

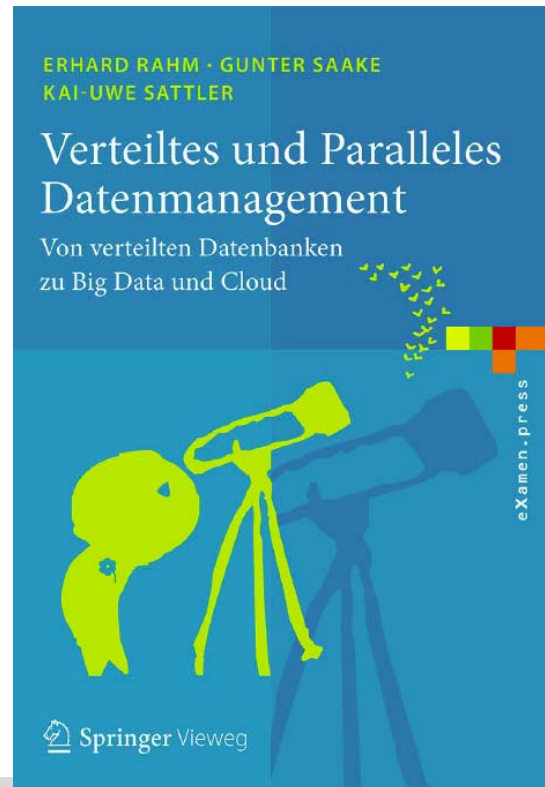
# Mehrrechner-Datenbanksysteme (Verteilte und Parallele DBS)

**Prof. Dr. E. Rahm**

Wintersemester 2015/2016

Universität Leipzig  
Institut für Informatik

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



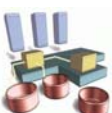
## DBS-Module

### ■ 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)
  - Seminar modul
- 10-202-2011 – Masterseminar Informatik
  - Masterarbeit

### ■ Bachelor-Studium

- 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 WS15/16

- Master-Module **Moderne Datenbanktechnologie** (zwei bzw. drei Veranstaltungen für kleines bzw. großes Modul):
  - **Mehrrechner-DBS**
  - NoSQL Datenbanken
  - Problemseminar *Privacy for Big Data*
  - Big Data Praktikum
- Bachelor-Modul **Realisierung von Informationssystemen**  
zwei Veranstaltungen aus den genannten
  - **Mehrrechner-DBS + X**
- **Seminarmodul**
  - Problemseminar *Privacy for Big Data*
- **Bachelorseminar / Masterseminar**
  - Vortrag über laufende Bachelor/Masterarbeit



## BACHELOR DBS-Profil

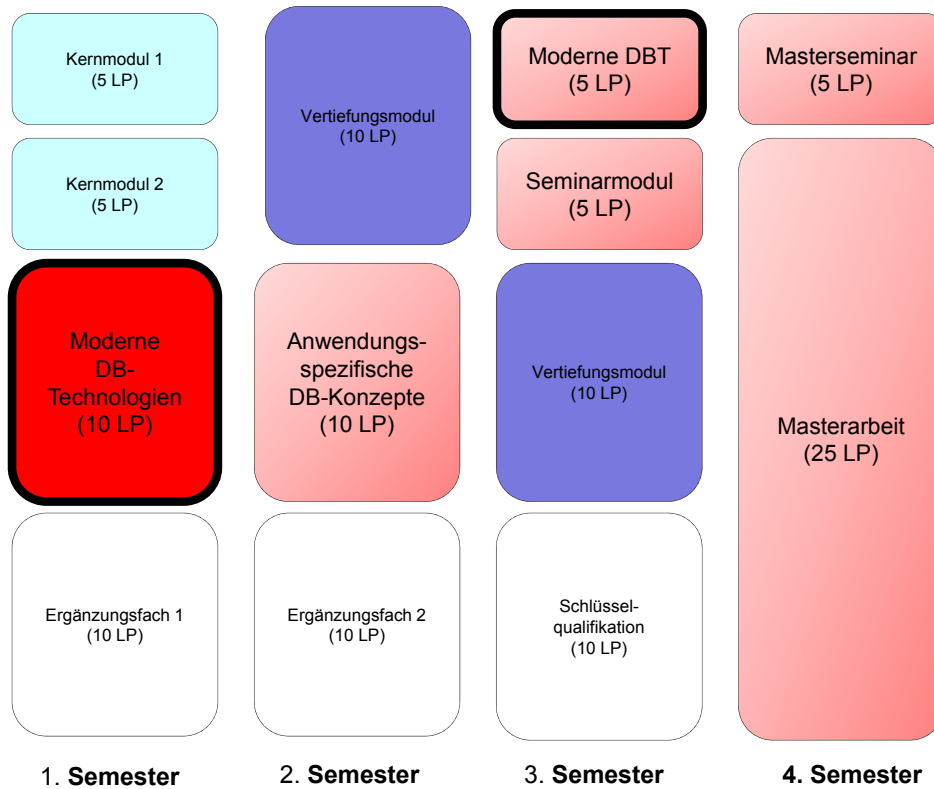
Algorithmen und Datenstrukturen 1 (5 LP)	Algorithmen und Datenstrukturen 2 (5 LP)	<b>DBS 1 (5 LP)</b>	<b>DBS 2 (5 LP)</b>	<b>Realisierung von IS (5 LP)</b>	Kernmodul 5 (5 LP)
Modellierung und Programmierung 1 (5 LP)	Modellierung und Programmierung 2 (5 LP)	Softwaretechnik (5 LP)	Softwarepraktikum (5 LP)	Seminarmodul (5 LP)	Bachelorseminar (5 LP)
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:

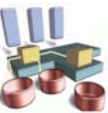
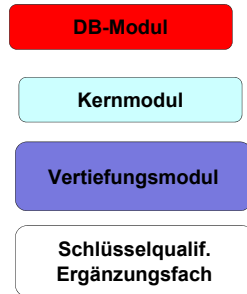
Inf-Pflichtmodul	MI-Modul
<b>DBS-Modul</b>	Mathematikmodul
	Schlüsselqualif. Ergänzungsfach



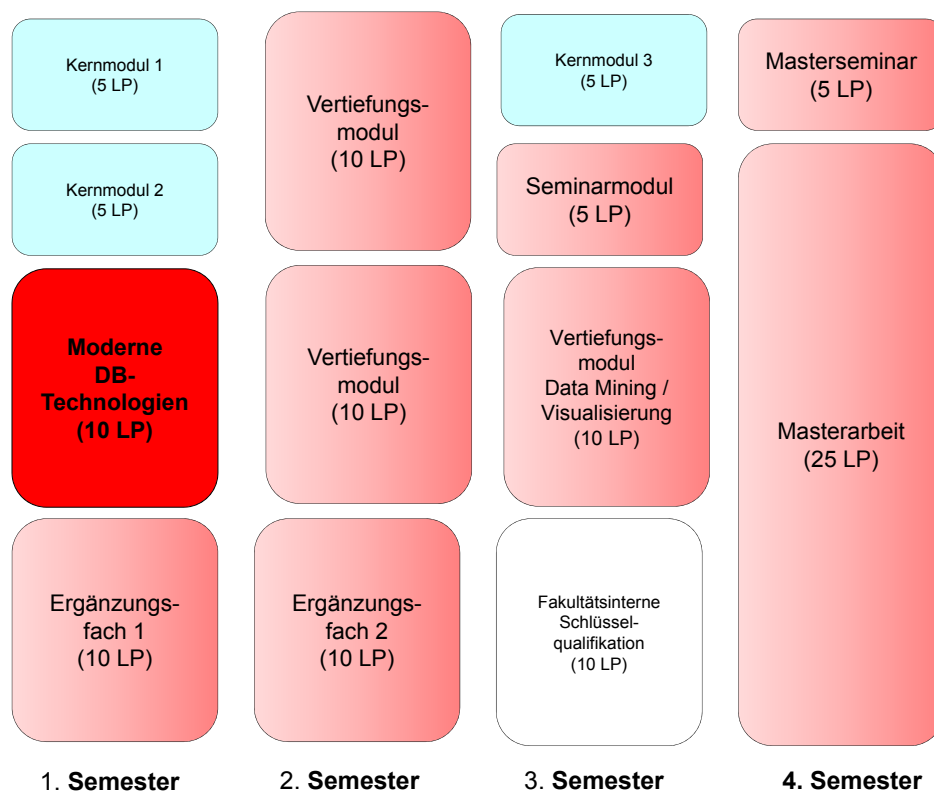
# Informatik Masterstudium **DBS-Profil**



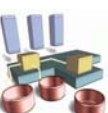
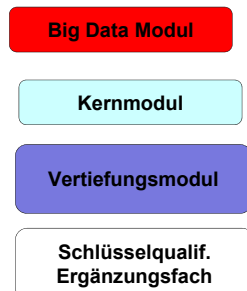
Legende:



# Masterstudium Informatik: **Schwerpunkt Big Data (NEU)**



Legende:



# Big Data Schwerpunkt: wählbare Module

## ■ Datenmanagement

– **Moderne Datenbanktechnologien (Pflichtmodul, 10 LP)**

– Anwendungsbezogene Datenbankkonzepte (10 LP)

*Die beiden Module können auch in einer Kernfachvariante (5 LP) belegt werden und beinhalten verschiedene Vorlesungen, Praktika und Seminare. Besonders relevant sind die Vorlesungen Cloud Datenmanagement, NoSQL-Datenbanken, Mehrrechner-DBS, Data Warehousing und Datenintegration sowie das Big Data Praktikum und Data Warehouse Praktikum.*

## ■ Data Mining / Visualisierung (je 10 LP), Wahlpflicht

– Neuroinspirierte Informationsverarbeitung

– Statistisches Lernen (inkl. R-Kurs)

– Visualisierung (inkl. Praktikum)

## ■ Web / Information Retrieval

– Textdatenbanken

– Fortgeschrittene Methoden des Information Retrieval

– Semantic Web

– Computational Advertising

## ■ Ergänzungsfächer

– Sequenzanalyse und Genomik

– Statistische Aspekte der Analyse molekularbiologischer und genetischer Daten

– Medizinische Bildaufnahme und Bildverarbeitung

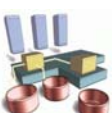
– Graphen und biologische Netzwerke

– Enterprise Computing



# Bachelor of Science (B.Sc.) Wirtschaftsinformatik

	10 Leistungspunkte		10 Leistungspunkte		10 Leistungspunkte	
1	Wirtschaftswissenschaften	Rechnungswesen	Wirtschaftsinformatik I	Strukturierte Programmierung	Recht für Wirtschaftswissenschaftler	Mathematik für Wirtschaftswissenschaftler
2	Mikroökonomik		Web Techniken I	Objektorientierte und generische Programmierung		
3	Statistik und Wahrscheinlichkeitsrechnung		Wirtschaftsinformatik II	Software-technik	Datenbanksysteme I	Algorithmen und Datenstrukturen I
4	Externes und internes Rechnungswesen			Entwicklung verteilter Anwendungen	Datenbanksysteme II	Algorithmen und Datenstrukturen II
5	Praktikum		Fakultätsübergreifende Schlüsselqualifikation oder Praktikum		Fachnahe Schlüsselqualifikationen oder Praktikum	
6	Wahlpflichtmodul(e)		Web Techniken II	Web Science	Bachelorarbeit	



# Master Wirtschaftsinformatik

Sem.	10 LP	10 LP	10 LP
1	Wahlpflichtmodule aus den Masterstudiengängen Betriebswirtschaftslehre bzw. Volkswirtschaftslehre	Operations Research	Wahlpflichtmodule aus den Masterstudiengängen Betriebswirtschaftslehre bzw. Volkswirtschaftslehre
2	Anwendungssysteme I	Wissensbasierte Systeme <b>und/oder</b> Computergrafik <b>und/oder</b> Datenbankpraktikum <b>und/oder</b> Textdatenbanken <b>und/oder</b> Linguistische Informatik <b>und/oder</b> Information Retrieval <b>und/oder</b> Grundlagen komplexer Systeme	Softwareengineering in frühen Phasen
3	Anwendungssysteme II	Moderne Datenbanktechnologien <b>oder</b> Integration und Architektur von Informationssystemen <b>oder</b> Softwaresystemfamilien	Wahlpflichtmodule aus den Masterstudiengängen Betriebswirtschaftslehre bzw. Volkswirtschaftslehre
4	Anwendungssysteme III <b>oder</b> Anwendungsbezogene Datenbankkonzepte		Masterarbeit



Wirtschaftsinformatik



## 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	SS
	Bio Data Management	Vertiefung	2	WS
	Ontologie-Management	Vertiefung	2	WS
	Cloud Data Management	Vertiefung	2	SS
	NoSQL-DB	Vertiefung	2	WS

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



# Vorläufiges Inhaltsverzeichnis

## ■ Einführung

- Anforderungen an Mehrrechner-Datenbanksysteme (Verteilte / Parallele DBS)
- Arten der Parallelität, Scaleup und Speedup

## ■ Klassifikation von Mehrrechner-DBS

## ■ VDBS: Schemaarchitektur, Katalogverwaltung

## ■ Datenverteilung in VDBS / PDBS (Fragmentierung, Allokation)

## ■ Verteilte / Parallele Query-Verarbeitung

## ■ Verteilte Transaktionsverwaltung (Commit-Protokolle, Synchronisation)

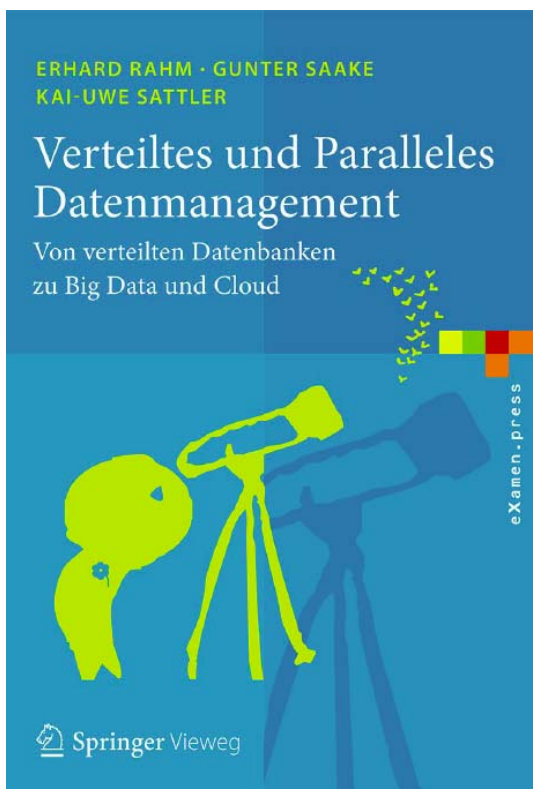
## ■ Replizierte DB

## ■ Shared-Disk-DBS

*Data Warehouses, Datenintegration, Cloud Data Management. NoSQL*  
-> eigene Vorlesungen



## Lehrbuch



## ■ Rahm/Saake/Sattler: Verteiltes und Paralleles Datenmanagement. Springer 2015, 35 Euro

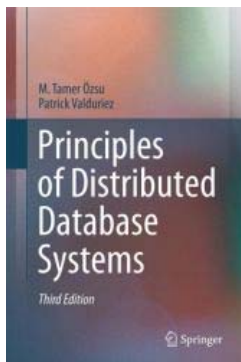
- E-Book online zugänglich im Uni-Netz
- mit Übungsaufgaben



# Zusatzliteratur



- E. Rahm: Mehrrechner-Datenbanksysteme.
  - Addison-Wesley 1994
  - Übungsaufgaben mit Musterlösungen
  - Online-Version verfügbar (PDF und HTML)



- M. T. Özsu, P. Valduriez: Principles of Distributed Database Systems.
  - 3<sup>rd</sup> edition, Springer-Verlag, 2011



# Online-Übungen

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

**LOTS** Leipzig Online-Test-System

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

Home Registrierung Impressum

**Login**

Username:

Password:

[Passwort vergessen?](#)

**System Info**

# Benutzer: 31

05.02.2007 12:15:21

**Gast Login**

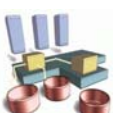
Sie können sich als Gast einloggen, um LOTS ohne vorherige Anmeldung zu testen. Der Gast Account ist auf 30 min Benutzung und im Funktionsumfang beschränkt. Bei weiterem Interesse sollten Sie sich als Gast [registrieren](#).

Viel Spass!

Ihr LOTS Team

**News**

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



# LOTS: Online-Übungen

## Aufgabe 3 (Nested-Block-Join)

0 von 2

Für den Gleichverbund zwischen R und S (je 100.000 Sätze, Blockungsfaktor 100) soll ein Nested-Block-Join genutzt werden. Welche Kombinationen zwischen verfügbarer Hauptspeichergröße M und Anzahl erreichbarer Plattenzugriffe (ohne Schreiben des Resultats) treffen zu?

- M=1001; 1 Million Plattenzugriffe
- M=5001; 2000 Plattenzugriffe
- M=1001; 2000 Plattenzugriffe
- M=501; 3000 Plattenzugriffe

Erklärung

Erklärung

Erklärung

## Aufgabe 4 (Hash-Join)

0 von 2

Markieren Sie die zutreffenden Aussagen.

- Durch die Nutzung von Bitvektoren lässt sich der Umfang der Hash-Tabellen kleiner halten
- Hash-Joins sollten nur eingesetzt werden, wenn die kleinere Eingabetabelle im Hauptspeicher gespeichert werden kann
- Hash-Joins können effektiv zur Realisierung von Mehr-Wege-Joins genutzt werden
- Hash-Join-Verfahren eignen sich zur Beantwortung von Equi-Join-Anfragen

Erklärung

Erklärung



# Forschung

## Database Group Leipzig

within the department of computer science

UNIVERSITÄT LEIPZIG

[help](#) | [register](#)

Homerahm.....login

### Contents

- ▶ Staff
- ▼ Research
  - ▣ Publications
  - ▶ Projects
  - ▣ Prototypes
  - ▣ Annual reports
  - ▣ Cooperations
  - ▶ Graduations
  - ▣ Colloquia
  - ▶ Conferences
- ▶ Study
- ▶ Service

## Research

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

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

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

[Graph-based Business Intelligence \(BIIIG\)](#)

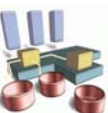
[Schema and Ontology Matching \(COMA++, GOMMA\), Ontology Merging \(ATOM\)](#)

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

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

[Bibliometric Analysis](#)

[Current prototypes and implementations](#)





# Research Projects

## ■ Datenintegration für Big Data

- **Dedoop**: Deduplication based on Hadoop
- Learning-based Link Discovery (DFG)
- Privacy-Preserving Record Linkage
- Schema and Ontology Matching (COMA, GOMMA, STROMA)
- Ontology Merging (ATOM)



## ■ Evolution von Ontologien und Annotationen (DFG)

## ■ **Gradoop**: Graph data management and analytics with Hadoop



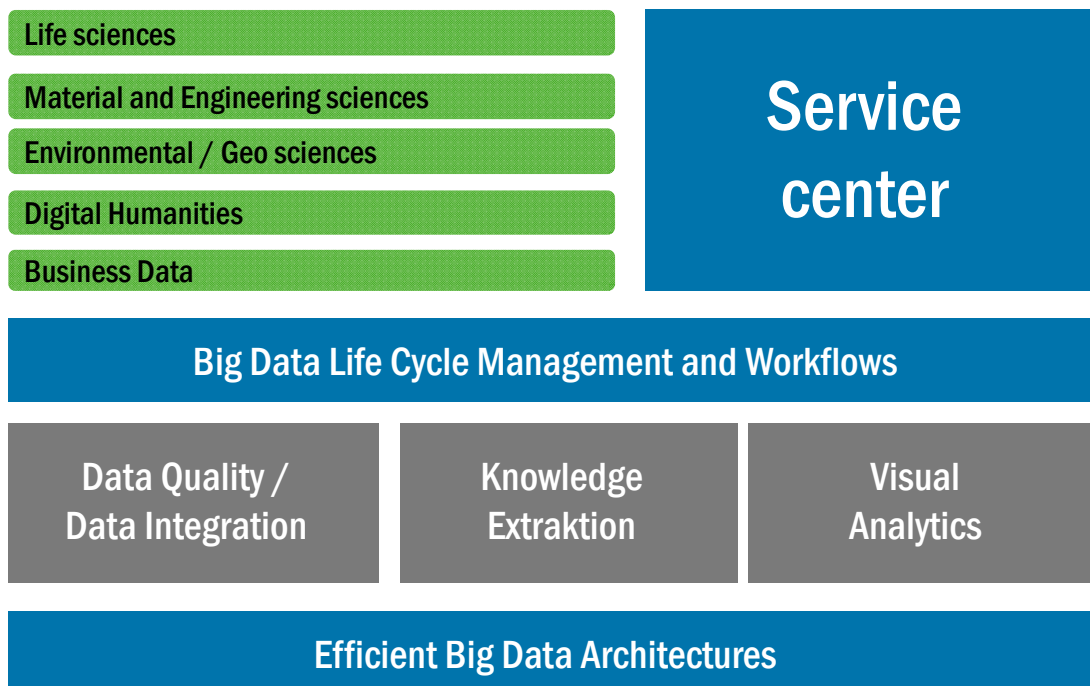
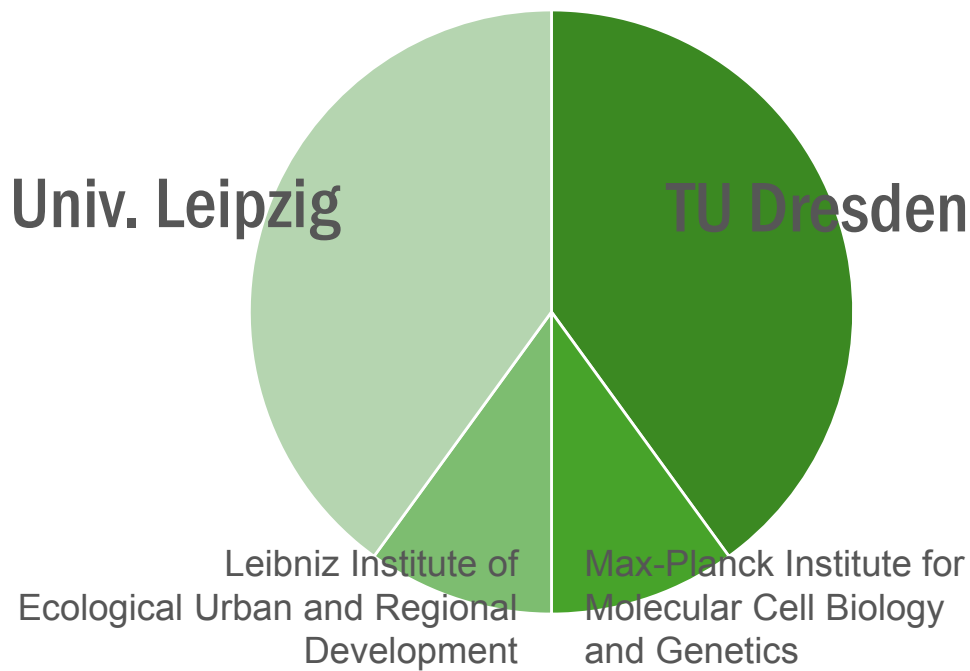
### Two Centers of Excellence for Big Data in Germany

- ScaDS Dresden/Leipzig
- Berlin Big Data Center (BBDC)

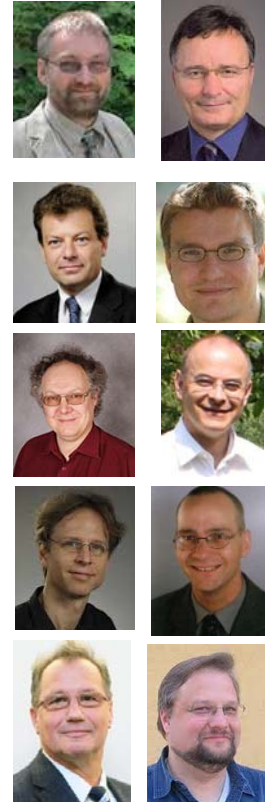
### ScaDS Dresden/Leipzig (Competence Center for Scalable Data Services and Solutions Dresden/Leipzig)

- scientific coordinators: Nagel (TUD), Rahm (UL)
- start: Oct. 2014
- duration: 4 years (option for 3 more years)

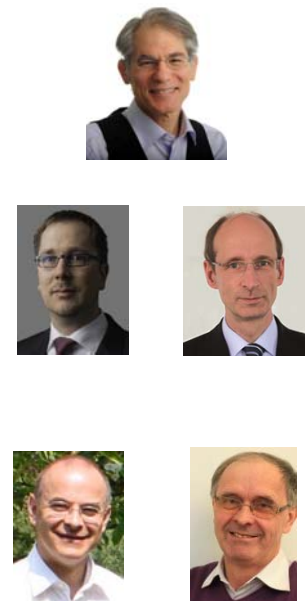




- Data-intensive computing **W.E. Nagel**
- Data quality / Data integration **E. Rahm**
- Databases **W. Lehner, E. Rahm**
- Knowledge extraction/Data mining  
**C. Rother, P. Stadler, G. Heyer**
- Visualization  
**S. Gumhold, G. Scheuermann**
- Service Engineering, Infrastructure  
**K.-P. Fähnrich, W.E. Nagel, M. Bogdan**



- Life sciences **G. Myers**
- Material / Engineering sciences **M. Gude**
- Environmental / Geo sciences **J. Schanze**
- Digital Humanities **G. Heyer**
- Business Data **B. Franczyk**



# Abschlussarbeiten / SHK-Jobs

- Wir suchen Studenten mit guten DB-Kenntnissen für
  - Bachelor- oder Masterarbeiten
  - SHK/WHK-Tätigkeiten
  
- Mitarbeit an aktuellen Forschungsthemen
  - des Lehrstuhls bzw
  - des BigData-Zentrums ScaDS

