

Студијски програми: Рачунарске науке				
Врста и ниво студија: мастер академске студије				
Назив предмета: Анализа великих података (ЦС709)				
Наставник (Име, средње слово, презиме): Срђан Шкрбић				
Статус предмета: изборни				
Број ЕСПБ: 6				
Услов: Напредно функционално програмирање				
Циљ предмета Управљање великом количином података дистрибуираним на кластеру користећи функционалне концепте представља раширену праксу у индустрији и једну од првих широко распрострањених индустријских употреба функционалних идеја. Циљ овог предмета је упознавање проблема присутних у анализи великих података и начина њиховог решавања користећи функционални програмски језик (на пример Scala) и најновије доступне алате за обраду великих количина података као што је Apache Spark.				
Исход предмета <i>Минимални:</i> На крају курса, се очекује да студенти схвате основне концепте и проблеме у анализи великих података и покажу способност да читају податке из постојећих извора података и учитају их у Apache Spark, као и да манипулишу тим подацима користећи програмски језик Scala. <i>Пожељни:</i> Поред горепомнутих вештина и знања, успешан студент ће научити како да искажу алгоритме за анализу великих података у функционалном стилу како да избегну „shuffles“ и прерачунавање у Spark-у. Успешан студент ће такође разумети кључне проблеме и њихова решења као и њихову примену у практичном раду.				
Садржај предмета <i>Теоријска настава</i> Овај курс се фокусира на дистрибуирано програмирање користећи Spark и Scala-у. Детаљно се приказује модел програмирања Spark-а, као и разлике у односу на од познате моделе програмирања, као што су дељене меморије паралелне колекције или секвенцијалне Scala колекције. Кроз практичне примере у Spark-у и Scala-у, студенти уче када треба да се разматрају питања везана за дистрибуцију података и латенцију мрежне комуникације и како се у датом окружењу могу ефикасно решавати проблеми са добрим перформансама. <i>Практична настава</i> Практична настава се концентрише на коришћење уведених алатки за анализу великих података на већем броју примера и студија у решавању класичних проблема у пракси.				
Литература 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.				
Број часова активне наставе				Остали часови 0
Предавања: 2	Вежбе: 2	Други облици наставе: 0	Студијски истраживачки рад: 0	
Методe извођења наставе На предавањима се користе класичне методе наставе уз коришћење пројектора и лаптопа, као и конекције на рачунарски кластер. Вежбе се састје од већег броја примера који потичу из практичне употребе и где се ставља нагласак на самостално решавање проблема уз употребу најновијих технологија и алата као и рачунарског кластера. Знање студената се испитује на два практична колоквијума, док се на усменом делу испита проверава разумевање теоријског дела градива.				

Оцена знања (максимални број поена 100)			
Предиспитне обавезе	поена	Завршни испит	поена
Два колоквијума	25, 25	усмени испит	50