

Course title:	Processing of Big Data using Apache Spark
Institute/Division:	F-3, Institute of Computer Science
Number of contact hours:	45 hours
Course duration:	1 semester
ECTS credits:	6
Course description:	Apache Spark is an open-source, fast and general-purpose cluster computing system. It provides high-level APIs in Java, Scala, Python and R and an optimized engine that supports general execution graphs. The students will learn to use some of the available in the Spark ecosystem rich set of higher-level tools including Spark SQL for SQL and structured data processing, MLlib for machine learning, GraphX for graph processing, and Spark Streaming.
Literature:	<ol style="list-style-type: none"> 1. https://spark.apache.org/docs/latest/ 2. http://docs.h2o.ai/h2o/latest-stable/h2o-docs/welcome.html 3. Many additional materials and links will be given to all students during the lectures.
Course type:	Lectures and exercises
Assessment method:	There will be few homeworks. Each one should be returned to the teacher up to 2 weeks after distribution).
Prerequisites:	None
Primary target group:	computer science students of the 3rd or 4th year
Lecturer:	Barbara Borowik, PhD
Contact person:	Barbara Borowik, PhD; bborowik@pk.edu.pl
Deadline for application:	15th of January