

Студијски програм : МАС ИТ			
Назив предмета: ОБРАДА ВЕЛИКИХ КОЛИЧИНА ПОДАТКА			
Наставник/наставници: Ана М. Капларевић-Малишић			
Гостујући професор: Dimitrios Fotiadis			
Статус предмета: Изборни			
Број ЕСПБ: 6			
Услов: Уписан одговарајући семестар			
Циљ предмета Упознавање са основама складиштења, обраде и анализе велике количине података коришћењем солидних математичких основа и актуелних софтверских алата.			
Исход предмета 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.			
Садржај предмета <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 рутине 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.			
Литература 1. Chodorow, Kristina, MongoDB: The Definitive Guide: Powerful and Scalable Data Storage, O'Reilly Media, Inc., 2013. 2. White, Tom, Hadoop: The definitive guide, O'Reilly Media, Inc., 2012. 3. Spark, Apache, Apache Spark: Lightning-fast cluster computing. URL http://spark.apache.org (2016).			
Број часова активне наставе	Теоријска настава: 30	Практична настава: 15+15	
Методe извођења наставе 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).			
Оцена знања (максимални број поена 100)			
Предиспитне обавезе	70 поена	Завршни испит	30 поена
активност у току предавања		усмени испит	30
практична настава	4		
пројекти			
колоквијум-и	18+18		
семинар-и	30		
Напомена: За извођење наставе на предмету доступан је Hadoop/HPC 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