

AIM 3 Advanced Information Management III – Scalable Data Science: Systems and Methods (4 IV) (Integrierte Veranstaltung/ Integrated Course)

0434 L 472

Content:

The module will focus on mainstream distributed processing platforms and paradigms and learn how to employ these to solve challenging big data problems using popular data mining methods. Students will learn how to implement and employ varying data mining algorithms, such as Naïve Bayes, K-Means Clustering, and PageRank on varying open-source systems (e.g., Apache Hadoop, Apache Flink). The last decade was marked by the digitalization of virtually all aspects of modern society. Today, businesses, government institutions, as well as science and engineering organizations, among others face an avalanche of digital data on a daily basis. In order to derive insight from all of this data, society needs individuals with a strong foundation in scalable data science. In this course students will learn about popular scalable data analysis systems and scalable data analytics methods and gain practical experience in conducting scalable data science.

This Integrated Course (Integrierte Veranstaltung, IV) consists of: (i) lectures on key concepts, (ii) practical theoretical & programming exercises, and (iii) student lead presentations (including literature search).

Active participation and contributions to all parts of this course are essential.

Target group:

This course targets **Master's students** focused on Database Systems and Information Management in Computer Science (Major: System Engineering), Computer Engineering (Major: Information Systems & Software Engineering), and Industrial Engineering. Compulsory Elective module for ERASMUS MUNDUS IT4BI, plus Compulsory for EIT-ICT Data Science (DS) and Compulsory Elective for EIT-ICT Cloud Computing and Services (CCS) Subject to space availability, Master's students in other academic programs may also enroll and satisfy elective module requirements.

Wahlpflichtmodul im **Masterstudiengang** Informatik/Studienschwerpunkt System Engineering, Technische Informatik/Studienschwerpunkte Informationssysteme & Software Engineering und im Masterstudiengang Wirtschaftsingenieurwesen (Studiengang IuK). Wahlpflichtmodul im ERASMUS MUNDUS IT4BI, sowie für EIT-ICT Cloud Computing and Services (CCS), Pflicht für EIT-ICT Data Science (DS).

Je nach Verfügbarkeit der Plätze können auch Studierende anderer Fachrichtungen als Wahlpflicht das Modul belegen.

Prerequisite:

Computer science topics addressed in TU Berlin modules in the Bachelor's curriculum, particularly, the database course ("Information Systems and Data Analysis") or the equivalent, as well as good Java programming skills are required. Basic knowledge in linear algebra, numerical analysis, probability, and statistics are strongly recommended. Furthermore, it is preferable if students have already completed (or are currently enrolled in) a machine-learning course. Since the course will be offered in English, fluency in English is also required.

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**

The portfolio exam is comprised of three parts, namely:

1. Homework (30 points)
2. in-class presentations (20 portfolio points), and
3. a written final exam ("Testat") (50 portfolio points).

The final grade will be computed according to the grading table 2 of faculty IV, according to German law, § 47 (2) AllgStuPO TU Berlin.

Short Comment:

The module can be completed within 1 semester.

The lab capacity limits this course to max. 30 participants.

Since 2014, this module is offered each summer and winter term.

For each topic during this course additional research papers and reports will be used.

Contact persons:

Juan Soto, Christoph Boden, Dr. Kaustubh Beedkar