GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	Techniques and Tools for Processing Big Data, PG_00063912		
Field of study	Informatics		
Date of commencement of studies	February 2025	Academic year of realisation of subject	2024/2025
Education level	second-cycle studies	Subject group	Optional subject group Specialty 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	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 Geoinformatics -> Faculty of Electronics, Telecommunications and Informatics		
Name and surname of lecturer (lecturers)	Subject supervisor	dr hab. inż. Emilia Lubecka	
	Teachers	dr hab. inż. Emilia Lubecka	

Lesson types and methods of instruction

Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
Number of study hours	15.0	0.0	15.0	15.0	0.0	45

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	45	8.0	47.0	100

Subject objectives

Theory and practice on large-scale data processing.

Learning outcomes

Course outcome	Subject outcome	Method of verification
[K7_U12] is able, to an increased extent, to analyze the operation of components and systems related to the field of study, as well as to measure their parameters and study their technical characteristics, and to plan and carry out experiments related to the field of study, including computer simulations, interpret the obtained results and draw conclusions	Student uses and processes large data sets.	[SU1] Assessment of task fulfilment
[K7_U07] can apply advanced methods of process and function support, specific to the field of study	Student is able do adequately process and export data for further analysis purposes in external programs.	[SU3] Assessment of ability to use knowledge gained from the subject
[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 acquaints with selected popular large-scale data processing tools.	[SW1] Assessment of factual knowledge
[K7_W08] knows and understands, to an increased extent, the fundamental dilemmas of modern civilisation, the main development trends of scientific disciplines relevant to the field of education	Student learns the current trends in computer science, particularly techniques and tools for large-scale data processing.	[SU2] Assessment of ability to analyse information

<table>
<tr><td>Subject contents</td><td>1. Architecture styles, including big data solution

2. Tools for large-scale data processing: Apache Hadoop and Spark

3. Scientific computation libraries for python: NumPy, SciPy

4. Clustering methods

5. High-performance computing (HPC)

6. Machine learning in large-scale data analysis

7. Code optimization and parallelization techniques (loops optimization, SIMD, openMP, MPI)</td></tr>
<tr><td>Prerequisites and co-requisites</td><td>Basic knowledge on python language.</td></tr>
<tr><td>Assessment methods and criteria</td><td>
<table>
<tr><th>Subject passing criteria</th><th>Passing threshold</th><th>Percentage of the final grade</th></tr>
<tr><td>Written exam</td><td>60.0%</td><td>40.0%</td></tr>
<tr><td>The task of semester</td><td>60.0%</td><td>30.0%</td></tr>
<tr><td>Laboratory exercises</td><td>60.0%</td><td>30.0%</td></tr>
<tr><td>The presence on lectures</td><td>0.0%</td><td>0.0%</td></tr>
</table>
</td></tr>
<tr><td>Recommended reading</td><td>
<table>
<tr><td>Basic literature</td><td>1. Big Data Demystified: How To Use Big Data, Data Science And Ai To Make Better Business Decisions And Gain Competitive Advantage, David Stephenson, Pearson, 2019.

2. "Big Data. Principles and best practices of scalable realtime data systems", Nathan Marz, James Warren, Simon and Schuster, 2015.

3. "Python for Programmers", Paul Deitel, Harvey Deitel, Pearson , 2019.</td></tr>
<tr><td>Supplementary literature</td><td>1. Hadoop framework documentation.

2. Spark framework documentation.</td></tr>
<tr><td>eResources addresses</td><td>Uzupełniające
Adresy na platformie eNauczanie:</td></tr>
</table>
</td></tr>
<tr><td>Example issues/ example questions/ tasks being completed</td><td>Sample question: How do you characterize Big Data?

Sample task: Implementation of program for processing and analyzing large data sets using Apache Spark platform.</td></tr>
<tr><td>Work placement</td><td>Not applicable</td></tr>
</table>

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