



**BTU** | ბიზნესისა და  
ტექნოლოგიების  
უნივერსიტეტი

# ღირი მონაცემები

## დიდი მონაცემები

ორი შთამბეჭდავი ფაქტი უნდა აღინიშნოს:

- 2012 წელს ყოველდღიურად იქმნებოდა 2,5 ეგზაბაიტი ( $2,5 \times 10^{18}$  ბაიტი) მოცულობის ციფრული კონტენტი<sup>1</sup>. დღესდღეობით, აღნიშნული რიცხვი კიდევ უფრო არის გაზრდილი;
- **მონაცემთა საერთაშორისო კორპორაციის (IDC - International Data Corporation)** ინფორმაციით, 2018 წლისთვის გენერირებული ციფრული მასალის მოცულობამ 33 ზეტაბაიტი შეადგინა ( $2,5 \times 10^{21}$  ბაიტი), ხოლო 2025 წლისთვის 175 ზეტაბაიტს მიაღწევს<sup>2</sup>.

გიგანტური ციფრული მასალის დაგროვებას ინტერნეტის და განსაკუთრებით, მსოფლიო აბლაბუდის (World Wide Web) სწრაფი ზრდა და განვითარება განაპირობებს. 2020 წლისთვის, რეალურ ცხოვრებაში მიმდინარე მოვლენების საკმაოდ დიდი ნაწილი სხვადასხვა ტიპის სენსორებით ფიქსირდება და ციფრულ ფორმატში გარდაიქმნება. ციფრული ინფორმაციის აქტიური გენერატორებია:

- მობილური მოწყობილობები (სმარტფონები, პლანშეტები, „ჭკვიანი“ საათები და ა.შ.);
- დაშორებული ზონდირების სისტემები (რაღარები, რადიომეტრები, ფოტომეტრები და სხვა, მათ შორის, დედამიწის ხელოვნური თანამგზავრები);
- კომპიუტერული პროგრამების მოქმედებათა ჟურნალები (ლოგები);
- ვიდეოკამერები და მიკროფონები;

<sup>1</sup> <http://www.ibm.com/big-data/us/en/>

<sup>2</sup> David Reinsel, John Gantz, John Rydning, The Digitization of the World - From Edge to Core, IDC White paper, November 2018, <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

- რადიოსაინდენტიფიკაციო (უკონტაქტო) ჩიპები (RFID);
- უკაბელო სენსორები;
- მსოფლიო აბლაბუდის სერვისები (მაგალითად, სოციალური ქსელები).

თითოეული ჩამოთვლილი წყაროდან მიღებული ინფორმაცია თითქოს, მხოლოდ მცირე პერიოდში არის სასარგებლო. ერთი შეხედვით, რატომ უნდა იყოს საინტერესო ადამიანის სმარტფონში „Google Maps“-აპლიკაციის და ლოკაციის სერვისის (GPS) ერთიანი მუშაობით მიღებული სტატისტიკა, მაგალითად: გასულ წელს 100-ჯერ იმყოფებოდა ერთ უბანში და 120-ჯერ სხვა უბანში, ან იმავე პერიოდში, ბეთჰოვენის 15 ნაწარმოებს მოუსმინა, ბახისას კი - მხოლოდ 12-ს. ამასთანავე, ციფრული ტექნოლოგიების მომხმარებლისთვის ნაკლებად საინტერესოა, ბოლო 5 წელიწადში ტანსაცმელი მეტი იყიდა თუ ფეხსაცმელი, ან სოციალურ ქსელებში მსახიობების სურათებზე მეტი „ლაიქი“ დააფიქსირა თუ ფეხბურთელების. თუმცა, აღნიშნულ მონაცემების ცხადი სახით დაგროვება არ ხდება, მხოლოდ მათი „დალექვა“ ციფრული ოკეანის ფსკერზე. ბოლო ათწლეულში გაირკვა, მოცემული ინფორმაცია თუ რიგითი მომხმარებლისთვის არა, ბიზნესისთვის ნამდვილი „საბადოა“ და მისი დამუშავება საქმეს მრავალი მიმართულებით წასწევს წინ.

**ამგვარად, შეიქმნა „დიდი მონაცემების“ (BigData) კონცეფცია, რომელიც სხვადასხვა წყაროებიდან დაგროვილი მონაცემების კონსოლიდაციას და მიღებული შედეგების ბიზნეს ანალიზის (BI - Business Intelligence) ამოცანებისთვის გამოყენებას გულისხმობს.**

ამასთანავე, მნიშვნელოვანია ე.წ. **3V-მოდელის** იდენტიფიცირება:

- „დიდი მონაცემების“ სივრცეში ნებისმიერი ტიპის ინფორმაცია მოძრაობს: ტექსტური, გრაფიკული, აუდიო, ვიდეო, ბინარული (ორობითი) და სხვა. მონაცემების ნაწილს მკაფიო სტრუქტურა გააჩნია (Structured), ნაწილი სანახევროდაა მოწესრიგებული (Semi-structured) და ნაწილიც საერთოდ ქაოტური სახით ინახება - „დიდი მონაცემების“ ამ მახასიათებელს **მრავალფეროვნება (Variety) ეწოდება;**

- „დიდი მონაცემები“ მართლაც დიდია და არ ჰგავს იმ ათასი ან თუნდაც 10 ათასი ჩანაწერის მომცველ ცხრილებს, რომლებთან კომფორტულ მუშაობასაც

საოფისე აპლიკაციების (Excel, Word) გარემოში ადამიანები არიან მიჩვეულნი. „დიდი მონაცემები“, როგორც წესი, კომპიუტერის გარე მეხსიერებაში ათობით ტერაბაიტს იკავებს - „დიდი მონაცემების“ ამ მახასიათებელს **მოცულობა (Volume) ეწოდება;**

■ „დიდი მონაცემებთან“ წვდომის დრო შეიძლება მკვეთრად განსხვავდებოდეს მათი განაწილებული შენახვის გამო - „დიდი მონაცემების“ ამ მახასიათებელს **სისწრაფე (Velocity) ეწოდება.**

შესაბამისად, „დიდი მონაცემებთან“ მომუშავე პროგრამული უზრუნველყოფის შექმნა, დანერგვა და გამოყენება ნებისმიერი ორგანიზაციისთვის ერთ-ერთ უდიდეს გამოწვევას წარმოადგენს. „დიდი მონაცემების“ აქტუალობა წლიდან წლამდე მატულობს, 2020 წლისთვის ბაზარზე დიდ მონაცემებთან მომუშავე ათობით მაღალი დონის პროგრამა არსებობს<sup>3</sup>. 2018 წლის ინფორმაციით, „დიდი მონაცემების“ აპლიკაციების გაყიდვა/მხარდაჭერით მიღებული შემოსავლების საერთო მოცულობა 42 მილიარდ დოლარს შეადგენს, ხოლო 2027 წლისთვის აღნიშნული მაჩვენებლის 103 მილიარდ დოლარამდე გაზრდას ვარაუდობენ.

„დიდი მონაცემების“ აპლიკაციების გამოყენების რამდენიმე უმნიშვნელოვანესი მიმართულება<sup>4</sup>:

■ **განათლება** - „დიდი მონაცემების“ დამუშავებით შესაძლებელია სასწავლო პროგრამის და მასალების დინამიური მართვა, რომელიც სტუდენტთა სასწავლო ისტორიების საფუძველზე თითოეული სტუდენტის ინდივიდუალური მოთხოვნების უკეთესად გათვალისწინების საშუალებას იძლევა;

■ **ჯანდაცვა** - „დიდი მონაცემები“ ცალსახად ამცირებს პაციენტებისთვის დიაგნოზის დასმის დროს და ამაღლებს დიაგნოზის ხარისხს, აუმჯობესებს დაავადებათა პრევენციას (ადრეულ ეტაპზე მათ გამომჟღავნებას), ასრულებს ამა თუ იმ გეოგრაფიულ რეგიონში ეპიდემიების გავრცელების პროგნოზირებას;

■ **ამინდის პროგნოზი** - სატელიტური თანამგზავრებიდან და მეტეოროლოგიური სენსორებიდან ათწლეულების განმავლობაში მიღებული „დიდი მონაცემების“ საფუძველზე შესაძლებელია ამინდის პროგნოზის

<sup>3</sup> <https://www.capterra.com/big-data-software/>

<sup>4</sup> <https://intellipaat.com/blog/7-big-data-examples-application-of-big-data-in-real-life/>



გაუმჯობესება, გლობალური დათბობის შესახებ უფრო მეტი და ხარისხიანი ინფორმაციის მიღება, ბუნებრივი კატასტროფების თვისებებში უკეთ გარკვევა და მათთან უფრო ეფექტურად გამკლავება.

დღესდღეობით, დიდი მონაცემები, ასევე, ეფექტურად გამოიყენება საცალო და საბითუმო ვაჭრობის, საბანკო, საფინანსო, სატრანსპორტო, სამთავრობო და სხვა მრავალ სფეროში.

მაგალითისთვის, საინტერესოა, „დიდი მონაცემების“ მართვა პროგრამა Zoomdata-ს გარემოში, რომლის ვებგვერდზე<sup>5</sup> შესვლით და ვებ-აპლიკაცია **1B rows - Big Data Analytics**-ის არჩევით ამერიკის შეერთებული შტატების ბაზარზე მომუშავე საცალო ვაჭრობის კომპანიის „დიდი მონაცემებს“ წარმოადგენს.

აპლიკაციისთვის ინფორმაციის წყაროს წარმოადგენს საცალო ვაჭრობის წერტილებიდან მიღებული 1 მილიარდი ჩანაწერი, რომლებიც თავს იყრიან Apache Hadoop-სისტემის მონაცემთა ბაზაში. მნიშვნელოვანია, რომ ჩანაწერთა სიმრავლის მიუხედავად, თითოეული დაფიქსირებული მოთხოვნა ძალიან სწრაფად მუშავდება, რაც „დიდი მონაცემთა“ ტექნოლოგიის უდიდეს უპირატესობას წარმოადგენს, ინფორმაციის შენახვის ტრადიციულ მეთოდებთან შედარებით.

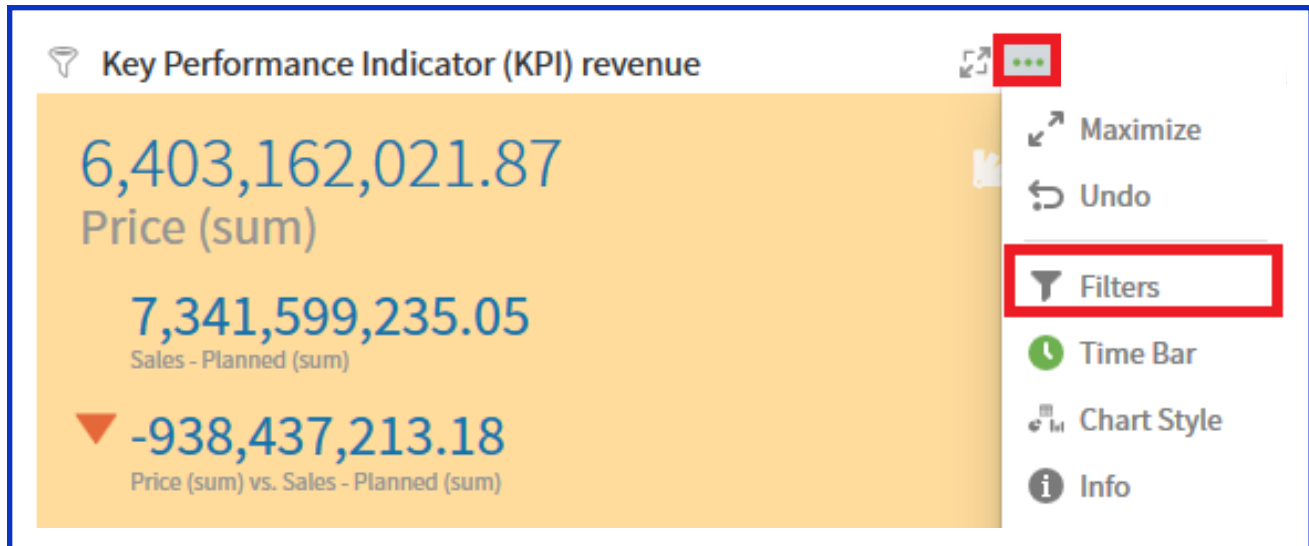
აპლიკაცია განკუთვნილია კომპანიის ბიზნეს-ანალიტიკოსებისთვის, რომლებიც მისი დახმარებით მრავალ საინტერესო შედეგს იღებენ. ინფორმაციის დახარისხება შესაძლებელია უამრავი სხვადასხვა კრიტერიუმით, მაგალითად, კონკრეტული შტატის რუქაზე დაწკაპებით. საყურადღებოა, რომ აპლიკაცია ანალიტიკურ უნარებსაც ფლობს. მაგალითად:

- აწვდის ანალიტიკოსს ინფორმაციას კონკრეტული ტიპის საქონლის ან კონკრეტულ შტატში გაყიდვების ტენდენციების (Trends) შესახებ - უმჯობესდება გაყიდვები თუ უარესდება;

- იქნა თუ არა მიღწეული გაყიდვების **სამიზნე ნიშნულები** (KPI - Key Performance Indicators) მთელი ქვეყნის, შტატ(ებ)ის, საგრაფო(ებ)ის თუ ცალკეული ქალაქ(ებ)ისთვის.

<sup>5</sup> <https://live.zoomdata.com/zoomdata/visualization>

ბოლო ფუნქციის გამოსაცდელად საჭიროა შემდეგი ქმედებების განხორციელება: Key Performance Indicator (KPI) Revenue-ფანჯარაში ღილაკ - „Show more..“ -ზე დაჭერა და ბრძანება Filter-ის არჩევა (იხ. სურათი #1).



**სურათი #1**

შემდგომ კი - ღილაკი Add Filter და ფილტრაცია შტატების მიხედვით - Customer State. შტატების სიაში შესაძლებელია ერთი ან მეტი შტატის არჩევა (მაგალითად, Georgia) და მთავარ ეკრანზე დაბრუნებისას დარწმუნდებით, რომ შტატში გაყიდვებს სამიზნე ნიშნულზე გასასვლელად 29 664 505 დოლარი და 85 ცენტი დააკლდა.

# პრაქტიკული დავალება

- მოიფიქრეთ, ვის და რა ტიპის სასარგებლო ანალიტიკური ინფორმაციის მიღება შეუძლია სოციალური ქსელების „დიდი მონაცემებიდან“;
- მოიფიქრეთ, როგორ შეიძლება გამოიყენოთ ავტობანებზე დაყენებული „ჭკვიანი“ ვიდეოკამერებიდან მიღებული „დიდი მონაცემები“.

**შემუშავებულია:** დ. გულუა

**BTU**

ბიზნესისა და  
ტექნოლოგიების  
უნივერსიტეტი

