Summary

This course is intended for students who want to understand modern large-scale data analysis systems and database systems. It covers a wide range of topics and technologies, and will prepare students to be able to build such systems as well as read and understand recent research publications.

Content

- Database systems
- Online analytics; data stream processing
- Column stores
- Decision support systems and data warehouses
- Large-scale data analytics infrastructure and systems
- Transaction processing. OLTP systems and concurrency control algorithms
- Distributed data management systems
- Query optimization; database tuning
- Logging and recovery
- Modern storage hierarchies

Learning Prerequisites

Required courses

- CS-322: Introduction to database systems
- CS-105: Introduction to object-oriented programming

Recommended courses

- CS-323: Operating systems
- CS-452: Foundations of software

Learning Outcomes

By the end of the course, the student must be able to:
• large databases
• Design big data analysis systems
• analysis algorithms
• Design big data analytics systems using state-of-the-art infrastructures for horizontal scaling, e.g., Spark
• Implement algorithms and data structures for streaming data analytics
• Decide between different storage models based on the offered optimizations enabled by each model and on the expected query workload
• Compare concurrency control algorithms, and algorithms for distributed data management
• Identify performance culprits, e.g., due to concurrency control

Teaching methods
Lectures, project, homework, exercises

Expected student activities
During the semester, the students are expected to:
• attend the lectures in order to ask questions and interact with the professor,
• attend the exercise sessions to solve and discuss exercises about the recently taught material,
• work on projects, which cover the practical side of the taught material,
• take a midterm,
• take a final exam,
• read scientific papers related to the course material

Assessment methods
• 70% exams
• 30% project

Supervision
Office hours: Yes
Others: Office hours on request. Q&A sessions in lectures and exercises.

Resources
Bibliography

Ressources en bibliothèque
• Database Management Systems / Ramakrishnan
• Readings in Database Systems / Hellerstein
• Mining of Massive Datasets / Rajaraman