Table 5.2. Course specification

Study program: Advanced Data Analytics in Business

Course title: Managing, Storage and Visualising Big Data

Teachers: Mirko Savić, Laslo Šereš, Lukas Cironis

Status of the course: Obligatory

Number of ECTS: 7

Condition: None

Goal of the course

The goal of the course is to give students an in-depth understanding of a wide range of fundamental Big Data Management systems. In particular, this course focuses on the "variety" of the 3Vs in big data, where how to store, index and query various types of data in a real-world application. Moreover, this course provides knowledge needed for solving big data management problems, which include data cleaning, data integration, data update, query processing and to learn students how to use tools for visualising big data for business analytics.

Learning outcome

The basic outcomes concern: 1) Problem Solving: Ability to model and implement efficient big data solutions for various application areas using appropriately selected tools and architectures. 2) Critical Analysis: Ability to analyse big data infrastructures and their components, to compare and evaluate them, and make appropriate design choices when solving real-world problems. 3) Communication: Ability to motivate and explain trade-offs in big data platform design and analysis in written and oral form. Student is capable of visualising big data in an effective way and to draw conclusions in the context of business analytics.

Content of the course

Theoretical part

1. *Big data in business*
2. *Managament of big data*
3. *Data types*
4. *Data collecting*
5. *Data cleansing*
6. *Data storing*
7. *NoSQL Data Stores*
8. *Data indexing*
9. *Data querying*
10. *Introduction to visualisation, Getting Started with ggplot2*
11. *Toolbox, The Grammar*
12. *Layers, Scales, Axes and Legends, Positioning, Themes*
13. *Data Analysis, Data Transformation*
14. *Modelling for visualisation*
15. *Programming with ggplot2.*

Practical part

Work on case studies based on real-world data using computer laboratory.

Literature

1. Marr, B. (2017). Data Strategy: How to Profit from a World of Big Data, Analytics and the Internet of Things, Kogan Page, ISBN-13: 978-0749479855
2. Kleppmann, M. (2017). Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems, O'Reilly Media, ISBN-13: 978-1449373320
3. Sullivan, D. (2015). NoSQL for Mere Mortals, Addison-Wesley Professional, ISBN-13: 978-0134023212
4. Wexler, S. (2017). The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios, Wiley, ISBN-13: 978-1119282716
5. Wickham, Hardley (2016), ggplot2-Elegant Graphics for Data Analysis, second edition. Springer, Houston, USA.
6. Chang, W. (2013), R Graphics Cookbook. O'Reilly, Sebastopol, Canada.

Number of hours of active teaching	Theoretical teaching: 3	Practical teaching: 2

Teaching methods

All lectures are conducted in computer lab.

Assessment (maximum number of points 100)

Pre-exam obligations	Points	Final exam	Points
Activities during semester	5	Written exam	15

Practical part	5	Oral exam	15
Colloquium (2 colloquiums times 20 points)	40	....	
Seminar paper	20		