Politechnika Krakowska
Wydział Informatyki
i Matematyki

Course title	Processing of Big Data using Apache Spark
Institute/Division	Faculty of Computer Science and Mathematics/ Department of Computer Science
Course code	F-1.BDAS
Erasmus subject code*	11.3 Informatics, Computer Science
Number of contact hours**	45 lecture hours (45h)
Course duration	1 semester (Spring)
ECTS credits	6
Course description (max 100 words)	Course description: Apache Spark is a unified analytics, open-source, fast, multi-language and general-purpose cluster computing system engine for large-scale data processing for executing data engineering, data science, and machine learning on clusters or on single-node machines. It provides high-level APIs in Java, Scala, R, Python and SQL. The students will learn to use some of the available in the Spark ecosystem rich set of higher-level tools and commands and will use Spark application library - called Sparkling Water, which provides rich H2O functionality. H2O is an open source, distributed, in-memory, fast, scalable, multi-language (R, Python, Scala, UI) machine learning platform.
Literature	**Pages with SPARK main documentation:** http://spark.apache.org/docs/latest/ ; https://kb.databricks.com/ ; https://docs.databricks.com/ ; https://www.databricks.com/ ; **Online free cloud platform:** https://databricks.com/try **Available machine learning guides and packages in SPARK:** https://spark.apache.org/docs/latest/ml-guide.html ; **H2O plaftorm:** https://spark.rstudio.com/guides/h2o.html ; https://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/glm.html ; https://docs.h2o.ai/ ; https://docs.h2o.ai/h2o/latest-stable/h2o-docs/index.html
Course type/organization	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
Contact person	Barbara Borowik, PhD; bborowik@pk.edu.pl
Remarks	

*please insert one of the following codes:
11.0 Mathematics, Informatics
11.1 Mathematics
11.2 Statistics
11.3 Informatics, Computer Science
11.4 Artificial Intelligence
11.5 Actuarial Science
11.9 Others Mathematics, Informatics

**1 lecture hour=45 minutes