

## დიდი მონაცემები - ნაკადების დამუშავება რეალურ დროში

### ანოტაცია

მას შემდეგ, რაც ინტერნეტმა მთელი სამყარო ციფრულ სივრცეში დააკავშირა, ორგანიზაციებისათვის ღირებული მონაცემების მოცულობა ძალიან გაიზარდა და კვლავაც იზრდება. ამ მოცულობის ინფორმაციის შენახვასა და დამუშავებას არსებული ტექნოლოგიები მეტწილად ახერხებდნენ იქამდე, სანამ არ გამოჩნდნენ სოციალური ქსელები, საძიებო ქსელები და ელექტრონული კომერციები, რომელთაც, მოკლედ რომ ვთქვათ, მონაცემთა ბუმი მოჰყვა. მომხმარებლებელთა მოთხოვნილებების დასაკმაყოფილებლად, ინოვაციურობისათვის თუ კონკურენციის დასაძლევად, კომპანიებს სჭირდებოდათ უზარმაზარი მოცულობის მონაცემის შეგროვება უამრავი სხვადასხვა წყაროსაგან, გაწმენდა, ტრანსფორმაცია, გაანალიზება და სწრაფი გადაწყვეტილებების მიღება, რაც არსებული საშუალებებით იოლი არ იყო. აქ პრობლემა მხოლოდ ამ ინფორმაციის მოცულობა არ იყო. ამ ფაქტორთან ერთად, ამ დიდ მონაცემთა მთავარი გამოწვევა გახლდა მათი არასტრუქტურირებული ბუნებაც და წარმოქმნის სისწრაფეც.

2005 წლიდან ოფიციალურად დამკვიდრდა ტერმინი დიდი მონაცემები – Big Data. იმავე წელსვე ღია პროგრამული უზრუნველყოფების სივრცეში შეიქმნა Hadoop-იც - დიდ მონაცემთა სამყაროს გული. მას შემდეგ ღია პროგრამული უზრუნველყოფების სივრცეში ძალიან აქტიურად მუშაობენ დიდ მონაცემების მიმართულებით. შეიქმნა უამრავი სპეციალიზირებული ტექნოლოგია და საშუალება, რომელთა წყალობითაც შესაძლებელია დიდ მონაცემთა დამუშავება, შენახვა, მართვა, ანალიზი და ა. შ. შედეგად, დღესდღეობით Big Data ეკოსისტემა ძალიან მრავალფეროვანია, არქიტექტურა კი – საკმაოდ კომპლექსური და მრავალი კომპონენტისგან შემდგარი.

წინამდებარე ნაშრომში ზოგადად არის მიმოხილული დიდ მონაცემები, მისი აქტუალობა, მთავარი კონცეფციები თუ პრინციპები, ძირითადი მახასიათებლები და არქიტექტურული თავისებურებები. საკვანძო ნაწილი კი ეხება Big Data სამყაროს ერთ-ერთ აქტუალურ მიმართულებას – მონაცემთა ნაკადების დამუშავება (თითქმის) რეალურ დროში და განხილულია მისი ორი ძირითადი არქიტექტურა. პრაქტიკული ნაშრომის სახით კი წარმოდგენილია მონაცემთა ნაკადების დამუშავებასთან დაკავშირებული კონკრეტული ამოცანის გადაჭრა პროგრამული უზრუნველყოფის კოდის სახით.

ანა ჯაფარიძე