

Implementierung von Datenbanksystemen 2 (IDBS2)

Sommersemester 2023

Prof. Dr. Erhard Rahm

Universität Leipzig

Institut für Informatik

<https://dbs.uni-leipzig.de/stud/2023ss/idbs2>



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, war Data Wrangling)
- 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 (kleines Modul)
- **10-202-2213 – Anwendungsbezogene Datenbankkonzepte (5 LP)**
- 10-202-2011 – Masterseminar Informatik
 - Masterarbeit

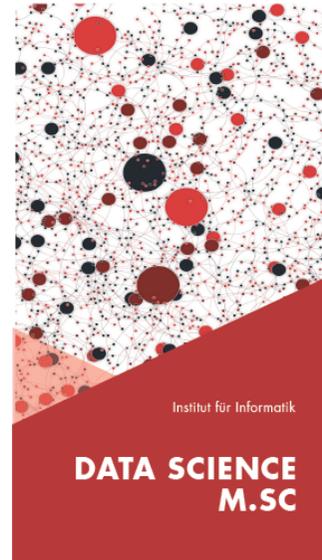


Mapping Module – Lehrveranstaltungen SS23

- Bachelor-Modul *Realisierung von Informationssystemen*,
Master-Module *Skalierbare Datenbanktechnologien 2* und
Moderne Datenbanktechnologie (5 LP):
 - IDBS2
 - NoSQL
- Aktuelle Trends in Data Science
 - SS23: Optimization for Machine Learning
 - beschränkt auf Master Data Science
- Bachelorseminar / Masterseminar
 - Vortrag über laufende Bachelor/Masterarbeit im Rahmen des DB-Oberseminars (23.-26. Mai im Harz)



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 Daten- management	Skalierbares Daten- management	Vertiefung	Mastersem. Data Science (5 LP)
Datenanalyse		Datenanalyse		Vertiefung	Masterarbeit (25 LP)
Ergänzung	Ergänzung	Ergänzung	Anwendung/ Ergänzung	Anwendung/ Ergänzung	



BEISPIELBELEGUNGEN DATA SCIENCE

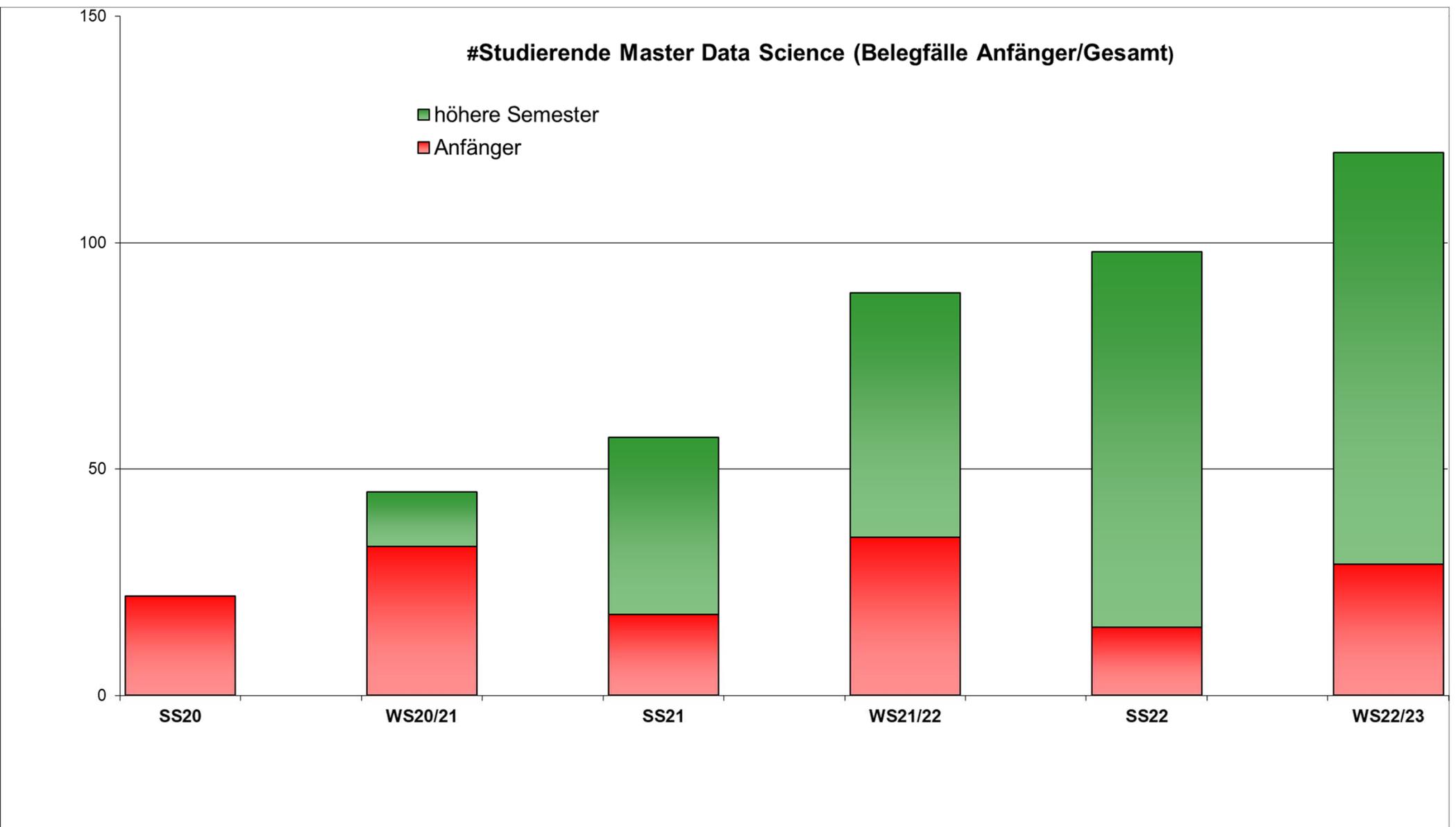
Schwerpunkt Big Data (Skalierbares Datenmgmt)

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)	Masterarbeit (25 LP)
Künstl. neuronale Netze u. maschin. Lernen		Multivariate Statistik und Data Mining		Mathem. Foundations of DS	Advanced NLP		
IT-Sicherheit	Prakt. IT-Sicherheit	Data Prep.& Cleaning	Aktuelle Trends in DS	Verfahren und Anwendungen in den Digital Humanities			

Schwerpunkt Datenanalyse

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

#STUDIERENDE MASTER DATA SCIENCE



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)



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

Crosscutting Topics

Responsible AI: Ethical and Societal Dimensions

Architectures / Scalability / Security

Leistungsbewertung

- Prüfungsklausur 60 Minuten
 - Präsenzklausur vstl. Im Juli 2023
 - überprüft konzeptionelles Wissen + Anwendungsfälle

- Klausurerfolg durch
 - intensives Vorlesungsstudium (Live/Video, Skript)
 - Online-Übungen (LOTS)
 - 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



Vorläufige Vorlesungsübersicht

1. Einführung: Transaktionsverwaltung, Integritätskontrolle

2. Synchronisation: Grundlagen, Sperrverfahren

- Mehrbenutzer-Anomalien
- Serialisierbarkeit
- Sperrverfahren: 2PL, Hierarchische Sperrverfahren
- Konsistenzstufen
- Deadlock-Behandlung

3. Synchronisation: Weitere Verfahren, Leistungsbewertung

- optimistische Verfahren
- Zeitstempel- und Mehrversionen-Verfahren
- Spezialverfahren für B*-Bäume und „High Traffic“-Elemente
- Leistungsanalyse und Lastkontrolle



Vorlesungsübersicht (2)

4. Logging und Recovery: Grundlagen

- Begriffe und Annahmen, Fehlermodell
- Logging-Verfahren
- Klassifikation von Recovery-Strategien

5. Crash- und Medien-Recovery

- Crash-Recovery
- Platten-Recovery

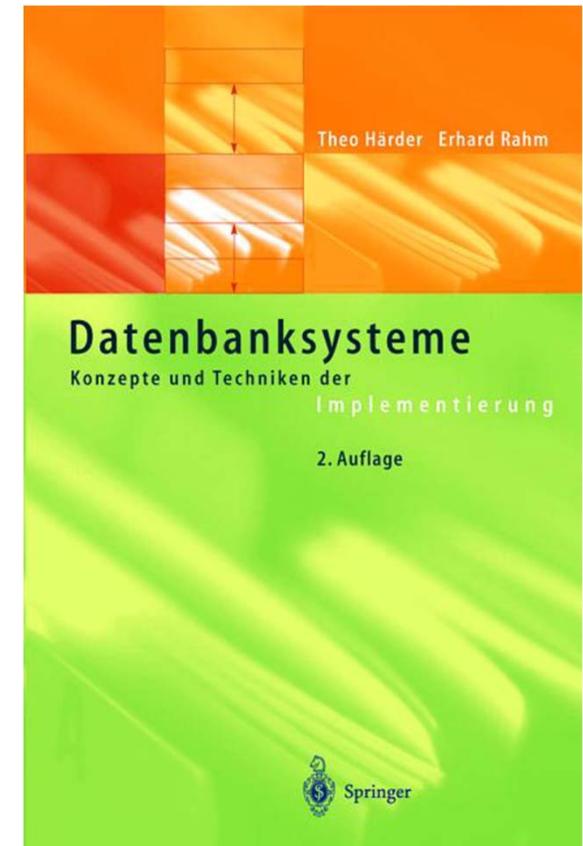
6. Transaktionskonzept: Weiterentwicklungen

- Geschachtelte Transaktionen
- Transaktionsketten (Sagas)



Literatur

- Härder, T., Rahm, E.: *Datenbanksysteme - Konzepte und Techniken der Implementierung*. Springer-Verlag, 2. Auflage 2001 (Kap. 1 und 13 online)
- weitere Lehrbücher
 - Saake, G., Heuer, A.: *Datenbanken: Implementierungstechniken*, MITP-Verlag, 2019
 - Garcia-Molina, H., Ullman, J.D., Widom, J.: *Database System Implementation*. Prentice Hall, 2000
 -
- Google Scholar
 - <http://scholar.google.com/> (Volltexte von Publikationen, Zitierungsangaben)



Online-Übungen

- **LOTS** (Leipzig Online Test System), <http://lots.uni-leipzig.de>
 - Kennung: [idbs#ss23](#)

The screenshot shows the homepage of the Leipzig Online-Test-System (LOTS). The header features the LOTS logo (a blue checkmark over the letters 'OTS') and the text 'Leipzig Online-Test-System'. To the right, it identifies the 'UNIVERSITÄT LEIPZIG' and the 'Institut für Informatik Abteilung Datenbanken'. A navigation bar includes 'Home', 'Registrierung', and 'Impressum'. The main content area is divided into three sections: 'Login', 'Gast Login', and 'System Info'. The 'Login' section has input fields for 'Username:' and 'Passwort:', a 'login' button, and a link for 'Passwort vergessen?'. The 'Gast Login' section explains that users can log in as guests for 30 minutes and provides a link to 'registrieren'. It includes a 'Viel Spass!' message, 'Ihr LOTS Team', and an 'als Gast einloggen' button. The 'System Info' section shows '# Benutzer: 31' and a clock icon with the date '05.02.2007' and time '12:15:21'. A 'News' section at the bottom prompts users to log in for personalized news.

