ПРИЛОЖЕНИЕ 6

Faculty: FACULTY OF TELECOMMUNICATIONS AND MANAGEMENT
Department: INFORMATION TECHNOLOGIES
Professional Field: 5.3 COMMUNICATION AND COMPUTER TECHNOLOGY
Specialty: SOFTWARE DEVELOPMENT
Educational-qualification degree: PROFESSIONAL BACHELOR

LEARNING COURSE DESCRIPTION

1. Course title: *Large Data Sets Management Systems*
2. Course code: *P.17.042.2*
3. Type of the course: *elective*
4. Level: *Professional Bachelor*
5. Year of study: *third*
6. Semester: *fifth*
7. Number of credits: *4.5*
8. Name of the lecturer: *Assoc. Prof. Dr. Krasimira Ivanova*
9. Course learning outcomes:
The aim of the course is to give students a systematic theoretical and practical knowledge of the basic principles and patterns in modern models and standards of large data sets, as well as to build knowledge and skills for analysis and extraction of information from them.
Students who have completed their training in the discipline must acquire:
Basic knowledge of: the challenges of processing large data sets; Big Data and Big Data processing technologies; methods and tools for distributed storing and processing of large data sets; Apache Hadoop and related projects; MapReduce technology; Google's cloud platform.
Practical skills for: data processing methods; "supervised" and "unsupervised" learning methods; discretization and clustering; Big Data analysis methods; extracting knowledge from streaming data; interconnection analysis; working with large data sets with frequent sets of elements.
Competencies: After successful completion of the course students should be able to: know the basic tasks and principles of working with large data sets and distributed DBMS; use algorithms and tools to extract patterns from large data sets.
10. Teaching methods:
Teaching methods: the course "*Large Data Sets Management Systems*" is conducted in the fifth semester, according to the current curriculum through lectures grouped by one lesson per week for 15 weeks and practical exercises grouped by two lessons weekly for 15 weeks.
Project-based learning: From the lectures the students get acquainted sequentially with the provided material. The study material is illustrated with examples and applications that facilitate the work of students in the exercises. The practical exercises are conducted in a computer room. Before each practical lesson, students are informed about the need for preliminary preparation on the topic.
11. Prerequisites:
Students should have prior knowledge in the disciplines: Fundamentals of Programming, Engineering Mathematics 2, Fundamentals of Operating Systems, Databases, Programming Workshop, Operating Systems Workshop, Algorithms and Data Structures, Digital Economics.
Courses using the discipline are: Serverless computing, Geographic information systems.
12. Course contents:
The course focuses on: Basic data types. Big Data - definition, phases of the life cycle of Big Data. Big Data storing and management systems - Haddop, Cloudera, MongoDB, Talend. Software tools for processing Big Data according to their purpose and performed functions. Key features of MapReduce technology. Physical organization of MapReduce. Phases Map and Reduce. Google's cloud platform. Data pretreatment methods. Dimension reduction and selection of characteristics. “Supervised” learning. Construction of a classification tree structure. Bayesian networks. Sampling of streaming data. “Unsupervised” learning. Clustering. Basic clustering strategies. Hierarchical clustering in Euclidean and non-Euclidean spaces. Efficiency. Clustering using the K-MEANS and CURE algorithms. Big Data analysis methods. Detection of similar items. Search by the nearest

neighbor method. Measures for distances. Euclidean, Jacquard and Heming distances. Extracting knowledge from streaming data. Filtering. Estimates. Relationship analysis. Page rank. Park, Chen and Yu algorithm, multistage algorithm and hashing algorithm. Basic algorithms for data analysis with frequent sets of elements. Market basket model. Associative rules. Selection of a training set. Working with associative rules. A-Priori algorithm.

13. Bibliography:

Basic literature:

1. Oracle. Introduction to Very Large Databases. In: Oracle Database Online Documentation 11g Release 1 (11.1) / Data Warehousing and Business Intelligence. 2021.
2. Preston Zhang. Practical Guide to Large Database Migration. 2021 CRC Press, Taylor & Francis. ISBN 9780367780074.
3. Alex Gorelik. The Enterprise Big Data Lake: Delivering the Promise of Big Data and Data Science. 2019. O'Reilly Media, Inc., USA. ISBN 978-1-491-93155-4.
4. Arshdeep Bahga & Vijay Madisetti. Big Data Analytics: A Hands-On Approach. 2019. Published by Arshdeep Bahga & Vijay Madisetti. ISBN: 978-1-949978-00-1.
5. Francesco Corea. An Introduction to Data: Everything You Need to Know About AI, Big Data and Data Science. 2019. Springer Nature Switzerland AG. ISBN 2197-6503 ISSN 2197-6511 (electronic).

Additional literature:

1. John Arundel and Justin Domingus. Cloud Native DevOps with Kubernetes Building, Deploying, and Scaling Modern Applications in the Cloud. 2019. O'Reilly Media, Inc. ISBN-13 978-1-492-04076-7.
2. Mamta Mittal, Valentina E. Balas, Lalit Mohan Goyal, Raghvendra Kumar (Editors). Big Data Processing Using Spark in Cloud. 2019. Springer Nature Singapore Pte Ltd. ISBN 978-981-13-0549-8 ISBN 978-981-13-0550-4 (eBook).
3. Bill Chambers and Matei Zaharia. Spark: The Definitive Guide. Big Data Processing Made Simple. 2018. Databricks, O'Reilly Media, Inc. ISBN-13: 978-1-491-91221-8.
4. Pierluigi Riti. Pro DevOps with Google Cloud Platform With Docker, Jenkins, and Kubernetes. 2018. Springer Science + Business Media Finance Inc. ISBN-13 (pbk): 978-1-4842-3896-7 ISBN-13 (electronic): 978-1-4842-3897-4.
5. Subhashini Chellappan, Dharanitharan Ganesan. Practical Apache Spark Using the Scala API. 2018. Springer Science + Business Media Finance IncISBN-13 (pbk): 978-1-4842-3651-2 ISBN-13 (electronic): 978-1-4842-3652-9.

14. Assessment methods and criteria:

The course ends with an ongoing assessment.

The final assessment of the course is formed on the basis of ongoing assessment of successful solving of tasks from two tests and activity during lectures and practical exercises.

15. Language of instruction: *Bulgarian*