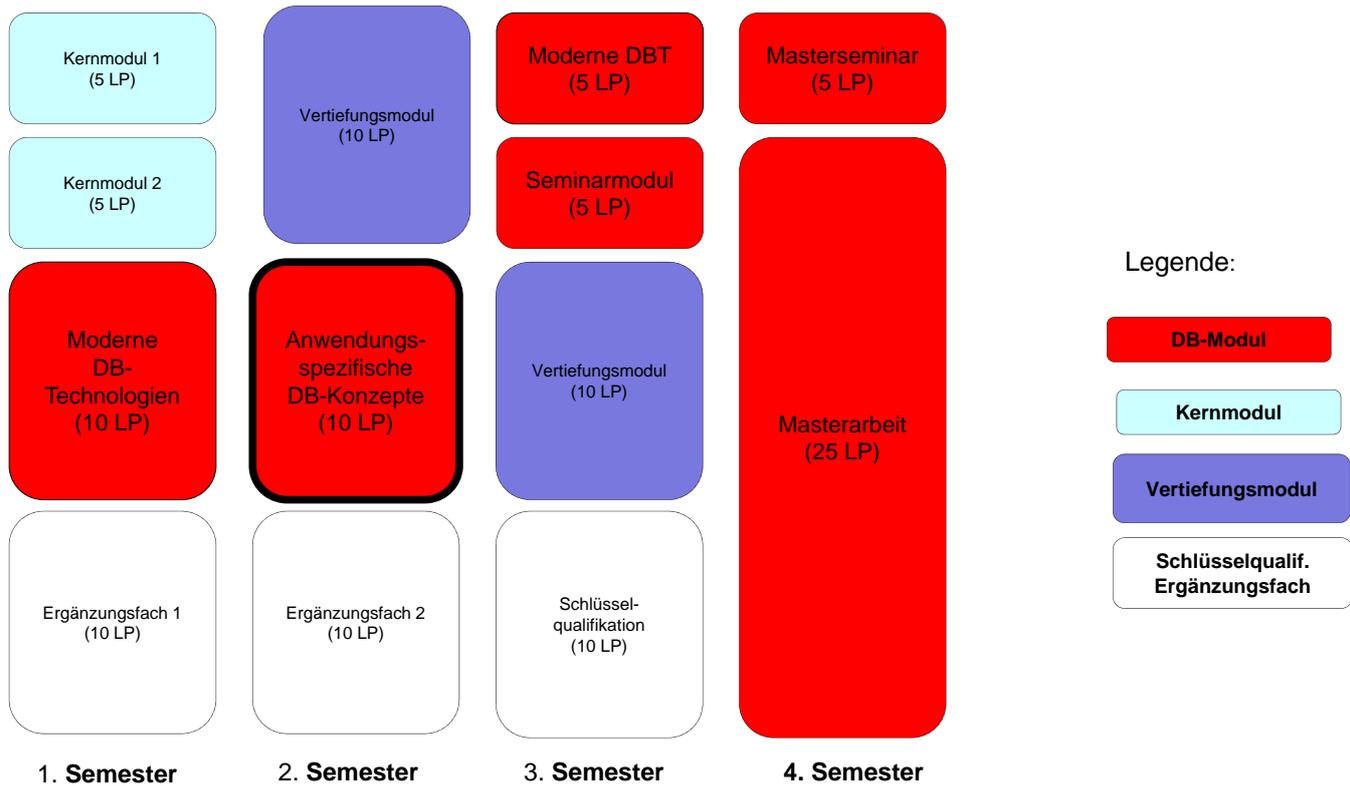
### ■ Bachelor-Studium Informatik

- 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**

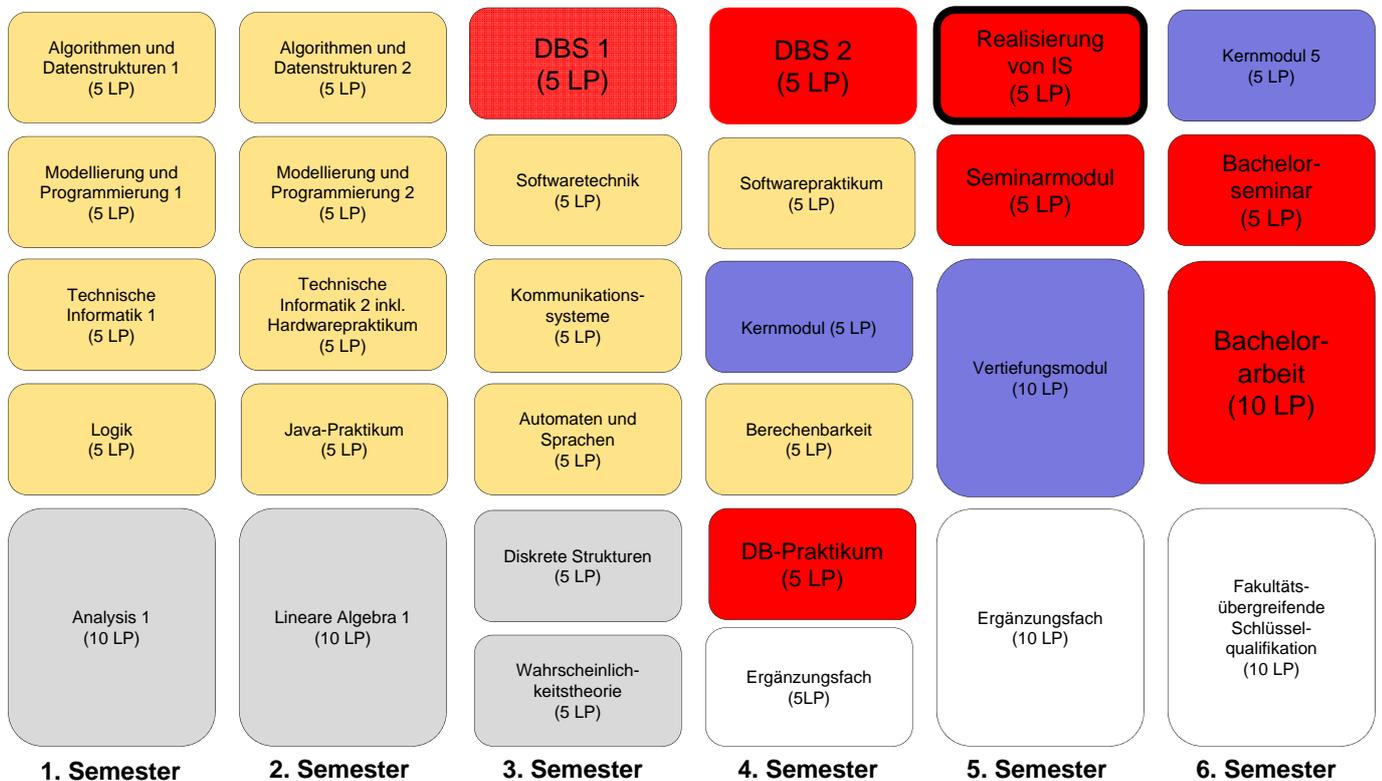
### ■ Anrechenbarkeit für Bachelor/Master Wirtschaftsinformatik und Master Bioinformatik



# Masterstudium DBS-Profil



# BACHELOR DBS-Profil



**Legende:**



# 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	SS
	Ontologie-Management	Vertiefung	2	WS
	Cloud Data Management	Vertiefung	2	WS
	NoSQL-DB	Vertiefung	2	SS

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

LV im SS14



SS14, © Prof. Dr. E. Rahm

## Mapping: Module – Lehrveranstaltungen SS14

- Master-Module “Anwendungsbezogene Datenbankkonzepte” (zwei bzw. drei Veranstaltungen für kleines bzw. großes Modul):
  - Data Warehousing
  - NoSQL-Datenbanken
  - Datenintegration
  - ggf. DBS2
- Bachelor-Modul “Realisierung von Informationssystemen”
  - Data Warehousing
  - NoSQL-Datenbanken oder Datenintegration
- Bachelorseminar / Masterseminar
  - Vortrag über laufende Bachelor/Masterarbeit
  - externes Seminar Uni-Außenstelle Zingst (Juni 2014)

SS14, © Prof. Dr. E. Rahm



# Leistungsbewertung

- Prüfungsklausur von 60 Minuten
  - Juli 2014
  - 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

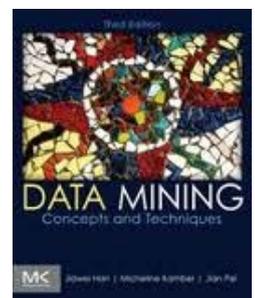
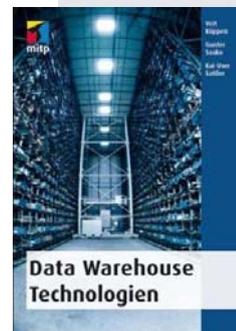
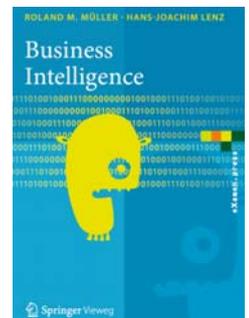
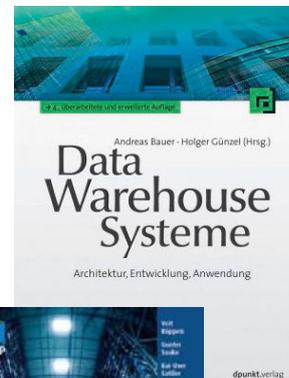
## 7. Data-Warehouse-Anwendung



## Literatur

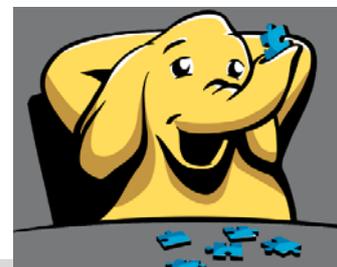
### ■ Buchauswahl

- 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 (FEVER, 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

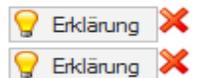


### Aufgabe 4 (Hash-Join)

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



# Lehrstuhl Datenbanksysteme

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



## Forschung

**Database Group Leipzig**  
within the department of computer science

UNIVERSITÄT LEIPZIG  
help | register

Home  search

**Contents**

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

**Research**

Publications (2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005, ...)

Entity Matching for Big Data

Graph-based Business Intelligence (BIIIG)

Evolution of ontologies and mappings (Schema Evolution bibliography)

Schema and Ontology Matching (COMA++, GOMMA), Ontology Merging (ATOM)

Bibliometric Analysis

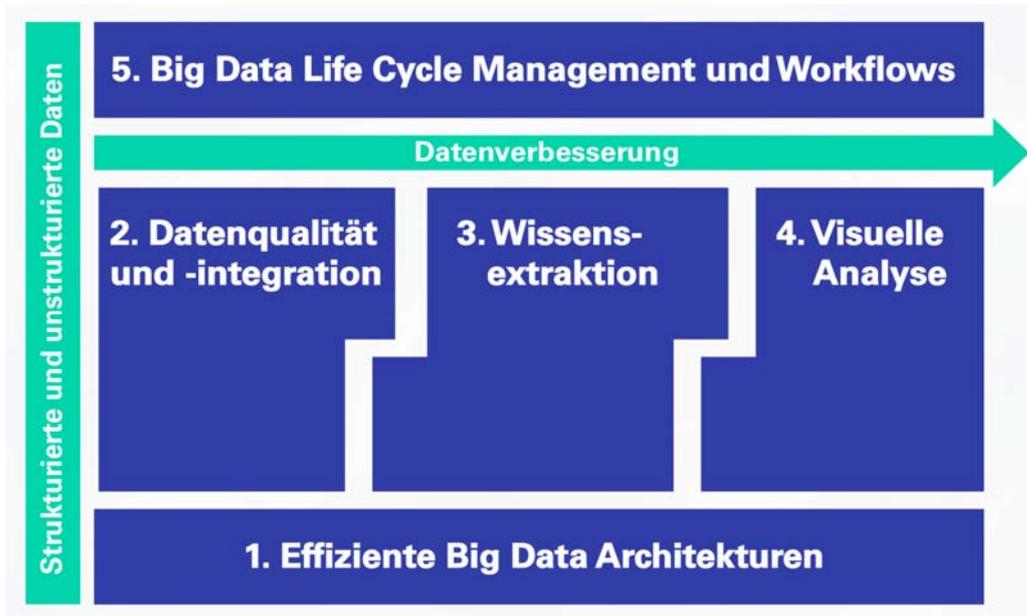
Current prototypes and implementations

Object Matching / Entity Resolution

Previous projects



- **ScaDS Dresden/Leipzig** (Competence Center for **Scalable Data Services** and **Solutions Dresden/Leipzig**)
- Forschungsschwerpunkte



## Abschlussarbeiten / SHK-Jobs

- Wir suchen Studenten mit guten DB-Kenntnissen für
  - Bachelor- oder Masterarbeiten
  - SHK/WHK-Tätigkeiten
- Mitarbeit an aktuellen Forschungsthemen des Lehrstuhls
- Kooperationen mit regionalen Unternehmen
  - eigene Spinoffs

