

Implementierung von Datenbanksystemen 1 (IDBS1)



Wintersemester 2020/21

Prof. Dr. Erhard Rahm

Universität Leipzig

Institut für Informatik

<http://dbs.uni-leipzig.de/stud/2020ws/idbs1>

DBS-Module Master

■ Master Data Science (neu)

- 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)

■ Master 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)
 - Seminarmodul
- 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
- Seminarmodul
- 10-201-2010 – Bachelorseminar Informatik
- Bachelorarbeit

Mapping Module – Lehrveranstaltungen WS20/21

- Master-Modul **Moderne Datenbanktechnologie** (10 LP)
 - **IDBS1**
 - Cloud und Big Data Management
 - Data Mining
- 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
- überzählige Vorlesung (z.B. IDBS1) kann auch in Modul im SS21 eingebracht werden

Masterstudium Data Science

- neuer Studiengang seit SS20
- Hauptinhalte
 - Skalierbares Datenmanagement („Big Data“), mind. 20 LP
 - Datenanalyse / Machine Learning, mind. 20 LP (davon mind. 10 LP Statistik)
 - Ergänzungs-/Anwendungsmodule
 - Anwendungsorientierung (Praktika, z.B. Bioinformatik/Digital Humanities)

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	Fortgeschr. Methoden Information Retrieval	Masterseminar (5 LP)
Künstl. neuronale Netze u. maschin. Lernen		Multivariate Statistik und Data Mining		Wissens- und Content Management	Masterarbeit (25 LP)
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)
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	



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



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. 2021
 - 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>

- Weitere Lehrbücher

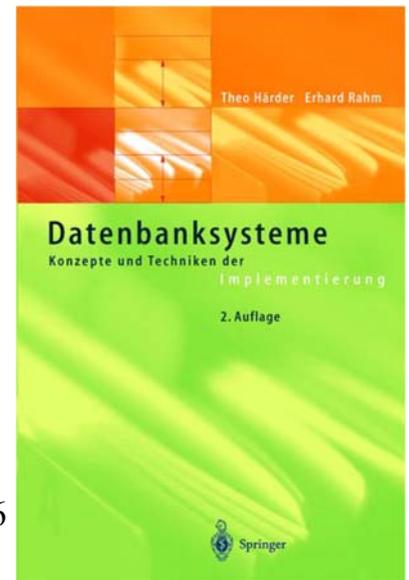
- Saake, Sattler, Heuer: *Datenbanken: Implementierungstechniken*, MITP-Verlag, 4. Auflage 2019
- Dittrich, J.: *Patterns in Data Management: A Flipped Textbook*. 2016
Website: Datenbankenlernen.de
- Garcia-Molina, H., Ullman, J.D., Widom, J.: *Database System Implementation*. Prentice Hall, 2000.

Teil von *Database System – The Complete Book*, 2nd ed. 2009, 1240 Seiten,

<https://people.inf.elte.hu/miqaai/elektroModulatorDva.pdf> (IDBS1 entspricht v.a. Kap. 13-16)

- Google Scholar

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



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

-
- Erklärung ✗
- Erklärung ✗
- Erklärung ✗

Aufgabe 4 (Hash-Join)

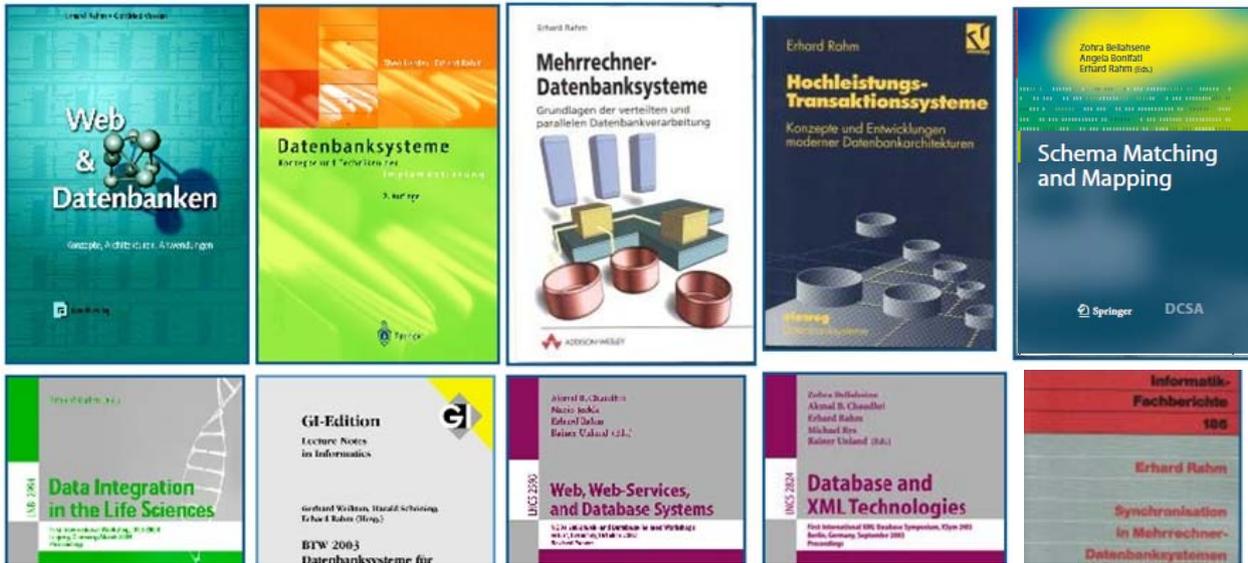
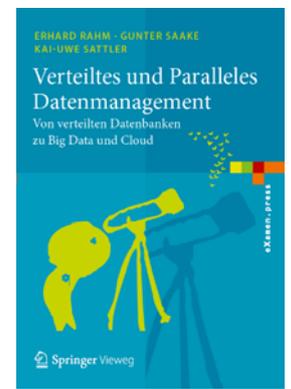
Markieren Sie die zutreffenden Aussagen.

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 Erklärung ✗
- Hash-Join-Verfahren eignen sich zur Beantwortung von Equi-Join-Anfragen Erklärung ✗

Lehrstuhl Datenbanken

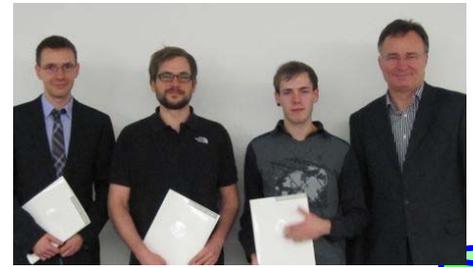
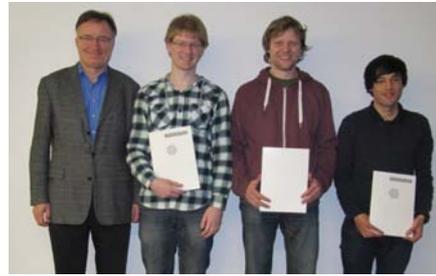
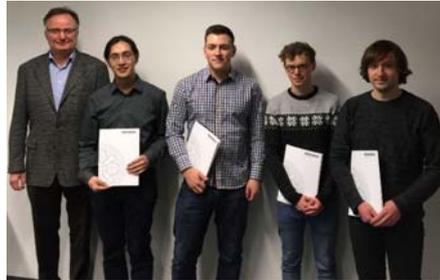
- seit 1994 am Institut für Informatik
- umfangreiches Lehrangebot
 - Vorlesungen, Praktika, Seminare
 - Online-Übungssystem LOTS
 - Eigene Lehrbücher: MRDBS, IDBS, ...



Oberseminare an Uni-Außenstelle Zingst – seit 2001



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



WS20/21, © Prof. Dr. E. Rahm

IDBS1

Forschung

🇩🇪 🇬🇧
Startseite login

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](#)

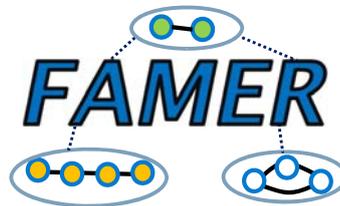
[Semantic annotations: ELISA project, LHA Annotation Linking](#)

[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](#)



WS20/21, © Prof. Dr. E. Rahm

IDBS1

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

- erweitert das seit 2014 etablierte Big-Data-Zentrum ScaDS Dresden/Leipzig
- seit Nov. 2019: KI- bzw. Data-Science-Zentrum ScaDS.AI
- Ko-Finanzierung durch Bund und Land Sachsen
- Direktoren: Nagle (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