

<b>Study Programme:</b> Computer Science – Master		
<b>Course Unit Title:</b> Big Data Analysis		
<b>Course Unit Code:</b> CS709		
<b>Name of Lecturer(s):</b> Danijela Boberić Krstićev		
<b>Type and Level of Studies:</b> Master Academic Degree		
<b>Course Status (compulsory/elective):</b> Elective		
<b>Semester (winter/summer):</b> Summer		
<b>Language of instruction:</b> Serbian (primary), English (secondary)		
<b>Mode of course unit delivery (face-to-face/distance learning):</b> Face-to-face		
<b>Number of ECTS Allocated:</b> 6		
<b>Prerequisites:</b> None		
<b>Course Aims:</b> The aim of this course is to introduce the problems inherent in the analysis of big data and ways of solving them using Python programming language and the latest available tools to handle large amounts of data like Apache Spark.		
<b>Learning Outcomes:</b> At the end of the course, students are expected to understand basic concepts and problems in the analysis of big data and demonstrate ability to read data from persistent storage and load it into Apache Spark, manipulate data with Spark and Python..		
<b>Syllabus:</b> <i>Theory</i> This course concentrates on distributed programming using Spark and Python. It covers Spark's programming model in detail, being careful to understand how and when it differs from familiar programming models, like shared-memory parallel collections. Through hands-on examples in Spark and Python, students learn when important issues related to distribution like latency and network communication should be considered and how they can be addressed effectively for improved performance.  <i>Practice</i> Practical classes concentrate on using introduced big data analysis tools on a number of examples and case studies in solving classical problems in the field.		
<b>Required Reading:</b> 1. Andy Konwinski, Holden Karau, Matei Zaharia, and Patrick Wendell, “Learning Spark: Lightning-Fast Big Data Analysis”, O’Reilly, 2015. 2. Penchikala, Srinu. Big data processing with apache spark. Lulu. com, 2018. 3. Rogel-Salazar, Jesus. Data Science and Analytics with Python. Chapman and Hall/CRC, 2018. 4. Viktor Mayer-Schönberger, Kenneth Cukier, “Big Data: A Revolution That Will Transform How We Live, Work, and Think”, Eamon Dolan/Mariner Books, 2013.		
<b>Weekly Contact Hours:</b> 5	<b>Lectures:</b> 3	<b>Practical work:</b> 2
<b>Teaching Methods:</b> During theoretical classes classical methods of teaching with the use of a projector are used to present stated topics. On practical classes, classical methods of teaching with the use of a projector and computers with appropriate software installed are used to practically master the skills of usage of suggested tools. A premise for successful practical classes is the existence of enough computers so that every student may work individually.		

<b>Knowledge Assessment (maximum of 100 points): 100</b>			
<b>Pre-exam obligations</b>	points	<b>Final exam</b>	points
Test 1	20	written exam	
Test 2	20	oral exam	50
The methods of knowledge assessment may differ; the table presents only some of the options: written exam, oral exam, project presentation, seminars, etc.			