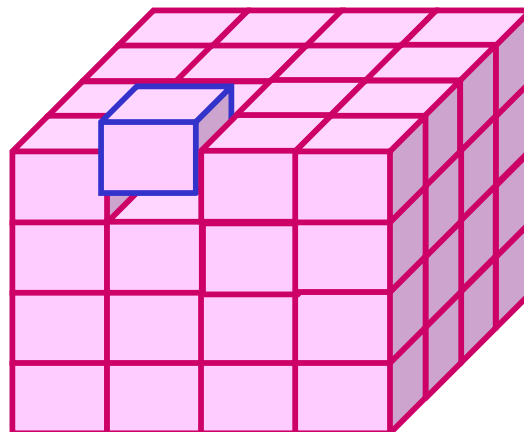


Data Warehousing

Sommersemester 2022

Prof. Dr. E. Rahm

Universität Leipzig
Institut für Informatik



<http://dbs.uni-leipzig.de/stud/2022ss/dwh>



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)

■ Master Informatik

- 10-202-2215 – Moderne Datenbanktechnologien (kleines Modul)
- 10-202-2216 – *Moderne Datenbanktechnologien (großes Modul)*
- 10-202-2213 – Anwendungsbezogene Datenbankkonzepte (kleines 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

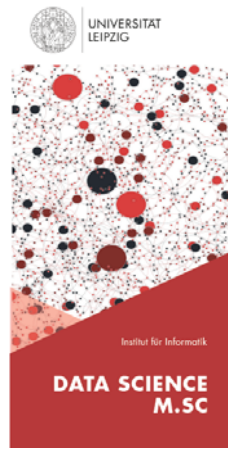


Mapping: Module – Lehrveranstaltungen SS22

- Master-Module SDBT2, Anwendungsbez. DBK (5 LP),
Bachelor-Modul Realisierung von IS
 - Data Warehousing
 - NoSQL-Datenbanken
- Bachelorseminar / Masterseminar
 - Vortrag über laufende Bachelor/Masterarbeit
 - extern an Uni-Außenstelle in Zingst (14.-18. Juni)



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

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	



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 und Big Data Management	Vertiefung	2	SS
	NoSQL-DB	Vertiefung	2	SS
	Data Mining	Vertiefung	2	WS
	Data Wrangling	Vertiefung	2+1	SS

Name	Typ	Sem.
DB-Praktikum	Praktikum	SS
Data-Warehouse-Praktikum	Praktikum	WS
Big Data Praktikum	Praktikum	SS
Problemseminar	Seminar	WS
Bachelorseminar	Seminar	SS+WS
Masterseminar	Seminar	SS+WS

LV im SS2022



Leistungsbewertung

- Prüfungsklausur von 60 Minuten
 - voraussichtl. Juli
 - überprüft konzeptionelles Wissen + Anwendungsfälle
- Klausurerfolg durch
 - Beherrschen der Vorlesungsinhalte
 - Nutzung der Online-Übungen
 - vertiefendes Literaturstudium



Online-Übungen

- **LOTS (Leipzig Online Test System)**, <http://lots.uni-leipzig.de>
 - Kennung zum Beitritt in Übungsgruppe s. DW-Webseite

The screenshot shows the homepage of the Leipzig Online-Test-System (LOTS). The header features the LOTS logo, the text 'Leipzig Online-Test-System', and the affiliation 'UNIVERSITÄT LEIPZIG Fakultät für Mathematik und Informatik Institut für Informatik Abteilung Datenbanken'. Navigation links include 'Home', 'Registrierung', and 'Impressum'. The main content area is divided into sections: 'Login' with input fields for 'Username:' and 'Passwort:', a 'login' button, and a link for 'Passwort vergessen?'; 'Gast Login' with a text block explaining guest access and a 'als Gast einloggen' button; 'System Info' showing '# Benutzer: 31' and a clock icon; and 'News' with a message: 'Bitte loggen Sie sich ein, um personalisierte News lesen zu können.'



Lernziele der Vorlesung

- **Überblicks- und Detailkenntnisse zum Aufbau, Betrieb, Funktionsweise und Einsatz von Data Warehouses**
 - Kenntnis von Architekturalternativen
 - Techniken zur Datenintegration / Datenbereinigung
 - Techniken zur Sicherstellung einer hohen Performanz
- **Modellierung von Data Warehouses**
- **Überblick zu Analyseverfahren**
 - OLAP, SQL-Erweiterungen
 - Data Mining
- **sachkundige Beurteilung von kommerziell verfügbaren Data-Warehouse-Lösungen**



Vorläufige Vorlesungsübersicht

1. Einführung

2. Architektur von Data Warehouse-Systemen

3. Mehrdimensionale Modellierung / Operationen

- MOLAP, ROLAP
- Sternschema und Varianten
- SQL-Erweiterungen für mehrdimensionale Auswertungen und Zeitreihen

4. ETL: Schemaintegration + Data Cleaning

- Metadaten-Integration, Schema Matching
- Datenbereinigung (Data Cleaning) / Instanz-Matching

5. Performance-Techniken

- Indexstrukturen
- Materialisierte Sichten
- Nutzung paralleler Datenbanken

6. Data Mining-Verfahren

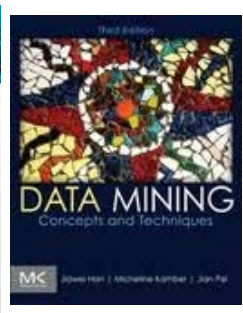
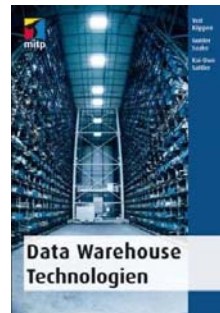
- Clusterung
- Assoziationsregeln (A-priori, FP tree)



Literatur

■ Buchauswahl

- Köppen/Saake/Sattler: Data Warehouse Technologien. 2. Auflage, Mitp 2014
- Bauer/Günzel (Hrsg.): *Data Warehouse Systeme*. 4. Auflage, dpunkt 2013
- Han / Kamber / Pei:
Data Mining: Concepts and Techniques.
3rd edition, Morgan Kaufmann 2011



<http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>

■ Eigene Forschungsarbeiten: <http://dbs.uni-leipzig.de>

- Schema Matching (COMA++)
- Data Cleaning / Instanz-Matching (FAMER)



Lehrstuhl Datenbanksysteme

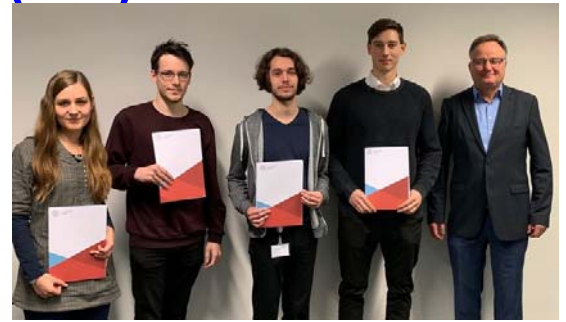
- seit 1994 am Institut für Informatik
- umfangreiches Lehrangebot
 - Vorlesungen, Praktika, Seminare
 - Online-Übungssystem LOTS
 - Eigene Lehrbücher: MRDBS, IDBS, ...



Oberseminare an Uni-Außenstelle Zingst – seit 2001



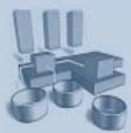
Auszeichnung von Top-Student(inn)en - seit 2008



SS2020, © Prof. Dr. E. Rahm



Forschung



Abteilung Datenbanken Leipzig

am Institut für Informatik

UNIVERSITÄT LEIPZIG

Hilfe | Registrieren



Startseite

rahm



login

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)

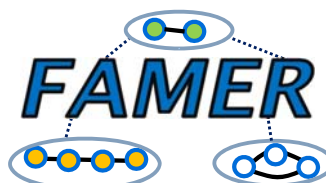
Benchmark datasets for entity resolution

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



SS2020, © Prof. Dr. E. Rahm

0-16



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)

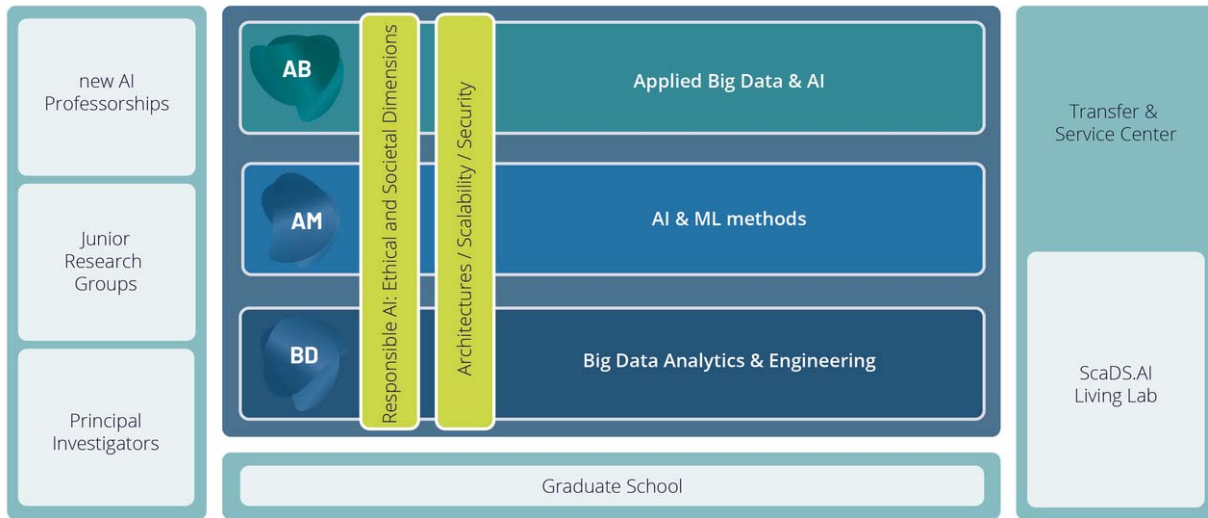


Neue Räumlichkeiten in Leipzig

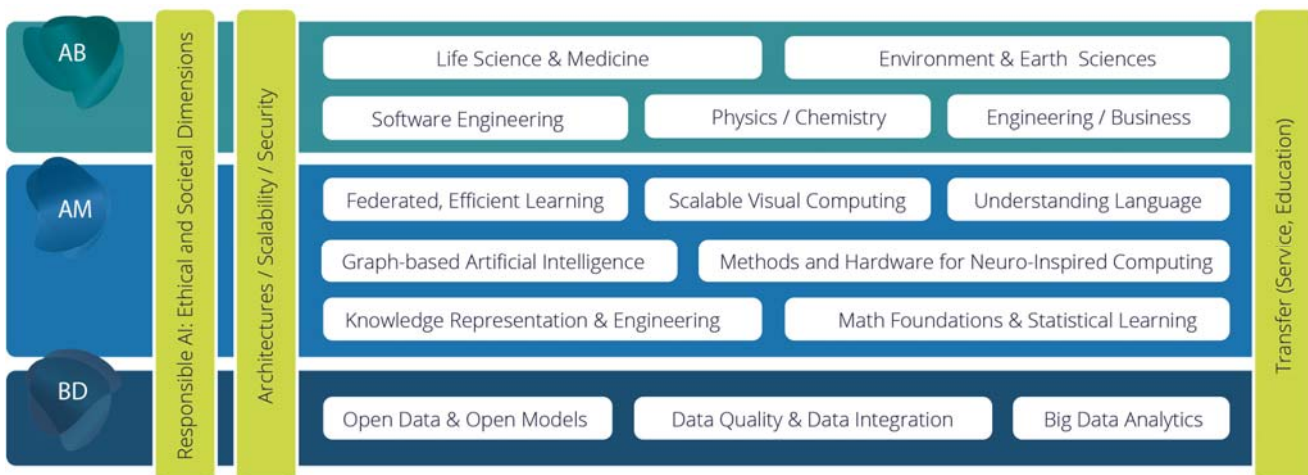
- **ScaDS.AI Leipzig** seit Dez. 2020 in Löhrs Carré (Humboldtstr. 25)
- Flächen für
 - Graduiertenschule
 - Living Lab
 - Transferzentrum
 - Seminarraum, Besprechungsräume
 - **KI-Professuren** (inkl. KI-Humboldt-Prof.)
- Umfang
 - ca. 2400 qm, 110 Arbeitsplätze



ScaDS.AI Grobstruktur



Forschungsthemen



KI-Professuren

- **8-12 neue KI-Professuren (4-6 für UL)**
- UL: 1. Professur (AI for Data Privacy & Security) ab April 2022 besetzt
 - **Prof. Erik Buchmann**
- KI-Humboldt-Professur
 - **Prof. Sayan Mukherjee** (von Duke Univ.)
 - Start im Mai 2022
- Ausgeschrieben sind bereits
 - W3 Multi-modal machine learning
 - W3 Large-Scale Knowledge Management
- Auch neu ab April 2022: Prof. Carsten Lutz (Wissensrepräsentation)

