

Course Specifications

Valid as from the academic year 2020-2021

Big Data (F000802)

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 45.0 h

Course offerings and teaching methods in academic year 2020-2021

A (semester 1)	English	Gent	teaching methods	hours
			seminar: practical PC room classes	37.5 h
			lecture	5.0 h
			group work	5.0 h
			online lecture	0.0 h
			online seminar: practical PC room classes	0.0 h

Lecturers in academic year 2020-2021

Van den Poel, Dirk EB23 lecturer-in-charge

Offered in the following programmes in 2020-2021

programme	crdts	offering
Master of Science in Teaching in Economics (main subject Business Economics)	6	A
Master of Science in Business Economics (main subject Accountancy)	6	A
Master of Science in Business Economics (main subject Corporate Finance)	6	A
Master of Science in Business Engineering (main subject Data Analytics)	6	A
Master of Science in Business Engineering (main subject Finance)	6	A
Master of Science in Business Economics (main subject Marketing)	6	A
Master of Science in Business Engineering (main subject Operations Management)	6	A
Exchange programme in Economics and Business Administration	6	A

Teaching languages

English

Keywords

Information Systems, Data Warehousing, SQL, Big Data, Apache Spark, Spark SQL, Python, PySpark, ML pipelines (Machine Learning).

Position of the course

The global objective of this course is to provide students with thorough theoretical as well as practical knowledge on the use and management of information. This knowledge can be of a strategic, a technical-analytical, as well as an organizational nature.

Contents

- 1 Importance of information management in general: which developments are at the basis of the increased importance of information use in marketing, which businesses or functions are marketing information intensive?
- 2 Data sources and data collection methods: which data sources are available to today's/tomorrow's data administrator, what is big data, how to deal with automated data collection methods such as scanning and internet? How to handle the nosql evolutions?
- 3 Building a marketing database: which principles are at the basis of building a good

- marketing database? How to build the structure (Entity Relationship Diagrams)?
- 4 Querying marketing databases: SQL (structured query language) programming language (in casu: Oracle SQL and Apache Spark/Spark SQL/Python) with exercises on large existing marketing information systems.
 - 5 Implementation/integration of MIS in the organisation: which traps are related to the process of implementing a MIS in the organisation, what are the principles of datawarehousing?

Each of these topics will be treated in-depth based upon a mixture of interactive class discussions, real-life cases.

In the big-data case NLP (natural language processing) and visualization will play a major role.

Initial competences

The "Data Mining" course.

Final competences

- 1 Understanding the structure and set-up of a database.
- 2 Mastering the programming language SQL (and NoSQL Big Data) to achieve optimal data and information management.
- 3 Building applications based on these data structures in order to make adequate conclusions for complex marketing problems.
- 4 Providing correct communication towards both technical and non-technical professionals.
- 5 Using a variety of external data sources (including new forms such as social media data) in an optimal way.
- 6 Applying Big Data algorithms through Spark ML Pipelines on case studies in order to build complex marketing decision models.
- 7 Validating own research results with existing literature of international top journals.
- 8 Executing a company case study in an international and interdisciplinary team that includes different levels of experience.
- 9 Presenting professionally about an advanced problem and its solution.

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

Group work, lecture, seminar: practical PC room classes, online lecture, online seminar: coached exercises, online seminar: practical PC room classes

Extra information on the teaching methods

Ex cathedra sessions as well as active class discussions of the different techniques and models with interactive exercises in the PC room.

Learning materials and price

- Oracle(TM) SQL course material
- Oracle Database 10g: Introduction to SQL
- Slides about Apache Spark/Spark SQL/Python & Big Data
- Cases
- Big Data book (free download of pdf): Learning Spark: Lightning-Fast Data Analytics, O'Reilly, 2nd Edition covering Apache Spark 3.0
- Journal article: van Dongen G. & Van den Poel D. (2020), Evaluating Stream Processing Frameworks, IEEE Transaction on Parallel and Distributed Systems

References

none

Course content-related study coaching

Numerous exercises are being solved during sessions. In addition, assignments (to be solved in teams) are handed out. Students will receive coaching in the process of solving the assignments and feedback afterwards (collectively, by team and individually). After a test about the programming language SQL, students will receive collective feedback.

Evaluation methods

continuous assessment

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

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

Examination methods in case of permanent evaluation

Written examination, oral examination, assignment

Possibilities of retake in case of permanent evaluation

examination during the second examination period is possible

Extra information on the examination methods

Written examination to investigate the extent to which students mastered the use of entity relationship diagrams, SQL query language. Team work using Apache Spark and the concept of Big Data.

Calculation of the examination mark

30% on SQL exercises/exam and 70% on Big Data/Apache Spark/Spark SQL/Python group assignment potentially adjusted by peer assessment. Of the 70% group assignment score 30% is based on individual questions and 40% is based on group performance.

To pass, a student should pass both parts of the evaluation. If a student does not pass for both parts and the score is 10/20 or more, the score will be reduced to 8/20.

Teams will be randomly selected during class to present their solutions to assignments. A maximum of two bonus points (of 20) can be earned in this way.