

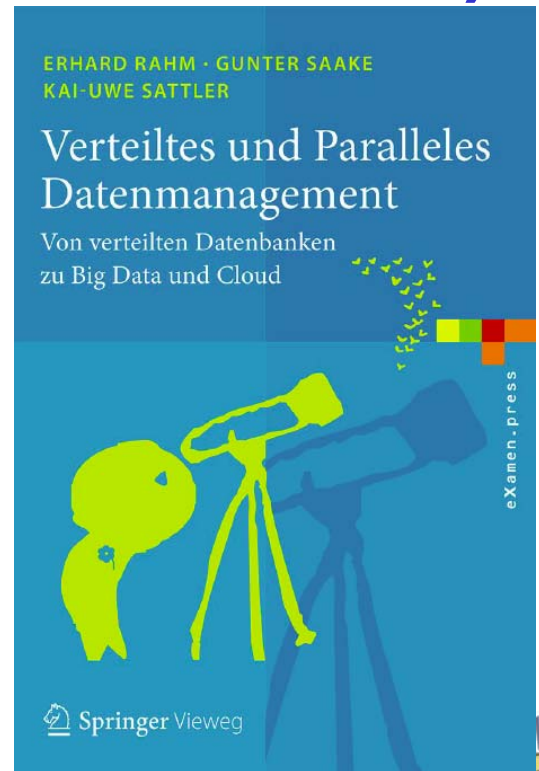
Mehrrechner-Datenbanksysteme (Verteilte und Parallele DBS)

Prof. Dr. E. Rahm

Wintersemester 2021/2022

Universität Leipzig
Institut für Informatik

<https://dbs.uni-leipzig.de/stud/2021ws/mrdbbs>



DBS-Module

■ 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
- Bachelorarbeit

■ Master-Studium Informatik

- **10-202-2215** – **Moderne Datenbanktechnologien (kleines Modul)**
- **10-202-2216** – **Moderne Datenbanktechnologien (großes Modul)**
- 10-202-2213 – Anwendungsbezogene Datenbankkonzepte (kleines Modul)
- 10-202-2011 – Masterseminar Informatik
- Masterarbeit

■ Master-Studium Data Science

- **10-INF-DS01** – **Skalierbare Datenbanktechnologien 1 (Pflichtmodul, 10 LP)**
- 10-INF-DS101 – Skalierbare Datenbanktechnologien 2 (5 LP)
- 10-INF-DS102 – Big Data Praktikum (5 LP)
- 10-INF-DS103 – Praktikum Data Warehousing und Data Mining (5 LP)
- 10-INF-DS301 – Aktuelle Trends in Data Science (5 LP)
- **10-INF-DS02** – **Masterseminar Data Science (5 LP)**
- Masterarbeit



Mapping Module – Lehrveranstaltungen WS21/22

- Master-Modul **Moderne Datenbanktechnologie** (10 LP)
 - **Mehrrechner-DBS**
 - Cloud und Big Data Management
 - Data Mining
- Bachelor-Modul *Realisierung von Informationssystemen* und Master-Modul *Moderne Datenbanktechnologie* (5 LP)
 - 2 Vorlesungen aus: **MRDBS**, Cloud/Big Data Management, Data Mining
- Data-Science-Modul *Skalierbare Datenbanktechnologien 1*
 - 2 Vorlesungen aus: **MRDBS**, Cloud/Big Data Management, Data Mining
 - Seminar: New Trends in Machine Learning and Data Analytics
- **Bachelorseminar / Masterseminar**
 - Vortrag über laufende Bachelor/Masterarbeit im Rahmen des DB-Oberseminars
- überzählige Vorlesung (z.B. MRDBS) kann auch in Modul im SS22 eingebracht werden

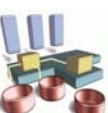


Bachelor Informatik mit DB-Profil

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	Diskrete Strukturen	Analysis	
2	Modellierung und Programmierung 2	Algorithmen u. Datenstrukturen 2	Java-Praktikum	Logik	Lineare Algebra	
3	Datenbanksysteme 1	Software-Technik	Softwaretechnik-Praktikum	Betriebs- und Kommunikationssysteme	Automaten und Sprachen	Wahrscheinlichkeitstheorie
4	Datenbanksysteme 2	Kernmodul 2	Techn. Inf. 2 / Hardware-Praktikum	Berechenbarkeit	DB-Praktikum	
5	Realisierung v IS	Seminarmodul	Vertiefungsmodul		Ergänzungsfach	
6	Kernmodul 4	Bachelorseminar	Bachelorarbeit		Schlüsselqualifikation	

Legende:

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



Masterstudium Data Science



- neuer Studiengang seit SS20
- Hauptinhalte
 - skalierbares Datenmanagement („Big Data“), mind. 20 LP
 - Datenanalyse / Machine Learning, mind. 20 LP
 - Ergänzungs-/Anwendungsmodule
 - viele Praktika möglich

1. Semester 2. Semester 3. Semester 4. Semester

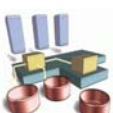
Skalierbare Datenbanktechnologien 1		Skalierbares Datenmanagement	Skalierbares Datenmanagement	Vertiefung	Mastersem. Data Science (5 LP)
Datenanalyse		Datenanalyse		Vertiefung	Masterarbeit (25 LP)
Ergänzung	Ergänzung	Ergänzung	Anwendung/ Ergänzung	Anwendung/ Ergänzung	



Beispielbelegungen Data Science

Wintersemester		Sommersemester		Wintersemester	4. Semester	
Skalierbare Datenbanktechnologien 1		SDBT2	Big Data Praktikum	Advanced Information Retrieval	Masterseminar (5 LP)	Masterarbeit (25 LP)
Künstl. neuronale Netze u. maschin. Lernen		Multivariate Statistik und Data Mining		Wissens- und Content Management		
IT-Sicherheit	Prakt. Data Wareh./ Mining	Textdatenbanken	Aktuelle Trends in DS	Verfahren und Anwendungen in den Digital Humanities		

Skalierbare Datenbanktechnologien 1		SDBT2	Big Data Praktikum	Text Mining	Masterseminar (5 LP)	Masterarbeit (25 LP)
Statistisches Lernen		Multivariate Statistik und Data Mining		Künstl. neuronale Netze u. maschin. Lernen		
Sequenzanalyse und Genomik		Grundl. komplexer Systeme	Aktuelle Trends in DS	Visualisierung		



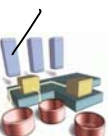
Leistungsbewertung

- Prüfungsklausur 60 Minuten
 - Feb. 2022
 - Präsenzklausur geplant
 - überprüft konzeptionelles Wissen + Anwendungsfälle
- Klausurerfolg durch
 - intensives Vorlesungsstudium
 - Online-Übungen (LOTS)
 - Literatur



Lernziele der Vorlesung MRDBS

- fortgeschrittene Kenntnisse der Funktionsweise von verteilten und parallelen Datenbank- und Datenmanagementsystemen
- Implementierungstechniken u.a. zu
 - Datenverteilung
 - Verteilte Anfrageverarbeitung
 - Verteilte Transaktionsverarbeitung
 - Replizierte Datenbanken
 - Blockchain-Systemen
- Verfahren relevant für verteilte relationale Datenbanken und NoSQL-Systeme



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

■ Block Chain und Verteilte Ledger-Systeme

■ *Shared-Disk-DBS*

Data Warehouses, Cloud & Big Data Management, NoSQL
-> eigene Vorlesungen



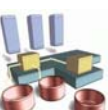
Lehrbuch



■ Rahm/Saake/Sattler:

Verteiltes und Paralleles Datenmanagement.
Springer 2015

- 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.
 - 4th edition, Springer-Verlag, 2020



Online-Übungen

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

LOTS Leipzig Online-Test-System

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

Home Registrierung Impressum

Login

Username:

Passwort:

[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)

0 von 2

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

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

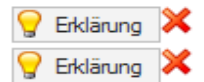


Aufgabe 4 (Hash-Join)

0 von 2

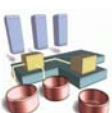
Markieren Sie die zutreffenden Aussagen.

- 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

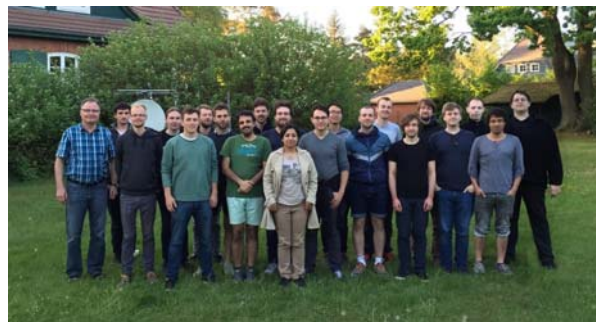
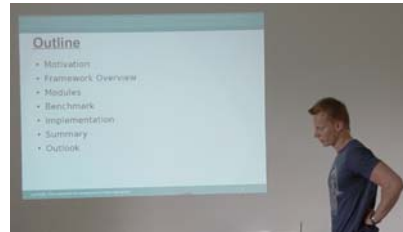


Lehrstuhl Datenbanksysteme

- 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



Inhalte

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

Forschung

[Publications \(2020, 2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, ...\)](#)

[FAst Multi-source Entity Resolution system \(FAMER\)](#)

[Benchmark datasets for entity resolution](#)

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

[Privacy-Preserving Record Linkage for Big Data](#)

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

[Entity Matching for Big Data \(Dedoop\), LOD Link Discovery](#)

[Semantic annotations: ELISA project, LHA Annotation Linking](#)

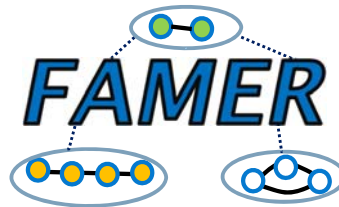
[Schema and Ontology Matching \(COMA++, GOMMA, STROMA\), SemRep repository, Ontology Merging \(ATOM\)](#)

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



Neue Publikationen

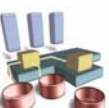
- [ERGAN: Generative Adversarial Networks](#)



Deutsche KI-Zentren

- KI-Strategie des Bundes beinhaltet Einrichtung von 5 Zentren für Künstliche Intelligenz (neben DFKI)

- Berlin (BIFOLD)
- Dortmund / Bonn (ML2R)
- **Dresden / Leipzig** (ScaDS.AI)
- München (MCML)
- Tübingen (tuebingen.ai)



SCADS.AI: Center for **Scalable Data Analytics** and **Artificial Intelligence**

- zunächst (2014-19) Big-Data-Zentrum ScaDS Dresden/Leipzig
- seit Nov. 2019: KI- bzw. Data-Science-Zentrum ScaDS.AI
- Ko-Finanzierung durch Bund und Land Sachsen
- Direktoren: Nagel (Dresden), Rahm (Leipzig)

■ Highlights ScaDS.AI

- 8 neue KI-Professuren, davon 4 an der Univ. Leipzig
- Forschung im Rahmen einer Graduiertenschule
- Demo and Living Lab

<https://scads.ai>



Grobstruktur

