

Tashkent University of Information
Technologies named after Muhammad al-
Khwarizmi


"Big Data - technology that
will change the world"


T.A.Khujakulov

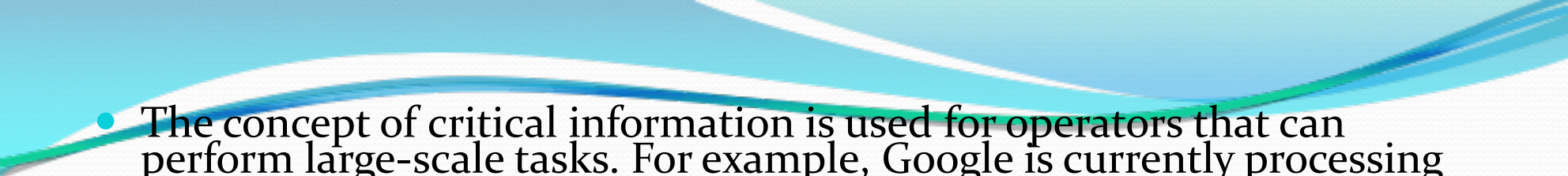



Big Data


- The number of data in the world has grown by more than one. According to IBS, in 2003, there were 5 exabytes of data (1 billion - 1 billion gigabytes). In 2008 it increased to 0.18 zettabytes (1 zettabyte = 1024 exabytes), by 1.76 zettabytes in 2011, and 4.4 Zettabs in 2013. In May 2015, the number of data collected in the world exceeded 6.5 times. By 2020, humans will generate 40 to 44 zettabytes of digital data.
- According to IBS experts, only 1.5% of the data collected in 2017 had access to information. Fortunately, nowadays there are huge data processing technologies in the world, with the help of which you can get useful, useful, and useful information from a vast array of data.


- 
- Big data is a large number of non-homogeneous and fast digital data that can not be processed by default. In some cases, it is understood that the processing of this data, along with the great data concept. Basically, the object of the analysis is called large data.
 - The term Big data was born in 2008. Nature journal editor Klifford Linch used a large data term in a special issue dedicated to the rapid growth of global data volumes. However, great information has already been made. Experts say that the flow of information, which is over 100 gb per day, is a great deal of data.
 - Analysis of large data helps to define the laws beyond human capacity. This will allow us to further improve our everyday life, government management, medicine, telecommunications, finance, transportation, manufacturing and other industries, and to find alternative solutions to these problems.


- 
- "Big Data", that is, the term "big data" has been used to refer to flows that flow to more than 100 gb per day. As a result of the sharp increase in the content of the information, it became widely accepted. This term is usually used for large amounts of data (terabytes, eczabytes, and petabytes).
 - The term was first published in the press in 2008, when the editor of Nature magazine, Cleftord Lynch, published an article in his journal on the development of science by using a large amount of information technology. Until 2009, this term was only available from the point of view of scientific analysis, but after publishing a few more articles on the subject in the press, Big Data became widely used.
 - In 2010, initial efforts to address the growing number of critical information issues have started. When using the huge flow of information, the software has been developed to reduce risk.
 - In 2011, large companies, such as Microsoft, Oracle, EMC, and IBM were interested in significant data - they were first to use Big Data development to succeed in their development strategies.

- 
- The concept of critical information is used for operators that can perform large-scale tasks. For example, Google is currently processing more than one petabyte. This figure is 100 times more than the materials in the US Library of Congress.
 - VVV - volume, speed, diversity
 - The term "substantial information" means a large number of information stored literally on any media carrier. Additionally, this amount is too large and can not be processed using conventional software or hardware products, and in some cases it is generally impossible.
 - "Big Data" is not just the data itself, but the technology of processing and handling them, the way to search for the information they need from a vast stream.
 - Basic principles - Volume, Velocity, Variety are closely related to this term. This is a process that depends on how fast the information is stored and how fast it is. Recently, these three basic principles have also been added to the concept of Value, which is the value of information. That is, it must be profitable and useful in theory or practice to justify the cost of storing and processing data.
 - Significant data source
 -

- 
- For example, a simple source of large data can include social networks - each profile or public page represents a small drop in the information ocean. No matter the amount of data stored on that profile, the interaction of each user must be as fast as possible.
 - Significant information accumulates in almost every aspect of human life. It can include any network related to human activities. These social media, both medical and banking, and the system of devices, the results of a large number of daily information, such as metrological information from astronomical observations, metrological information, and Earth exploration equipment and information may be.
 - Real-time information obtained from different monitoring systems is provided to the company or server. Television and radio broadcasts, mobile operators' call base - each communicate with each other at the minimal level, but all of these information is aggregated into a solid source.
 - Sophisticated data technology can not be imagined by research and business. In addition, they have begun to occupy public administration - everywhere there is a need for more effective storage and manipulation of information.

- 
- In practice
 - World-renowned companies (Facebook, Google, and social giants such as IBM, as well as MasterCard, VISA and Bank of America) have their own customers everywhere on the globe. Active introduction of "Big Data" technology into the modern world market has begun. For example, IBM uses a sophisticated data method in money transfer operations. With this technique, more than 15% fraudulent transactions (fraud) were detected in the transactions, and the amount of secured funds increased by 60%. Also, problems with the system have been solved - they have been reduced to half.
 - VISA also uses Big Data technology to track fraud in one or another operation. As a result, they save about \$ 2 billion a year from "leakage".
 - The German Ministry of Labor has introduced a large amount of information into the unemployment pension scheme, costing up to \$ 10 billion. to the euro. Almost every fifth of these citizens have been deprived of their right to benefit from this pension.

- 
- Problems
 - The key problem is that the information gathered over the decades is still very important and open to any system. Another major problem is the cost of processing them. This can be done by expensive equipment, and the cost of wages for qualified professionals serving a large amount of information. Obviously, the equipment requires regular updates, which should not lose the efficiency of the business while increasing the amount of data.
 - The third problem relates to a large number of information that needs to be processed. For example, researches provide results not only in 2-3 times, but also in many studies because it is very difficult to distinguish between data and objective evaluations for real-world effects.
 - Problems with information loss. Preventive measures require no one-off simple backup data, at least 2-3 spare parts of storage. However, increasing volume makes backups harder - IT professionals are trying to find the optimal solution to this problem.

- 
- In summary, we can say that we can not hide the technology. Big Data is changing the world, gradually entering our city, our home, and our gadgets. How fast the technology gets on the planet - it's hard to say. But one thing is clear - you can keep the tradition as old as Bob Kelso says in the "Clinic" episode.