SYLLABUS

1. Information regarding the programme

1.1 Higher education institution	**Babeş Bolyai University**
1.2 Faculty	**Faculty of Biology and Geology**
1.3 Department	**Department of Molecular Biology and Biotechnology**
1.4 Field of study	**Biology**
1.5 Study cycle	**Master**
1.6 Study programme / Qualification	**Bioinformatics applied in life sciences**

2. Information regarding the discipline

2.1 Name of the discipline (en) (ro)			**Big Data Processing and Applications** **Procesare şi Aplicaţii Big Data**				
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	**Compulsory**
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					42
Additional documentation (in libraries, on electronic platforms, field documentation)					42
Preparation for seminars/labs, homework, papers, portfolios and essays					42
Tutorship					10
Evaluations					4
Other activities:					-

3.7 Total individual study hours	140
3.8 Total hours per semester	196
3.9 Number of ECTS credits	8

4. Prerequisites (if necessary)

4.1. curriculum	
4.2. competencies	• Basic knowledge of data analytics • Basic knowledge of data visualization • Programming skills

5. Conditions (if necessary)

5.1. for the course	• Room with video projector
5.2. for the seminar /lab activities	• Lab with computers • Big data software installed.

	• High level programming language environment

6. Specific competencies acquired

Professional competencies	C3.2 Identification and explanation of basic computer models appropriate for the application domain C3.3 Use of computer science and mathematical models and instruments to solve problems specific for the application domain C3.4 Analysis of data and models C5.3 Use of methodologies and database design environments for particular problems
Transversal competencies	CT1. Application of efficient work rules and responsible attitudes towards the scientific domain, for the creative exploitation of one's own potential according to the principles and rules of professional ethics CT2. Efficient conduct of activities organized in an interdisciplinary group and development of empathic capacity of interpersonal communication, networking and collaboration with diverse groups CT3. Use of efficient methods and techniques for learning, information, research and development of abilities for knowledge exploitation, for adapting to the needs of a dynamic society and for communication in a widely used foreign language.

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		
3. Big Data Architecture		
4. Batch Layer		
5. Serving Layer - part I		
6. Serving Layer - part II		
7. Speed Layer - part I		
8. Speed Layer - part II		
9. Data Ingestion		

10. NoSQL Solutions for Big Data		
11. Data Visualization		
12. Big Data Case Studies		
13. Big Data Research Essays Presentation		
14. Big Data Research Essays Presentation		

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 initiatives: local industry, European Data Science Academy (EDSA, https://edsa-project.eu/),EU projects such as LETHE (https://cordis.europa.eu/project/id/101017405), FARE (https://cordis.europa.eu/project/id/853566), the Human Brain Project (https://www.humanbrainproject.eu/en/), SoBigData (http://project.sobigdata.eu/), etc.

- Collaboration with the IT industry: invited lectures with real-life use cases, semester project topics, equipment (e.g., 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 -to be able to model a problem from a specific application field relying on emergent Big Data technologies - to be able to apply these principles in real-life use cases	Written exam / Evaluation of a research essay	50%
10.5 Seminar/lab activities	- to be able to propose viable creative solutions to real-life big data challenges from various application domains - to be able to consume (query, analyze)Big Data in order to derive information relevant to use cases from various application domains - to demonstrate critical thinking - to successfully perform individual and team-based tasks	Semester project	50%
10.6Minimum 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
19.06.2024	Lect. Dr. Ioana-Georgiana Ciuciu	Lect. Dr. Ioana-Georgiana Ciuciu

Date of approval	Signature of the head of department
19.06.2024	Conf. dr. Adrian Sterca