



## Subject card

Subject name and code	Big Data processing frameworks, PG_00045325						
Field of study	Data Engineering						
Date of commencement of studies	October 2020	Academic year of realisation of subject			2022/2023		
Education level	first-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			at the university		
Year of study	3	Language of instruction			English		
Semester of study	6	ECTS credits			5.0		
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: 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		6.0		59.0	125
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		
	[K6_W05] Knows and understands programming models and evolution of related languages. Knows the methods of analysing and designing information systems and the modeling languages used in them, as well as the basic object-oriented programming platforms.		Student knows design patterns for distributed processing with MapReduce.		[SW2] Assessment of knowledge contained in presentation		
	[K6_U03] analyses problems and creates appropriate models, data structures and algorithms (including heuristic and numerical ones), assesses their computational complexity, estimates errors of the received solutions		Student is able to choose appropriate tools to solve a problem.		[SU2] Assessment of ability to analyse information		
	[K6_W07] Knows the methods of information processing, storage, extraction of data stored in various models including: relational, graph and document ones		Student knows how to use Apache Spark and Hadoop to process Big Data in parallel and how to use HDFS to store Big Data.		[SW2] Assessment of knowledge contained in presentation		
[K6_U06] Independently solves complex engineering tasks using literature, materials and devices, prepares extensive documentation of the developed solution using appropriate description techniques.		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.		[SU4] Assessment of ability to use methods and tools			
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> </ol>						
Prerequisites and co-requisites	Programming in Java and Python.						

Assessment methods and criteria	Subject passing criteria	Passing threshold	Percentage of the final grade
	lab exam	40.0%	40.0%
	exam	40.0%	30.0%
	project	40.0%	30.0%
Recommended reading	Basic literature	<ol style="list-style-type: none"> <li><a href="http://hadoop.apache.org/">http://hadoop.apache.org/</a></li> <li><a href="http://storm.apache.org/">http://storm.apache.org/</a></li> <li><a href="http://spark.apache.org/">http://spark.apache.org/</a></li> </ol>	
	Supplementary literature	<ol style="list-style-type: none"> <li>Hwang, K., Dongarra, J., Fox, G.: Distributed and Cloud Computing: From Parallel Processing to the Internet of Things. Morgan Kaufmann, 2011</li> <li>Karau, H., Konwinski, A., Wendell, P., Zaharia, M.: Learning Spark: Lightning-Fast Big Data Analysis. O'Reilly, 2015</li> <li>Erl, T., Puttini, R., Mahmood, Z.: Cloud Computing: Concepts, Technology, and Architecture. Prentice Hall, 2013</li> <li>Miner, D., Shook, A.: MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems. O'Reilly, 2012</li> </ol>	
	eResources addresses	Podstawowe <a href="https://enauczenie.pg.edu.pl/moodle/course/view.php?id=25440">https://enauczenie.pg.edu.pl/moodle/course/view.php?id=25440</a> - Supplementary materials on Moodle.	
Example issues/ example questions/ tasks being completed			
Work placement	Not applicable		