

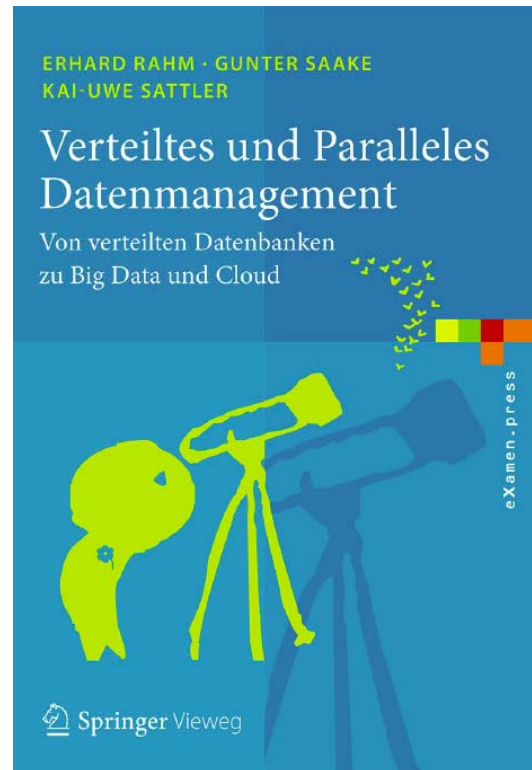
Mehrrechner-Datenbanksysteme (Verteilte und Parallele DBS)

Prof. Dr. E. Rahm

Wintersemester 2017/2018

Universität Leipzig
Institut für Informatik

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



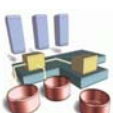
DBS-Module

■ Master-Studium




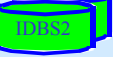







- 10-202-2215 – Moderne Datenbanktechnologien (Kleines Modul)
- 10-202-2216 – Moderne Datenbanktechnologien (Großes Modul)
- 10-202-2213 – Anwendungsbezogene Datenbankkonzepte (Kleines Modul)
- 10-202-2214 – Anwendungsbezogene Datenbankkonzepte (Großes Modul)
 - Seminar modul
- 10-202-2011 – Masterseminar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

■ Bachelor-Studium

- 10-201-2211 – Datenbanksysteme 1
- 10-201-2212 – Datenbanksysteme 2
- 10-201-2210 – Datenbankpraktikum
- 10-201-2224 – Realisierung von Informationssystemen
 - Seminar modul
- 10-201-2010 – Bachelorseminar Informatik
 - Bachelorarbeit



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

Name	Typ	Sem.
DB-Praktikum	Praktikum	SS
Data-Warehouse-Praktikum	Praktikum	WS
Big Data Praktikum	Praktikum	SS
Problem-seminar	Seminar	WS
Bachelor-seminar	Seminar	SS+WS
Master-seminar	Seminar	SS+WS

LV im WS2017/18



Mapping Module – Lehrveranstaltungen WS17/18

- Master-Module **Moderne Datenbanktechnologie** (zwei bzw. drei Veranstaltungen für kleines bzw. großes Modul):
 - **Mehrrechner-DBS**
 - Cloud Data Management
 - Forschungsseminar *Deep Learning* (max. 20 TN)
 - Data Warehouse Praktikum (max. 18 TN)
- Bachelor-Modul **Realisierung von Informationssystemen**
 - **Mehrrechner-DBS**
 - Cloud Data Management
- **Seminarmodul**
 - Forschungsseminar *Deep Learning*
- **Bachelorseminar / Masterseminar**
 - Vortrag über laufende Bachelor/Masterarbeit



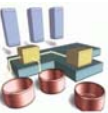
Masterstudium Informatik

Sem	5 LP	5LP	10 LP	10 LP
1.	Kernmodul I	Kernmodul II	Vertiefungsmodul I	Ergänzungsfach
2.	Kernmodul III	Seminarmodul	Vertiefungsmodul II	Ergänzungsfach
3.	Vertiefungsmodul III		Vertiefungsmodul IV	Fakultätsinterne Schlüsselqualifikation
4.	Masterseminar	Masterarbeit		

■ mögliche Schwerpunkte (Ausweisung im Zeugnis)

- Big Data
- Medizinische Informatik

■ individuelle Schwerpunktbildung, zB. Datenmanagement/DB



Master Informatik mit DB-Profil

Sem	5 LP	5LP	10 LP	10 LP
1.	Kernmodul I	Kernmodul II	Moderne DB-Technologien	Ergänzungsfach
2.	Vertiefungsmodul III		Anwendungsspezifische DB-Konzepte	Ergänzungsfach
3.	Moderne DB-Technologien	Seminarmodul	Vertiefungsmodul IV	Fakultätsinterne Schlüsselqualifikation
4.	Masterseminar	Masterarbeit		

DB-Modul

- *Moderne DBT und Anwendungsspez. DB-Konzepte* können in verschiedenen Semestern in großer und kleiner Ausführung belegt werden
 - falls disjunkte Untersetzung vorliegt



Master Informatik mit Schwerpunkt Big Data

Sem	5 LP	5LP	10 LP	10 LP
1.	Kernmodul I	Kernmodul II	Moderne DB-Technologien	Ergänzungsfach
2.	Kernmodul III	Seminarmodul	Big Data Vertiefung	Ergänzungsfach
3.	Big Data Vertiefung		Vertiefungsmodul zu Data Mining / Visualisierung	Fakultätsinterne Schlüsselqualifikation
4.	Masterseminar	Masterarbeit		

Big Data Modul

- obligatorisches Vertiefungsmodul *Moderne DB-Technologien*
- weitere Kern/Vertiefungsmodule: Anwend.spez. DBK, neuroinspirierte Informationsverarbeitung, statistisches Lernen, Visualisierung, Information Retrieval, Semantic Web, Computational Advertising
- Ergänzungsfächer z.B. aus Bioinformatik oder Digital Humanities



Bachelor Informatik mit DB-Profil

Sem.	5 LP	5 LP	5 LP	5 LP	5 LP	5 LP
1	Modellierung und Programmierung 1	Algorithmen u. Datenstrukturen 1	Technische Informatik 1	Logik	Analysis	
2	Modellierung und Programmierung 2	Algorithmen u. Datenstrukturen 2	TI 2 + HW-Praktikum	Java-Praktikum	Lineare Algebra	
3	Datenbanksysteme 1	Software-Technik	Kommunikationssysteme	Automaten und Sprachen	Diskrete Strukturen	Wahrscheinlichkeitstheorie
4	Datenbanksysteme 2	Softwaretechnik-Praktikum	Kernmodul 2	Berechenbarkeit	DB-Praktikum	
5	Realisierung v IS	Seminarmodul	Vertiefungsmodul		Ergänzungsfach	
6	Kernmodul 4	Bachelorseminar	Bachelorarbeit		Schlüsselqualifikation	

Legende:

Praktische Inf.	Technische Inf.	Wahlmodule
DBS-Modul	Theoretische Inf.	Schlüsselqualif. Ergänzungsfach
	Mathematikmodul	



Master Wirtschaftsinformatik

Sem.	10 LP	10 LP	10 LP
1	Wahlpflichtmodule aus den Masterstudiengängen Betriebswirtschaftslehre bzw. Volkswirtschaftslehre	Integration und Architektur von Anwendungssystemen	Wahlpflichtmodule aus den Masterstudiengängen Betriebswirtschaftslehre bzw. Volkswirtschaftslehre
2	Anwendungssysteme I Modellierung und Management von Geschäftsprozessen	Zusammenstellung aus: a) den Modulen des Masterstudiengang Informatik: Neuroinspirierte Informationsverarbeitung, Ausgewählte Verfahren des Mobile Advertising, Grundlagen komplexer Systeme, Textdatenbanken b) den Wahlpflichtmodulen der Masterstudiengänge BWL/VWL c) den Modulen der Wirtschaftsinformatik: Architekturen und Systeme für Geschäftsprozesse, Softwaresystemfamilien und -produktlinien, Business Innovation	Advanced Software Engineering
3	Anwendungssysteme II Überbetriebliche Anwendungssysteme	Vertiefungsmodul aus > Service Science, > Computational Advertising > Mobile Peer-to Peer Systeme, > Wissensrepräsentation	Wahlpflichtmodule aus den Masterstudiengängen Betriebswirtschaftslehre bzw. Volkswirtschaftslehre
4	Vertiefungsmodul aus > Anwendungsbezogene Datenbankkonzepte, > Betriebliche Informationssysteme oder > ein Wahlpflichtmodul aus 2. Semester a) oder c)	Masterarbeit	



Wirtschaftsinformatik



Vorläufiges Inhaltsverzeichnis

■ Einführung

- Anforderungen an Mehrrechner-Datenbanksysteme (Verteilte / Parallele DBS)
- Arten der Parallelität, Scaleup und Speedup

■ Klassifikation von Mehrrechner-DBS

■ VDBS: Schemaarchitektur, Katalogverwaltung

■ Datenverteilung in VDBS / PDBS (Fragmentierung, Allokation)

■ Verteilte / Parallele Query-Verarbeitung

■ Verteilte Transaktionsverwaltung (Commit-Protokolle, Synchronisation)

■ Replizierte DB

■ Shared-Disk-DBS

*Data Warehouses, Datenintegration, Cloud Data Management. NoSQL
-> eigene Vorlesungen*



Lehrbuch



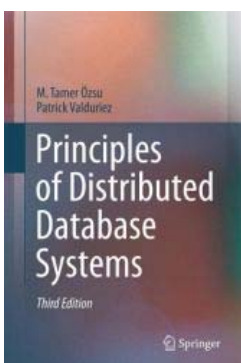
- Rahm/Saake/Sattler: *Verteiltes und Paralleles Datenmanagement*. Springer 2015, 35 Euro
 - E-Book online zugänglich im Uni-Netz
 - mit Übungsaufgaben



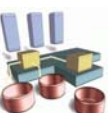
Zusatzliteratur



- E. Rahm: *Mehrrechner-Datenbanksysteme*.
 - Addison-Wesley 1994
 - Übungsaufgaben mit Musterlösungen
 - Online-Version verfügbar (PDF und HTML)




- M. T. Özsu, P. Valduriez: *Principles of Distributed Database Systems*.
 - 3rd edition, Springer-Verlag, 2011



Online-Übungen

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



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

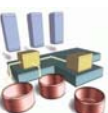
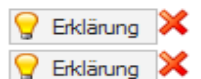


Aufgabe 4 (Hash-Join)

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

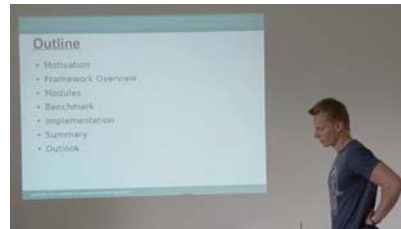


Lehrstuhl Datenbanksysteme

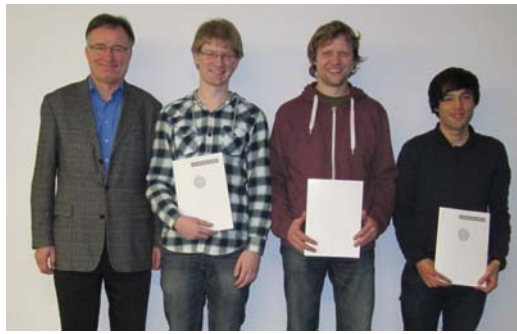
- seit 1994 am Institut für Informatik
- umfangreiches Lehrangebot
 - Vorlesungen, Praktika, Seminare
 - Online-Übungssystem LOTS
 - Eigene Lehrbücher: MRDBS, IDBS, ...



Oberseminare an Uni-Außenstelle Zingst – seit 2001



Auszeichnung von Top-Student(inn)en - seit 2008



WS17/18, © Prof. Dr. E. Rahm

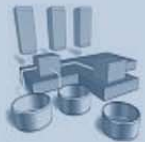
0 - 17



Forschung

Abteilung Datenbanken Leipzig

am Institut für Informatik



Suchen

UNIVERSITÄT LEIPZIG

Hilfe | Registrieren



Startseite

login

Inhalte

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

Forschung

Publications (2017, 2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005, ...)

Big Data Center ScaDS Dresden/Leipzig

Graph-based data analysis (GRADOOP, BIIIG)

Privacy-Preserving Record Linkage for Big Data

Entity Matching for Big Data (Dedoop)

LOD Link Discovery

Evolution of Semantic Annotations (ELISA)

Schema and Ontology Matching (COMA++, GOMMA), Ontolog



WS17/18, © Prof. Dr. E. Rahm

0 - 18



E. Rahm

Lebenswissenschaften

Werkstoff- und Ingenieurwissenschaften

Umwelt- /Verkehrswissenschaften

Digital Humanities

Business Data

Service-
zentrum

W. E. Nagel

Big Data Life Cycle Management und Workflows

Datenqualität /
DatenintegrationWissens-
extraktionVisuelle
Analyse

Effiziente Big Data Architekturen



W. Lehner



B. Franczyk



M. Bogdan



P. Stadler



G. Heyer



G. Scheuermann



S. Gumhold



Alle angebotenen Bachelorarbeiten können sowohl in Deutsch als auch in Englisch verfasst und betreut werden. Wir freuen uns über Studenten mit Interesse an Big Data. In Absprache sind auch weitere nicht genannte Themen möglich.

Open Theses

- Bachelor/Master Thesis (Leipzig): Parallel sampling and layouting of large graphs (BIGGR)
- Bachelor's Thesis (Leipzig): From Relational to Graph Data with Gradoop
- Master-Thesis (Dresden): Entwicklung eines interaktiven Explorationswerkzeuges für regionale Wasserhaushaltsuntersuchungen
- Master-Arbeit (Leipzig/Dresden) Visualisierung von regionenbasierten Hydrologiedaten
- Master's Thesis (Leipzig): Apache Beam and the price of unification
- Bachelor's or Master's Thesis (Leipzig): Dockerizing Big Data Software
- Bachelor's or Master's Thesis (Leipzig): Evaluation on the reproducibility of performance measurements on clusters in the cloud
- Bachelor Thesis (Leipzig): Development of a Big Data Business Models Framework
- Master Thesis (Leipzig): Using Time Series Analysis for automatized generation of phase recognition rules
- Bachelor's or Master's Thesis (Leipzig): Distributed Percentile Calculation for Floating Point Numbers in NoSQL Databases
- Bachelor's or Master's Thesis (Leipzig): Linking and Duplicate Detection in Big Graphs – blocking approaches in distributed linking
- Bachelor's or Master's Thesis (Leipzig): Massspectrometry data management and workflows
- Bachelor's or Master's Thesis (Leipzig): Integration of Graph-Analytics Operators within the KNIME-Analytics Platform
- Bachelor's or Master's Thesis (Leipzig): Efficient storage and retrieval of temporal graph data
- Bachelor's or Master's Thesis (Leipzig): Visualization-driven graph data reduction

