



Subject card

Subject name and code	Advanced nonrelational databases, PG_00045386						
Field of study	Data Engineering						
Date of commencement of studies	October 2021		Academic year of realisation of subject		2023/2024		
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		blended-learning		
Year of study	3		Language of instruction		English		
Semester of study	5		ECTS credits		6.0		
Learning profile	general academic profile		Assessment form		assessment		
Conducting unit	Department of Software Engineering -> Faculty of Electronics, Telecommunications and Informatics						
Name and surname of lecturer (lecturers)	Subject supervisor		dr inż. Teresa Zawadzka				
	Teachers		dr inż. Teresa Zawadzka dr inż. Wojciech Waloszek dr inż. Michał Zawadzki				
Lesson types and methods of instruction	Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
	Number of study hours	15.0	0.0	30.0	30.0	0.0	75
	E-learning hours included: 15.0						
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	75		5.0		70.0	150
Subject objectives	The aim of the course is to familiarize students with analytical data processing in a distributed Hadoop environment.						
Learning outcomes	Course outcome		Subject outcome		Method of verification		
	[K6_W07] Knows the methods of information processing, storage, extraction of data stored in various models including: relational, graph and document ones		The student knows the methods of processing analytical queries in the Hadoop environment.		[SW1] Assessment of factual knowledge		
	[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		The student analyzes the problem of building a data warehouse in a distributed environment and models this warehouse in an appropriate tool.		[SU4] Assessment of ability to use methods and tools [SU3] Assessment of ability to use knowledge gained from the subject [SU2] Assessment of ability to analyse information [SU1] Assessment of task fulfilment		
	[K6_W04] Knows the architecture of computers, operating system processes, file systems, text processing programs, disk and ram memories management rules. Knows the problems of sharing the state, presentation and transformation of information in a distributed system, hypermedia technologies and related services, the architecture of interactive distributed simulation and agent interaction methods.		The student knows the Hadoop file system and arhitecture of analytical tools for processing data in this system stored.		[SW1] Assessment of factual knowledge		

Subject contents	1. Introduction to the Hadoop system.		
	2. HIVE as a data warehouse.		
	3. HBase as a data warehouse.		
Prerequisites and co-requisites	Knowledge of relational databases.		
	Basics of the Linux system.		
Assessment methods and criteria	Subject passing criteria	Passing threshold	Percentage of the final grade
	Lecture	50.0%	33.0%
	Laboratories	50.0%	33.0%
	Project	50.0%	34.0%
Recommended reading	Basic literature	1. Apache Hive Cookbook, Hanish Bansal, Saurabh Chauhan, Shrey Mehrotra, Publisher: Packt Publishing, Release Date: April 2016 2. Apache Kylin, http://kylin.apache.org/docs/ 3. Apachy Hive, https://hive.apache.org/	
	Supplementary literature	Tools' documentation	
	eResources addresses	Adresy na platformie eNauczanie: Advanced Nonrelational Databases DE 2023/2024 - Moodle ID: 30892 https://enauczanie.pg.edu.pl/moodle/course/view.php?id=30892	
Example issues/ example questions/ tasks being completed	1. Build a data warehouse in the Hive tool		
	2. Set some tables as external.		
	3. Design at least two partitions.		
Work placement	Not applicable		