SYLLABUS

1. Information regarding the programme

1.1 Higher education institution	**Babes-Bolyai University**
1.2 Faculty	**Mathematics and Computer Science**
1.3 Department	**Computer Science**
1.4 Field of study	**Computer Science**
1.5 Study cycle	**Undergraduate (3rd year bachelor)**
1.6 Study programme / Qualification	

2. Information regarding the discipline

2.1 Name of the discipline	**Introduction to Big Data Analysis**						
2.2 Course coordinator	**Ciuciu Ioana**						
2.3 Seminar coordinator	**Ciuciu Ioana**						
2.4. Year of study	**3**	2.5 Semester	**5**	2.6. Type of evaluation	**E**	2.7 Type of discipline	**optional**

3. Total estimated time (hours/semester of didactic activities)

3.1 Hours per week	3	Of which: 3.2 course	2	3.3 seminar/laboratory	1
3.4 Total hours in the curriculum	42	Of which: 3.5 course	28	3.6 seminar/laboratory	14
Time allotment:					hours
Learning using manual, course support, bibliography, course notes					
Additional documentation (in libraries, on electronic platforms, field documentation)					
Preparation for seminars/labs, homework, papers, portfolios and essays					
Tutorship					
Evaluations					
Other activities:					-

3.7 Total individual study hours	
3.8 Total hours per semester	
3.9 Number of ECTS credits	4

4. Prerequisites (if necessary)

4.1. curriculum	
4.2. competencies	

5. Conditions (if necessary)

5.1. for the course	
5.2. for the seminar /lab activities	

6. Specific competencies acquired

Professional competencies	• Use of non-traditional databases for storing and processing large amounts of data • Advanced querying over distributed information resources • Evaluation, testing and validation with real-world data
Transversal competencies	• Methods and algorithms for data processing and analysis applied to Big Data • Multidisciplinary competencies spanning various application sectors (e.g., life sciences and bioinformatics, telco, media, finance, security, health, energy, etc.) • Data Science competencies, combining data analyst and data specialist- specific competencies (e.g., competencies from the fields of mathematics, statistics, information science, computer science, databases, machine learning, data mining, visualization, etc.)

7. Objectives of the discipline (outcome of the acquired competencies)

7.1 General objective of the discipline	• Handling (extremely) large amounts of digital data in various formats (text, video, financial, medical, etc.)
7.2 Specific objective of the discipline	• Enable the use of novel algorithms, software infrastructures and methodologies for the purpose of handling (store, retrieve, analyze) large amounts of data • Provide decision support over large volumes of data • Enable the creation of applications and services for various business domains based on the results of big data analysis.

8. Content

8.1 Course	Teaching methods	Remarks
• Introduction to Data Science and Big Data	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
• Programming tools for data science and data visualization	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
• Basics of NoSQL data management solutions	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
• Basic analytics	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
• Basic Machine Learning	• Interactive exposure	

	• Explanation • Conversation • Didactical demonstration	
• Basic Data Mining	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
• Big Data Architecture	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
• Big Data Analysis	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
• Overview of Apache Mahout, Spark, Storm & Shark	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
• Information Systems concepts	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
• Fundamentals of Data Visualization	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
• Introduction to Semantics and Linked Data for Data Science	• Interactive exposure • Explanation • Conversation • Didactical demonstration	

Bibliography

Frontiers in Big Data Analysis, The National Academies Press, Washington, prepublication draft
V. Agneeswaran, Big Data Analytics Beyond Hadoop, Pearson Education, 2014
T. White, Hadoop: The Definitive Guide, O'Reilly, 2009
D. Miner, *A. Sh*ook, MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems, O'Reilly, 2012
P. K. Janert, Data Analysis with Open Source Tools, O'Reilly, 2010
Q. E. McCallum, Bad Data Handbook: Cleaning Up The Data So You Can Get Back To Work, O'Reilly, 2012
O'Reilly Radar Team, Big Data Now: Current Perspectives from O'Reilly Radar, 2011
S.T. Allen, Storm Applied, 2015
M. Hamstra, Learning Spark, 2014
M. Barlow, Real-Time Big Data Analytics: Emerging Architecture, O'Reilly Media, 2013
J. Janssens, Data Science at the Command Line: Facing the Future with Time-Tested Tools, O'Reilly, 2014

T. Ojeda et al., Practical Data Science Cookbook, 2014
Data Science and Big Data Analytics, EMC Education Services, 2014
R. Morisson, Big Data Now, 2014
G. De Francisci Morales, Big Data and the Web: Algorithms for Data Intensive Scalable Computing IMT Institute for Advanced Studies, 2012
K Asanivik et al., The Landscape of Parallel Computing Research: A View from Berkeley, 2006
J. Dean, Big Data, Data Mining and Machine Learning: Value Creation for Business Leaders and Practitioners, Wiley, 2014
R. Glass and s. Callahan, The Big Data-Driven Business: How to Use Big Data to Win Customers, Beat Competitors, and Boost Profits, Wiley, 2014
D.L. Herben, Big Data, Big Analytics: Emerging Business Intelligence, 2014
A. M. Paganoni and P. Secchi, Advances in Complex Data Modeling and Computational Methods in Statistics, Springer, 2014

8.2 Seminar / laboratory	Teaching methods	Remarks
Semester project organized with groups of 3-4 students	• Team work • Individual work • Periodic meetings with the lab responsible • Periodic deliverables • Project groups will be monitored via a project wiki managed by the course/lab responsible	

Bibliography

http://mahout.apache.org/
http://www.tutorialspoint.com/mahout/mahout_introduction.htm
http://spark.apache.org/documentation.html
http://shark.cs.berkeley.edu/
http://spark.apache.org/
http://nosql-database.org/
https://www.mongodb.com/nosql-explained

9. Corroborating the content of the discipline with the expectations of the epistemic community, professional associations and representative employers within the field of the program

Synergies with various local and EU innitiatives: local industry, European Data Science Academy (EDSA), EU projects such as FERARI, LIFT, LOD2, Open Data Monitor, Data Publishing through the Cloud, Trendminder, Web Observatory, etc.

10. Evaluation

Project-based evaluation	• Project groups will present and demonstrate their semester project		
	•		
Written exam	• Evaluation of he theoretical aspects • Evaluation of the targeted competencies		

	via exercises		

Date	Signature of course coordinator	Signature of seminar coordinator
20 Nov. 2015	I. Ciuciu	I. Ciuciu

Date of approval	Signature of the head of department