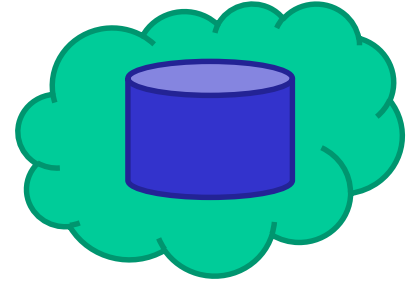


# Cloud Data Management

## Kapitel 0: Organisatorisches

**Dr. Anika Groß**  
**Wintersemester 2016**

**Universität Leipzig**  
**Institut für Informatik**  
**<http://dbs.uni-leipzig.de>**



# Organisatorisches

- ~12 Vorlesungstermine
  - Donnerstag, 9:15-10:45 Uhr, HS 19
  - Start: 13.10.2016
- Anmeldung zur Vorlesung und Prüfung bis **16.10.2016** via Almaweb
- Prüfung
  - Klausur am Ende des Semesters (konzeptionelles Wissen + Anwendungsfälle), 60 min
- Webseite mit allen wichtigen Infos:
  - [http://dbs.uni-leipzig.de/de/study/ws\\_2016\\_17/cdm](http://dbs.uni-leipzig.de/de/study/ws_2016_17/cdm)
- Übung
  - Thema: voraussichtl. MapReduce/Hadoop, Spark
  - Freiwillige Bearbeitung/Abgabe/Testat
  - Bei korrekter Bearbeitung können (voraussichtl.) bis zu 5 Punkte verdient werden, die zu den Klausurpunkten hinzugerechnet werden
  - Auch ohne Zusatzpunkte sind in der Klausur 100% der Punkte erreichbar



# DBS Module

- Master-Studium

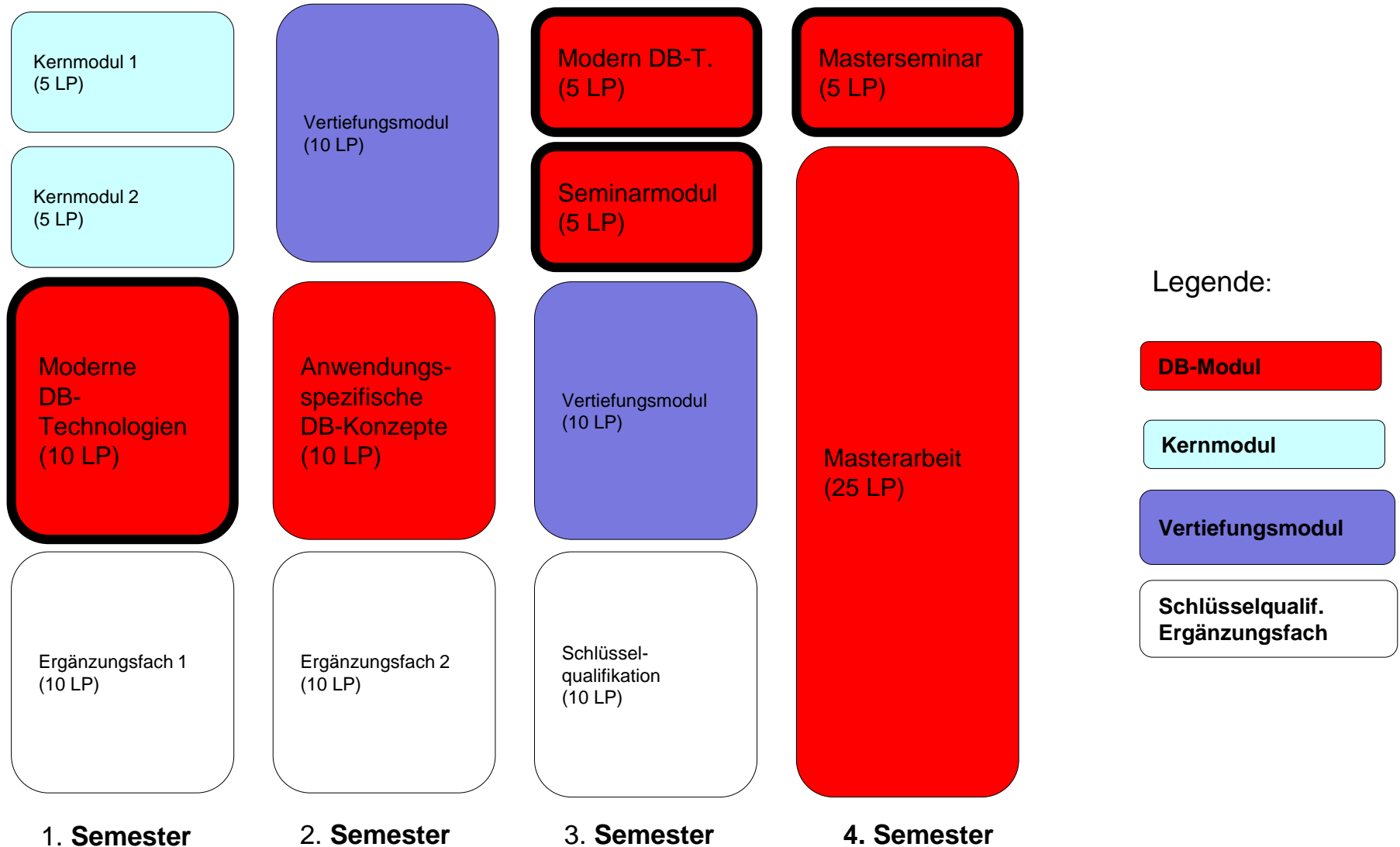
- 10-202-2215 – Moderne Datenbanktechnologien (Kleines Modul)
- 10-202-2216 – Moderne Datenbanktechnologien (Großes Modul)
- 10-202-2213 – Anwendungsbezogene Datenbankkonzepte (Kleines Modul)
- 10-202-2214 – Anwendungsbezogene Datenbankkonzepte (Großes Modul)
- 10-202-2011 – Masterseminar Informatik

- Bachelor-Studium

- 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



# Masterstudium **DBS Profil**



# BACHELOR DBS-Profil

Algorithmen und Datenstrukturen 1 (5 LP)	Algorithmen und Datenstrukturen 2 (5 LP)	<b>DBS 1 (5 LP)</b>	<b>DBS 2 (5 LP)</b>	<b>Realisierung von IS (5 LP)</b>	Kernmodul 5 (5 LP)
Modellierung und Programmierung 1 (5 LP)	Modellierung und Programmierung 2 (5 LP)	Softwaretechnik (5 LP)	Softwarepraktikum (5 LP)	<b>Seminarmodul (5 LP)</b>	<b>Bachelorseminar (5 LP)</b>
Technische Informatik 1 (5 LP)	Technische Informatik 2 inkl. Hardwarepraktikum (5 LP)	Kommunikationssysteme (5 LP)	Kernmodul (5 LP)	Vertiefungsmodul (10 LP)	<b>Bachelorarbeit (10 LP)</b>
Logik (5 LP)	Java-Praktikum (5 LP)	Automaten und Sprachen (5 LP)	Berechenbarkeit (5 LP)		
Analysis 1 (10 LP)	Lineare Algebra 1 (10 LP)	Diskrete Strukturen (5 LP)	<b>DB-Praktikum (5 LP)</b>	Ergänzungsfach (10 LP)	Fakultätsübergreifende Schlüsselqualifikation (10 LP)
		Wahrscheinlichkeitstheorie (5 LP)	Ergänzungsfach (5LP)		
<b>1. Semester</b>	<b>2. Semester</b>	<b>3. Semester</b>	<b>4. Semester</b>	<b>5. Semester</b>	<b>6. Semester</b>

Legende:

Inf-Pflichtmodul

MI-Modul

**DBS-Modul**

Mathematikmodul

Schlüsselqualif. Ergänzungsfach



# Module - Lehrveranstaltungen WS16/17 (1)

- Kleines Master-Modul “Moderne Datenbanktechnologien”  
(zwei Veranstaltungen, davon mind. eine Vorlesung)
  - Vorlesung “Implementierung von Datenbanksystemen 1”
  - Vorlesung “Cloud Data Management”
  - Data Warehouse Praktikum
  - Forschungsseminar “Big Data - Streaming”
- Großes Master-Modul “Moderne Datenbanktechnologien”  
(Pflichtmodul im Schwerpunkt “Big Data” / Vertiefungsmodul);  
(zwei Vorlesungen + Praktikum / Seminar)
  - Vorlesung “Implementierung von Datenbanksystemen 1”
  - Vorlesung “Cloud Data Management”
  - Data Warehouse Praktikum
  - Forschungsseminar “Big Data - Streaming”



# Module - Lehrveranstaltungen WS16/17 (2)

- Bachelor-Modul “Realisierung von Informationssystemen” (2 Vorlesungen)
  - Vorlesung Implementierung von Datenbanksystemen 1
  - Vorlesung Cloud Data Management
- Seminar modul “Forschungsseminar Datenbanken”
  - Forschungsseminar “Big Data - Streaming”
- Bachelorseminar / Masterseminar
  - Vortrag über laufende Bachelor/Masterarbeit
  - Nächste Termine: Dezember 2016, SS 2017 (voraussichtl. Juni/Juli in Zingst)



# Master Informatik: neuer Schwerpunkt *Big Data*

- Zunehmende Bedeutung der Verwaltung und Analyse riesiger Datenmengen ("Big Data") in Unternehmen und Wissenschaft
  - Datenmanagement: skalierbare Verwaltung, schnelle Auswertungen...
  - Analyse + Vorhersage: Data Mining, maschinelles Lernen, statistische Verfahren, ...
  - Visualisierung
- Hoher Bedarf an Informatikern mit Kenntnissen im Bereich "Big Data" bzw. "Data Science"

1. Semester	2 Semester	3. Semester	4. Semester
Kernmodul I 5 LP	Kernmodul III 5 LP	<b>Vertiefungsmodul III im Bereich Data Mining / Visualisierung 10 LP</b>	<b>Masterarbeit / Masterseminar im Bereich Big Data</b> 25 LP / 5LP
Kernmodul II 5 LP	Seminarmodul 5 LP		
<b>Moderne Datenbank- technologien 10 LP</b>	Vertiefungsmodul II 10 LP	Vertiefungsmodul IV 10 LP	
Ergänzungsfach I 10 LP	Ergänzungsfach II 10 LP	Fakultätsinterne Schlüsselqualifikation 10 LP	





# Schwerpunkt Big Data - Module

- **Beschreibung**
  - [Beschreibung auf den Seiten des Studienbüros](#)
  - [Detailbeschreibung \(pdf\)](#)
- **Datenmanagement**
  - **10-202-2216: Moderne Datenbanktechnologien (Pflichtmodul, 10 LP)**
  - 10-202-2214/10-202-2213: Anwendungsbezogene Datenbankkonzepte
  - auch als Kernfachvariante (5 LP)
  - Verschiedene Vorlesungen, Praktika und Seminare, insb:
    - Cloud Data Management, NoSQL-Datenbanken, Mehrrechner-DBS, Data Warehousing, Datenintegration, Big Data Praktikum, Data Warehouse Praktikum, ...
- **Data Mining / Visualisierung (je 10 LP, davon ein Wahlpflichtmodul)**
  - 10-202-2201: Visualisierung (inkl. Praktikum)
  - 09-INF-BI01: Statistisches Lernen
  - 10-202-2104: Neuroinspirierte Informationsverarbeitung
  - Wahlpflichtmodul, d.h. es muss mind. eines der drei Module belegt werden.



# Schwerpunkt Big Data - Module

- **Web / Information Retrieval**
  - 10-202-2322: Textdatenbanken
  - 10-202-2314: Fortgeschrittene Methoden des Information Retrieval
  - 10-202-2309: Semantic Web
  - 10-202-2120: Computational Advertising (ohne/ mit Prak.)
  
- **Empfehlungen für Ergänzungsfächer**
  - 10-202-2207: Sequenzanalyse und Genomik
  - 09-202-2413: Statistische Aspekte der Analyse molekularbiologischer und genetischer Daten / Statistische Analyse von High-Throughput-Daten
  - 10-202-2204: Medizinische Bildaufnahme und Bildverarbeitung
  - 10-202-2205: Graphen und biologische Netzwerke
  - Enterprise Computing



# Online Übungen

- LOTS (Leipzig Online Test System), <http://lots.uni-leipzig.de>

The screenshot shows the homepage of the Leipzig Online Test System (LOTS). The header features the text 'Leipzig Online Test System' and 'UNIVERSITÄT LEIPZIG' with the affiliation 'Institut für Informatik Abteilung DBS'. 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 contains input fields for 'Username:' and 'Passwort:' with a 'login' button. The 'Gast Login' section provides information for guest users, including a 'als Gast einloggen' button. The 'System Info' section displays the number of users and the current date and time.

**Leipzig Online Test System** **UNIVERSITÄT LEIPZIG**  
Institut für Informatik  
Abteilung DBS

**Home** **Registrierung** **Impressum**

**Login**

Username:

Passwort:

**login**

**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 [registrieren](#).

Viel Spass!

Ihr LOTS Team

**als Gast einloggen**

**System Info**

# Benutzer: 2  
12.10.2005  
15:27:05

**News**

Bitte loggen Sie sich ein, um personalisierte News lesen zu können.



# LOTS: Online Übungen

## CDM: Cloud Computing

### Aufgabe 1 (Cloud Computing)

Bewerten Sie die folgenden Aussagen zum Thema Cloud Computing auf ihre Korrektheit!

- Cloud Computing bezeichnet die externe Bereitstellung von IT-Infrastrukturen sowie Applikations-Hosting ueber das Internet (bzw. Intranet).
- Virtualisierung bezeichnet die dynamische Anpassung der IT-Ressourcen an den Bedarf (z.B. einer Web-Anwendung)
- Platform as a Service kann auch als Utility Computing bezeichnet werden.
- Software as a Service meint das Bereitstellen von Web-Anwendungen zur Nutzung, z.B. standardisierter Office-Anwendungen.

### Aufgabe 2 (Cloud Data Center)

Bewerten Sie die folgenden Aussagen zu Data Centern auf ihre Richtigkeit!

- Cloud Data Center verfolgen einen "scale up"-Ansatz.



# Forschung

- <http://dbs.uni-leipzig.de/de/research>



## Abteilung Datenbanken Leipzig

am Institut für Informatik

UNIVERSITÄT LEIPZIG

Hilfe | Registrieren



Startseite

login

### Inhalte

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

### Forschung

[Publications \(2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005, ...\)](#)

[Big Data Center ScaDS Dresden/Leipzig](#)

[Graph-based data analysis \(GRADOOP, BIIIG\)](#)

[Entity Matching for Big Data \(Dedoop\)](#)

[LOD Link Discovery](#)

[Evolution of Semantic Annotations \(ELISA\)](#)

[Schema and Ontology Matching \(COMA++, GOMMA\), Ontology Merging \(ATOM\)](#)

[Semantic ontology matching \(STROMA mapping enrichment, SemRep repository\)](#)

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

[Web Data Integration \(WDI\) Lab](#)

[Object Matching / Entity Resolution](#)

[Bibliometric Analysis](#)

[Current prototypes and implementations](#)



### Neue Publikationen

- [Graph Mining for Complex Data Analytics](#)
- [Holistic Entity Clustering for Linked Data](#)
- [Scalable privacy-preserving linking of](#)

# Forschungsgebiete / Projekte

- Big Data Zentrum ScaDS Dresden/Leipzig
- Themen u.a. Big Data / Cloud Data Management, z.B.
  - Entity Matching for Big Data (Dedoop)
  - Privacy preserving record linkage
  - Holistic entity clustering
- ...
- Graph-based data analysis (BIIG, Gradoop)
- Datenintegration
  - Schema and Ontology Matching (COMA++, GOMMA, STROMA, SemRep)
  - Object Matching / Entity Resolution
  - Evolution of ontologies and mappings
  - DFG-Projekt: “ELISA - Evolution of Semantic Annotations”
  - DFG Projekt: “Learning-based Link Discovery”
  - BMPF-Projekt: “Leipzig Health Atlas“ u.a. Kooperation mit IMISE, IZBI und dem LIFE-Projekt...



# Abschlussarbeiten / SHK

- Wir suchen ständig interessierte Studenten für
  - SHK/WHK Tätigkeiten
  - Bachelor- oder Masterarbeiten
- zur Mitarbeit an aktuellen Forschungsthemen des Lehrstuhls, sowie gemeinsamen Themen mit dem Big Data Zentrum ScaDS Dresden/Leipzig
  - [Auswahl mögl. Themen](#)
- Kooperationen mit (regionalen) Unternehmen, z.B.
- Anfragen per Formular oder persönlich
  - [http://dbs.uni-leipzig.de/de/service/anfrage\\_abschlussarbeit](http://dbs.uni-leipzig.de/de/service/anfrage_abschlussarbeit)

