GHENT UNIVERSITY

Course Specifications

Valid as from the academic year 2024-2025

Big Data Technology (E018240)

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

Credits 4.0 **Study time 120 h**

Course offerings and teaching methods in academic year 2024-2025

A (semester 1)	English	Gent	practical	20.0h
			lecture	20.0h

Lecturers in academic year 2024-2025

De Witte, Dieter	TWO6	lecturer-in-charge
Mannens, Erik	TWO6	co-lecturer

Offered in the following programmes in 2024-2025	**crdts**	**offering**
Bridging Programme Master of Science in Bioinformatics(main subject Engineering)	4	A
Master of Science in Bioinformatics(main subject Engineering)	4	A
Master of Science in Industrial Engineering and Operations Research(main subject Manufacturing and Supply Chain Engineering)	4	A
Master of Science in Industrial Engineering and Operations Research(main subject Sustainable Mobility Analytics)	4	A
Master of Science in Industrial Engineering and Operations Research(main subject Transport and Mobility Engineering)	4	A
Master of Science in Computer Science Engineering	4	A
Master of Science in Computer Science Engineering	4	A
Master of Science in Industrial Engineering and Operations Research	4	A

Teaching languages

English

Keywords

Big Data platforms & architecture, interactive data visualizations, knowledge graphs, FAIR data, Open Data, stream management, information retrieval, recommender systems.

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 Engineering. They will learn how to manage and visualize Big Data.

Contents

- Data collection & Open Data
  - Web scraping
  - Data formats
- Big Data Systems
  - Spark & Hadoop
  - MapReduce programming model
  - Cloud AI & pre-trained models
- Knowledge Graphs & FAIR
  - FAIR data principles
  - Linked Data
  - RDF databases
  - Data spaces
  - GDPR basics
- Data Visualization 101
  - Human Perception

    - Design principles
    - Interaction
    - Data journalism
  - Data processing architectures
    - Stream management system (messenging, processing)
    - Building blocks and principles of Big Data architectures
    - Lambda & Kappa architecture
    - Modern designs: Microservices, Data Mesh and Data Virtualization
  - Information retrieval
    - Inverted indexing
    - Query matching
    - Link Analysis
    - Image Retrieval
    - Impact of LLMs on retrieval
  - Recommender system
    - Neighborhood-based
    - Latent factor model
    - Evaluation

  - Guest Lectures from Belgian Big Data companies

Initial competences

- basic programming skills
  - Experience with Python (passed the course Informatics E015041 or an equivalent course)
  - Experience with Object Oriented Programming (passed the course Computer Programming E017210 or an equivalent course)
- 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 Handle datasets with multiple challenging dimensions (size, format, quality, ...)
10 Dealing with high-velocity data via messaging and stream processing
11 Overcoming data heterogeneity through semantic technologies

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

Study material

Type: Handbook

Name: Mining of Massive Datasets (3rd edition)
Indicative price: Free or paid by faculty
Optional: no
Language : English
Author : Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman
ISBN : 978-1-13905-845-2
Number of Pages : 315
Online Available : Yes
Available in the Library : No

Available through Student Association : No
Usability and Lifetime within the Course Unit : regularly
Usability and Lifetime within the Study Programme : one-time
Usability and Lifetime after the Study Programme : not

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.
  - examination during the second examination period is possible in modified form. The weight of the assignment will correspond to the workload for all labs and the project ~ 72 hours.

Calculation of the examination mark

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.

The weights for calculating the examination mark are as follows: 40% oral exam, 20% project, 40% labs.