

Big Data Science (E018210)

Due to Covid 19, the education and evaluation methods may vary from the information displayed in the schedules and course details. Any changes will be communicated on Ufora.

Course size (nominal values; actual values may depend on programme)
Credits 6.0 Study time 180 h Contact hrs 60.0 h

Course offerings and teaching methods in academic year 2020-2021

A (semester 2)	English	Gent	lecture	30.0 h
			practicum	30.0 h

Lecturers in academic year 2020-2021

Mannens, Erik	TW06	lecturer-in-charge
Dimou, Anastasia	TW06	co-lecturer
Verborgh, Ruben	TW06	co-lecturer

Offered in the following programmes in 2020-2021

	crdts	offering
Bridging Programme Master of Science in Bioinformatics (main subject Engineering)	6	A
Bridging Programme Master of Science in Industrial Engineering and Operations Research	6	A
Bridging Programme Master of Science in Industrial Engineering and Operations Research	6	A
Master of Science in Business Engineering (main subject Data Analytics)	6	A
Master of Science in Bioinformatics (main subject Engineering)	6	A
Master of Science in Business Engineering (main subject Operations Management)	6	A
Master of Science in Industrial Engineering and Operations Research	6	A
Master of Science in Computer Science Engineering	6	A
Master of Science in Computer Science Engineering	6	A
Master of Science in Industrial Engineering and Operations Research	6	A
Exchange Programme in Computer Science (master's level)	6	A

Teaching languages

English

Keywords

Big Data platforms, interactive data visualizations, Semantic Web technologies, large-scale data mining, large-scale machine learning, 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, practicum

Extra information on the teaching methods

Lectures, practicums, project

Learning materials and price

annotated slides, articles, and book chapters (freely available online)

References

- Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, Jeffrey Ullman,

ISBN: 978-1-107-07723-2

- LAMBA Book

Course content-related study coaching

Evaluation methods

end-of-term evaluation and continuous assessment

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

Open book examination, oral examination

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

Open book examination, oral examination

Examination methods in case of permanent evaluation

Assignment, skills test, report

Possibilities of retake in case of permanent evaluation

examination during the second examination period is possible in modified form

Extra information on the examination methods

- Periodical evaluation
 - oral exam with written preparation, during which the Internet and all course materials can be accessed
- Non-periodical evaluation
 - graded lab session reports in groups
 - graded project reports

Calculation of the examination mark

The final grade is the average score of the two parts (exam and labs).

In case the score for one part is more than 7/20 but less than 10/20, the final grade is capped at 9/20.

In case the score for one part is 7/20 or less, the final grade is capped at 7/20.