GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	Big Data processing frameworks, PG_00045325		
Field of study	Data Engineering		
Date of commencement of studies	October 2022	Academic year of realisation of subject	2024/2025
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	Polish
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_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 HDFSto store Big Data.	[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 ia able to choose appropriate tools to solve a problem.	[SU2] Assessment of ability to analyse information
[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_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

1. Introduction to Big Data and cloud computing
2. Apache Hadoop
3. Apache Storm
4. Apache Spark
5. Keras

Prerequisites and co-requisites

Programming in Java and Python.

<table>
<tr><td>Assessment methods and criteria</td><td colspan="2"><table>
<tr><th>Subject passing criteria</th><th>Passing threshold</th><th>Percentage of the final grade</th></tr>
<tr><td>lab exam</td><td>40.0%</td><td>40.0%</td></tr>
<tr><td>exam</td><td>40.0%</td><td>30.0%</td></tr>
<tr><td>project</td><td>40.0%</td><td>30.0%</td></tr>
</table></td></tr>
<tr><td rowspan="3">Recommended reading</td><td>Basic literature</td><td>1. http://hadoop.apache.org/
2. http://storm.apache.org/
3. http://spark.apache.org/</td></tr>
<tr><td>Supplementary literature</td><td>1. Hwang, K., Dongarra, J., Fox, G.: Distributed and Cloud Computing: From Parallel Processing to the Internet of Things. Morgan Kaufmann, 2011
2. Karau, H., Konwinski, A., Wendell, P., Zaharia, M.: Learning Spark: Lightning-Fast Big Data Analysis. O’Reilly, 2015
3. Erl, T., Puttini, R., Mahmood, Z.: Cloud Computing: Concepts, Technology, and Architecture. Prentice Hall, 2013
4. Miner, D., Shook, A.: MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems. O’Reilly, 2012</td></tr>
<tr><td>eResources addresses</td><td></td></tr>
<tr><td>Example issues/ example questions/ tasks being completed</td><td colspan="2"></td></tr>
<tr><td>Work placement</td><td colspan="2">Not applicable</td></tr>
</table>