GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	Big Data processing frameworks, PG_00045325		
Field of study	Data Engineering, Data Engineering		
Date of commencement of studies	October 2026	Academic year of realisation of subject	2028/2029
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 -> Faculties of Gdańsk University of Technology		
Name and surname of lecturer (lecturers)	Subject supervisor	dr Adam Przybyłek	
	Teachers	dr Adam Przybyłek	

Lesson types

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_U05] develops innovative solutions for data analysis and processing, using appropriate methods and tools	Student knows how to use Apache Spark and Hadoop to process Big Data in parallel and how to use HDFSto store Big Data.	[SU1] Assessment of task fulfilment [SU4] Assessment of ability to use methods and tools
[K6_U04] formulates logical solutions to complex or unstructured problems	Student knows design patterns for distributed processing with MapReduce.	[SU3] Assessment of ability to use knowledge gained from the subject
[K6_W02] demonstrates advanced preparation in methods and techniques for formulating and solving 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.	[SW3] Assessment of knowledge contained in written work and projects

<table>
<tr><td>Subject contents</td><td>Course content – lecture
1. The Apache Hadoop Framework

2. The Hadoop Distributed File System

3. The Map/Reduce Paradigm

4. Introduction to Big Data and Cloud Computing

5. The Apache Spark Framework

6. The Apache Storm Framework

7. Process Mining
Course content – laboratory
Hadoop Labs:
Lab 1,2: MapReduce fundamentals
Lab 3-5: Advanced MapReduce analytics
Lab 6,7: Iterative algorithms in MapReduce

Apache Spark Labs:
Lab 8,9: Spark RDDs
Lab 10,11: Spark DataFrames and Spark SQL
Lab 12,13: Spark MLlib
Lab 14: Spark Structured Streaming
Course content – project
Hands-on exercises and guided tutorials covering practical skills in big data technologies, data engineering, and data science workflows</td></tr>
<tr><td>Prerequisites and co-requisites</td><td>Programming in Java and Python.</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>labs</td><td>40.0%</td><td>40.0%</td></tr>
<tr><td>lecture</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>Recommended reading</td><td>
<table>
<tr><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. OReilly, 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. OReilly, 2012</td></tr>
<tr><td>eResources addresses</td><td></td></tr>
</table>
</td></tr>
<tr><td>Example issues/ example questions/ tasks being completed</td><td></td></tr>
<tr><td>Practical activites within the subject</td><td>Not applicable</td></tr>
</table>

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