

Тема: Складиштење и анализа великих података са применом у хидроинформационим системима

Наставник: Бобан Стојановић

У информационим технологијама под великим подацима се подразумевају скупови података који расту толико великим брзинама да са њима није могуће радити коришћењем стандардних алата за управљање базама података. Тешкоће у раду се јављају приликом прикупљања, смештања, претраге, дељења, анализе и визуелизације. Тренутни лимити у раду су реда величина терабајта, егзабајта и зетабајта података. Овакви проблеми се јављају у области метеорологије, генетике, финансија, итд. Раст скупова података је нарочито изазван све једноставнијим и учесталијим прикупљањем података помоћу разних мобилних уређаја и сензора, тако да је 90% светских података креирано управо у последње две године.

Задатак овог рада је испитивање могућности управљања великим подацима коришћењем масовне паралелизације на десетинама, стотинама, па чак и хиљадама сервера. Посебна пажња биће посвећена *Apache Hadoop* софтверском окружењу за дистрибуирану обраду великих скупова података на кластерима рачунара, коришћењем једноставног програмског модела. Истраживања која су предмет овог рада ће бити вршена у *Microsoft развојном центру* у Београду и примењена на хидроинформационе системе за управљање водним ресурсима у Србији и Републици Српској.

Литература

1. T. White, Hadoop, *The Definitive Guide*, O'Reilly Media, 2009.
2. D. Kuznetzky, *What is „Big Data“?*, ZDNet, <http://blogs.zdnet.com/virtualization/?p=1708>.
3. A. Vance, *Start-Up Goes After Big Data With Hadoop Helper*, New York Times Blog, 2010, <http://bits.blogs.nytimes.com/2010/04/22/start-up-goes-after-big-data-with-hadoop-helper/?dbk>.