

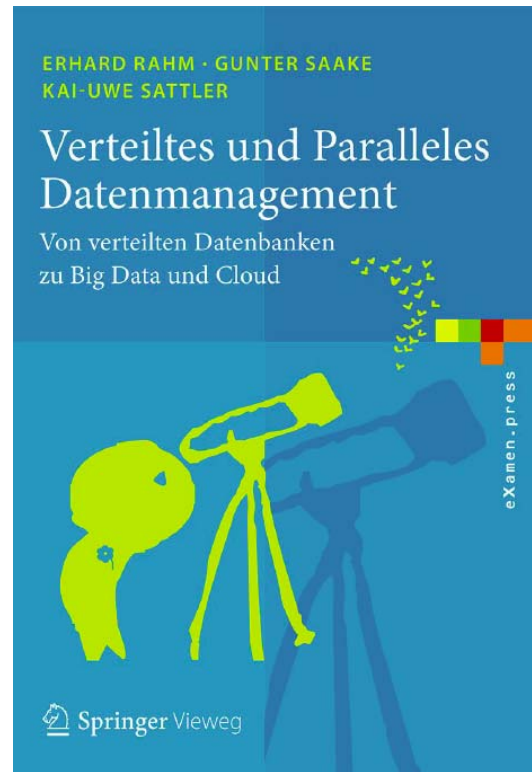
# Mehrrechner-Datenbanksysteme (Verteilte und Parallele DBS)

**Prof. Dr. E. Rahm**

Wintersemester 2019/2020

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

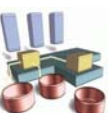
Name	Typ	Sem.
DB-Praktikum	Praktikum	SS
Praktikum Data Warehousing & Data Mining	Praktikum	WS
Big Data Praktikum	Praktikum	SS
Seminarmodul	Seminar	WS
Bachelorseminar	Seminar	SS+WS
Masterseminar	Seminar	SS+WS

LV im WS2019/20



## Mapping Module – Lehrveranstaltungen WS19/20

- Bachelor-Modul **Realisierung von Informationssystemen** und kleines Master-Modul **Moderne Datenbanktechnologie** (5 LP)
  - **Mehrrechner-DBS**
  - Data Mining
- zusätzlich bei großem Master-Module **Moderne DBT** (10 LP)
  - Seminar *Trends in Machine Learning and Data Analysis* (max. 20 TN) oder
  - Praktikum *Data Warehousing und Data Mining* (max. 20 TN)
- **Seminarmodul**
  - Forschungsseminar *Trends in Machine Learning and Data Analysis*
- **Bachelorseminar / Masterseminar**
  - Vortrag über laufende Bachelor/Masterarbeit



# 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	<b>Realisierung v IS</b>	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



# 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		

## ■ noch mögliche Schwerpunkte (Ausweisung im Zeugnis)

- Big Data
- Medizinische Informatik

## ■ individuelle Schwerpunktbildung, zB. Datenmanagement/DB



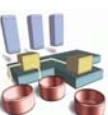
# Master Informatik mit DB-Profil

Sem	5 LP	5LP	10 LP	10 LP
1.	Kernmodul I	Kernmodul II	<b>Moderne DB-Technologien</b>	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	<b>Moderne DB-Technologien</b>	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
- Ergänzungsfächer z.B. aus Bioinformatik oder Digital Humanities



## Master Data Science (ab 2020)

- Hauptinhalte
  - Skalierbares Datenmanagement („Big Data“), mind. 20 LP
  - Datenanalyse / Machine Learning, mind. 20 LP darunter mind. 10 LP Statistik-orientiert
  - Ergänzungs-/Anwendungsmodule

1. Semester

2. Semester

3. Semester

4. Semester

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



# Master Data Science (2)

## ■ DBS-Module im Masterstudiengang Data Science

Master Data Science	Entsprechung Master Informatik
Skalierbare Datenbanktechnologien 1 (10 LP)	Moderne Datenbanktechnologien (10 LP)
Skalierbare Datenbanktechnologien 2 (5 LP)	Anwendungsspezifische DB-Konzepte (5 LP)
Big Data Praktikum (5 LP)	Teil von Anw. DBK (10 LP)
Praktikum Data Warehousing und Data Mining (5 LP)	Teil von MDBT (10 LP)
Aktuelle Trends in Data Science (Ringvorlesung)	



## Beispielbelegungen Data Science

1. Semester		2. Semester		3. Semester		4. Semester	
Skalierbare Datenbanktechnologien 1		SDBT2	Big Data Praktikum	Fortgeschr. Methoden 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		



# 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

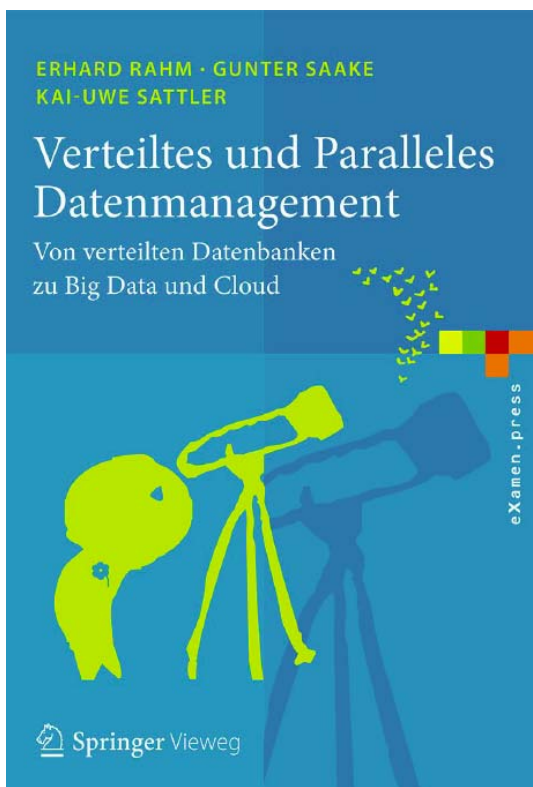
## ■ Shared-Disk-DBS

## ■ Block Chain und Verteilte Ledger-Systeme

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



## Lehrbuch



## ■ Rahm/Saake/Sattler:

*Verteiltes und Paralleles Datenmanagement.*  
Springer 2015

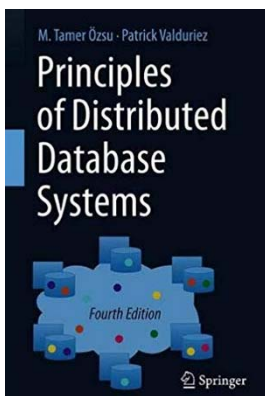
- E-Book online zugänglich im Uni-Netz
- mit Übungsaufgaben



# 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.
  - 4<sup>th</sup> edition, Springer-Verlag, Nov. 2019 (to appear)



# 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:

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)

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

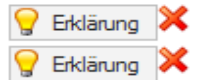


## Aufgabe 4 (Hash-Join)

0 von 2

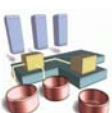
Markieren Sie die zutreffenden Aussagen.

- 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

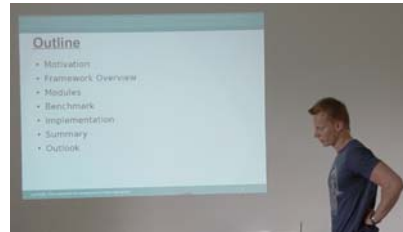


## Lehrstuhl Datenbanksysteme

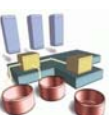
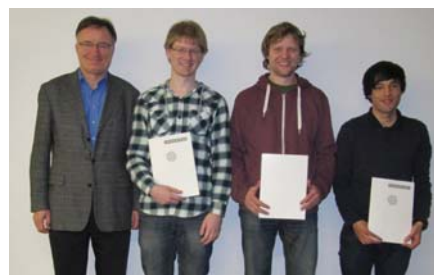
- seit 1994 am Institut für Informatik
- umfangreiches Lehrangebot
  - Vorlesungen, Praktika, Seminare
  - Online-Übungssystem LOTS
  - Eigene Lehrbücher: MRDBS, IDBS, ...



# Oberseminare an Uni-Außenstelle Zingst – seit 2001



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



# Forschung

## Inhalte

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

## Forschung

Publications (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

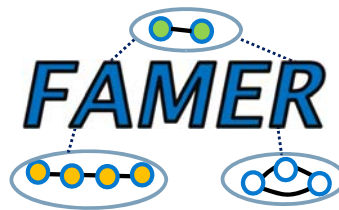
Schema and Ontology Matching (COMA++, GOMMA, STROMA), SemRep repository, Ontology Merging (ATOM)

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



## Neue Publikationen

- DBpedia FlexiFusion The Best of Wikipedia > Wikidata > Your



WS19/20, © Prof. Dr. E. Rahm

0 - 21



BIG DATA KOMPETENZZENTRUM

UNIVERSITÄT  
LEIPZIG

zwei deutsche Kompetenzzentren für Big Data seit Okt. 2014  
(BMBF-Wettbewerb)

- ScaDS Dresden/Leipzig
- Berlin Big Data Center (BBDC)

ScaDS Dresden/Leipzig (Competence Center for  
Scalable Data Services and Solutions Dresden/Leipzig)

- wissenschaftliche Koordinatoren: Nagel (TUD), Rahm (UL)



### Application Areas

Life Science & E-Health

Business Applications

Environmental Sciences

Matter / Energy / Chemistry

Material Sciences/Engineering

Digital Humanities

Service  
Center

### Visual Analytics

Scalable Visual Analytics

Immersive Visual Interaction

### Big Data Integration & Analytics

Big Data Integration

Data Analytics

### Scalable and Secure Data Platforms

Scalable Architectures

Hardware-based Data Security

23

- **SCADS.AI:** Center for **Scalable Data Analytics** and **Artificial Intelligence**
- zunächst Ausbau von ScaDS (ab Nov. 2019), anschließend Verstetigung
- Ko-Finanzierung durch Bund und Land Sachsen
- Förderung im Rahmen der KI-Strategie der Bundesregierung
- **Highlights ScaDS.AI**
  - 8 neue KI-Professuren
  - Forschung im Rahmen einer Graduiertenschule
  - Demo and Living Lab

24

