GHENT UNIVERSITY

Course Specifications

Valid in the academic year 2023-2024

Big Data Science (E018210)

Course size *(nominal values; actual values may depend on programme)*

Credits 6.0 **Study time 180 h**

Course offerings and teaching methods in academic year 2023-2024

A (semester 2)	English	Gent	lecture	30.0h

Lecturers in academic year 2023-2024

Offered in the following programmes in 2023-2024	crdts	offering

Teaching languages

English

Keywords

Big Data platforms & architecture, interactive data visualizations, Semantic Web technologies, stream management, streaming algorithms, recommender systems, technology ethics & privacy

Position of the course

The main purpose of this course is to let the students gain hands-on experience with the most important concepts of End-to-End Big Data Science. They will learn how to manage, analyze and visualize Big Data and think about its societal impact.

Contents

AI Ethics
- Ethics
- Privacy & GDPR
- Societal Challenges

Big Data Systems
- Big Data Management Systems
- Batch vs. Stream AI
- Cloud vs. Edge AI

Big Data Analytics
- Algorithms on the web (PageRank, Adwords, ...)
- Machine Learning (Recommender Systems, Classification, Regression, ...)
- Scalable Data Mining (Dimensionality Reduction, Clustering, ...)

Knowledge Graphs
- Graph Theory
- FAIR, Open and Linked Data
- Decentralised & Federated Querying

Stream Processing
- Stream Management Systems
- Heterogeneous Stream Pre-processing
- Stream Mining

Interactive Data Visualizations
- Human Perception

- Design Principles & Interaction
- Dashboard Frameworks

Guest Lectures from Belgian Big Data companies

Initial competences

- Basic programming skills
  - In particular, experience with Python, Java, and JavaScript is advantageous (yet, with some extra effort, not a necessity)
- Elementary understanding about basic data formats
  - CSV, TSV, etc.
- Linear algebra, introductory course on statistics

Final competences

1 Understanding the possibilities and limitations of Big Data technology
2 Understanding the components of Big Data systems
3 Understanding the industry applications of Big Data
4 Combining Big Data components into a system architecture to meet specific product needs
5 Understanding the Big Data life cycle
6 Cleaning Big Data for production use
7 Visually and non-visually exploring Big Data
8 Creating interactive dashboards over Big Data
9 Setting up and performing scalable data mining
10 Applying machine learning algorithms to Big Data
11 Handle datasets with multiple challenging dimensions (size, format, quality, ...)
12 Dealing with high-velocity data via messaging and stream processing
13 Overcoming data heterogeneity through semantic technologies
14 Arguing the ethical and privacy aspects of large-scale data processing

Conditions for credit contract

Access to this course unit via a credit contract is determined after successful competences assessment

Conditions for exam contract

This course unit cannot be taken via an exam contract

Teaching methods

Lecture, Practical

Extra information on the teaching methods

Lectures, practicums, project

Learning materials and price

annotated slides, articles, and book chapters (freely available online), and for some lectures a syllabus by the lecturer will be provided.

References

Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, Jeffrey Ullman, ISBN: 978-1-107-07723-2

Course content-related study coaching

Assessment moments

end-of-term and continuous assessment

Examination methods in case of periodic assessment during the first examination period

Oral assessment

Examination methods in case of periodic assessment during the second examination period

Oral assessment

Examination methods in case of permanent assessment

Oral assessment, Skills test, Assignment

Possibilities of retake in case of permanent assessment

examination during the second examination period is possible in modified form

Extra information on the examination methods

- Periodical evaluation
  - Oral exam consists of 2 parts:
    - part I: Q&A with lecturer, with short preparation time (open book and open internet)
    - part II: open question where the student demonstrates his/her knowledge of the course on a new problem (open book and open internet, large time window to record answer)
- Non-periodical evaluation
  - graded lab session reports in groups
  - graded project reports, oral defense, and pitch deck.

Calculation of the examination mark

- The final grade consists of 3 parts: labs, project and oral exam
- The weight of the labs is 40%, for the project it is 20%, and for the oral exam it's 40%.
- The student needs a 7/20 score for the permanent evaluation (labs + project) as well as for the exam. If the student obtains less than 7/20 for one of the parts (permanent evaluation versus exam), the student can no longer obtain a pass mark for the course as a whole, in which case the final mark will be capped at 7/20.
- For the labs the students are graded per group, if there is an unfair division of effort the lecturer can opt for distinct marks.
- Fair project contributions are obtained by using peer assessment.