

Наставно-научном већу Електротехничког факултета у Београду

Комисија II степена Електротехничког факултета у Београду именовала нас је за чланове Комисије за преглед и оцену мастер рада кандидата Хасана Рецовића под насловом “Методe и тестирање сложене обраде података у дата центрима”. Након прегледа рада подносимо следећи

Извештај

1. Биографски подаци

Хасан К. Рецовић је рођен 10. фебруара 1990. године у Београду. Похађао је одличним успехом основну школу “Дуле Караклајић” и Средњу електротехничку школу “Никола Тесла” у Београду.

Електротехнички факултет Универзитета у Београду уписао је 2009. године. Студирао је на одсеку: Телекомуникације и информационе технологије, смер: Радио комуникације, и дипломирао је 9. септембра 2013. године са просечном оценом 8,61 и оценом 10 на дипломском раду “Испитивање отпорности сајта - Академија инжењерских наука Србије”. Дипломске академске – мастер студије уписао је септембра 2013. године на Електротехничком факултету у Београду и положио све испите са просечном оценом 10,00.

2. Предмет мастер рада

Потреба за складиштењем и обрадом велике количине података стално расте. Предмет истраживања је Хадуп опен сорс софтвер чија је главна намена управо анализа велике количине података. Рад се бави описом функционисања сервиса Хадуп екосистема. Врши се тестирање имплементације, конфигурације и извршавања апликације у циљу утврђивања ефикасности датог система.

3. Основни подаци о мастер раду

Мастер рад кандидата Хасана Рецовића “Методe и тестирање сложене обраде података у дата центрима” садржи 72 стране штампаног текста, 40 слика, 2 табеле, и 9 референци. Мастер рад је подељен у неколико целина које се баве начином функционисања, имплементацијом и тестирањем рада Хадуп софтвера.

4. Садржај и анализа рада

Прво поглавље представља увод у којем је описан значај и потенцијал Хадупа у обради великих количина података. Такође наведен је и циљ и предмет тезе.

У другом поглављу представљене су основе Хадупа, односно која је намена система, који су основни делови и како те компоненте функционишу. Циљ овог поглавља је да се стекне уопштена слика о Хадупу, где он може да се примени, над каквим подацима и који је основни концепт његовог функционисања.

Треће поглавље описује три главне компоненте Хадупа: HDFS (*Hadoop Distributed File System*), *MapReduce* и YARN (*Yet Another Resource Negotiator*). HDFS је фајл систем Хадуп кластера и дефинише начин на који се подаци дистрибуирају и складиште по чворовима кластера. Познавање начина на који су подаци смештени јако је битно за разумевање рада осталих сервиса, тако да је овој компоненти дата посебна пажња приликом описивања. *MapReduce* је радни оквир и интерфејс за писање апликација за обраду података. У овом поглављу описане су фазе рада *MapReduce* апликације и како се извршава паралелна обрада података на више чворова. Трећа компонента YARN, одваја управљање ресурса кластера од *MapReduce* радног оквира, што отвара нове могућности за развијање и имплементирање нових сервиса чије управљање и обрада података може у потпуности да се разликује од *MapReduce* апликације. Свака од ових компоненти је јако битна за функционисање система и познавање начина рада увелико помаже приликом пројектовања и прилагођавања конкретној намени. Из тог разлога ово поглавље описује бројне појединости на које треба обратити пажњу.

Имплементација Хадуп кластера и тестирање апликације представљени су у четвртој поглављу. Наведени су основни кораци како конфигурирати и покренути Хадуп кластер на виртуалним машинама. Такође анализирано је убацивање и управљање подацима у HDFS. У овом поглављу дат је и опис писања и покретања апликације написане у *Pig* апликацији. У овом програмском језику могу се писати апликације за обраду података на једноставнији начин али мана је мања флексибилност и прилагођеност рада у кластеру у односу на *MapReduce*. У овом поглављу извршено је тестирање две апликације које имају исти циљ обраде али једна је написана у *MapReduce* апликацији а друга у *Pig* апликацији. Анализирано су резултати реда апликација и перформансе њиховог извршавања.

У петом поглављу, дат је осврт на целокупан рад, резимирани су резултати конфигурације и тестирања система. Овде је изложен закључак мастер тезе и предложен је правац у којем треба да се развија Хадуп софтвер како би постао ефикаснији и прилагођен већем броју захтева за обрадом различитих типова података.

5. Закључак и предлог

У мастер раду кандидата Хасан Рецовића, основни допринос јесте приказ поступка конфигурације и начина рада Хадуп екосистема. Пружена је анализа рада апликација у овом окружењу и предности које доноси обрада података у паралели. Предложен је даљи смер за развој Хадупа. Такође, овај рад ће омогућити свим заинтересованим читаоцима да се упознају са основама рада и конфигурације овог система и да имају добру основу за прилагођавају имплементацију својим потребама.

На основу свега изложеног чланови Комисије предлажу Комисији II степена Електротехничког факултета у Београду да рад Хасана Рецовића под насловом "Методe и тестирања сложене обраде података у дата центрима" прихвати као мастер тезу и да кандидату одобри јавну усмену одбрану.

Београд, 20.11.2014.

Чланови комисије


др Александра Смиљанић, ван. проф.


др Милан Бјелица, ван. проф.