

IDB-PRA – Implementation of a Database Engine (4 PRA)

0434 L 468

Content:

In the project, students will implement components of a relational database system and get hands-on experience with a parallel data processing platform. The actual components implemented may vary each year, but will include parsing, query optimizer, execution engine, index structures and storage system. The global data volume is increasing dramatically each year. Understanding how to store, process and manage these huge amounts of data efficiently is a key requirement for software engineers and data analysts in the modern IT world. This project (following the corresponding lecture topics of IDB -Database Internals & Scalable Data Processing) will teach students both the fundamentals of data processing in traditional single-node database systems and how to scale out these techniques to huge amounts of data in large-scale, distributed environments. During this implementation project, in which students will get hands-on experience with important data processing techniques by implementing several components of a relational database system and by using parallel programming platforms like Apache Hadoop or Nephel/PACT.

Lectures are accompanied by exercises in small groups to practically rehearse the theory taught in the lectures. In the project, the students will be split in teams and under self-control will be implement some components of a database system, with the goal to have a running demonstrator at the end of the semester.

Target group:

This course addresses **master students** with a focus on database systems and information management after the first (master) term in "Informatik", "Technische Informatik", "Wirtschaftsingenieurwesen". (If capacity is available, it will be open also for other faculties).

Moreover, it is a **compulsory course for the ERASMUS MUNDUS programme IT4BI**.

(Wahlpflichtmodul im Masterstudiengang Informatik/ Studienschwerpunkt System Engineering, Technische Informatik/ Studienschwerpunkt Informationssysteme und im Masterstudiengang Wirtschaftsingenieurwesen (Studiengang IuK). Bei ausreichenden Kapazitäten auch als Wahlpflichtmodul in anderen Studiengängen oder Schwerpunkten wählbar.)

Prerequisite:

This course is the base course for master students with focus on database systems and information management and should be attended in the first semester of the master program. In contrast to the introduction of database systems (MPGI5/DBS), which looks database systems from an application programmers point of view, this class focuses on the internals of database systems. To participate, students are required to have successfully completed a Bachelor in computer science with a focus on database systems (participation in the Datenbankpraktikum, Datenbankprojekt). Knowledge of data modeling, relational algebra, and SQL as well as a very good command of Java, or possibly C/C++/C#, programming is required to participate in the course.

Registration:

Students are required to register via the DIMA course registration tool before the start of the first lecture (<http://www.dima.tu-berlin.de>). Within the first six weeks after commencement of the lecture, students will have to register for the course at **QISPOS (university examination protocol tool) and **ISIS (course organization tool)** in addition to the registration at the DIMA course registration tool.**

Contributions:

Prüfungsform: Portfolioprüfung

Benotung: benotet.

Studienleistung	Punkte
Presentation/Demonstration of the implementation project	30
Successful completion of the implementation project	70

Short Comment:

Das Modul kann in 1 Semester(n) abgeschlossen werden.

Das Modul ist auf 60 Teilnehmer begrenzt.

The module can be completed in 1 term.

The lab capacity limits this course to max. 60 participants.

Contact persons:

Jonas Traub, Dr. Sebastian Breß, Tuan Quang Dam