Study programme(s): Computer Science (CS)	
Level: master	
Course title: Big Data Analysis (CS709)	
Lecturer: Srđan M. Škrbić	
Status: elective	
ECTS: 6	
Requirements: Advanced functional programming	

Learning objectives

Manipulating big data distributed over a cluster using functional concepts is rampant in industry, and is arguably one of the first widespread industrial uses of functional ideas. The aim of this course is to introduce the problems inherent in the analysis of big data and ways of solving them using functional programming language (eg Scala) and the latest available tools to handle large amounts of data like Apache Spark.

Learning outcomes

Minimal: 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 Scala.

Desirable: At the end of the course, beside the skills mentioned above, a successful student will learn how to express algorithms for data analysis in a functional style, recognize how to avoid shuffles and recomputation in Spark. The successful student will also understand the key issues and their solutions and applications in practical work.

Syllabus

Theoretical instruction

This course concentrates on distributed programming using Spark and Scala. 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 or sequential Scala collections. Through hands-on examples in Spark and Scala, 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.

Practical instruction

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.

Literature

1. Andy Konwinski, Holden Karau, Matei Zaharia, and Patrick Wendell, "Learning Spark: Lightning-Fast Big Data Analysis", O'Reilly, 2015.
2. Bill Venners, Lex Spoon, and Martin Odersky, "Programming in Scala: A Comprehensive Step-by-Step Guide", Artima, 2011.
3. Viktor Mayer-Schönberger, Kenneth Cukier, "Big Data: A Revolution That Will Transform How We Live, Work, and Think", Eamon Dolan/Mariner Books, 2013.

Weekly teaching load				Other:
Lectures: 2	Exercises: 0	Other forms of teaching: 2	Student research:	

Teaching methodology

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.

Grading method (maximal number of points 100)			
Pre-exam obligations	**points**	**Final exam**	**points**

| Two colloquiums | **25, 25** | Oral exam | *50* |