MUHAMMAD AL-XORAZMIY
NOMIDAGI TATU FARG'ONA FILIALI

FERGANA BRANCH OF TUIT
NAMED AFTER MUHAMMAD AL-KHORAZMI

"AL-FARG'ONIY AVLODLARI"

ELEKTRON ILMIY JURNALI | ELECTRONIC SCIENTIFIC JOURNAL

TA'LIMDAGI ILMIY, OMMABOP VA ILMIY TADQIQOT ISHLARI

4-SON 1(8)
2024-YIL

TATU, FARG'ONA
O'ZBEKISTON

ISSN 2181-4252

O‘ZBEKISTON RESPUBLIKASI RAQAMLI TEXNOLOGIYALAR VAZIRLIGI

MUHAMMAD AL-XORAZMIY NOMIDAGI TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI FARG‘ONA FILIALI

Muassis: Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti Farg‘ona filiali.
Chop etish tili: O‘zbek, ingliz, rus.
Jurnal texnika fanlariga ixtisoslashgan bo‘lib, barcha shu sohadagi matematika, fizika, axborot texnologiyalari yo‘nalishida maqolalar chop etib boradi.

Учредитель: Ферганский филиал Ташкентского университета информационных технологий имени Мухаммада ал-Хоразми.
Язык издания: узбекский, английский, русский.
Журнал специализируется на технических науках и публикует статьи в области математики, физики и информационных технологий.

Founder: Fergana branch of the Tashkent University of Information Technologies named after Muhammad al-Khorazmi.
Language of publication: Uzbek, English, Russian.
The magazine specializes in technical sciences and publishes articles in the field of mathematics, physics, and information technology.

2024 yil, Tom 1, №4
Vol.1, Iss.4, 2024 y

ELEKTRON ILMIY JURNALI **ELECTRONIC SCIENTIFIC JOURNAL**

«Al-Farg‘oniy avlodlari» («The descendants of al-Fargani», «Potomki al-Fergani») O‘zbekiston Respublikasi Prezidenti administratsiyasi huzuridagi Axborot va ommaviy kommunikatsiyalar agentligida 2022-yil 21 dekabrda 054493-son bilan ro‘yxatdan o‘tgan.

Jurnal OAK Rayosatining 2023-yil 30 sentabrdagi 343-sonli qarori bilan Texnika fanlari yo‘nalishida milliy nashrlar ro‘yxatiga kiritilgan.

Tahririyat manzili:
151100, Farg‘ona sh.,
Aeroport ko‘chasi 17-uy,
202A-xona
Tel: (+99899) 998-01-42
e-mail: info@al-fargoniy.uz

FARG’ONA - 2024 YIL

Jurnal quyidagi bazalarda indekslanadi:

MUNDARIJA | ОГЛАВЛЕНИЕ | TABLE OF CONTENTS

MUNDARIJA | ОГЛАВЛЕНИЕ | TABLE OF CONTENTS

MUNDARIJA | ОГЛАВЛЕНИЕ | TABLE OF CONTENTS

THE USE OF BIG DATA IN THE DIGITAL ECONOMY

Okhunov Dilshod Mamatjonovich,
associate Professor of the Department of "Economics and vocational education" of the Ferghana branch of the Tashkent University of information technologies named after Muhammad al-Khorezmi
e-mail: mamatdilshod@rambler.ru

Okhunov Mamatjon Xamidovich,
associate Professor of the Ferghana Polytechnic Institute
e-mail: mamatdilshod@rambler.ru

Azizov IskandarAbdusalim ugli,
Assistant professor at the department of "Economics and vocational education" of the Ferghana branch of the Tashkent University of information technologies named after Muhammad al-Khorezmi
e-mail: mamatdilshod@rambler.ru

Ismoilzhonov Abdullokh Farrukhbek ugli,
student of group 667-24 of the Tashkent University of Information Technologies named after Muhammad al-Khorezmiy
e-mail: mamatdilshod@rambler.ru

Abstract: The term "Big Data" refers to data sets whose size exceeds the capabilities of typical databases for recording, storing, managing and analyzing information. The article discusses the main components of big data and their essence.

Keywords: Big Data, information, traditional database, big data database, Internet, specialized DataSet

Introduction.

Companies that use an IT platform of their own development to conduct their activities can be conditionally called digital. In addition to their own IT platform, digital companies can also use mobile workplaces, cloud HR management services, cloud accounting, etc., promote their goods and services online, carry out transactions or sell their products online. Digital companies can be represented in any sector of the economy: finance, oil, energy, transport, communications, etc.

The penetration of information technology, on the one hand, gives digital companies an advantage in the speed of providing services, their quality and price, etc., but on the other hand, cybersecurity risks are growing. Various technical solutions are used to minimize cybersecurity risks. The cybersecurity technical solutions market consists of 12 segments: from network security to anti-fraud solutions. However, there is no such segment in the market as protection against internal intruders (insider attacks). Insider protection solutions themselves exist, but due to their dampness and inefficiency, they are not combined into a segment. The greatest risk is posed by insider attacks, which are implemented as data leaks from Big Data repositories.

Instagram Facebook likes, or uploading photos to Instagram, we produce data every day: through Google search, Facebook likes, or Instagram uploads. This flow of information is fundamentally changing the

way we interact with each other and with the world. Big data is an opportunity, it is a powerful tool that can be used to solve problems and raise questions. It is already clear that big data will change our lives, the only question is what will become a familiar everyday life - a paradise on Earth or a dystopia [1].

Literature Review.

When writing this work, scientific and educational literature, articles in electronic journals, as well as articles in periodicals were used.

To reveal the concept of big data, its parameters and characteristics, the works of V. Mayer-Schoenberger, K. Kukier, N. Martz, D. Warren and numerous Internet sources were used. In these sources, the concepts of big data, their main features, ways of processing them, ways of analyzing them, areas of application and various examples were analyzed.

The concept of "Big Data" refers to the exact sciences, but, like many concepts, it does not have a clear definition. The authors interpret this concept in different ways. The concept of "big data" originated in the days of mainframes and related scientific computing, as high-tech computing has always been complex and is usually inextricably linked to the need to process large amounts of information. One of the first to use the term "Big Data" was John Mashey, a consultant to technology companies and trustee of the Computer History Museum in Mountain View, California in 1998. When introducing the term "Big Data", John Mashey did not pursue the goal of popularizing this term among representatives of the high-tech community. "I used one label for a number of questions, and I wanted the simplest and shortest phrase to say that the boundaries of computing continue to expand," said John Mashi. The term "Big Data", which John Mashey did not attach much importance to in 1998, along with other technologies has become a catalyst for economic growth in the 21st century.

In 2001 Doug Laney has published the results of his research, in which he identified three key parameters of big data: Volume, Velocity, and Variety. Later, the number of key parameters increased to five: volume, velocity, variety, veracity, and value. Currently, the following key parameters have been added: viability (viability), variability (variability), visualization (visualization). All these parameters are dynamic. For example, the first parameter is large volumes, if we look at the time scale, then those volumes that were stored in traditional databases (from gigabytes to terabytes) a few years ago could conditionally be attributed to "big data" [2].

The avalanche-like growth of information led to the need for digitization of information, and starting from the end of the 20th century, digital information began to displace analog information. In 2012, the share of analog and digital information equaled and this year became the frontier of digitalization.

Experts have already turned all our habits, interests and everyday joys into arrays of numbers - now it remains to figure out how to dispose of this data. We used to look at small data and think what it meant, what it meant, to try to understand the world, but now we have a lot more data, more than we could have mined before. And when we have a large data corpus, we can do things that we couldn't do when the data was smaller. Big data is important and big data is a novelty, and if you think about it, the only way for our planet to cope with global problems: to feed people, provide medical care, provide them with energy and electricity, cope with global warming is the effective use of big data [3].

Information (data) has gone the way from something stationary, static to something current and dynamic. So the disk found in Crete for 4000 years is heavy, does not store much information and this information is unchanged. On the contrary, all the files that Edward Stowden from the US National Security Agency took are placed on a memory card the size of a fingernail, and it can be distributed at the speed of light.

Materials and methods.

The term "Big Data" refers to data sets whose size exceeds the capabilities of typical databases for recording, storing, managing and analyzing information. A comparison of traditional databases and big data databases is presented in Table 1.

Table 1

Traditional database and big data database

Characteristic	Traditional databases	Databases of "big data"
The amount of information	From gigabytes to terabytes (from 1e+9 bytes to 1e+12 bytes)	From petabytes to exabytes (from 1e+15 bytes to 1e+18 bytes)
Data type	Structured	Semi-structured, unstructured.
Storage method	Centralized	Decentralized
Data storage and processing model	Vertical model	Horizontal model
The relationship of data	Strong	Weak

The phrase Big Data gained its popularity after the publication of an article in a special issue of the journal Nature on September 3, 2008. The issue of the journal was devoted to the material on the phenomenon of explosive growth in the volume and variety of processed data and technological prospects [4].

Since 2012, the term "big data" has become a fashionable topic. This term is advertised because it is a very important tool through which society progresses. In today's world, many companies view big data as an opportunity to gain absolute control over their competitors.

In 2013, the definition of the term "Big Data" was included in the Oxford English Dictionary. The translation of the definition can be interpreted as follows: "Data of a very large size, as a rule, in the sense that they present serious difficulties in logistical support for manipulation and management, as well as the direction of calculations using this type of data" [5].

From the materials of the free encyclopedia, "Big Data" is the designation of structured and unstructured data of huge volumes and significant diversity, efficiently processed by horizontally scalable software tools that appeared in the late 2000s and alternative to traditional database management systems and Business Intelligence solutions.

Despite its widespread use and already solid consolidation in modern digital society, the concept of big data remains one of the most controversial in science. The concept of "Big Data" implies working with information of a huge volume and diverse composition, which is very often updated and located in different sources [6,7].

Results.

Let $U = \{u1,\ldots um\}$ be the hosts of the network. Let the property P correspond to the processing of information technology in time □>$T0$. Let the object O be a set of hosts allocated by the provider for the implementation of information technology. When technology repeats, many objects $\{O^+\}$ are allocated, on which delays are observed □>$T0$ in the implementation of information technology, and many objects $\{O^-\}$ for which □□$T0$. I.e., the so-called "flickering" error occurs in the network.

Suppose that the reason for the o property of P is searched for using the intersection of objects from the set $\{O^+\}$. An empirical reason is a subobject of o if o does not occur in the set $\{O^-\}$. Let a new object O be added to the set of objects, and let o □ O, but information technology is implemented in such a way that □□$T0$. Thus, there is a falsification of the empirical reason.

In turn, revenue from the sale of software and services in the global big data market is projected to grow from $42 billion in 2018 to $103 billion in 2027, reaching a cumulative annual growth rate (CAGR) of 10.48% (Fig. 2).

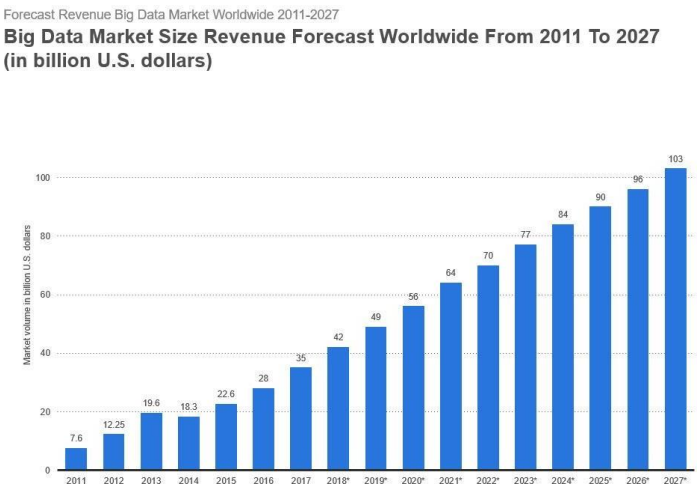

Fig. 2. Big Data Market Forecast

According to an Accenture study, 79% of business leaders agree that companies that do not use big data will lose their competitive position and may be on the verge of extinction. Moreover, 83% of managers have implemented Big Data projects in order to increase the company's competitive advantage in the market.

Big data technologies actively take advantage of smart technologies. Thus, 59% of executives say that Big Data in their company will be improved through the use of AI.

In all definitions, it can be seen that "Big Data" is a complex concept that combines:

- direct data;
- a set of technologies for working with data (a technology without which a modern business cannot compete in the market);
- A new perspective, a new paradigm in data science;
- A global phenomenon, the facts of the environment.

Today, the term Big Data is usually used to refer not only to the data arrays themselves, but also to the tools for processing them and the potential benefits that can be obtained as a result of painstaking analysis.

The sources of big data. We live in the information age - the age of the digital economy. An enormously large number of sources generate data. Conventionally, big data sources can be divided into internal and external. Internal sources: ERP, classifiers, CRM. External sources: social networks, Internet, specialized, DataSet.

We are increasingly surrounding ourselves with a network of satellites, scanners, cameras and other recording devices that create huge arrays of unstructured data - these are continuously incoming data from measuring devices, events from radio frequency identifiers, etc. A typical example of big data is information coming from various physical installations, for example, the Large Hadron Collider, which produces a huge amount of data, and does it all the time (about a petabyte of data per second, 150 zetabytes of information per year, this is more than from all other sources in the world). The installation continuously outputs large amounts of data, and scientists use them to solve many tasks in parallel.

There are currently over 400 Earth remote sensing satellites in operation. The data of many satellite systems are publicly available, and this data availability transforms satellite systems from "observational" to "measuring" ones. It is planned that by 2026, more than 1,000 systems will be operating in orbit. In Chile, by 2025, it is planned to complete the construction of the world's largest telescope, which will be able to receive information about 40 billion objects (over 60 thousand Petabytes). The Gaia Space Telescope, launched into orbit in 2013, collects data to create a three-dimensional catalog of a billion astronomical objects. Every day, more than 30 satellites transmit environmental monitoring data in excess of 8 terabytes. It is expected that after the start of operation of the radio telescope (Square Kilometer Array (SKA) - the largest astronomical project of our time), it will generate a huge amount of raw data: about 1 exobyte per day, which corresponds to today's daily traffic volume of the world Wide Internet or the total memory capacity of 15 million 64 GB iPads. After data compression, the daily amount of radio telescope information can be reduced to 1 petabyte. The billions of terabytes of data that SKA will generate annually far exceed modern databases of genomics and climatology.

Discussions.

The authors of the study, who presented their report in the journal PLoS Biology, state that by 2025, geneticists will have data on the genomes of 100 million to 2 billion people [8]. To store this amount of data, you need from 2 to 40 exabytes of space. The amount of data required to store information about a single genome is 30 times the size of the genome itself. The authors of the study concluded that the collected genetic information exceeds the projected annual requirements for storing data on the YouTube portal, which will need 1-2 exabytes by 2025 (Table 2).

Table 2

Projected largest databases by 2025

Phases	Astronomy	Genomics	Twitter	YouTube
Collecting	25 ZB/per year	1 ZB/per year	0.5-15 mill. tweets per year	500-900 mill. hours per year
Keeping	50-100 ZB/per year	2-40 ZB/per year	1-17 RV/per year	1-2 ZB/per year

The New York Stock Exchange generates about a terabyte of data per day. The storage capacity of the Facebook social network increases by about 500 terabytes every day. The Internet Archive project grows by 20 terabytes per month.

Conclusions.

The emergence of big data in the public space is due to the fact that this data has affected almost all people, not just the scientific community, where such tasks have been solved for a long time. Big Data technologies entered the public sphere when it came to a very specific number - the number of inhabitants of the planet. We all generate data using our mobile phones, sensors, social media, purchase transactions and GPS signals. These are information flows from the Internet of Things with sensory data, log files, media with audio and video files, call centers with call logs and from internal information of enterprises and organizations, from the fields of medicine and bioinformatics, from astronomical observations, etc.

References

1. Chernyak L. "Big data" - new theory and practice // Open systems. - 2021. - № 10.

2. AAPOR report on "Big Data" / L. Japek, F. Kreiter, M. Berg [et al.] / American Association of Public Opinion Researchers. - M., 2019.

3. Korytnikova N.V. Online Big Data as a source of analytical information in online research // Socis. - 2023. - No. 8. - pp. 14-24.

4. Tolstova Yu.N. Sociology and computer technologies // Socis. - 2022. - No. 8. - pp. 3-13.

5. D.M. Okhunov, M.H. Okhunov, M.U. Akbarova. A general methodology for evaluating and selecting components of automated systems. - CAD and modeling in modern electronics: a collection of scientific papers of the III International scientific and practical conference. Bryansk, 2019, pp. 54-58.

6. Dilshod Okhunov, Mamatjon Okhunov, Mukaddas Akbarova. Method of calculation of system reliability on the basis of construction of the logical function of the system. E3S Web of Conferences 139, (2019)/ RSES 2019.

7. D.Okhunov, S.Semenov, S. Gulyamov, D.Okhunova, M.Okhunov. Tools to support the Development and Promotion of Innovative Projects. SHS web of Conferences 100, 01008(2021) ISCSAI 2021, https://doi.org/10.1051/SHSconf/202110001008

8. S.V.Zapechnikov. Information security of open systems. In 2 volumes. Vol.1 – Threats, vulnerabilities, attacks and approaches to protection/ S.V.Zapechnikov. – M.:GLT, 2017. – 536s.