University of Gdańsk

Subject card

Subject name and code	Implementation of Big Data Solutions, PG_00128796		
Field of study	Informatics and Econometrics		
Date of commencement of studies	October 2023	Academic year of realisation of subject	2024/2025
Education level	Master's studies	Subject group	Obligatory subject group in the field of study
Mode of study	full-time studies	Mode of delivery	at the university
Year of study	2	Language of instruction	Polish
Semester of study	3	ECTS credits	6.0
Learning profile	academic	Assessment form	credit
Conducting unit	Department of Business Informatics -> Faculty of Management -> Rector		
Name and surname of lecturer (lecturers)	Subject supervisor	dr Jacek Maślankowski	
	Teachers	dr Jacek Maślankowski	

Lesson types

Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
Number of study hours	15.0	0.0	45.0	0.0	0.0	60

E-learning hours included: 0.0

Learning activity and number of study hours

Learning activity	Participation in didactic classes included in study plan	Participation in consultation hours	Self-study	SUM
Number of study hours	60	0.0	0.0	60

Subject objectives

Familiarizing students with a comprehensive approach to acquiring and processing large data sets.Preparing students to create Big Data solutions.

Learning outcomes			
	Course outcome	Subject outcome	Method of verification
	[IiEMU2_U04] The student is able to plan, design, and program information systems at an advanced level, supporting the operation of business entities.	designs web scraping scripts, creates NoSQL queries.	[SU2] presentation/project/paper/ report
	[IiEMU2_K01] The student understands the need for continuous completion and deepening of acquired knowledge. The student inspires and organizes others' learning processes.	adheres to the principles of teamwork in the process of implementing systems, adheres to the principles related to data quality, willingly takes on challenges related to changing user requirements.	[SK2] presentation/project/paper/ report
	[IiEMU2_U02] The student is able to proficiently acquire detailed information about economic processes and phenomena through direct observation, planned experimentation or database queries, as well as collect and process it using modern information technology tools.	designs web scraping scripts, creates NoSQL queries.	[SU2] presentation/project/paper/ report
	[IiEMU2_K02] The student can communicate freely with the public on specialised topics in the field of computer science and econometrics in and outside the workplace, communicate his knowledge and share his skills through various media. Culturally participates in discussions, is not afraid to ask questions and knows how to give constructive criticism.	adheres to the principles of teamwork in the process of implementing systems, adheres to the principles related to data quality, willingly takes on challenges related to changing user requirements.	[SK2] presentation/project/paper/ report
	[IiEMU2_W04] The student has an in-depth knowledge of advanced mathematical, statistical, econometric and IT methods that enable the acquisition, processing and analysis of data reflecting the functioning and growth of the national economy and its components, as well as the phenomena and processes occurring in their environment.	understands the purpose of creating and using Big Data systems, critically evaluates Big Data solutions.	[SW4] test/exam - oral or written
	[IiEMU2_W05] The student has an in-depth knowledge of socio-economic data sources, their databases and how to create them.	understands the purpose of creating and using Big Data systems, critically evaluates Big Data solutions.	[SW4] test/exam - oral or written
	[IiEMU2_K03] The student is able to communicate freely with the public inside and outside the workplace, transfer his knowledge and share his skills through various media.	adheres to the principles of teamwork in the process of implementing systems, adheres to the principles related to data quality, willingly takes on challenges related to changing user requirements.	[SK2] presentation/project/paper/ report

Subject contents	LectureIntroduction to Big Data, types of data, data division, classifications and technologiesWeb scraping techniques, generic and dedicated web scraping, legal conditions of web scrapingNoSQL databases - collections and documents - creating, saving and downloading dataOverview of Big Data analytical tools, libraries supporting data processingApache Hadoop ecosystemPractical application of Data Mining, Text Mining, Web MiningSupervised and unsupervised machine learningExercisesWeb scraping methods - using the Python language to automatically download data from the InternetMachine learning methods - supervised and unsupervised learning, using text and numeric setsText mining methods - automatic extraction of valuable information from text setsCollecting large data sets - NoSQL databases, saving the content of websites, creating and selecting queriesWorking with Open Data data, using APIsProcessing data from various file formats - JSON, CSV and XMLProcessing large sets data in Apache Hadoop and Apache Spark - PySpark application, MapReduce algorithms: WordCount analysis, HDFS - Hadoop Distributed File SystemApplications dedicated to web scrapingCase studies of Big Data solutions implementation
Prerequisites and co-requisites	Familiarizing students with a comprehensive approach to acquiring and processing large data sets.Preparing students to create Big Data solutions.

<table>
<tr><td>Assessment methods and criteria</td><td colspan="2"><table>
<tr><th>Subject passing criteria</th><th>Passing threshold</th><th>Percentage of the final grade</th></tr>
<tr><td>Exam - test</td><td>50.01%</td><td>30.0%</td></tr>
<tr><td>Student activity during classes</td><td>50.01%</td><td>10.0%</td></tr>
<tr><td>Project - Big Data solutions</td><td>50.01%</td><td>60.0%</td></tr>
</table></td></tr>
<tr><td rowspan="3">Recommended reading</td><td>Basic literature</td><td>Deitel P., Deitel H., Python dla programistów. Big Data i AI. Studia przypadków, Helion, 2020

Documentation Apache Hadoop and Spark: http://hadoop.apache.org, http://spark.apache.org; Python: http://python.org,

Materials on pe.ug.edu.pl</td></tr>
<tr><td>Supplementary literature</td><td>Glass, R., Callahan, S., (2015) The Big Data-Driven Business: How to Use Big Data to Win Customers, Beat Competitors, and Boost Profits, John Wiley & Sons

Mayer-Schonberger, V., Cukier, K., (2013) Big Data: A Revolution That Will Transform How We Live, Work, and Think, Eamon Dolan/Houghton Mifflin Harcourt</td></tr>
<tr><td>eResources addresses</td><td></td></tr>
<tr><td>Example issues/ example questions/ tasks being completed</td><td colspan="2">Test a test form in the form of open-ended and multiple choice questions, verifying knowledge of theoretical issues related to Big Data. Big Data system project, including collecting and processing large data sets. Student activity during classes points earned for correctly solving the given problem issues.</td></tr>
<tr><td>Work placement</td><td colspan="2">Not applicable</td></tr>
</table>

Document generated electronically. Does not require a seal or signature.