SYLLABUS

1. Information regarding the programme

1.1 Higher education institution	**Babeş-Bolyai University of Cluj-Napoca**
1.2 Faculty	**Faculty of Mathematics and Computer Science**
1.3 Department	**Department of Computer Science**
1.4 Field of study	**Computer Science**
1.5 Study cycle	**Master**
1.6 Study programme / Qualification	**Software Engineering**

2. Information regarding the discipline

2.1 Name of the discipline (en) (ro)			**Big Data Processing and Applications**				
2.2 Course coordinator			**Lect. Dr. Ioana-Georgiana Ciuciu**				
2.3 Seminar coordinator			**Lect. Dr. Ioana-Georgiana Ciuciu**				
2.4. Year of study	**2**	2.5 Semester	**3**	2.6. Type of evaluation	**E**	2.7 Type of discipline	**Optional**
2.8 Code of the discipline		MME8158					

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

3.1 Hours per week	4	Of which: 3.2 course	2	3.3 seminar/laboratory	2
3.4 Total hours in the curriculum	56	Of which: 3.5 course	28	3.6 seminar/laboratory	28
Time allotment:					hours
Learning using manual, course support, bibliography, course notes					35
Additional documentation (in libraries, on electronic platforms, field documentation)					45
Preparation for seminars/labs, homework, papers, portfolios and essays					47
Tutorship					15
Evaluations					16
Other activities:					-

3.7 Total individual study hours	119
3.8 Total hours per semester	175
3.9 Number of ECTS credits	7

4. Prerequisites (if necessary)

4.1. curriculum	
4.2. competencies	• Basic knowledge of data analytics, preferably

	• Basic knowledge of data visualization, preferably • Programming skills

5. Conditions (if necessary)

5.1. for the course	• Room with video projector
5.2. for the seminar /lab activities	• Room with computers as needed; • Big Data software installed • High level programming language environment

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 • Learning to conduct incipient research in the field of Big 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 engineer- specific competencies (e.g., competencies from the fields of mathematics, statistics, information science, computer science, databases, machine learning, data mining, visualization, etc.) • Manifest responsible attitudes towards the scientific and didactic fields

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 processing (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
1. Introduction to Data Science and Big Data	Exposure, description, explanation, examples,	

	case studies	
2. Industrial Standards for Data Mining Projects	Exposure, description, explanation, examples, case studies	
3. Big Data Architecture	Exposure, description, explanation, examples, case studies	
4. Batch Layer	Exposure, description, explanation, examples, case studies	
5. Serving Layer - part I	Exposure, description, explanation, examples, case studies	
6. Serving Layer - part II	Exposure, description, explanation, examples, case studies	
7. Speed Layer - part I	Exposure, description, explanation, examples, case studies	
8. Speed Layer - part II	Exposure, description, explanation, examples, case studies	
9. Data Ingestion	Exposure, description, explanation, examples, case studies	
10. NoSQL Solutions for Big Data	Exposure, description, explanation, examples, case studies	
11. Data Visualization	Exposure, description, explanation, examples, case studies	
12. Big Data Case Studies	Exposure, description, explanation, examples, case studies	
13. Big Data Research Essays Presentation	Exposure, description, explanation, examples, case studies	
14. Big Data Research Essays Presentation	Exposure, description, explanation, examples, case studies	

Bibliography

N. Marz, J. Warren, Big Data. Principles and Best Practices of scalable real-time systems, Manning Publications, 2015
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. Shook, 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 about 2- 3 students (depending on the requirements and the equipment needed) Team work will be autonomous (focus on creativity and critical thinking) Technical tutorials will be provided to support student work around the most important aspects of Big Data storage and processing (e.g., Hadoop shell, PySpark, Data Ingestion with Apache Sqoop, NoSQL, etc.)	Research-informed Learning Tutorial-based Problem-solving approach Team work Big Data solutions for concrete problems and case studies	Groups will be monitored via a project wiki managed with the course/lab the responsible The lab takes place every two weeks and takes two hours
Bibliography Same as for the course		

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. Collaboration with the IT industry (e.g., Robert Bosch): invited lectures with real-life use cases, semester project topics, equipment (smart sensors).

10. Evaluation

Type of activity	10.1 Evaluation criteria	10.2 Evaluation methods	10.3 Share in the grade (%)
10.4 Course	- to be familiar with the main concepts of the domain	Written exam/ Evaluation of a research essay	50%
	- to be able to apply these principles in real-life use cases		
10.5 Seminar/lab activities	- to be able to propose viable creative solutions to real-life big data challenges - critical thinking - individual/team-based research work	Semester project	50%
10.6 Minimum performance standards			
➢ A minimum grade of 5 (on a scale from 1 to 10) is necessary for the written exam, the practical work and the research essay ➢ The lab attendance is compulsory at a rate of 90%, according to the decision of the Computer Science Department Council (http://www.cs.ubbcluj.ro/wp-content/uploads/Hotarare-CDI-15.03.2017.pdf)			

Date	Signature of course coordinator	Signature of seminar coordinator
...4 May 2020........	Lect. Dr. Ioana-Georgiana Ciuciu	Lect. Dr. Ioana-Georgiana Ciuciu

Date of approval	Signature of the head of department
....	Lect. Dr. Adrian Sterca