

## მონაცემთა შენახვა-დამუშავების განაწილებული სისტემები

გია სურგულაძე, ბექა კახელი, გიორგი მაისურაძე

საქართველოს ტექნიკური უნივერსიტეტი

### რეზიუმე

განხილულია ინფორმაციის შენახვა-დამუშავების თანამედროვე ტექნოლოგიები, მონაცემთა დამუშავების განაწილებული გარემოს ეფექტურობა და მნიშვნელობა. აღწერილია თანამედროვე ტექნოლოგიები და მათი პრობლემური მხარეები, ტექნოლოგიური და სისტემური მოდელები, რომელთა გამოყენებითაც შესაძლებელია ინფორმაციის შენახვა-დამუშავების ერთიანი განაწილებული ეკოსისტემის ფორმირება. დეტალურად არის წარმოდგენილი ამ ეკოსისტემის ერთ-ერთი ყველაზე მნიშვნელოვანი მოდული, რომელზეც ხდება მთლიანი სისტემის დაშენება.

**საკვანძო სიტყვები:** განაწილებული სისტემები. ინფორმაციის შენახვა-დამუშავება. პარალელური დამუშავება. მაღალი წვდომადობა.

### 1. შესავალი

მონაცემთა შენახვა-დამუშავების განაწილებული სისტემა გულისხმობს რიგი აპარატურული და პროგრამული ელემენტების ურთიერთდაკავშირებით მიღებულ ერთიან სისტემას, რომელიც წარმოადგენს სხვადასხვა ტექნოლოგიების გაერთიანებას. აღნიშნული ტექნოლოგიები გამოიყენება როგორც მონაცემების შესანახად, ასევე მათი გადაცემა დამუშავებისთვის, ეს მიმართულებებია ქსელური მარშუტიზატორები, ინფორმაციის მატარებლები და გამოთვლითი რესურსი. აღნიშნულ ნაშრომში განხილული იქნება სამივე მიმართულების ტექნოლოგიური შესაძლებლობები და მათი ნაკლოვანებები.

აღნიშნული ნაშრომის მიზანია რაც შეიძლება დაბალი ღირებულების ტექნოლოგიების გამოყენება სასურველი მიზნის მისაღწევად. ამისთვის საჭირო იქნება ზემოთ აღნიშნული მიმართულებებიდან ავარჯიოთ გარკვეული ტექნოლოგიები, რომლებსაც გამოვიყენებთ პროექტის ფარგლებში.

როგორც მოგეხსენებათ, ინფორმაციული ტექნოლოგიების განვითარების ტემპი ძალიან მაღალია. დღითი-დღე ურო მეტად დახვეწილი და გაუმჯობესებული ტექნოლოგიები იქმნება, რაც საშუალებას გვაძლევს გავაკეთოთ ისეთი პროექტები, რომლებიც აქამდე მხოლოდ თეორიულად იყო შესაძლებელი.

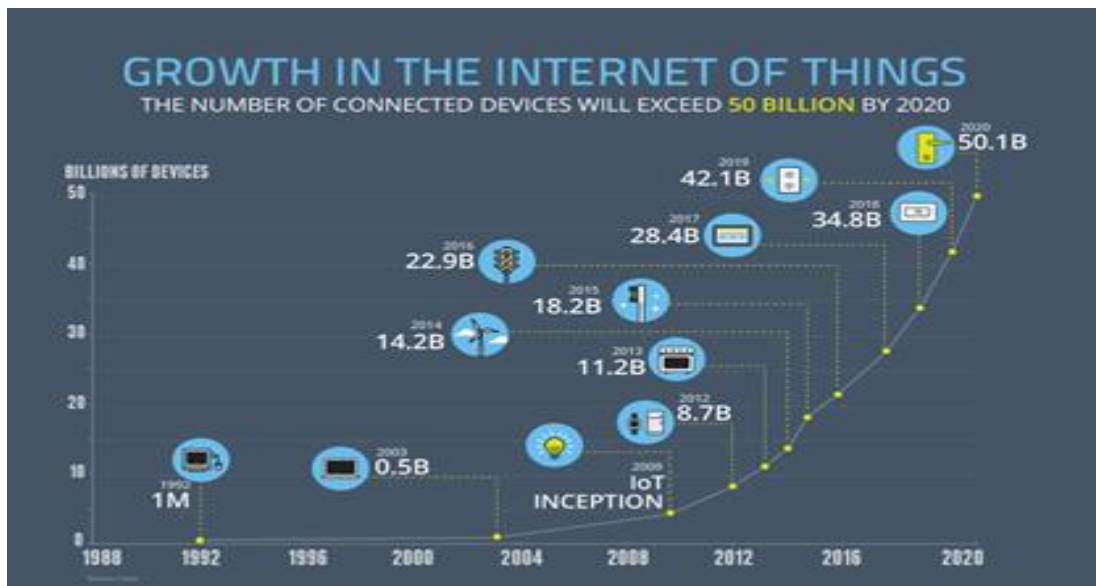
განვიხილოთ გარკვეული აპარატურული და პროგრამული ტექნოლოგიები, როგორცაა Linux ოპერაციული სისტემა, ქსელური მარშუტიზატორები, აპარატურული მოწყობილობები (Raspberry Pi 3, Odroid C2, Udo, Banana Pi, UP Board). ამ ტექნოლოგიების გაერთიანებით შევძლებთ დავაპროექტოთ ერთიანი მონაცემთა შენახვა-დამუშავების განაწილებული სისტემა, რომელიც საშუალებას მოგვცემს, გავაერთიანოთ ამ მოწყობილობების, როგორც გამოთვლითი, ასევე მონაცემების შენახვის რესურსი. მსგავს ტექნოლოგიებს იყენებენ როგორც რელაციური, ასევე არარელაციური ბაზების

დაპროექტებისას. მსგავსი ტექნოლოგიები აქვთ ისეთ დიდ ორგანიზაციებს, როგორცაა: Oracle, IBM, Google, Amazon, Facebook და ა.შ.

მსგავსი ტექნოლოგიების გამოყენება ხშირად დაკავშირებულია დიდ თანხებთან, რადგან გასათვალისწინებელი ხდება ბევრი ფაქტორი, მაგალითად: აპარატურისთვის შენობა ნაგებობის გამოყოფა, უწყვეტი კვების წყაროს გამართვა, გაგრილების სისტემა, აპარატურის მომსახურე პერსონალი და სხვ., რაც საკმაოდ დიდი პრობლემაა დაბალბიუჯეტური ორგანიზაციებისათვის.

## 2. ძირითადი ნაწილი

ინფორმაციის შენახვა-დამუშავება თანამედროვე ეპოქის პრობლემაა, რადგან დასამუშავებელი ინფორმაცია დღითი დღე უფრო მატულობს. ყოველდღიურად იზრდება ინტერნეტ მომხმარებლების რაოდენობა, თანამედროვე ტექნოლოგიების დახმარებით შესაძლებელი გახდა ქსელში ჩართულიყო არა მარტო კომპიუტერები, არამედ ყველა საყოფაცხოვრებო მოწყობილობა (ნახ.1) [1].

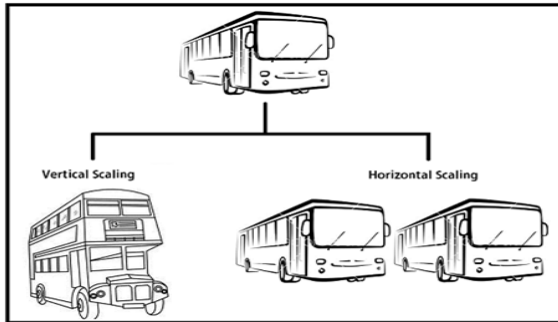


ნახ.1.

ნახაზზე ნაჩვენებია ქსელში ჩართული მოწყობილობების რაოდენობის ზრდა დროის მიხედვით. ეს რაოდენობა 2020 წლისთვის 50 ბილიონ მოწყობილობამდე გაიზრდება. შესაბამისად გაიზრდება ამ მოწყობილობებიდან გენერირებული მონაცემები, რომლებიც ორგანიზაცია „NASSCOM“ ის კვლევის თანახმად ზეტაბაიტებს მიაღწევს [1].

აღნიშნული მონაცემების დამუშავება რელაციური ბაზების მოდელით თითქმის შეუძლებელია, რადგან აღნიშნული მოდელით საჭირო იქნება მონაცემების მატებასთან ერთად გავზარდოთ აპარატურული რესურსი (ექსტენსიონალური მიდგომა), რაც არის ვერტიკალური ზრდის მოდელი. აღნიშნული მოდელი კი შეზღუდულია, რადგან არსებული ტექნოლოგიების განვითარება მონაცემების ზრდასთან შედარებით ბევრად ნაკლებია. შესაბამისად, არსებული ინფორმაციის დასამუშავებლად საჭიროა ალტერნატიული გზების მოძიება (ინტენსიონალური მიდგომა).

## 2.1. ჰორიზონტალური ზრდა



ნახ.2

ჰორიზონტალური ზრდა არის ის ალტერნატიული გზა, რომელიც საშუალებას გვაძლევს დავამუშაოთ დიდი მონაცემები (BigData). ეს არქიტექტურა მონაცემთა დამუშავებისათვის ერთდროულად იყენებს რამდენიმე სერვერს [2,3]. ინფორმაცია ნაწილდება სერვერებზე და, შესაბამისად, მისი დამუშავების სიჩქარე ბევრად მაღალია (ნახ.2).

ნახაზზე ნაჩვენებია ვერტიკალური და ჰორიზონტალური ზრდის მაგალითები, სადაც ნათლად ჩანს ჰორიზონტალური ზრდის უპირატესობა. იგი მდგომარეობს იმაში, რომ როდესაც დაგჭირდება მეტი გამოთვლითი სიმძლავრე, არ მოგვიწევს მთელი არქიტექტურის შეცვლა. საკმარისი იქნება კომპიუტერული რესურსის, სერვერის დამატება. რაც ბევრად მოსახერხებელია ვერტიკალურ ზრდასთან შედარებით. ვერტიკალური ზრდა ხდება სერვერის მონაცემების დეტალების გამოცვლით ან დამატებით, რასაც ბევრი პრობლემა ახლავს თან. კერძოდ, უნდა მოხდეს არსებული აპარატურის გაძლიერების შესაძლებლობის გათვალისწინება და ახალი დეტალების თავსებადობის შემოწმება რასაც საბოლოოდ ერთ კონკრეტულ ბრენდამდე მივყავართ, რის შედეგადაც ჩვენ ვხდებით კონკრეტულ მწარმოებელზე დამოკიდებულნი. ხშირ შემთხვევაში შეიძლება მთლიანი აპარატურის გამოცვლაც გახდეს საჭირო.

თანამედროვე ინფორმაციული ტექნოლოგიები, რომლებიც ორიენტირებულია ჰორიზონტალური ზრდისკენ, საშუალებას გვაძლევს დავივიწყოთ აპარატურის თავსებადობის მსგავსი პრობლემები. მათი საშუალებით შესაძლებელია მონაცემთა შენახვა-დამუშავების ერთიანი სისტემის შექმნა სხვადასხვა მწარმოებლის და მოდელის აპარატურისგან. ეს ყველაფერი საერთო ჯამში ბევრად უფრო მოსახერხებელი და დაბალ ბიუჯეტურია.

## 2.2. განაწილებული სისტემები

მონაცემთა შენახვა-დამუშავების თანამედროვე განაწილებული სისტემებია „Apache Hadoop“ და „MongoDB“ [3]. ისინი საშუალებას გვაძლევს ჩვენი საჭიროებისამებრ გამოვიყენოთ აპარატურული რესურსი. იმისათვის, რომ აღნიშნულმა სისტემებმა ფუნქციონირება შეძლოს, აუცილებელია შემდეგი წინაპირობები:

- კომპიუტერული ტექნიკა, რომელზეც მოხდება ოპერაციული სისტემის გამართვა;
- ოპერაციული სისტემა (რეკომენდებულია Linux ოპერაციული სისტემა);
- ქსელური ინფრასტრუქტურა: კომპიუტერული ტექნიკა რომელზეც გამართული სისტემები უნდა ხედავდეს ერთმანეთს და შეეძლოს მონაცემების გაცვლა.

აღნიშნული პირობების დაკმაყოფილება შესაძლებელია სხვადასხვა მწარმოებლის ტექნოლოგიებით, რაც ბევრად მოსახერხებელია და გვაძლევს იმის საშუალებას გამოვიყენოთ ამ სისტემების ისეთი შესაძლებლობები როგორცაა: ინფორმაციის

პარარელურ რეჟიმში შენახვა-დამუშავება, სარეზერვო ასლების ავტომატური გენერირება, საჭიროებისამებრ რესურსის ვერტიკალური ზრდა, ონლაინ რეპორტიინგი და სხვ.

ინფორმაციის შენახვა-დამუშავების განაწილებული სისტემის ერთ-ერთი ყველაზე მნიშვნელოვანი კომპონენტია *ოპერაციული სისტემა*. განვიხილოთ Unix-ის ტიპის ოპერაციული სისტემა (Linux), რომელსაც გამოვიყენებთ პროექტის ფარგლებში [4-6]. Linux არის თავისუფლად გავრცელებადი GPL ლიცენზიაზე დამყარებული ოპერაციული სისტემა, რომელიც შეიქმნა UNIX ოპერაციული სისტემის მსგავსად. მისი თავდაპირველი ვერსიის შემქმნელია ჰელსინკის უნივერსიტეტის სტუდენტი ლინუს ტორვალდსი (Linus Torvalds) [7], რის შემდეგაც მის განვითარებაზე ზრუნვა განაგრძეს UNIX პროგრამისტთა და ინტერნეტის ქსელში მომუშავე ენთუზიასტების ჯგუფებმა. მასში არ არის გამოყენებული AT&T UNIX-ის პროგრამული კოდი. Linux-ის მეტი წილი პროგრამები შექმნილია GNU Free Software Foundation პროექტის ჩარჩოში, კემბრიჯში, მასაჩუსეტსში, თუმცა მის შექმნაში ასევე მონაწილეობა მიიღეს თითქმის მთელი მსოფლიოს პროგრამისტებმა. აღსანიშნავია ისიც, რომ Linux-ის ბაზაზე შექმნილი პროგრამების უმეტესობა არის ღია პროგრამული კოდით (open source) და ამასთან უფასოა.

ლინუსს ოპერაციული სისტემის ბირთვი შედგება 22 მილიონი ხაზის მოცულობის კოდისგან, რაც საკმაოდ მცირე რესურსს მოითხოვს: პროცესორი 400 Mhz, ოპერატიული მეხსიერება 256 MB, მეხსიერების ზომა მყარ დისკზე 10 MB [8]. ეს ოპერაციულ სისტემას ბევრად უფრო მოსახერხებელს ხდის. მომხმარებელს ოპერაციული სისტემის დაყენებისას შეუძლია დააყენოს მხოლოდ მინიმალური პაკეტი, რის შემდეგაც დააყენებს მხოლოდ იმ სისტემებს რომლებიც მას სჭირდება. შესაბამისად ოპერაციული სისტემის ზომა მინიმალურია, რაც ამცირებს რისკებს და ზრდის სისტემის საიმედოობას. ლინუსს ამ მახასიათებლების გამო ხშირად იყენებენ შემდეგი მიზნებისთვის: web server , ftp server, http/https server, proxy server, firewall gateway, database server, distributoin systems და ა.შ.

ლინუსის ასევე გამოირჩევა თავდაცვითი მექანიზმებით როგორცაა „Firewall“. ესაა ყველაზე ეფექტური დაცვის საშუალება, რომელიც იცავს მომხმარებელს სხვადასხვა შეტევისგან. იგი მართავს და აკონტროლებს ქსელის ტრაფიკს და აღმოფხვრის არასანქცინირებულ წვდომას. ეს მექანიზმი გამოცდილი მომხმარებლის ხელში უფრო მეტად დაცულს და საიმედოს ხდის Linux ოპერაციულ სისტემას. აღსანიშნავია ის, რომ ტექნოლოგიების განვითარებასთან ერთად ვითარდება ლინუსის დაცვის მექანიზმებიც.

### 3. დასკვნა

თანამედროვე სამყაროში ინფორმაციის რაოდენობის ექსპონენციალურმა ზრდამ, საჭირო გახადა დიდი მოცულობის მონაცემების შენახვა-დამუშავება რეალურ დროში, რაც ტრადიციული რელაციური მონაცემთა ბაზებისათვის გარკვეულ სირთულეებს წარმოადგენს. დასამუშავებელი ინფორმაციის ზრდასთან ერთად საჭირო ხდება უფრო და უფრო ძლიერი აპარატურა. გარკვეული ზღვრის შემდეგ, აპარატურის გაძლიერება თითქმის შეუძლებელია. ამგვარად, სპეციალისტები მივიდნენ იმ დასკვნამდე, რომ ინფორმაციის დამუშავების მეთოდოლოგია საჭიროებდა ცვლილებას, საჭირო გახდა ისეთი სისტემების შექმნა რომლებიც შეძლებდა ინფორმაციის პარალელურ რეჟიმში დამუშავებას და ამასთან დააკმაყოფილებდა შემდეგ მოთხოვნებს: მაღალი საიმედოობა,

ინფორმაციის შენახვა-დამუშავება, პარალელური დამუშავება, მაღალი წვდომადობა. თანამედროვე ინფორმაციული ტექნოლოგიების განვითარების ტემპი ძალიან მაღალია. დღითიდღე უფრო მეტად დახვეწილი და გაუმჯობესებული ტექნოლოგიები იქმნება, რაც საშუალებას გვაძლევს, გავაკეთოთ ისეთი პროექტები, რომლებიც აქამდე მხოლოდ თეორიულად იყო შესაძლებელი.

### ლიტერატურა – References – Литература:

1. <http://www.nasscom.in/iot-revolution-making>
2. სურგულაძე გ., კვიციანი გ., კახელი ბ. (2016). NoSQL მონაცემთა ბაზები: განვითარების პერსპექტივები და პრობლემები მართვის საინფორმაციო სისტემებში. სტუ, შრ.კრ. „მას“, N 2(22). გვ.230-239.
3. სურგულაძე გ., კვიციანი გ. (2016). შესავალი NoSQL მონაცემთა ბაზებში (MongoDB). ISBN 978-9941-0-9642-6. სტუ. „IT კონსალტინგის ცენტრი“. თბ., -152 გვ.
4. Linux Online. (2008) Linux Logos and Mascots <http://wayback.archive.org/web/201008-15085106/http://www.linux.org/info/logos.html>
5. webopedia. <http://www.webopedia.com/TERM/U/UNIX.html>
6. Nemeth E., Snyder G., Hein Trent R., Whaley B. (2011). UNIX and Linux System Administration Handbook 4th Edition. Prentice Hall. USA.
7. Is Linux Kernel Free Software. <https://www.kernel.org/category/faq.html>
8. Cobbaut P. (2015). Linux Fundamentals. <http://linux-training.be/linuxfun.pdf>

### DATA STORAGE AND PROCESSING DISTRIBUTED SYSTEMS

Surguladze Gia, Kakheli Beka, Maisuradze Giorgi

Georgian Technical University

#### Summary

The subject of my current research refers to the Distributed Large-scale Data Storage and Processing in Low-cost Computer Technologies Environment. The objective of the work is to research worldwide technical tendencies on designing distributed ecosystems with modern information technologies. It is considered a modern technologies of distributed data processing environment, their problematic sides and the importance of data storage and processing efficiency.

### РАСПРЕДЕЛЕННЫЕ СИСТЕМЫ ДЛЯ ХРАНЕНИЯ И ОБРАБОТКИ ДАННЫХ

Сургуладзе Г., Кахели Б., Маисурадзе Г.

Грузинский Технический Университет

#### Резюме

Рассматриваются современные технологии хранения и обработки информации, вопросы эффективности и значимости распределенной среды обработки данных. Описываются современные технологии и их проблемные стороны, технологические и системные модели, с использованием которых возможно формирование единой распределенной экосистемы хранения-обработки информации. Детально представлен один наиболее важный модуль этой экосистемы, на которой производится надстройка целой системы.