

ROC - Research Oriented Course on Data Science and Engineering Systems and Technologies (2+2 VL + UE)

0434 L 479

(Integrierte Veranstaltung/ Integrated Course)

Content:

The central focus of this module is on contemporary research methodology (CRM), data management technologies, and current research challenges. After an initial presentation on CRM, including scientific reading, writing, and presenting, in subsequent lectures, students will read about foundational data management methods/technologies and offer a presentation, which will then be followed by an instructor led presentation addressing related advanced topics. Topics of discussion, include data storage and indexing, specification and compilation of data analysis programs, query optimization and self-tuning, adaptive methods, processing data science pipelines as well as responsible data management. The course will also include a lab component, where students analyze and evaluate discussed methods, technologies, and settings in a methodical and scientific way. Big Data (BD) and Machine Learning (ML) are key drivers underlying the current wave of innovation in artificial intelligence and data science. Indeed, these drivers have had a profound impact on both the economy and the sciences. This course targets research-oriented students who aim to pursue a PhD in Big Data Management or Data Science and Engineering Systems and Technologies. Upon completion of this course, students will have learned about contemporary research methodology, gained both theoretical and practical skills in data management and big data technologies, and be attuned to today's major research challenges in scalable data management and processing.

The course is designed to principally impart technical skills (30%), method skills (40%), systems skills (10%), and social skills (20%). This Integrated Course (Integrierte Veranstaltung, IV) consists of: (i) lectures on key concepts, (ii) discussions, (iii) student lead presentations (including literature search), and (iv) a practical project element. Active participation and contributions to all parts of this course are essential.

Target group:

This course targets research-oriented **Bachelor's and Master's students** interested in focusing on Database Systems and Information Management in Computer Science (Major: System Engineering), Computer Engineering (Major: Information Systems and Software Engineering), and Industrial Engineering, as well as students pursuing the Data Science and Engineering Master's Track.

Prerequisite:

Desirable prerequisites for participation in the courses:

Computer science topics addressed in TU Berlin modules in the Bachelor's curriculum, particularly, both ISDA (Information Systems and Data Analysis) and DBPRA (Practical Database Systems Lab) or their equivalents, as well as good programming skills in C, Java, and SQL are all required. Additionally, an undergraduate course in linear algebra, probability, and statistics. This course will be offered in English. Thus, fluency in English is also required.

Registration:

Prior to the start of the first lecture, students must register in the DIMA Course Registration Tool: <http://www.dima.tu-berlin.de>. In addition, students must register both in **ISIS (the course organization tool) -and- **QISPOS (the TU Berlin Examination Management Tool)** within the first six weeks (**30.11.2019**) of the current semester.**

Contributions:

The portfolio exam (worth 100 points) is comprised of three parts:

(i) technology presentation (30 points), (ii) experimentation presentation (30 points), and (iii) a final report (40 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:

This module can be completed in one semester.

The lecture will take place in **EN 742**.

Contact persons:

Prof. Dr. Volker Markl