MD-2068, CHISINAU, 9/7 STUDENTILOR STR, PHONE: 022 50-99-63, www.utm.md

Big Data Technologies

1. Course/Module information

Faculty	Computers, Informatics, and Microelectronics				
Chair/department	Informatics and Systems Engineering				
Study cycle	CYCLE II, Master's Studies				
Study program	**Data Science**				
Year of study	**Semester**	**Evaluation type**	**Formative category**	**Optionality category**	**ECTS credits**
I	II	E	F	O	5

2. Estimated total time

Total hours in the curriculum	Including				
	Auditory hours		Individual work		
	Lecture	Practical work	Term paper	Study of theoretical material	Application development
150	20	20	---	110	-

3. Prerequisites for access to the course/module

According to the curriculum plan	The study is based on the knowledge accumulated by students in the following subjects: probability theory, mathematical statistics, programming methods, web technologies, and databases. Additionally, master's students should be familiar with computer architecture, at least with one operating system (Windows, Linux, MacOS), as well as some programming languages such as C#, C++, HTML, PHP, and Python.
According to competencies	Master's students must possess the following knowledge, competencies, and skills to master the academic discipline "Big Data Technologies": data analysis methods, application of modern Database Management Systems, including MySQL and MS SQL Server, multidimensional data modeling, modeling and application of Operational Databases and Data Warehouses, OLTP, OLAP, and BI technologies (e.g., VS, Power BI), as well as a certain level of data exploration using Data Mining tools. Additionally, master's students should be able to work in teams, effectively use ICT for solving applied professional problems, and apply mathematical modeling skills to socio-economic phenomena, environmental issues, technical fields, as well as domains such as finance and banking, education, healthcare, insurance, and marketing. The main outcomes of the "Big Data Technologies" course include applying knowledge, competencies, and skills in Data Engineering activities, as well as in the study of related disciplines such as "Data Analysis," "Data Science," "Big Data Analytics," and "Business Informatics," among others.

4. Conditions for conducting the educational process

Lecture	The theoretical material is presented in the lecture hall, combining the use of a projector and computer as well as a traditional whiteboard. Phone calls are not tolerated during the lecture. Throughout the course, students will take three generalizing quick tests on previously covered material, specifically after the 3rd, 5th, and final lecture sessions. The conditions for these tests are provided on the MOODLE platform.

Laboratory/seminar/practical	For completed tasks, students must prepare reports in digital format and, if required, in hard copy (printed), following the conditions specified in the methodological guidelines. The deadline for submitting laboratory or practical work, whether individual or group-based, is determined at the lecturer's discretion. For late submissions, students will be penalized by 1 (one) point per week of delay.

5. Specific competencies acquired

Professional competencies	**CPM1:** System architecture design and development **CPM2:** Monitoring technological trends. Innovation. Sustainable development **CPM4:** Staff development
Transversal competencies	**CTM1:** Autonomy and responsibility **CTM2:** Social interaction **CPM5:** Process improvement

6. Course/Module objectives

General objective	**The acquisition of knowledge, development of skills, practical abilities, and competencies in Data Engineering/Big Data Analytics aims to facilitate their application** to solving a wide range of challenges within the national economy. This is achieved through the simulation of processes and phenomena in fields relevant to future ICT specialists trained at the **Faculty of Computers, Informatics, and Microelectronics**, under the **Data Science** study program. ***Specifically, upon completing this course, master's students should be able to:*** • conduct studies and evaluations on the implementation and efficiency of Big Data technologies and tools within enterprises (Business Big Data Management); • implement and apply analytical and decision-support tools based on Big Data for Decision Management; • develop new models for an enterprise's information infrastructure, considering the capabilities of Big Data technologies (Model Management).
Specific objectives	•to acquire and gain new knowledge regarding the use of the technologies and tools underlying the "Big Data Technologies" course, necessary for activities in related fields such as data engineering, data science, and artificial intelligence; •to familiarize with the application of emerging ICT fields, such as artificial intelligence, Big Data analytics, cloud computing, augmented and virtual reality, in innovation and deep research activities on socio-economic phenomena, in order to be prepared to apply emerging technological trends in transforming research and production models, addressing current challenges; •to develop new capabilities for analyzing and using Big Data technologies, from the perspective of data utilization efficiency, when working on a concrete informational project related to the decision-making process efficiency of an entity, where Big Data is assumed to provide real support; •to develop new capabilities for applying distributed data processing principles and utilizing high-performance computing architectures in this context; •to develop the ability to adapt solutions based on data-driven approaches for specific problems within a particular field of applicability; •to develop the ability to plan and organize research activities in a field by organizing group work; to understand responsibilities and adhere to professional ethics, as well as data security and confidentiality rules. ***At the knowledge level:***

	• to acquire knowledge and competencies to use the technologies and tools derived from the essence and specifics of the "Big Data Technologies" course in data engineering, data science, and business activities; • to develop interdisciplinary competencies by using languages such as Python, platforms like Power BI, Visual Studio, cloud technologies, and cloud platforms such as AWS, Azure, Google Cloud, for Big Data analysis and analytics; • to develop competencies that provide an understanding of business architecture, the impact of introducing new IT technologies, including Big Data technologies, on enterprise management efficiency, and the transformation of the value chain; • to develop management competencies in the field of implementing Big Data systems and services based on Big Data technologies; • to develop and acquire new knowledge and competencies in the use of effective learning, information, and specialization methods and techniques; the ability to select and critically analyze bibliographic resources; the commitment to a continuous professional development plan through the ongoing updating of knowledge and skills in alignment with the scientific and technological advancements in the field. ***At the skills level:*** • to develop practical skills and abilities in the processes of using technologies and tools specific to the field of "Big Data Technologies" for data engineering, data science, and business activities; • to develop new capabilities to operate with fundamental concepts from mathematical, computer science, and econometric modeling, as well as from statistical analysis, thereby acquiring skills that are applied in practical research or production contexts; • to develop practical skills and acquire new competencies, both in data engineering (the practice of designing and building systems for collecting, storing, and analyzing large-scale data) and analytics (statistical modeling, supervised and unsupervised learning methods, machine learning and deep learning techniques), enabling future specialists to apply the accumulated knowledge to solve diverse problems in the national economy by simulating processes and phenomena in fields relevant to future ICT specialists; • to develop skills in using platforms and technologies specific to processing large volumes of data and implementing scalable applications; • to acquire mathematical and technological knowledge and skills for selecting, evaluating, analyzing, and using Big Data tools and technologies; • to acquire research knowledge and skills in big data analysis, stochastic optimization, predictive modeling, forecasting, enterprise data management, business analysis, and economic and mathematical modeling.

7. Course/Module content

Syllabus of teaching activities	Number of hours
Course topics	
TP1 **Introduction to Big Data:** definition of Big Data and the reasons for its emergence. Examples of business opportunities and Big Data. Characteristics of Big Data. Types of Big Data. The difference between Big Data and regular data. (distributed structure of Big Data). Managing and analyzing Big Data using microservices technology. Areas of application of Big Data. Advantages of applying Big Data.	**2**

TP2 **Big Data Technologies, TBD:** ***data collection*** (during the collection process, data is subject to cleaning); ***data storage*** (Hadoop, MongoDB, etc., DWH/ETL vs ELT, Data Lake (procedure: extraction, loading into the database/DWH, and transformation into a format suitable for current tasks), DBMS (relational/SQL, MS SQL Server, VS, Power BI, non-relational/NoSQL, MongoDB, Atlas/Compass MongoDB)); ***data processing*** (MapReduce technology, Hadoop, Apache Spark): ***data analysis/processing/analytics*** (BI technologies/Data analytics/Data Mining, Microsoft Azure).	**2**
TP3 **TBD and limitations of traditional Big Data processing:** Data integrity issue. Scalability challenge. ACID transactions. Data scaling and storage across multiple levels: CAP theorem; BASE theorem. NoSQL paradigm. Classification of NoSQL databases.	**2**
TP4 **TBD and high-performance computing:** distributed computing across multiple servers, the MapReduce computing paradigm. The Apache Hadoop project and its ecosystem. Apache Spark and its components.	**2**
TP5 **TBD and data repositories:** Data management. Security. Data quality (Data discovery, Data auditing, Data provenance, Data exploration). The difference between data lakes and data warehouses.	**2**
TP6 **TBD and Cloud Computing:** Cloud service providers - AWS (Amazon), Azure, Google Cloud. Features. Cloud service providers – Data Lake. Definition. Basic concepts of Data Lake. Data Lake architecture. Benefits and risks of using Data Lake in TBD.	**2**
TP7 **TBD and DWH + Data Lake or what is LakeHouse:** Components of LakeHouse (Delta Tables, Work Catalog). Delta Lake. Delta Lake architecture. How Delta Lake differs from traditional Big Data processing? Why is Delta Lake important? The need for Delta Lake. Azure Delta Lake. What is ETL and why is it important? And ELT? Differences. Uses. Hybrid Data Lake. Advantages. How do we utilize the advantages of Data Warehouse with Data Lake in TBD?	**2**
TP8 **TBD and Azure Databricks:** Introduction to Azure Databricks. Spark architecture. Data manipulation in Azure Databricks. Streaming data processing with Azure Databricks structured streaming. Integration of Azure Databricks with Azure Synapse. Azure Databricks Lakehouse ETL and TBD.	**2**
TP9 **TBD and Big Data Applications:** Big Data Analytics. Converting unstructured data into actionable data, knowledge, and wisdom. DWH in Big Data, Big Data management tools in Big Data Analytics.	**2**
TP10 **TBD and the NoSQL paradigm:** NoSQL Database Management System. NoSQL vs SQL — which type of database is more efficient for Big Data applications. Tools for storing Big Data.	**2**
Total lectures:	**20**

8. Content of the practical works

Syllabus of teaching activities	**Number of hours**
Laboratory works topics	
LP1. **Big Data Analytics – basic stages: Data** collection and integration; Data processing; Data cleaning and transformation; Data analysis. Using Big Data Analytics in: **Descriptive analysis** (information) ("what happened"); **Diagnostic analysis** ("why it happened"); **Predictive analysis** ("what will happen next"); **Prescriptive analysis** ("how to improve it");	**4**
LP2. **Big Data Analytics technologies and tools, including:** for storage, exploitation, analysis, and visualization / **Data storage and integration** (MongoDB, etc); **Data Mining** (VS, RapidMiner, ElasticSearch, etc): **Big Data Analytics** (Spark, lp. R, Python, etc); **Data visualization (**Excel, Tableau, MongoDB Charts, Plotly, etc**).**	**4**
LP3. **Solving challenges in Big Data analysis:** (Data storage and processing; Data quality; Data security; Correct tools and techniques) using Atlas MongoDB, Azure, and their numerous easy-to-use technologies. Big Data Analytics in BI and Data Science. Azure Synapse Analytics.	**4**
LP4. **Big Data Analytics applications in business:** (Risk management; Understanding market competitors; Identifying trends and patterns; Using Big Data analysis in retail; Creating new products; Managing human resources, in DSS, etc.)	**4**

LP5. **Complex analytical methods:** classification of analysis tasks: Text, Data, Web, Social Mining. Applying machine learning in analysis. K-means and C-means clustering, classification. Logistic regression, associations, Apriori algorithm. **Text analysis:** Search engines: Lucene, Solr, ElasticSearch. Word2Vec and Glove algorithms.	**4**
Total practical works:	**20**

9. Using generative AI

Permission to use	The use of generative AI in assignments and projects is permitted, provided that students adhere to the following rules: • Generative AI may be used to generate ideas, text structures, or code, but all generated materials must be reviewed and adjusted by the student to ensure that they meet academic requirements. • Any use of generative AI must be declared in the appendix section of each paper, using the phrase: "During the preparation of this paper, the author used [NAME OF TOOL / SERVICE] for the purpose of [REASON]. After using this tool / service, the author reviewed and edited the content as necessary and assumes full responsibility for the content of the paper."
Restrictions to use	Students ***MUSTN'T consider generative AI as a reliable source of information***, as it does not provide clear references or documented sources. • ***Direct citation of AI-generated content*** in academic papers as if it were a primary source ***isn't permitted.*** • Activities in which the use of **generative AI is prohibited** are specified by the teacher and are usually ***intermediate and final assessments*** or that don't involve professional competence development activities.

10. Bibliographic references

Main	1. **Michael Z. Zgurovsky Yuriy P. Zaychenko,** Big Data: Conceptual Analysis and Applications, SSN 2197-6503 ISSN 2197-6511 (electronic) Studies in Big Data ISBN 978-3-030-14297-1 ISBN 978-3-030-14298-8 (eBook) https://doi.org/10.1007/978-3-030-14298-8 Library of Congress Control Number: 2019933181 © Springer Nature Switzerland AG 2020 2. **Zeng Deze, Huan Huang, Rui Hou, Seungmin Rho, Naveen Chilamkurti,** Big Data Technologies and Applications, 10th EAI International Conference, BDTA 2020 and 13th EAI International Conference on Wireless Internet, WiCON 2020 Virtual Event, December 11, 2020. ISSN 1867-8211 ISSN 1867-822X (electronic) Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering ISBN 978-3-030-72801-4 ISBN 978-3-030-72802-1 (eBook) https://doi.org/10.1007/978-3-030-72802-1 3. **Sherif Sakr, Albert Y. Zomaya,** Encyclopedia of Big Data Technologies, ISBN 978-3-319-77524-1 ISBN 978-3-319-77525-8 (eBook) ISBN 978-3-319-77526-5 (print and electronic bundle), Library of Congress Control Number: 2018960889 © Springer Nature Switzerland AG 2019 https://doi.org/10.1007/978-3-319-77525-8 4. **S. Ejaz Ahmed,** Big and Complex Data Analysis Methodologies and Applications, SSN 1431-1968 Contributions to Statistics ISBN 978-3-319-41572-7 ISBN 978-3-319-41573-4 (eBook) DOI 10.1007/978-3-319-41573-4 Library of Congress Control Number: 2017930198 © Springer International Publishing AG 2017 5. **Big Data Analytics with Spark,** Copyright © 2015 by Mohammed Guller, ISBN-13 (pbk): 978-1-4842-0965-3, ISBN-13 (electronic): 978-1-4842-0964-6, www.apress.com/source-code/, OR www.it-ebooks.info 6. **Thomas Erl, Wajid Khattak and Paul Buhler, Big Data Fundamentals, Concepts, Drivers & Techniques,** Copyright © 2016 Arcitura Education Inc., ISBN-13: 978-0-13-429107-9 7. **Raul Estrada Isaac Ruiz*, Big Data SMACK*: *A Guide to Apache Spark, Mesos, Akka, Cassandra, and Kafka,*** Mexico City Mexico City, ISBN-13 (pbk): 978-1-4842-2174-7 ISBN-13

(electronic): 978-1-4842-2175-4, DOI 10.1007/978-1-4842-2175-4, Library of Congress Control Number: 2016954634, Copyright © 2016 by Raul Estrada and Isaac Ruiz

8. **Wes McKinney, Python for Data Analysis,** Copyright © 2013 Wes McKinney. All rights reserved. Printed in the United States of America. Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.
9. **John Wiley & Sons, Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data,** 10475 Crosspoint Boulevard, Indianapolis, IN 46256, www.wiley.com Copyright © 2015 by John Wiley & Sons, Inc., Indianapolis, Indiana, Published simultaneously in Canada ISBN: 978-1-118-87613-8, ISBN: 978-1-118-87622-0 (ebk) ISBN: 978-1-118-87605-3 (ebk)
10. **Peters Morgan, Data analysis from scratch with Python. Step by step guide ©** Copyright 2016 by AI Sciences LLC, all rights reserved. First Printing, 2016, Edited by Davies Company Ebook Converted and Cover by Pixels Studio Publised by AI Sciences LLC, ISBN-13: 978-1721942817, ISBN-10: 1721942815
11. **Building Big Data and Analytics Solutions in the Cloud** International Technical Support Organization December 2014
12. **Big Data Principles and Paradigms, Edited by Rajkumar Buyya The University of Melbourne and Manjrasoft Pty Ltd, Australia Rodrigo N. Calheiros** The University of Melbourne, Australia Amir Vahid Dastjerdi The University of Melbourne, Australia, Morgan Kaufmann is an imprint of Elsevier 50 Hampshire Street, 5th Floor, Cambridge, MA 02139, USA Copyright © 2016 Elsevier Inc. All rights reserved.496 pag
13. **10 Best Big Data Books in 2023 [Learn Big Data ASAP]**
https://azure.microsoft.com/en-us/products
https://realtoughcandy.com/best-big-data-books/
14. **20 Best Big Data Books for Beginners**
https://bookauthority.org/books/beginner-big-data-books
15. **Must Read Books for Beginners on Big Data, Hadoop and Apache Spark**
https://www.analyticsvidhya.com/blog/2015/10/books-big-data-hadoop-apache-spark/
https://dokumen.pub/big-data-concepts-technology-and-architecture-9781119701828.html
16. **MIcrosoft Aure for students**
https://engineering.buffalo.edu/computer-science-engineering/information-for-students/information-technology/software-distribution/microsoft-azure-for-students.html
https://learn.microsoft.com/en-us/azure/education-hub/azure-dev-tools-teaching/azure-students-program
https://azure.microsoft.com/en-us/resources/students?activetab=pivot:githubtab
https://learn.microsoft.com/en-us/azure/education-hub/azure-dev-tools-teaching/program-faq
https://www.techopedia.com/definition/26434/azure-service-platform
17. **Big Data Analytics - Statistical Methods.**
https://www.tutorialspoint.com/big_data_analytics/statistical_methods.htm
18. **Майер-Шенбергер, В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим** / Виктор Майер-Шенбергер, Кеннет Кукьер ; пер. с англ. Инны Гайдюк. — М. : Манн, Иванов и Фербер, 2014. — 240
19. **Гусева Е. Н. Теория вероятностей и математическая статистика** : Учебники и учебные пособия для ВУЗов [Электронный ресурс] - Москва : ФЛИНТА , 2016 - 220 – Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=83543
20. **Гутова С. Г., Алтемерова О. А. Теория вероятностей и математическая статистика:** Учебники и учебные пособия для ВУЗов [Электронный ресурс] - Кемерово : Кемеровский государственный университет , 2016 - 216 - Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=481538
21. **Кочетков Евгений Семенович. Теория вероятностей и математическая статистика:** Учебник [Электронный ресурс] : Форум , 2018 - 240 - Режим доступа: http://znanium.com/go.php?id=944923
22. **Миркин Б. Г. Введение в анализ данных.** Учебник и практикум[Электронный ресурс] : М.:Издательство Юрайт , 2019 - 174 - Режим доступа: https://biblioonline.ru/book/vvedenie-v-analiz-dannyh-432851

Supplementary	1. **Белько Иван Васильевич.** Теория вероятностей, математическая статистика, математическое программирование : Учебное пособие [Электронный ресурс] , 2016 - 299 - Режим доступа: http://znanium.com/go.php?id=542521 2. **Волкова Полина Андреевна.** Статистическая обработка данных в учебно-исследовательских работах : Учебное пособие [Электронный ресурс] : Форум , 2019 - 96 - Режим доступа: http://znanium.com/go.php?id=1030246 3. **Непомнящая Наталья Васильевна.** Статистика: общая теория статистики, экономическая статистика. Практикум : Учебное пособие [Электронный ресурс] , 2015 - 376 -Режим доступа: http://znanium.com/go.php?id=549841 4. **Основы теории вероятностей и математической статистики** [Электронный ресурс] , 2018 - 90 - Режим доступа: https://lib.rucont.ru/efd/684477

11. Evaluation

Periodic		**Current**	**Individual study**	**Project/thesis**	**Exam**
PE 1	**PE 2**				
15%	**15%**	**15%**	**15%**	**-**	**40%**

Minimum performance standards:

1. Attendance and activity in lectures and laboratory/practical/seminar work;
2. Achieving the minimum grade of "5" for each of the tests/assessments and individual or group practical works;
3. Demonstrating in the final examination paper the theoretical knowledge, competencies, and practical skills required for the use of Big Data Technologies in the development of a specific project, regarding the Course Final Paper.

12. Evaluation criteria

Activity	**Evaluation components**	**Evaluation method, Evaluation criteria**	**Weight in the final grade of the activity**	**Weight in the course evaluation**
Full-time education				
Periodic evaluation I	Theoretical content, topics 1-3	Test on MOODLE	100%	**15%**
Periodic evaluation II	Theoretical content, topics 4-5	Test on MOODLE	100%	**15%**
Current evaluation	Practical activity	Discussions during seminars Test on MOODLE	50%	**15%**
File completed with Reports for each Case Study under discussion			50%	**15%**
Individual study	Research on the topic **Course final project.**	Presentation/public speech	100%	
Final evaluation	Theoretical and practical content	Oral exam. **Course final project presentation.** Grading according to grading scale	100%	**40%**