

Implementierung von Datenbanksystemen 1 (IDBS1)



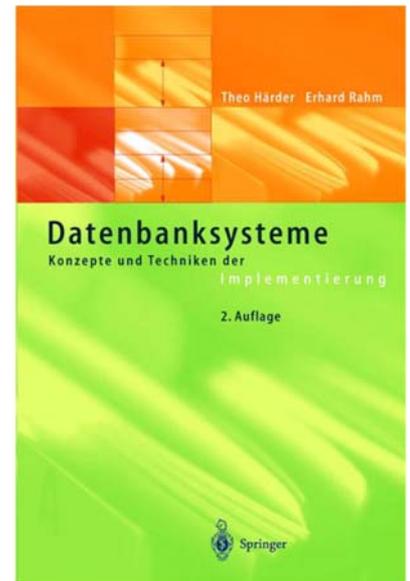
Wintersemester 2022/23

Prof. Dr. Erhard Rahm

Universität Leipzig

Institut für Informatik

<http://dbs.uni-leipzig.de/stud/2022ws/idbs1>



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-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

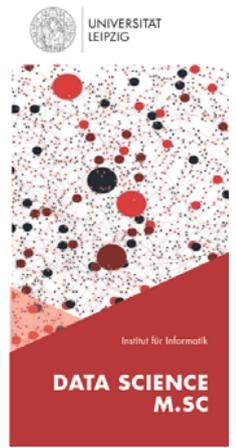
DBS-Module Bachelor

- 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

Mapping Module – Lehrveranstaltungen WS22/23

- Bachelor-Modul *Realisierung von Informationssystemen* und Master-Modul *Moderne Datenbanktechnologie (5 LP)*
 - 2 Vorlesungen aus: **IDBS1**, Cloud/Big Data Management, Data Mining
- Data-Science-Modul *Skalierbare Datenbanktechnologien (SDBT) 1*
 - 2 Vorlesungen aus: **IDBS1**, 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

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

Schwerpunkt Big Data (Skalierbares Datenmgmt)

Wintersemester Sommersemester Wintersemester 4. Semester

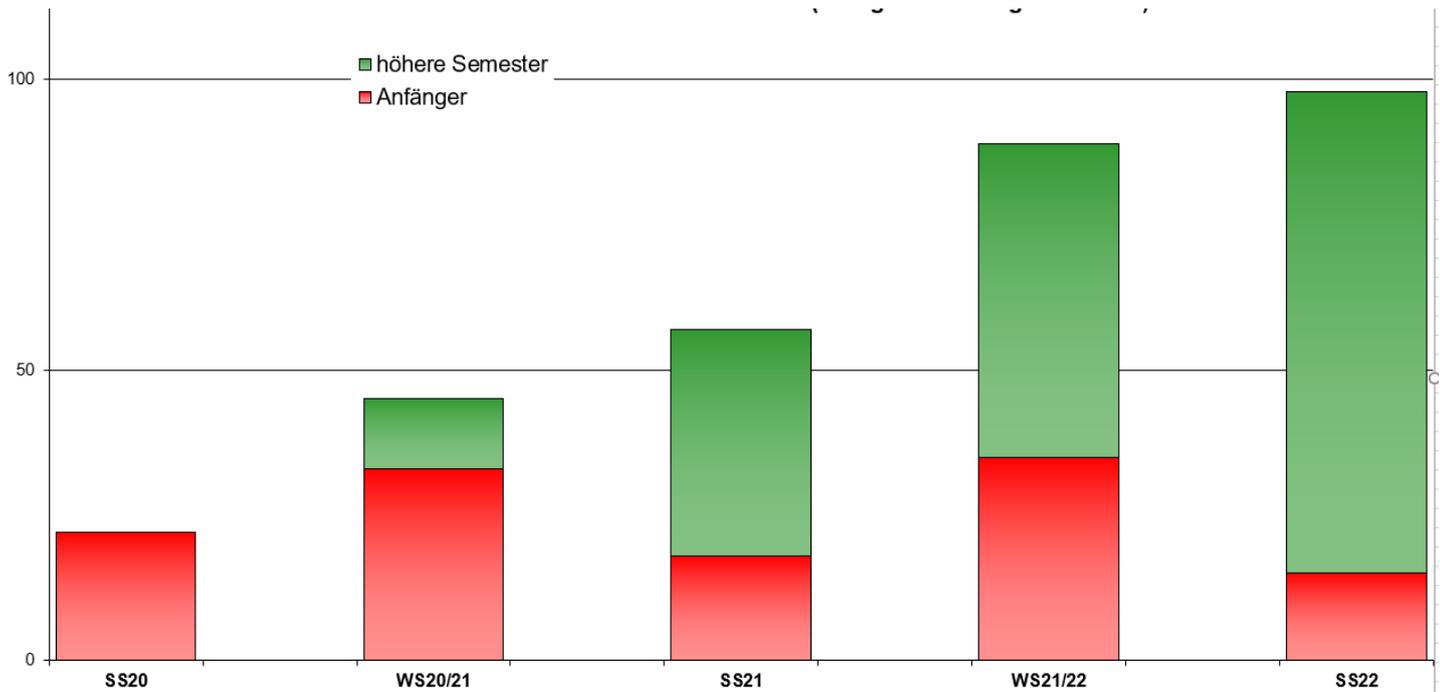
Skalierbare Datenbanktechnologien 1		SDBT2	Big Data Praktikum	Advanced Information Retrieval	Masterseminar (5 LP)
Künstl. neuronale Netze u. maschin. Lernen		Multivariate Statistik und Data Mining		Cyber- und IT-Sicherheit	Masterarbeit (25 LP)
IT-Sicherheit	Prakt. Data Wareh./ Mining	Datenschutz von A bis Z	Aktuelle Trends in DS	Verfahren und Anwendungen in den Digital Humanities	

Schwerpunkt Datenanalyse

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



#STUDIERENDE MASTER DATA SCIENCE



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



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
	Cloud und Big Data Management	Vertiefung	2	WS
	NoSQL-DB	Vertiefung	2	SS
	Data Mining	Vertiefung	2	WS

Name	Typ	Sem.
DB-Praktikum	Praktikum	SS
Praktikum Data Warehousing & Data Mining	Praktikum	WS
Big Data Praktikum	Praktikum	SS
Data-Science-Seminar	Seminar	WS
Bachelor-seminar	Seminar	SS+WS
Masterseminar	Seminar	SS+WS

LV im WS2020/21

Leistungsbewertung

- Prüfungsklausur 60 Minuten (Präsenzmodus geplant)
 - Feb. 2023
 - überprüft konzeptionelles Wissen + Anwendungsfälle
- Klausurerfolg durch
 - Vorlesungsteilnahme und –nachbearbeitung
 - Online-Übungen
 - Literatur

Lernziele der Vorlesung IDBS

- fundierte Kenntnisse der Funktionsweise von DBS und zu *Big Data Engineering*
- Implementierungstechniken u.a. zur
 - Sicherstellung einer hohen Performanz/Skalierbarkeit auf großen Datenmengen
 - Datensicherheit
- **IDBS1**: Verfahren zur Externspeicher-Nutzung, Verwaltung von Pufferspeichern, Indexstrukturen, Anfrageoptimierung ...
- **IDBS2**: Verfahren zur Transaktionsverwaltung: Synchronisation (Concurrency Control), Logging/Archivierung, Recovery
- tiefergehende Kenntnisse wichtig für DB-Administration sowie generell für anspruchsvolle DB-Nutzung

Vorlesungsübersicht IDBS1

1. Einführung

- Architektur von DBS: Schichtenmodelle
- Tuning von DBS

2. E/A-Architekturen und Speicherhierarchien

- Speichertechnologien
- Disk-Arrays
- nicht-flüchtige Halbleiterspeicher

3. DBS-Pufferverwaltung

- Ersetzungsverfahren

4. Satzverwaltung

- Zuordnung Sätze – Seiten, Clusterung, BLOBs
- Satzadressierung
- Column Stores

Vorlesungsübersicht (2)

5. Indexstrukturen

- Mehrweg-Bäume
- externe Hash-Verfahren
- mehrdimensionale Zugriffspfade
- Bitlisten-Indizes
- Text-Indexierung

6. Algorithmen zur Implementierung relationaler Operationen

- Selektion
- Verbund / Join
- Sortierung

7. Anfrageoptimierung

- Übersetzung von DB-Anweisungen
- Optimierung
- EXPLAIN

Literatur

- Härder, T., Rahm, E.: *Datenbanksysteme - Konzepte und Techniken der Implementierung*. Springer-Verlag, 2. Auflage 2001 (Kap. 1 und 13 online)

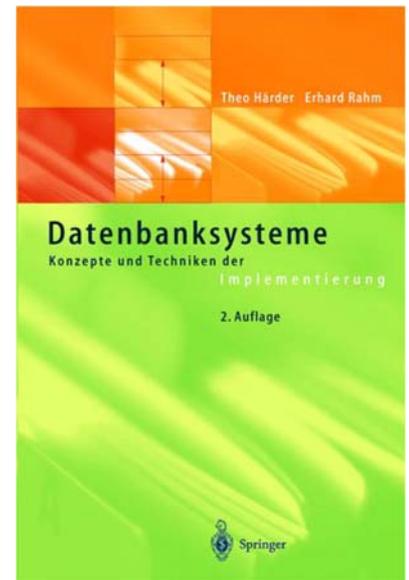
<http://dbs.uni-leipzig.de/buecher/DBSI-Buch/inhalt.html>

- Weiteres Lehrbuch

- Saake, Sattler, Heuer: *Datenbanken: Implementierungstechniken*, MITP-Verlag, 4. Auflage 2019

- Google Scholar

- <http://scholar.google.com/> (Volltexte von Publikationen, Zitierungsangaben)



Online-Übungen

- LOTS (Leipzig Online Test System), <http://lots.uni-leipzig.de>
Kennung: idbs1#ws22



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



 Erklärung

 Erklärung

 Erklärung

Aufgabe 4 (Hash-Join)

Markieren Sie die zutreffenden Aussagen.

0 von 2

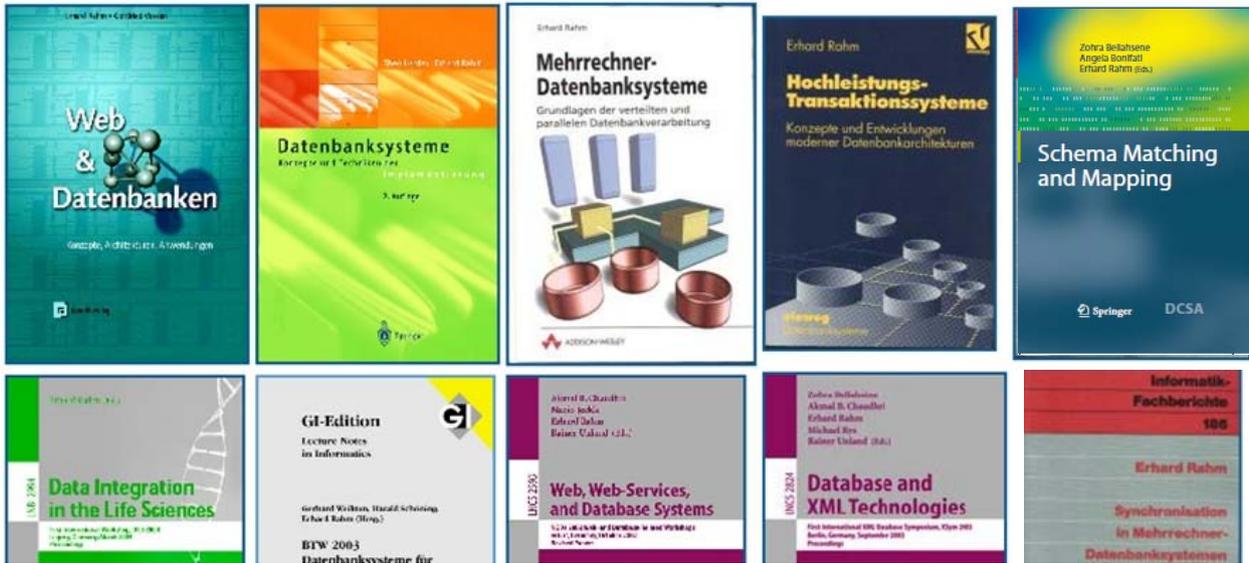
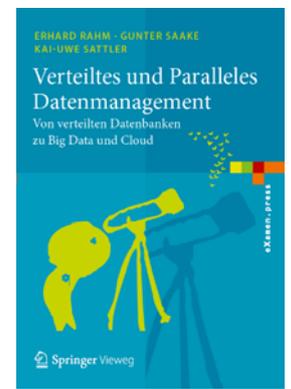
- Durch die Nutzung von Bitvektoren lässt sich der Umfang der Hash-Tabellen kleiner halten
- Hash-Joins sollten nur eingesetzt werden, wenn die kleinere Eingabetabelle im Hauptspeicher gespeichert werden kann
- Hash-Joins können effektiv zur Realisierung von Mehr-Wege-Joins genutzt werden
- Hash-Join-Verfahren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

 Erklärung

 Erklärung

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



Forschung

Inhalte

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

Forschung

Publications (2022,2021,2020, 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

Data science Center ScaDS.AI 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 [🔗](#))

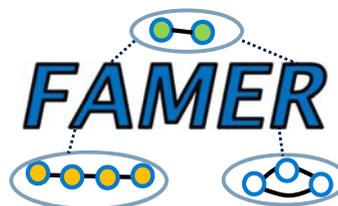
Current prototypes and implementations

Previous projects



Neue Publikationen

- D-TOUR : Detour-based point of interest detection in privacy-sensitive trajectories
- Enhancing Cross-lingual Biomedical Concept Normalization Using



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

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>

Research Areas



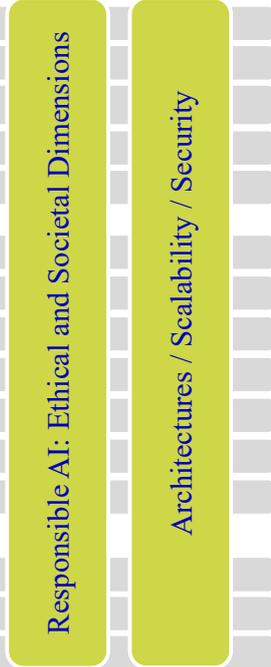
Topic Areas

- 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

Crosscutting Topics



Living Lab: KI erleben

We

- *Demonstrate*
- *Co-innovate*
- *Organize Events*
- *Build a Meeting point* for AI research and transfer



- GI ist **gemeinnützige Fachgesellschaft** mit rund 20.000 Mitgliedern an Hochschulen, in Unternehmen, in der Verwaltung, an Schulen und in der Freiberuflichkeit.
- Die GI bietet ein **fachlich einmaliges, persönliches Netzwerk** für den Wissens- und Erfahrungsaustausch mit anderen Studierenden, Doktoranden und Professorinnen und Professoren.
- Wir als Informatikerinnen und Informatiker sind die einzigen, die die Technik richtig kennen und deshalb Verantwortung haben. Diese Verantwortung muss gebündelt und richtig adressiert werden. Das macht die GI.

Wie Studierende die GI nutzen

- Spezielle Nachwuchstagungen, DevCamps, Wettbewerbe, SKILL
- Publikationen selbst veröffentlichen (z.B. in der SKILL)
- viele Publikationen frei lesen - <https://dl.gi.de/>
- Begegnungen in lockerer Atmosphäre mit Personen, die einem im Studium weiterhelfen
- Mitarbeit in den Hochschulgruppen und weiteren GI-Gremien
- Vergünstigte Tickets der Deutschen Bahn zu GI-Veranstaltungen
- **Beitragsfrei für Studierende!**

<https://gi.de/mitgliedschaft/mitglied-werden/studierende-auszubildende>