

Advanced Information Management II – Management of Data Streams (4 IV) (Integrierte Veranstaltung/ Integrated Course)

0434 L 471

Content:

Through the technological advances in the last few years more and more applications are being created that constantly generate data which is only relevant for a certain time frame. Because of this, this type of application has to be able to handle various streams of data. You will gain conceptual, methodological and practical skills in the area of processing data streams, by using examples from various application areas.

In recent years, advances in hardware technology have facilitated new ways of collecting data continuously. In many applications such as for instance network monitoring, the volume of such data is so large that it may be impossible to store the data on disk. Furthermore, even when the data can be stored, the volume of the incoming data may be so large that it may be impossible to process any particular record more than once. Therefore, many database operations and data analysis algorithms such as for instance filtering, indexing, classification and clustering become significantly more challenging in this context.

The course has the following main topics:

- Basic conceptual understanding and terminology of data streams management, introduction to data streams, the difference to classical data management, examples (telephone networks, automotive electronics, avionics, medical, transport management, building monitoring, etc.)
- Basic concepts of technical information systems, modeling of data streams
- Data sources, requirements elicitation, requirements structuring, requirements of data stream management systems (DSMS)
- Reference architecture of a DSMS, architecture modeling
- Modeling of the functionality, logical architecture. Description on technical architecture, interface definition, behavior modeling
- Data streams processing: Windowing, The Sliding-Window Computation Model and Results
- Synopsis Construction in Data Streams (Sampling, Wavelets, Sketches and Histograms)
- Filtering, counting in data streams
- Data streams analysis: Classification & Clustering
- Data processing in sensor networks, resource utilization, transmission and transfer costs
- Modeling examples (automotive electronics, avionics). Prototype Systems (Aurora, STREAM, TelegraphCQ). Frameworks (Flink, Spark, Storm, Samza, SAMOA)

AIM-2 (Advanced Information Management 2 - Management of Data Streams) consists of 4 teaching elements:

- lectures, in parallel integrated phases of student presentations (individual work) and
- labs work (group projects plus home work), plus intensive study of literature.

Target Group:

Master students

Prerequisite:

BSc in Computer Science / Computer Engineering (or similar); good knowledge in 'Databases & Information Systems', 'Software Engineering', and Mathematics

Registration:

Before semester start, you must announce your interest to the "DIMA Anmeldetool" ('DIMA pre-semester registration tool')! (<http://anmeldung.dima.tu-berlin.de>)

Within six weeks (**30.11.2018**) after commencement of the class (but before the first result is to be delivered), you must officially register in **QISPOS (TUB examination protocol system')** in addition to the registration at the DIMA course registration tool.

Contributions:

The exam of this module consists of 4 portfolio elements:

3 'deliverable assessments' ("Ergebnisprüfungen"): seminar talk, seminar report, and project report plus one 'learning process review' ("Lernprozess-Evaluierung"): project (lab & home) work evaluation

Short Comment:

This module can be completed in one semester.

The maximum capacity of students is 30.

Contact persons:

Dr. Alexander Borusan

Wed 14 – 18 in E 020 (starts at Oct 17th 2018)

Wed 14 – 18 in E-N 719 (from Oct 24th 2018)

4 SWS/6 ECTS