

FACULTY OF FUNDAMENTAL PROBLEMS OF TECHNOLOGY	
SUBJECT CARD	
Course name in Polish	Algorytmy Big Data
Course name in English	Big Data Algorithms
Course language	English
Field of study	Big Data Analytics
Degree of study	II degree
Type of course (obligatory)	
Educational effects according to Act of Senate no. 505/23/2016-2020: BDA2-W01, W02, W03, W04, BDA2-U01, U03, U06, BDA-K01, K02, K03	
Subject code	

	Lecture	Laboratory	Classes
Number of hours of organized classes in University (ZZU)	30	15	30
Number of hours of total student workload (CNPS)	60	60	60
Form of crediting		solving programming tasks	Final test
For a group of courses final course mark	X		
Number of ECTS points	2	2	2
including number of ECTS points for practical (P) classes	1	2	2
including number of ECTS points for direct teacher-student contact (BK) classes	1	1	1

PREREQUISITES RELATING TO KNOWLEDGE, SKILLS AND OTHER COMPETENCES
Basic knowledge of programming in Spark and advanced topics of linear algebra

SUBJECT OBJECTIVES	
C1	Understanding model MapReduce
C1	Knowledge of tools for analysis of Big Data
C2	Understanding the notion of random graph

SUBJECT EDUCATIONAL EFFECTS

Relating to knowledge:

PEK_W01 – knows the MapReduce paradigm (rel. to BDA2_W04)

PEK_W02 – knows the notion of Page Rank (rel. to BDA2_W02)

PEK_W03 – knows the notion of random graph and social networks (rel. to BDA2_W01, W04)

PEK_W04 – knows the notion of frequent items (rel. to BDA2_W03)

Relating to skills:

PEK_U01 – can apply MapReduce paradigm for programming (rel. to BDA2_U01, BDA2_U03)

PEK_U02 – knows tools for Big Data Analysis (rel. to BDA2_U06)

PEK_U03 – can find frequent items (rel. to BDA2_U06)

Relating to social competences:

PEK_K01 – can extract and transform very big collections of data (rel. to BDA2_K01, K02, K03)

PROGRAM CONTENTS

PROGRAM CONTENTS		
Form of classes – lecture		Number of hours
L 1	Word – count problem. Introduction to Big Data	2
L 2	Model Map Reduce	6
L 3	Matrix operations in Map Reduce	2
L 4	Link Analysis and Google page rank	4
L 5	Frequent Itemsets	4
L 6	Programming with Spark	4
L 7	Data stream with Spark	2
L 8	Random graphs	2
L 8	Properties of social networks	2
L 8	Graph processing with Spark	2
		30
Form of classes - classes		Numer of hours
Cl 1	Computational complexity	6
Cl 2	Map Reduce	4
Cl 3	Algorithms on graphs	6
Cl 4	Graphs and second eigenvalue	2
Cl 5	Random graphs	6
Cl 6	Parameters of graphs	4
Cl 7	Final test	2
	Total hours	30

	Form of classes - laboratory	Number of hours
Lab 1	Word – count problem	2
Lab 2	MapReduce with Scala	3
Lab 3	Spark – basic applications	2
Lab 4	Spark – big data transformations	3
Lab 5	Spark – recursion	2
Lab 7	Frequent items	3
	Total hours	15

TEACHING TOOLS USED	
N1	Lecture using board and computer presentations
N2	Solving exercises with students
N3	Solving programming tasks

EVALUATION OF ACHIEVED SUBJECT EDUCATIONAL EFFECTS		
Evaluation: F – forming (partial) C – concluding	Educational effect number	Way of evaluating achievement of educational effects
F1	PEK_W01-W04 PEK_K01	Test
F2	PEK_U01-U03	Solutions of programming tasks
C=F1+F2		

PRIMARY AND SECONDARY LITERATURE
<u>PRIMARY LITERATURE:</u> [1] J. Leskovec, A. Rajaraman, J. D. Ullman, <i>Mining of Massive Datasets</i>, book.pdf, 2016 [2] Tyler Akidau, Slava Chernyak, Reuven Lax, <i>Streaming Systems. The What, Where, When, and How of Large-Scale Data Processing</i>, O'Reilly Media, July 2018 <u>SECONDARY LITERATURE:</u> [1] Martin Odersky, <i>Programming in Scala</i>, Artima Press, 2016 [2] Misra, J., Gries, David, <i>Finding repeated elements</i>, Science of Computer Programming, 2 (2): 143–152.

SUBJECT SUPERVISOR
Prof. dr hab. Jacek Cichoń, jacek.cichon@pwr.edu.pl Dr Jakub Lemiesz, jakub.lemiesz@pwr.edu.pl

