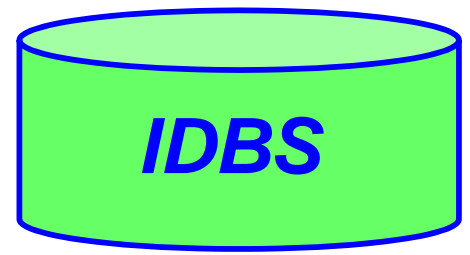


Implementierung von Datenbanksystemen 1 (IDBS1)



Wintersemester 2016/17

Prof. Dr. Erhard Rahm

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

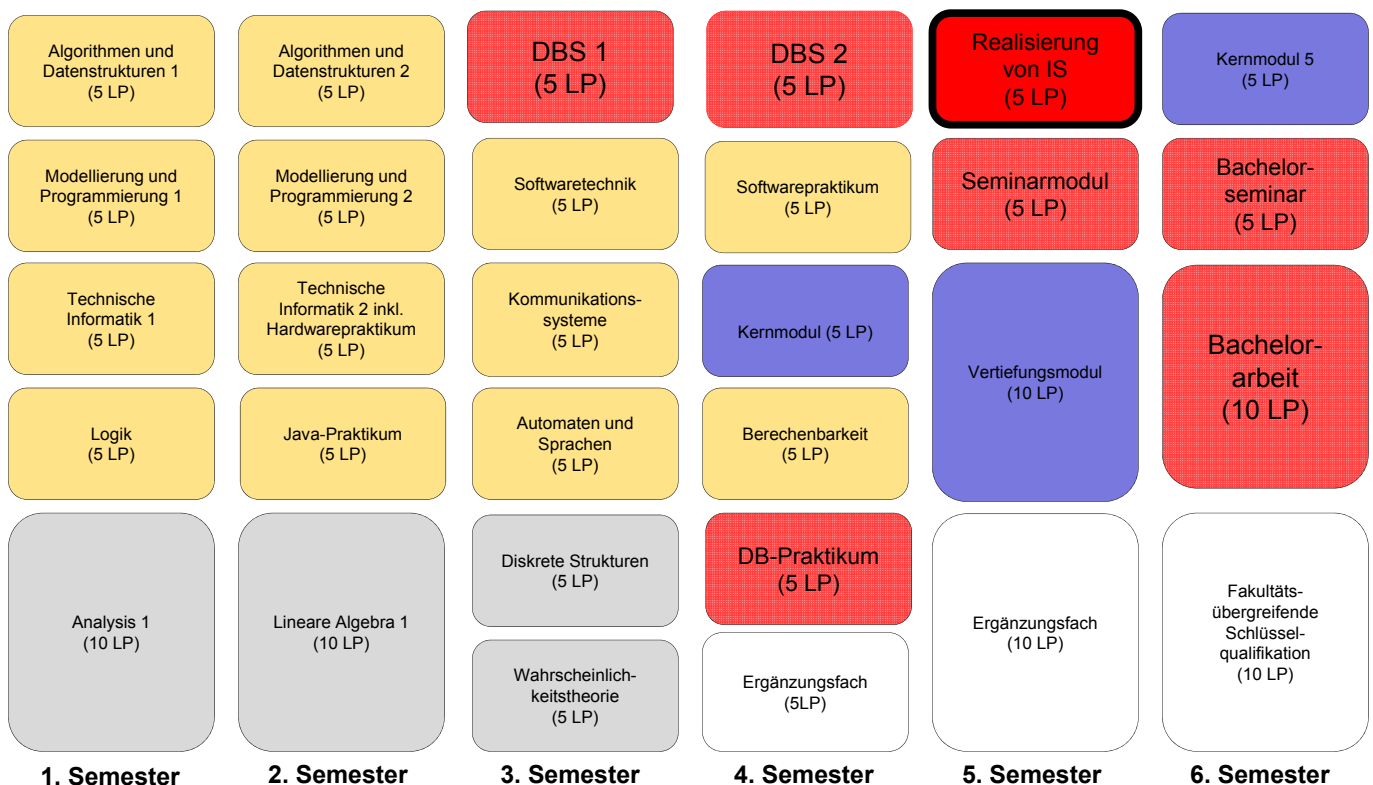


Mapping Module – Lehrveranstaltungen WS16/17

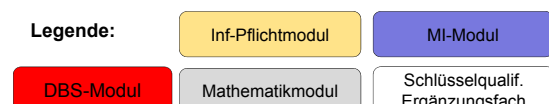
- Master-Module **Moderne Datenbanktechnologie** (zwei bzw. drei Veranstaltungen für kleines bzw. großes Modul):
 - IDBS1
 - Cloud Data Management
 - Problemseminar *Big Data Streaming*
 - Data Warehouse Praktikum
- Bachelor-Modul **Realisierung von Informationssystemen**
zwei Veranstaltungen aus den genannten
 - IDBS1 + X
- Seminar-Modul
 - Problemseminar *Big Data Streaming*
- Bachelorseminar / Masterseminar
 - Vortrag über laufende Bachelor/Masterarbeit



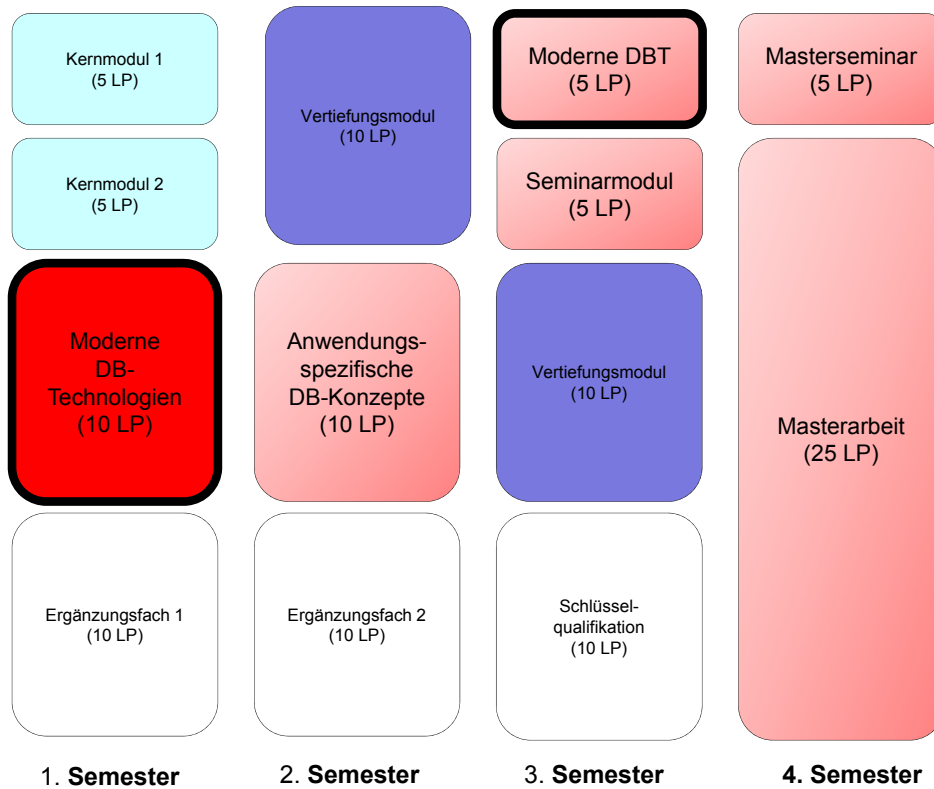
BACHELOR DBS-Profil



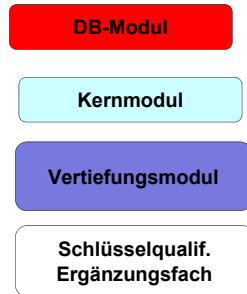
Legende:



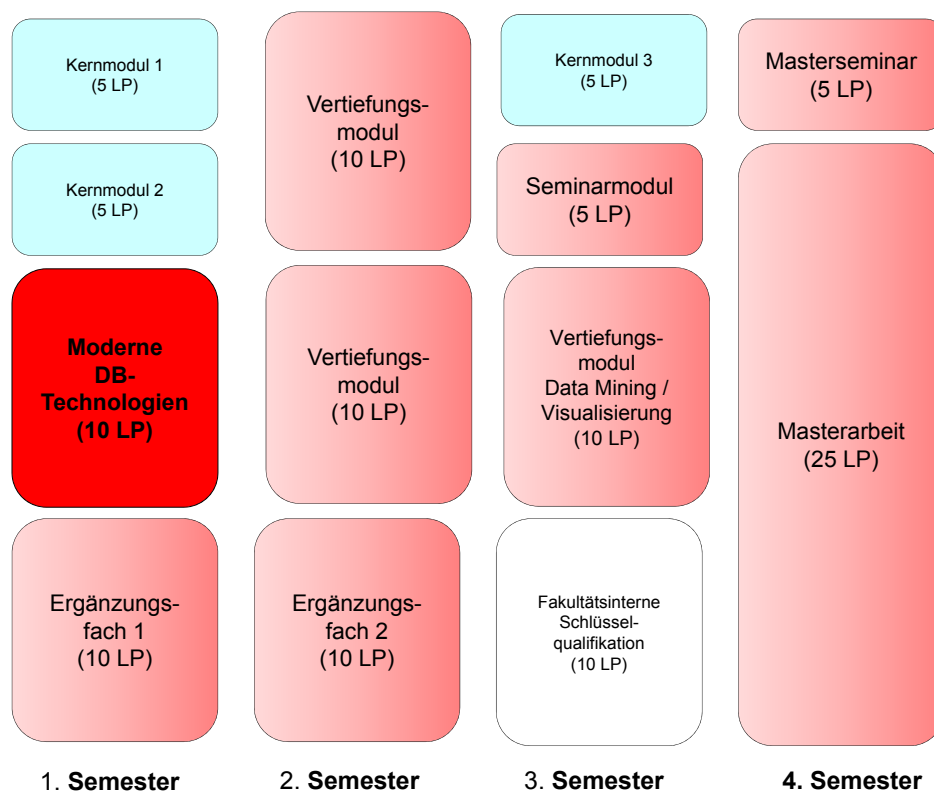
Informatik Masterstudium **DBS-Profil**



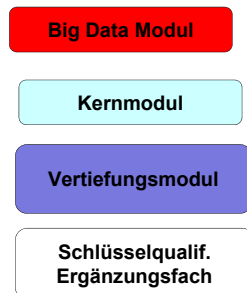
Legende:



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

- Fortgeschrittene Methoden des Information Retrieval
- Semantic Web
- Textdatenbanken (5 LP)
- Computational Advertising (5 LP)

■ Ergänzungsfächer

- Sequenzanalyse und Genomik
- Statistische Aspekte der Analyse molekularbiologischer und genetischer Daten
- Medizinische Bildaufnahme und Bildverarbeitung
- Graphen und biologische Netzwerke
- Alternativ: Module zu Digital Humanities












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 und/oder Computergrafik und/oder Datenbankpraktikum und/oder Textdatenbanken und/oder Linguistische Informatik und/oder Information Retrieval und/oder Grundlagen komplexer Systeme	Softwareengineering in frühen Phasen
3	Anwendungssysteme II	Moderne Datenbanktechnologien oder Integration und Architektur von Informationssystemen oder Softwaresystemfamilien	Wahlpflichtmodule aus den Masterstudiengängen Betriebswirtschaftslehre bzw. Volkswirtschaftslehre
4	Anwendungssysteme III oder 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	WS
	NoSQL-DB	Vertiefung	2	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 WS2016/17



Leistungsbewertung

- Prüfungsklausur von 60 Minuten
 - Feb. 2017
 - überprüft konzeptionelles Wissen + Anwendungsfälle
- Klausurerfolg durch
 - Vorlesungsteilnahme und –nachbearbeitung
 - Online-Übungen
 - Literatur

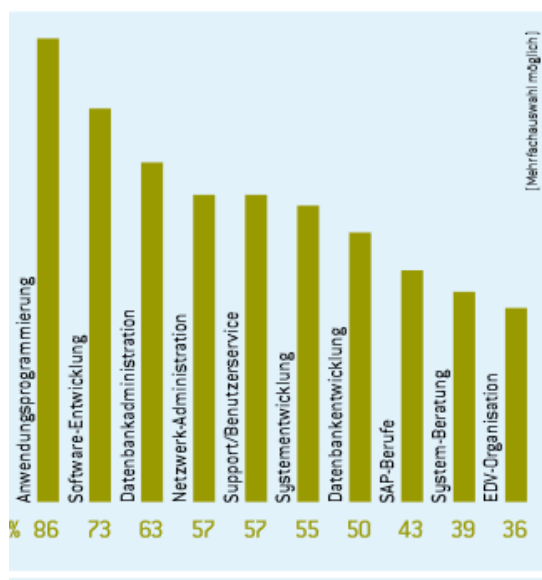


Lernziele der Vorlesung IDBS

- fundierte Kenntnisse der Funktionsweise von DBS
- Implementierungstechniken u.a. zur
 - Sicherstellung einer hohen Performanz
 - Datensicherheit
- **IDBS1**: Verfahren zur Externspeicher-Nutzung, Verwaltung von Pufferspeichern, Indexstrukturen, Anfrageoptimierung ...
- **IDBS2**: Verfahren zur Transaktionsverwaltung: Synchronisation (Concurrency Control), Logging/Archivierung, Recovery
- tiefgehende Kenntnisse wichtig für DB-Administration sowie generell für anspruchsvolle DB-Nutzung
- sachkundige Beurteilung von kommerziell verfügbaren DBS
- Verfahren nicht nur für Datenbanksysteme relevant (-> Big Data, Web-/ Applikations-Server, Datei-Management,...)



Umfrage: Wo setzen Unternehmen Informatiker ein?



Informatiker starten bei den Unternehmen am häufigsten in der **Anwendungsprogrammierung** (86 Prozent) und der **Software-Entwicklung** (73 Prozent). Ebenfalls wichtige Einsatzfelder für IT-Spezialisten in den befragten Unternehmen sind die Datenbank- und Netzwerk-Administration sowie der Bereich Support und Benutzerservice. Der Bereich Multimedia spielt als Einsatzfeld für Informatiker eine deutlich geringere Rolle.

staufenbiel
**JOB
TRENDS
2015**



Vorlesungsübersicht IDBS1

1. Einführung

- Architektur von DBS: Schichtenmodelle
- Tuning von DBS

2. E/A-Architekturen und Speicherhierarchien

- Speichertechnologien
- Disk-Arrays
- nicht-flüchtige Halbleiterspeicher

3. DBS-Pufferverwaltung

- Ersetzungsverfahren

4. Satzverwaltung

- Zuordnung Sätze – Seiten, Clusterung, BLOBs
- Satzadressierung
- Column Stores

Vorlesungsübersicht (2)

5. Indexstrukturen

- Mehrweg-Bäume
- externe Hash-Verfahren
- mehrdimensionale Zugriffspfade
- Bitlisten-Indizes
- Text-Indexierung

6. Algorithmen zur Implementierung relationaler Operationen

- Selektion
- Verbund / Join
- Sortierung

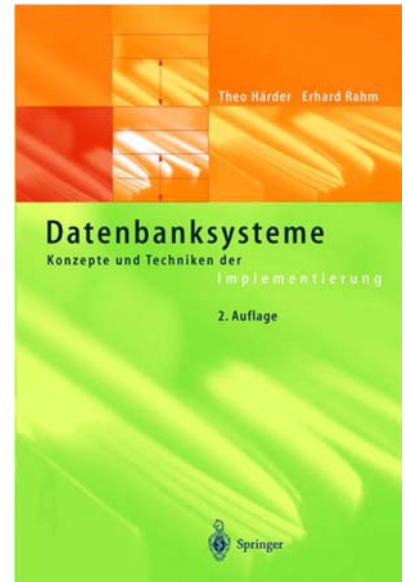
7. Anfrageoptimierung

- Übersetzung von DB-Anweisungen
- Optimierung
- EXPLAIN

Literatur

- Härder, T., Rahm, E.: *Datenbanksysteme - Konzepte und Techniken der Implementierung*. Springer-Verlag, 2. Auflage 2001 (Kap. 1 und 13 online)

<http://dbs.uni-leipzig.de/buecher/DBSI-Buch/inhalt.html>



- Weitere Lehrbücher

- Dittrich, J.: *Patterns in Data Management: A Flipped Textbook*. 2016
Website: Datenbankenlernen.de
- Saake, Sattler, Heuer: *Datenbanken: Implementierungstechniken*, MITP-Verlag, 2011
- Garcia-Molina, H., Ullman, J.D., Widom, J.: *Database System Implementation*. Prentice Hall, 2000

- Forschungsergebnisse aus Konferenzen / Zeitschriften

- DBLP-Portal: <http://dblp.uni-trier.de>
- Google Scholar <http://scholar.google.com/> (Volltexte, Zitierungsangaben)

Online-Übungen

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

LOTS 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.

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 HauptspeichergroÙe 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

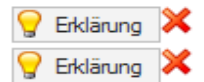


Aufgabe 4 (Hash-Join)

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

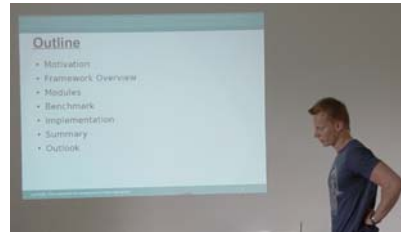


Lehrstuhl Datenbanksysteme

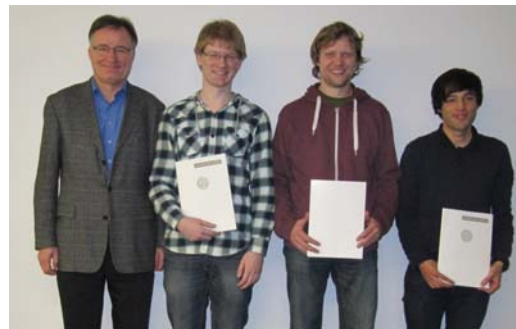
- seit 1994 am Institut für Informatik
- umfangreiches Lehrangebot
 - Vorlesungen, Praktika, Seminare
 - Online-Übungssystem LOTS
 - Eigene Lehrbücher: MRDBS, IDBS, ...

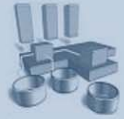


Oberseminare an Uni-Außenstelle Zingst – seit 2001



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





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](#)



Neue Publikationen

- [Holistic Entity Clustering for Linked Data](#)



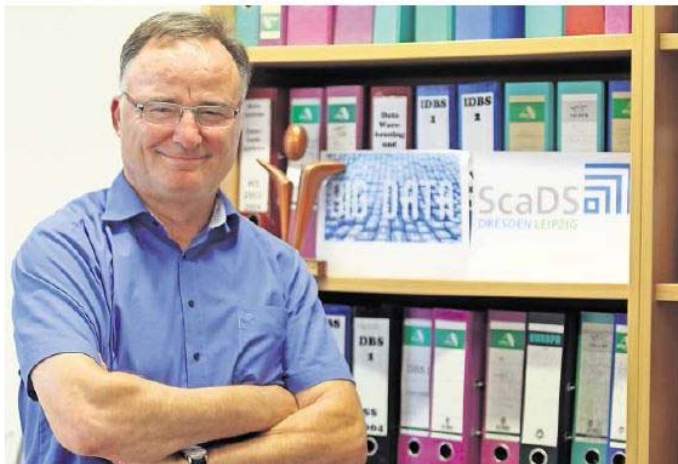
Galaxy und Sirius: Big-Data-Zentrum an der Uni mit mehr Computerpower

1,3 Millionen Euro für neue Server plus Zubehör / Professor Rahm: „Viel Performance durch Parallelverarbeitung“

VON MARIO BECK

Das technische Rückgrat des von der Leipziger Universität zusammen mit der TU Dresden betriebenen Big-Data-Kompetenzzentrums (Scads) wird weiter gestärkt. Zur bisher installierten Hardware kommt jetzt ein als Galaxy-Cluster bezeichneter Komplex mit 90 Servern. 60 davon sind an der Alma mater schon installiert und gehen nach und nach ans Netz. 30 Server sind für Dresden bestimmt, werden aber auch vom hiesigen Uni-Rechenzentrum (URZ) aus gesteuert. Über exakt 1080 Prozessoren verfügt das Galaxy-Gesamtsystem, die Festspeicherkapazität erreicht 2,16 Petabyte – ein Petabyte entspricht einer Million Gigabyte. „Das ist schon eine richtige Hausnummer“, sagt Erhard Rahm, der an der Uni die Professur für Datenbanken innehat und am Scads einer der zwei Koordinatoren ist.

Für dieses Jahr wird am URZ noch die Lieferung des Sirius-Clusters mit zwei Memory-Servern erwartet, die alleine in ihren Hauptspeichern je sechs Terabyte Daten bunkern können. Solche Mengen fallen beispielsweise bei der Erforschung von Genomen oder Krankheitsmarkern an. Beides sind Felder, in die sich die im Scads vereinten Experten unter anderem einbringen. Vor zwei Jahren war das Big-Data-Zentrum als Gemeinschaftsprojekt



Professor Erhard Rahm leitet seitens der Leipziger Uni das gemeinsam mit der TU Dresden betriebene Big-Data-Kompetenzzentrum.

Foto: André Kempner

der Unis in Leipzig und Dresden mit einem Fördervolumen von 5,6 Millionen Euro an den Start gegangen. Galaxy, Sirius und Zubehör schlagen jetzt mit rund 1,3 Millionen Euro zu Buche.

Zum Scads-Konsortium gehören noch das Leibniz-Institut für ökologische Raumentwicklung und das Max-Planck-Institut

für Zellbiologie und Genetik, die in Dresden ansässig sind. Hinzu gesellen sich eine Vielzahl assoziierter Partner aus Wissenschaft und Wirtschaft, darunter die Leipziger Data Virtuality GmbH und das Logistik-Netzwerk Leipzig-Halle.

Die neuen Galaxy-Server, die sowohl für die Grundlagen- als auch die ange-

wandte Forschung eingespannt werden, sind in einer Shared-Nothing-Architektur konfiguriert. Rahm: „Im Verbund liegt ihre Leistungsstärke. Die Performance kommt aus der Parallelverarbeitung.“ Aufgesetzt ist dabei meist eine Software-Technik namens Hadoop, auf deren Basis Rahm mit seinem Team ein Vorhaben mit der Kennung Gradoop vorantreibt. Es zielt auf die schnelle Analyse vernetzter Daten, die – wie bei Facebook – von hoher Beziehungsdynamik geprägt sind. Mit den Themen Datenintegration, Wissensextraktion und Visualisierung hat das Scads drei feste Standbeine und strahlt stark aus: In die Lebens- und Umweltwissenschaften, die Werkstoffkunde oder die digitalen Geisteswissenschaften. Der gute Ruf des Zentrums hat sich schon international herumgesprochen und sorgt für Echo.

Zur jüngsten Big-Data-Sommerschule an der Uni war der Andrang aus aller Welt groß. Rund 120 Teilnehmer kamen – und Rahm ließ wegen seiner Verpflichtungen bei der Summerschool eine Einladung in die USA sausen. Er war zum 25. Geburtstag des Microsoft-Research-Centers (MRC) gebeten worden, um dort den Outstanding Collaborator Award entgegenzunehmen. 32 Auserwählte aus aller Herren Länder, die in der Vergangenheit als Gastforscher am MRC gewirkt und dabei die nachhaltigsten Spuren hinterlassen hatten, bekamen die Ehrung.

