



Subject card

Subject name and code	Big Data processing frameworks, PG_00048043						
Field of study	Informatics						
Date of commencement of studies	February 2023	Academic year of realisation of subject			2022/2023		
Education level	second-cycle studies	Subject group			Optional subject group Subject group related to scientific research in the field of study		
Mode of study	Full-time studies	Mode of delivery			blended-learning		
Year of study	1	Language of instruction			Polish		
Semester of study	1	ECTS credits			4.0		
Learning profile	general academic profile	Assessment form			exam		
Conducting unit	Department of Software Engineering -> Faculty of Electronics, Telecommunications and Informatics						
Name and surname of lecturer (lecturers)	Subject supervisor	dr Adam Przybyłek					
	Teachers	dr Adam Przybyłek					
Lesson types and methods of instruction	Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
	Number of study hours	15.0	0.0	30.0	15.0	0.0	60
	E-learning hours included: 12.0						
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		8.0		32.0	100
Subject objectives	The aim of the course is to introduce students to the foundations of Big Data Systems. The course covers 3 frameworks for easily writing applications which process vast amounts of data in-parallel on large clusters of commodity hardware in a reliable, fault-tolerant manner.						
Learning outcomes	Course outcome	Subject outcome			Method of verification		
	[K7_W06] Knows and understands, to an increased extent, the basic processes taking place in the life cycle of devices, facilities and technical systems.	Students know how to use Apache Spark and Hadoop to process Big Data in parallel.			[SU4] Assessment of ability to use methods and tools		
	[K7_W03] Knows and understands, to an increased extent, the construction and operating principles of components and systems related to the field of study, including theories, methods and complex relationships between them and selected specific issues - appropriate for the curriculum.	Student knows how data and computation is distributed across a Hadoop cluster.			[SW1] Assessment of factual knowledge		
	[K7_W42] Knows and understands, to an increased extent, the principles and trends in the analysis and design of local and distributed IT systems and the basics of computer modeling and computerization of complex cognitive and decision-making processes.	A student who has completed the course can choose appropriate tools to solve a problem.			[SW2] Assessment of knowledge contained in presentation		
	[K7_K02] is ready to provide critical evaluation of received content and to acknowledge the importance of knowledge in solving cognitive and practical problems	Student is able to formulate a research problem, apply the appropriate methods, solve the problem and properly interpret the results. Student is also able to critically evaluate the results.			[SK5] Assessment of ability to solve problems that arise in practice		

Subject contents	<ol style="list-style-type: none"> <li>1. Introduction to Big Data and cloud computing</li> <li>2. Apache Hadoop</li> <li>3. Apache Storm</li> <li>4. Apache Spark</li> <li>5. Machine Learning</li> <li>6. Platforma Trusted Analytics</li> </ol>		
Prerequisites and co-requisites	Programming in Java and Python		
Assessment methods and criteria	Subject passing criteria	Passing threshold	Percentage of the final grade
	egzam	50.0%	40.0%
	project	50.0%	30.0%
	lab sessions	50.0%	30.0%
Recommended reading	Basic literature	<ol style="list-style-type: none"> <li>1. <a href="http://hadoop.apache.org/">http://hadoop.apache.org/</a></li> <li>2. <a href="http://storm.apache.org/">http://storm.apache.org/</a></li> <li>3. <a href="http://spark.apache.org/">http://spark.apache.org/</a></li> </ol>	
	Supplementary literature	<ol style="list-style-type: none"> <li>1. Hwang, K., Dongarra, J., Fox, G.: Distributed and Cloud Computing: From Parallel Processing to the Internet of Things. Morgan Kaufmann, 2011</li> <li>2. Karau, H., Konwinski, A., Wendell, P., Zaharia, M.: Learning Spark: Lightning-Fast Big Data Analysis. O'Reilly, 2015</li> <li>3. Erl, T., Puttini, R., Mahmood, Z.: Cloud Computing: Concepts, Technology, and Architecture. Prentice Hall, 2013</li> <li>4. Miner, D., Shook, A.: MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems. O'Reilly, 2012</li> </ol>	
	eResources addresses	Adresy na platformie eNauczanie:	
Example issues/ example questions/ tasks being completed			
Work placement	Not applicable		