



Research Oriented Course (ROC) on Data Science and Engineering Systems and Technologies

Module title: Research Oriented Course (ROC) on Data Science and Engineering Systems and Technologies	Credits: 9	Responsible person: Markl, Volker
Website: https://www.dima.tu-berlin.de/menue/teaching/roc	Office: EN 7	Contact person: Markl, Volker
	Display language: Englisch	E-mail address: volker.markl@tu-berlin.de

Learning Outcomes

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, including scientific reading, writing, presenting, prototyping and experimental design, 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 (20%), method skills (40%), systems skills (20%), and social skills (20%).

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, presenting, prototyping and experimental design, 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.

In an accompanying lab component, students will prototype and evaluate discussed methods, technologies, and settings in a methodical and scientific way, and produce a scientific report on their findings.

Module Components

Course Name	Type	Number	Cycle	SWS
Research Oriented Course (ROC) on Data Science and Engineering Systems and Technologies	IV	0434 L 479	WS	6

Workload and Credit Points

Research Oriented Course (ROC) on Data Science and Engineering Systems and Technologies (Integrierte Veranstaltung)	Multiplier	Hours	Total
Plenary Sessions	15.0	4.0h	60.0h
Lab Course (Programming)	15.0	2.0h	30.0h
Lab Course (System Setup)	15.0	2.0h	30.0h
Preparation (including Reading, Literature Search, and Presentations)	15.0	2.0h	30.0h
Lab Course (Experimental Setup)	15.0	2.0h	30.0h
Report	15.0	2.0h	30.0h
Lab Course (Performance Evaluation)	15.0	4.0h	60.0h
			270.0h

The Workload of the module sums up to 270.0 Hours. Therefore the module contains 9 Credits.

Description of Teaching and Learning Methods

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 systems research project including (1) system setup, (2) prototyping, (3) experimental design, and (4) performance evaluation as well as (v) creating a presentation and report on the findings. Active participation and contributions to all parts of this course are essential.

Requirements for participation and examination

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. Knowledge of master's level coursework in database technology (DBT) and advanced information management (AIM) is necessary. This course will be offered in English. Thus,

fluency in English is also required.

Mandatory requirements for the module test application:

No information

Module completion

Grading:	Type of exam:	Language:
graded	Portfolio examination 100 points in total	English

Grading scale:

Note:	1.0	1.3	1.7	2.0	2.3	2.7	3.0	3.3	3.7	4.0
Punkte:	95.0	90.0	85.0	80.0	75.0	70.0	65.0	60.0	55.0	50.0

Test description:

The portfolio exam (worth 100 points) is comprised of four parts: (i) technology presentation (20 points), (ii) a quiz on database technology and research methodology (30 points), (iii) performance evaluation presentation, and (iv) a final report (30 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.

Test elements	Categorie	Points	Duration/Extent
Technology Presentation (Deliverable Assessment)	oral	20	30 min. / about 30 slides
Experimentation Presentation (Deliverable Assessment)	oral	20	30 min. / about 30 slides
Written Mid-term Test/Quiz (Examination)	written	30	max 75 minutes
Final Report (Deliverable Assessment)	written	30	about 12 Pages, conference style

Duration of the Module

This module can be completed in one semester.

Maximum Number of Participants

The maximum capacity of students is 8

Registration Procedures

Prior to the start of the first lecture, students must register themselves 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 of the current semester.

Recommended reading, Lecture notes

Lecture notes:

unavailable

Electronical lecture notes :

unavailable

Recommended literature:

Readings in Database Systems, 5th Edition, Peter Bailis, Joseph M. Hellerstein, Michael Stonebraker, editors, <http://www.redbook.io/>
Various Research Papers, made available during the first lecture

Raj Jain: The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling (Wiley Professional Computing), 1991

Supplementary reading material may be assigned to complement course lectures.

Assigned Degree Programs

This module is used in the following modulelists:

Computer Engineering (Master of Science)

StuPO 2015

Modullisten der Semester: WS 2020/21

Computer Science (Informatik) (Master of Science)

StuPO 2015

Modullisten der Semester: WS 2020/21

Elektrotechnik (Master of Science)

StuPO 2015

Modullisten der Semester: WS 2020/21

Informatik (Bachelor of Science)

StuPO 2015

Modullisten der Semester: WS 2020/21

Information Systems Management (Wirtschaftsinformatik) (Master of Science)

StuPO 2017

Modullisten der Semester: WS 2020/21

Miscellaneous

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.