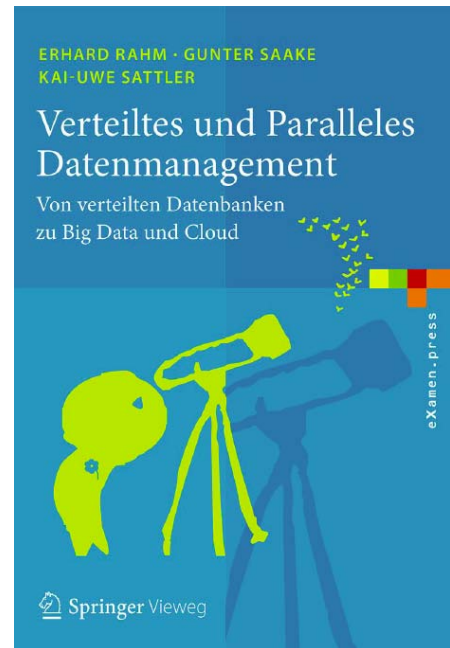


Mehrrechner-Datenbanksysteme (Verteilte und Parallele DBS)

Prof. Dr. E. Rahm

Sommersemester 2024

Universität Leipzig
Institut für Informatik



<https://dbs.uni-leipzig.de/study/semester/ss-2024/mrdbbs/>



DBS-Module Bachelor

- 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



DBS-Module Master

■ Master 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-DS104 – Data Preparation & Cleaning (5 LP)
- 10-INF-DS301 – Aktuelle Trends in Data Science (5 LP)
- 10-INF-DS02 – Masterseminar Data Science (5 LP)
 - Masterarbeit

■ Master Informatik

- 10-202-2215 – Moderne Datenbanktechnologien (5 LP)
- **10-202-2213 – Anwendungsbezogene Datenbankkonzepte (5 LP)**
- 10-202-2011 – Masterseminar Informatik
 - Masterarbeit



Mapping Module – Lehrveranstaltungen SS24

- Bachelor-Modul *Realisierung von Informationssystemen*, Master-Module *Skalierbare Datenbanktechnologien 2* und *Anwendungsbezogene Datenbankkonzepte (5 LP)*:

- **MRDBS**
- NoSQL

- **Aktuelle Trends in Data Science**

- SS24: Bio-Image Data Science (Dr. Robert Haase)

- **Bachelorseminar / Masterseminar**

- Vortrag über laufende Bachelor/Masterarbeit im Rahmen des DB-Oberseminars (21.-24. Mai im Erzgebirge)



Masterstudium Data Science



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

Schwerpunkt Skalierbares Datenmanagement

1. Semester

2. Semester

3. Semester

4. Semester

Skalierbare Datenbanktechnologien 1		SDBT2	Big Data Praktikum	Prakt. Data Wareh./ Mining	Medical Data Science	Masterseminar (5 LP)
Künstl. neuronale Netze u. maschin. Lernen		Multivariate Statistik und Data Mining		Mathem. Foundations of DS	Advanced Statistics	Masterarbeit (25 LP)
IT-Sicherheit	Prakt. IT-Sicherheit	Data Prep. & Cleaning	Aktuelle Trends in DS	Software Engineering für KI-Systeme		

Schwerpunkt Datenanalyse

Skalierbare Datenbanktechnologien 1		SDBT2	Big Data Praktikum	Mathem. Foundations of DS	Advanced Statistics	Masterseminar (5 LP)
Künstl. neuronale Netze u. maschin. Lernen		Multivariate Statistik und Data Mining		Mediz. Bildverarbeitung u. Bildaufnahme		Masterarbeit (25 LP)
Sequenzanalyse und Genomik		Probabilistic ML	Aktuelle Trends in DS	Wiss. Visualisierung		



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)



Forschungsthemen



Crosscutting Topics

Research Areas

Applied AI & Big Data

AI Algorithms & Methods

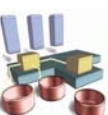
Big Data Analytics & Engineering

Topic Areas

- Life Science & Medicine
- Environment & Earth Sciences
- Software Engineering
- Physics / Chemistry
- Engineering / Business
- Understanding Language
- Methods and Hardware for Neuro-Inspired Computing
- Graph-based Artificial Intelligence
- Knowledge Representation & Engineering
- Scalable Visual Computing
- Federated, Efficient Learning
- Math Foundations & Statistical Learning
- Big Data Analytics
- Open Data & Open Models
- Data Quality & Data Integration

Responsible AI: Ethical and Societal Dimensions

Architectures / Scalability / Security



Leistungsbewertung

- Prüfungsklausur 60 Minuten
 - Juli 2024
 - 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

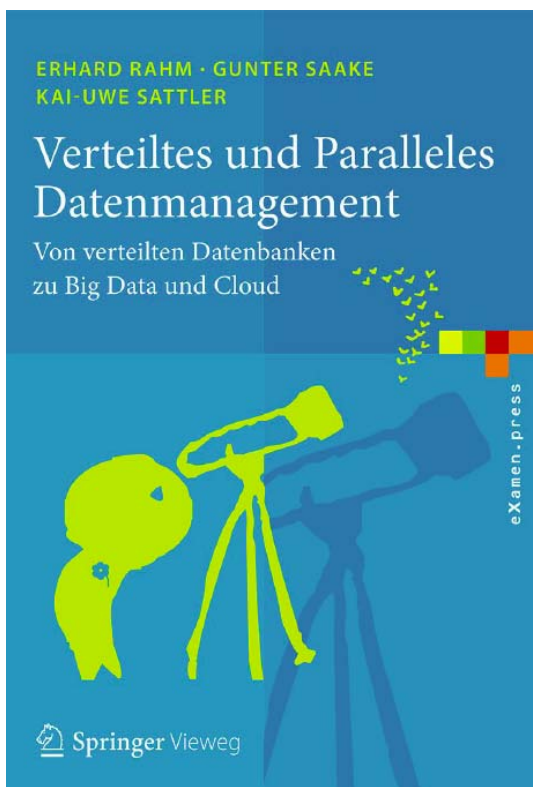
■ Block Chain und Verteilte Ledger-Systeme

■ *Shared-Disk-DBS*

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



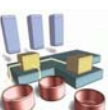
Lehrbuch



■ 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 Hauptspeichergroße 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



Erklärung

Erklärung

Erklärung

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

Erklärung

Erklärung

