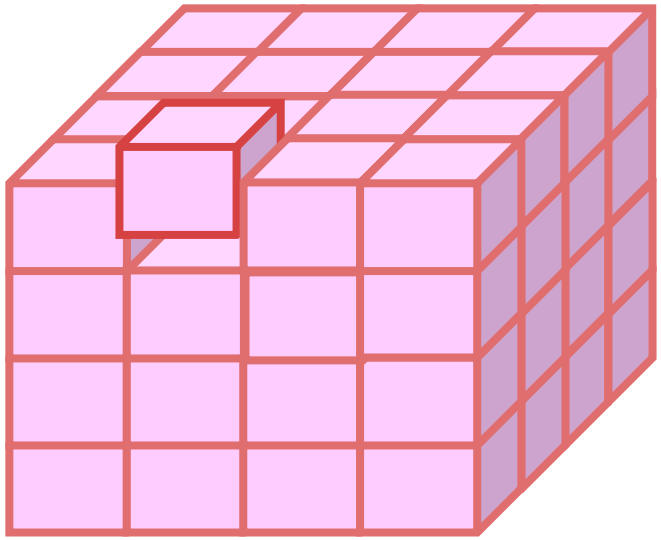


UNIVERSITÄT
LEIPZIG

DATA WAREHOUSING



Kapitel 0: Einführung

Victor Christen
Universität Leipzig
Institut für Informatik
Abteilung Datenbanken

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-DS104 – Data Preparation & Cleaning (5 LP)
  - 10-202-2012 - Aktuelle Trends der Informatik (5 LP)
    - Physik-basiertes Maschinelles Lernen

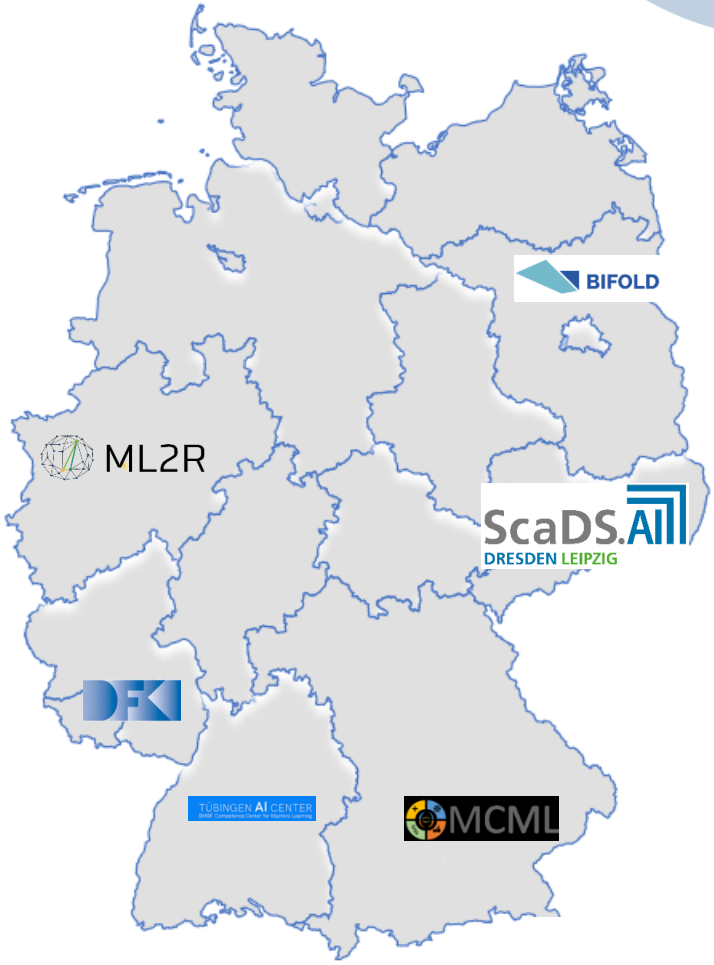
- Master Informatik
  - 10-202-2215 – Moderne Datenbanktechnologien (5 LP)
  - 10-202-2213 – Anwendungsbezogene Datenbankkonzepte (5 LP)

DBS-MODULE BACHELOR

- 10-201-2211 – Datenbanksysteme 1
- 10-201-2212 – Datenbanksysteme 2
- 10-201-2210 – Datenbankpraktikum
- 10-201-2224 – Realisierung von Informationssystemen

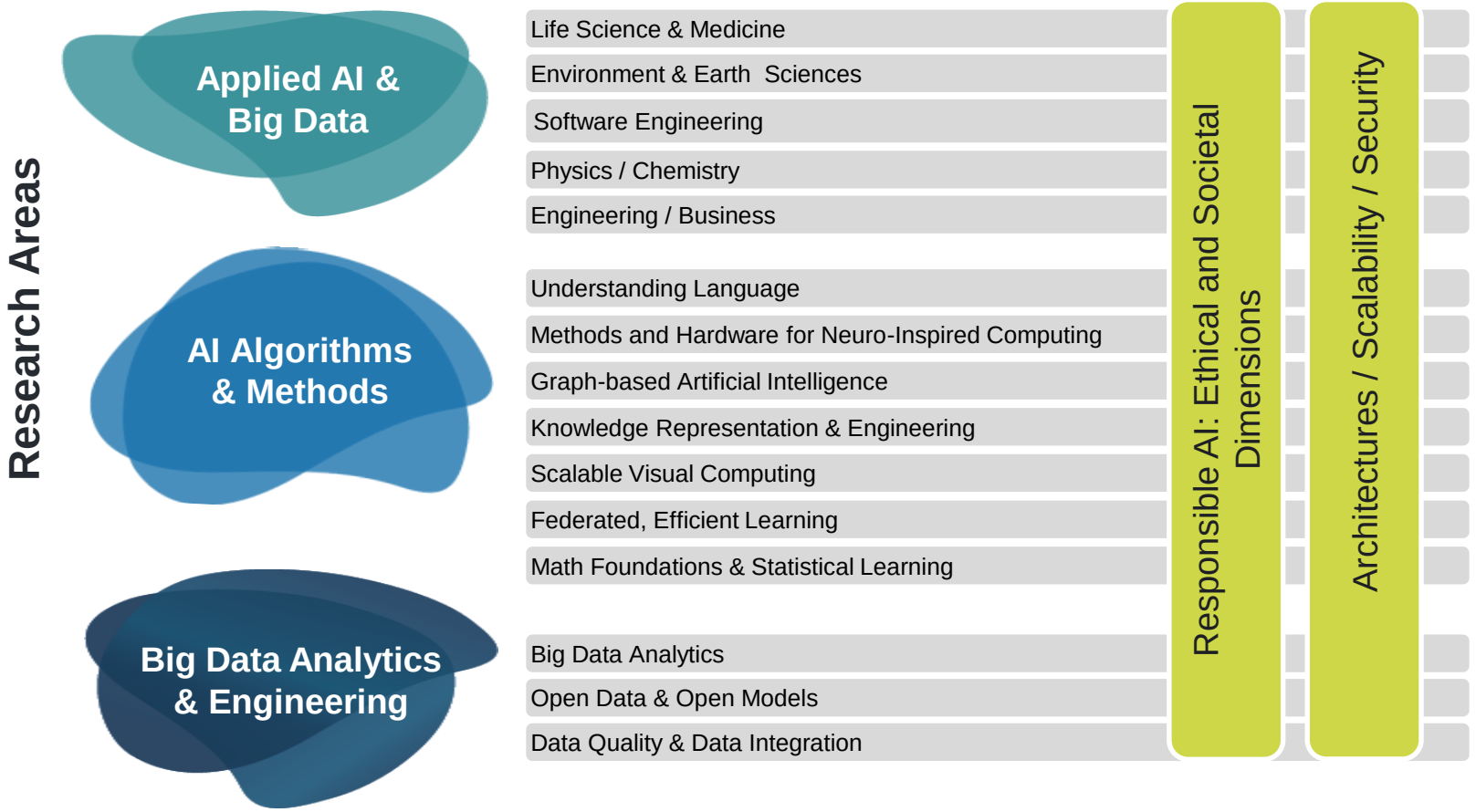
DEUTSCHE KI-ZENTREN

- 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

ScaDS.AI DRESDEN LEIPZIG



LEISTUNGSBEWERTUNG

- Prüfungsklausur von 60 Minuten
  - voraussichtl. Anfang Februar 2027 und Ende März 2027
  - überprüft konzeptionelles Wissen + Anwendungsfälle

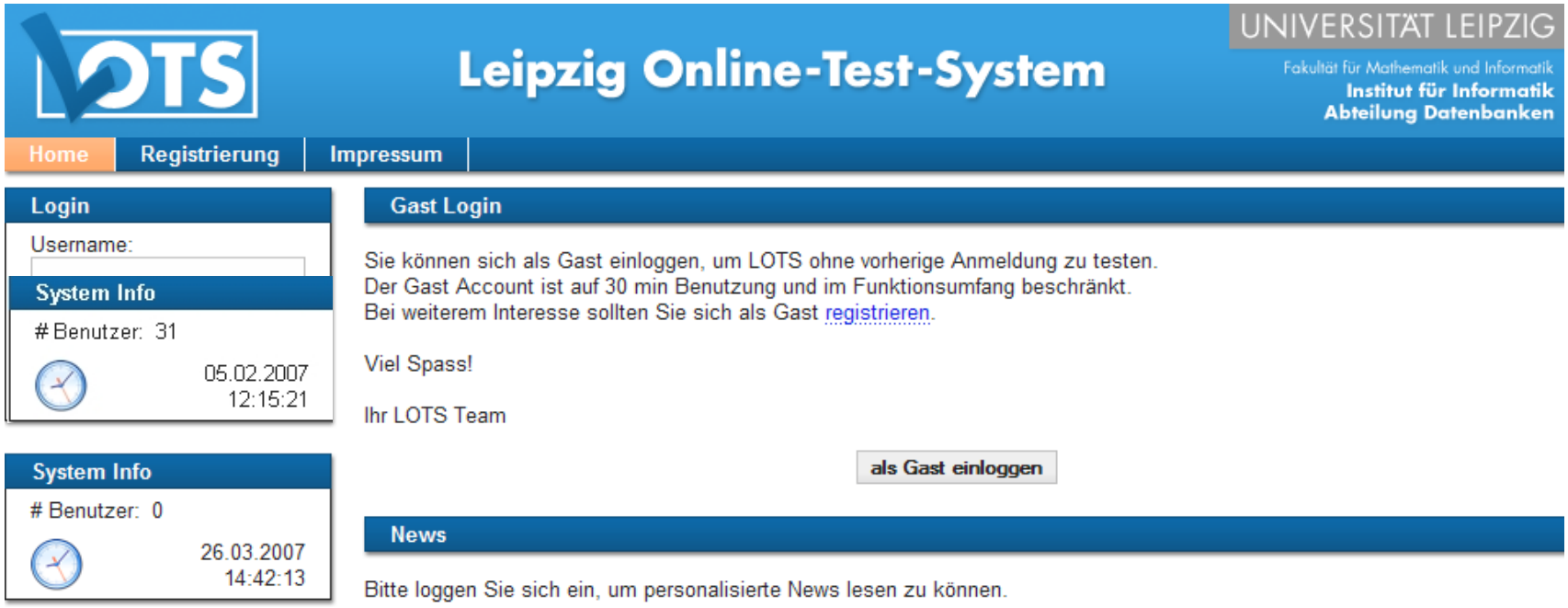
- Klausurerfolg durch
  - Beherrschen der Vorlesungsinhalte
  - Nutzung der Online-Übungen
  - vertiefendes Literaturstudium

FLIPPED-CLASSROOM

- Initial selbstständige Aufbereitung der Kapitel mithilfe der Skripte und der Aufzeichnungen
- 2/3 wöchige “Verständnisvorlesung” bietet Raum für:
  - Zusammenfassung des Kapitels und Auszug von “komplexen” Sachverhalten
  - Inhaltliche Fragen (bestenfalls bereits im Moodle-Kurs stellen)

ONLINE-ÜBUNGEN

- LOTS (Leipzig Online Test System), http://lots.uni-leipzig.de
  - dwh#2627 Kennung zum Beitritt in Übungsgruppe



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

VORLESUNGSÜBERSICHT

- Einführung
- Architektur von Data Warehouse-Systemen
- Mehrdimensionale Modellierung / Operationen
  - MOLAP, ROLAP
  - Sternschema und Varianten
  - SQL-Erweiterungen für mehrdimensionale Auswertungen und Zeitreihen
- ETL: Schemaintegration + Data Cleaning
  - Metadaten-Integration, Schema Matching
  - Datenbereinigung (Data Cleaning) / Instanz-Matching
- Performance-Techniken
  - Indexstrukturen
  - Materialisierte Sichten
  - Nutzung paralleler Datenbanken
- 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)

