

Subject card

Subject name and code	Implementing Big Data Solutions, PG_00177527						
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	full-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	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		4.0		61.0	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		
	[liEMU2_U11] Can collaborate effectively in teams and assume leadership roles		Is able to cooperate and manage team workflow during the design and implementation of Big Data systems, demonstrating the ability to coordinate tasks related to web scraping, data stream processing in Apache Spark/Hadoop, and storage in NoSQL databases.		[SU2] presentation/project/paper/report		
	[liEMU2_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		
	[liEMU2_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		

Subject contents	Lecture: Introduction to Big Data, types of data, data division, classifications and technologies Web scraping techniques, generic and dedicated web scraping, legal conditions of web scraping NoSQL databases - collections and documents - creating, saving and downloading data Overview of Big Data analytical tools, libraries supporting data processing Apache Hadoop ecosystem Practical application of Data Mining, Text Mining, Web Mining Supervised and unsupervised machine learning Exercises Web scraping methods - using the Python language to automatically download data from the Internet Machine learning methods - supervised and unsupervised learning, using text and numeric sets Text mining methods - automatic extraction of valuable information from text sets Collecting large data sets - NoSQL databases, saving the content of websites, creating and selecting queries Working with Open Data data, using APIs Processing data from various file formats - JSON, CSV and XML Processing large sets data in Apache Hadoop and Apache Spark - PySpark application, MapReduce algorithms: WordCount analysis, HDFS - Hadoop Distributed File System Applications dedicated to web scraping Case 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.		
Assessment methods and criteria		Subject passing criteria	Passing threshold
		Exam - test	51.0%
		Project - Big Data solutions	60.0%
		Student activity during classes	10.0%
Recommended reading	Basic literature		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
	Supplementary literature		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
	eResources addresses		
Example issues/ example questions/ tasks being completed	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.		
Work placement	Not applicable		

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