



<b>Студијски програм:</b> Мастер академске студије информатике					
<b>Назив предмета:</b> <b>Обрада великих количина података</b>					
<b>Статус предмета:</b> Изборни на свим модулима мастер академских студија информатике					
<b>Број ЕСПБ:</b> 6					
<b>Услов:</b> Уписан одговарајући семестар					
<b>Циљ предмета</b> Упознавање са основама складиштења, обраде и анализе велике количине података коришћењем солидних математичких основа и актуелних софтверских алата.					
<b>Исход предмета</b> Студент познаје и разуме математичку базу, стандарде и технологије које се баве складиштењем, обрадом и анализом масовних података. Способан је да самостално креира апликацију за подршку одлучивању која као улаз користи велику количину података.					
<b>Садржај предмета</b> <i>Теоријска настава</i> Математичке основе. Веза са класичним паралелним програмирањем дељене и дистрибуиране меморије. Научне и пословне апликације које укључују обраду велике количине података. Концепт MapReduce, његове предности и ограничења. Скалирање у NoSQL базама података као што су MongoDB, Cassandra и Neo4j. Оптимизације и хеуристике над великом количином података. Кластер рачунарство помоћу алата Apache Hadoop, Apache Spark и Hive. Напредна примена Apache Spark-а помоћу библиотека MLlib и SparkSQL. Стриминг апликације коришћењем редова за поруке, као и Apache Kafka. Концепти виртуелизације и контејнеризације. <i>Практична настава</i> Креирање илустративних апликација применом актуелних <i>Big Data</i> технологија. Самостални развој и креирање кластер апликација како употребом базног RDD приступа у <i>Spark</i> -у, тако и позивањем рутина из библиотека. Скриптовање помоћу програмских језика Python и R. Мерење и оптимизација перформанси <i>MapReduce</i> апликација. Стриминг апликације помоћу редова за поруке и Apache Kafka оквира. Израда плана скалирања. Администрација кластера, контејнеризација, Docker и Kubernetes.					
<b>Литература</b> <ol style="list-style-type: none"><li>Chodorow, Kristina, MongoDB: The Definitive Guide: Powerful and Scalable Data Storage, O'Reilly Media, Inc., 2013.</li><li>White, Tom, Hadoop: The definitive guide, O'Reilly Media, Inc., 2012.</li><li>Spark, Apache, Apache Spark: Lightning-fast cluster computing. URL <a href="http://spark.apache.org">http://spark.apache.org</a> (2016).</li></ol>					
<b>Број часова</b>	<b>активне наставе</b>	<b>Теоријска настава:</b>	<b>2</b>	<b>Практична настава:</b>	<b>1+1</b>
<b>Методe извођења наставе</b> У настави/учењу предмета примењују се методе активног учења/наставе. Поред ex-catedra предавања лекција и лабораторијских вежби, примењују се интерактивне методе учења у учионици, као и појединачне и тимске самосталне активности студената ван учионице (у рачунарском центру, библиотеци, кући). Интерактивно учење се примењује у виду самосталног рада појединца, кооперативног и колаборативног учења, учења базираног на проблему, тимског рада и изради групних или тимских пројеката (семинара).					
<b>Оцена знања (максимални број поена 100)</b>					
<b>Предиспитне обавезе</b>	<b>70</b>	<b>поена</b>	<b>Завршни испит</b>	<b>30</b>	<b>поена</b>
практична настава	4		писмени испит		
колоквијуми	18+18		усмени испит	30	
семинар	30				
<b>Напомена:</b> Како је документовано у Табели 10.2, за извођење наставе на предмету доступан је Hadoop/НРС кластер од 6 чворова и то: 1 x Fujitsu PRIMERGY RX2540 M1 , 2 Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40GHz, 32 GB, 8 TB 5 x HP ProLiant DL165 G7, 2 X AMD Opteron(TM) Processor 6272, 16 GB, 300GB					