University of Gdańsk

Subject card

Subject name and code	Implementing Big Data Solutions, PG_00178741		
Field of study	Informatics and Econometrics		
Date of commencement of studies	October 2026	Academic year of realisation of subject	2027/2028
Education level	Master's studies	Subject group	Obligatory subject group in the field of study Optional subject group Subject group related to scientific research in the field of study
Mode of study	part-time studies	Mode of delivery	at the university
Year of study	2	Language of instruction	Polish
Semester of study	4	ECTS credits	5.0
Learning profile	academic	Assessment form	exam
Conducting unit	Department of Business Informatics -> Faculty of Management -> Rector		
Name and surname of lecturer (lecturers)	Subject supervisor	dr Patrycja Krauze-Maślankowska	
	Teachers		

Lesson types

Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
Number of study hours	8.0	0.0	24.0	0.0	0.0	32

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	32	2.0	91.0	125

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_U12] Can adapt, design, create, and operate IT systems that support business entities	is able to design web scraping scripts and open data retrieval	[SU2] presentation/project/paper/ report
[IiEMU2_U03] Is able to obtain and verify data from properly selected sources and to collect, process, and visualize it using modern econometrics, informatics or statistics tools	is able to use advanced forms of large-scale databases	[SU2] presentation/project/paper/ report
[IiEMU2_W05] Possesses advanced knowledge and understanding of informatics, statistics, and econometrics techniques and tools used to acquire, process, or visualise data to aid in decision-making and verify research hypotheses	understands the purpose of creating and using Big Data systems	[SW4] test/exam - oral or written

<table>
<tr><td>Subject contents</td><td>Lecture:

Introduction 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</td></tr>
<tr><td>Prerequisites and co-requisites</td><td>Familiarizing students with a comprehensive approach to acquiring and processing large data sets.Preparing students to create Big Data solutions.</td></tr>
<tr><td>Assessment methods and criteria</td><td>
<table>
<tr><th>Subject passing criteria</th><th>Passing threshold</th><th>Percentage of the final grade</th></tr>
<tr><td>Student activity during classes</td><td>51.0%</td><td>10.0%</td></tr>
<tr><td>Project - Big Data solutions</td><td>51.0%</td><td>60.0%</td></tr>
<tr><td>Exam - test</td><td>51.0%</td><td>30.0%</td></tr>
</table>
</td></tr>
<tr><td>Recommended reading</td><td>
<table>
<tr><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>
</table>
</td></tr>
<tr><td>Example issues/ example questions/ tasks being completed</td><td>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>Not applicable</td></tr>
</table>

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