ВЫСШАЯ ШКОЛА УПРАВЛЕНИЯ И ИННОВАЦИЙ МГУ
ИМЕНИ М.В.ЛОМОНОСОВА

UNIVERSIADE IN INNOVATION.
Qualifying stage

CASE
«INNOVATIVE TECHNOLOGIES IN THE FIELD OF DATA STORAGE»

The rapid development of information technology leads to an increase in the importance of the problem of information storage. The volume of information used each year exceeds the forecasts of previous years and the growth rate continues to increase. Thus, it is necessary not only to improve existing technologies, but also to develop qualitatively new ones.

The collection of data and its processing began 8 thousand years ago with the emergence of culture and writing. Over time, the tasks of collecting information and areas of application expanded; this ceased to be the prerogative of individual structures. But the information became too much both for storage and for processing. It took 8 years to process the US Census in 1880, which led Herman Hollerith to create the first statistical tabulator in 1888.

Since then, the speed of data processing has constantly increased, new inventions have repeatedly surpassed previous ones in terms of performance.

Today we can say that we are in the information age, in which information communication technologies based on electronics and genetic engineering are the main tool for achieving the goal. In the field of data processing, mankind has achieved impressive results - supercomputers that solve complex problems and even quantum computing are no longer just a dream of futurists, but a modern reality. For example, at the end of 2021, scientists from the Russian Quantum Center patented a new quantum processor architecture based on qudits, which has analogues only in three countries, and on its basis by the end of 2024, Russia is going to build a universal quantum computer with cloud access. Until 2024, Rosatom plans to allocate more than 23 billion rubles for its creation. [1]. Research in the field of artificial intelligence is also ongoing and requires ever-greater speeds of information processing. About 99,000 searches are performed every second on Google [2], and according to conservative estimates, 148 GB is the amount of data that one Internet user generates on average per day [3]. By some estimates, 90% of all the data accumulated in the world was created in two years, and according to forecasts, the volume of data will increase by about 40% annually. In 2020, about 64.2 zettabytes of data were created, which is 314% more than in 2015. A significant proportion of this data is passively collected data obtained from everyday digital operations [4]. By 2025, according to various estimates, there will be up to 200 zettabytes of data in the world [5]. Consequently, the need for information storage and processing will increase even more.

A natural question that arises is where and how to store such volumes of information. Data center technology is constantly evolving. The capacity of HDD drives is increasing, SSD drives based on NVMe protocols are getting faster. Much of the effort goes into improving solid state drives, but hard drives have room for improvement.

In the field of modern data storage solutions in 2019, for the first time in history, the Russian company Yadro became the leader in terms of the volume of supplied capacity [6]. The company is engaged in the development and production of data processing and storage systems - high-performance Vesnin servers and the Tatlin family of data storage systems. Thus, domestic manufacturers have a decent solution in the field of modern data storage systems, but given the growth rate of data in the world, it becomes obvious that existing technologies will not be able to fully solve the problems of storing and processing information in the long term. We need a new technological breakthrough.

One of the most incredible technologies is the storage of data in DNA. Unlike the optical and magnetic forms of data storage technologies used today, DNA does not represent data in binary units (1s and 0s). The DNA helix consists of many alternating pairs of four unique bases - adenine (A), cytosine (C), guanine (G) and thymine (T). Based on these grounds, information is encoded and stored. These molecular chains will then be converted into binary numbers if necessary. Experiments on the practical storage of DNA data have been going on for several years, with proposals for methods covering both living organisms such as bacteria and the use of synthesized DNA. In November 2021, a team of researchers achieved the first synthesis of arbitrary DNA sequences with nanoscale element sizes that can provide high enough density to be used as DNA storage [7].

A solid-state emerging technology in this area is five-dimensional (5D) optical storage, and the University of Southampton team has been working on it for more than a decade. The data is recorded using a laser that emits incredibly short but powerful pulses of light, creating tiny structures in the glass that are measured at the nanoscale. These structures contain information about the intensity and polarization of the laser beam, in addition to their three spatial dimensions, which is why scientists call them the 5D data store. Despite the lack of compactness, which theoretically can be achieved with DNA recording, five-dimensional optical storage devices can withstand up to 1000 degrees Celsius and can be stored for 13.8 billion years at room temperature without deteriorating quality [8].

A natural question that arises is where and how to store such volumes of information. Data center technology is constantly evolving. The capacity of HDD drives is increasing, SSD drives based on NVMe protocols are getting faster. Much of the effort goes into improving solid state drives, but hard drives have room for improvement.

One of the most incredible technologies is the storage of data in DNA. Unlike the optical and magnetic forms of data storage technologies used today, DNA does not represent data in binary units (1s and 0s). The DNA helix consists of many alternating pairs of four unique bases - adenine (A), cytosine (C), guanine (G) and thymine (T). Based on these grounds, information is encoded and stored. These molecular chains will then be converted into binary numbers if necessary. Experiments on the practical storage of DNA data have been going on for several years, with proposals for methods covering both living organisms such as bacteria and the use of synthesized DNA. In November 2021, a team of researchers achieved the first synthesis of arbitrary DNA sequences with nanoscale element sizes that can provide high enough density to be used as DNA storage [7].

A solid-state emerging technology in this area is five-dimensional (5D) optical storage, and the University of Southampton team has been working on it for more than a decade. The data is recorded using a laser that emits incredibly short but powerful pulses of light, creating tiny structures in the glass that are measured at the nanoscale. These structures contain information about the intensity and polarization of the laser beam, in addition to their three spatial dimensions, which is why scientists call them the 5D data store. Despite the lack of compactness, which theoretically can be achieved with DNA recording, five-dimensional optical storage devices can withstand up to 1000 degrees Celsius and can be stored for 13.8 billion years at room temperature without deteriorating quality [8].

Another innovative technology is the data storage method using fluorescent dyes, which can be much more durable and cost effective than traditional storage methods. Harvard chemists have shown that digital data can be stored in mixtures of fluorescent dye molecules applied with an inkjet printer. The printed data can be read using a microscope, which determines the wavelength of light emitted by each dye and allows the dyes in the mixture to be distinguished. The presence or absence of a dye can represent a value in binary, the basis of modern computing. The researchers compare this technology to the concept of DNA storage

and argue that it has the advantage that molecules are simply mixed rather than complex structures are synthesized, so the speed is much higher [9].

However, all these technologies are still under development and will not be able to be used everywhere for a long time. Moreover, it remains to be seen which of these technologies may eventually become a solution to the ever-increasing needs for information processing and storage.

Despite the attractiveness of new technologies, it must be remembered that new storage technologies have many requirements in the information age. In addition to the ability to store a huge amount of data per unit volume, the ability to read and process this data quickly, without loss, must be maintained. The safety of data in various storage and transportation conditions is also an important condition for choosing one or another technology. The storage system must be fault tolerant, and the technology itself must be affordable, because the transition to new technologies will only happen when it is economically viable.

Sources:

1. Интернет-издание в сфере высоких технологий в России и странах СНГ «CNews», «Ученые из России запатентовали новую архитектуру квантового процессора», URL: https://www.cnews.ru/news/top/2022-05-26_uchenye_iz_rossii_zapatentovali
2. Открытый счетчик активности в Интернете «Internet Live Stats», Url: https://www.internetlivestats.com/
3. Интернет-платформа «FinancesOnline», «53 Important Statistics About How Much Data Is Created Every Day» Url: https://financesonline.com/how-much-data-is-created-every-day/
4. Организация объединенных наций, «Инновации в сфере данных для целей развития» Url: https://www.un.org/ru/global-issues/big-data-for-sustainable-development
5. Ведущий мировой исследователь мировой киберэкономики «Cybersecurity Ventures», «The World Will Store 200 Zettabytes Of Data By 2025», 08.06.2020 Url: https://cybersecurityventures.com/the-world-will-store-200-zettabytes-of-data-by-2025/
6. Интернет-издание «ComNews», «Рынок СХД на подъеме», 09.01.20, URL: https://www.comnews.ru/content/203813/2020-01-09/2020-w02/rynok-skhd-podeme
7. Bichlien H. Nguyen et all. «Scaling DNA data storage with nanoscale electrode wells» SCIENCE ADVANCES, 24.11.21 Url: https://www.science.org/doi/10.1126/sciadv.abi6714#tab-contributors
8. Интернет-издание «Engadget», «'5D' storage could fit 500TB on a CD-sized glass disc», 01.11.2021, Url: https://www.engadget.com/5-d-storage-could-fit-500-tb-on-cd-sized-disc-095039455.html
9. Интернет издание «The Harvard Crimson», «Harvard Chemists Use Dyes as a New Data Storage Method», Url: https://www.thecrimson.com/article/2021/10/21/dye-storage-method-explored/

Exercise:

1. Conduct a comparative analysis of the data storage technologies presented in the task. Which of them do you think is the most promising? Justify your answer.

2. Analyze one of the three proposed technologies in more detail in terms of prospects for use and current problems.

Case solution requirements

The case solution should be presented in the form of two files:

1) Presentation (.pdf or .pptx format) with the main provisions of the decision and conclusions (no more than 15 slides);
2) Text file (.pdf or .docx format) with additional information (no more than 1 page A4 12 font): calculations, analytical data, links to information sources.

The presentation and the text file must contain different materials.
Case solution files must be submitted by March 20, 2022 23:59 Moscow time at the following email address: olympicmsu@mail.ru.

Main evaluation criteria

The following criteria will be used when scoring the case solution:

• The quality of the analysis carried out and the validity of the conclusions drawn.

• Logic and structure of presentation.

• The quality of presentation design.

• Non-standard thinking when developing a solution.

• Accounting for modern economic features and conditions.