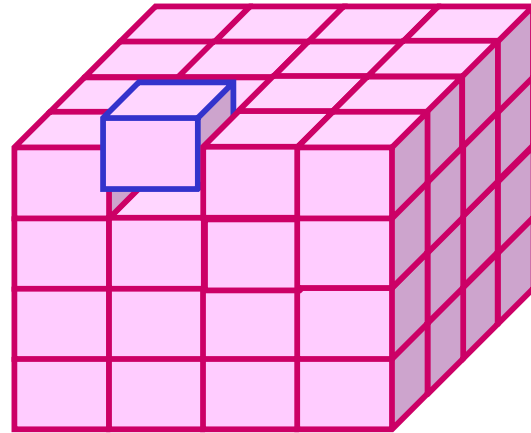


Data Warehousing

Wintersemester 2024/25

Prof. Dr. E. Rahm

Universität Leipzig
Institut für Informatik



<http://dbs.uni-leipzig.de/study/semester/ws-2024/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 (5 LP)
- 10-202-2213 – Anwendungsbezogene Datenbankkonzepte (5 LP)
- 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
- Seminar modul
- 10-201-2010 – Bachelorseminar Informatik
- Bachelorarbeit



Mapping: Module – Lehrveranstaltungen SS22

- **Master-Modul Skalierbare Datenbanktechnologien 1**
 - Data Warehousing
 - Data Mining
 - Seminar „Trends in Data Analytics and Machine Learning“
- **Master-Module Moderne Datenbanktechnologien, Bachelor-Modul Realisierung von IS**
 - Data Warehousing
 - Data Mining
- **Bachelorseminar / Masterseminar**
 - Vortrag über laufende Bachelor/Masterarbeit



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 Skalierbares Datenmanagement

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 Deep Learning		
IT-Sicherheit	Prakt. IT-Sicherheit	Data Prep.& Cleaning	Aktuelle Trends in DS	KI und Ethik	Research Data Mgmt		

Schwerpunkt Datenanalyse

Skalierbare Datenbanktechnologien 1		SDBT2	Big Data Praktikum	Mathem. Foundations of DS	Advanced Deep Learning	Masterseminar (5 LP)	Masterarbeit (25 LP)
Künstl. neuronale Netze u. maschin. Lernen		Multivariate Statistik und Data Mining		Wiss. Visualisierung			
Sequenzanalyse und Genomik		Probabilistic ML	Aktuelle Trends in DS	Language Technologies			

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)



Forschungsthemen



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 von 60 Minuten
 - voraussichtl. Feb. 2025
 - 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



Leipzig Online-Test-System

UNIVERSITÄT LEIPZIG
Fakultät für Mathematik und Informatik
Institut für Informatik
Abteilung Datenbanken

[Home](#) | [Registrierung](#) | [Impressum](#)

Login

Username:

Passwort:

[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.



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

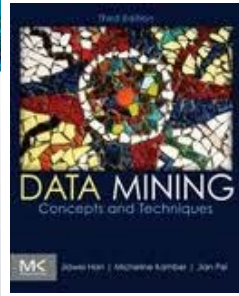
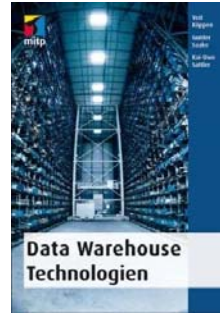
- 1. Einführung**
- 2. Architektur von Data Warehouse-Systemen**
- 3. Mehrdimensionale Modellierung / Operationen**
 - MOLAP, ROLAP
 - Sternschema und Varianten
 - SQL-Erweiterungen für mehrdimensionale Auswertungen und Zeitreihen
- 4. ETL: Schemaintegration + Data Cleaning**
 - Metadaten-Integration, 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>
- Serra: *Deciphering Data Architectures*. O'Reilly 2024



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

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

