PROGRESS IN DATA STORING

Verhasau V.M.

Belarusian State University of Informatics and Radioelectronics,
Minsk, Republic of Belarus

Subbotkina I. G. – Associate Professor

The history of data storing systems, the path of their progress, the main types of storage and their capacity are analyzed in this paper. The latest and the most popular technologies for data storing, such as cloud storage, USB flash drives, HDDs, SSDs are also considered. Much attention is given to data storage development, including DNA storage.

It is a well-known fact that storing information is very important in different fields of IT. Big Data, Artificial Intelligence, Deep Learning require a lot of space to store data. At present, the world data capacity is about 295 exabytes (1018 bytes). In 2020 people shared about 44 zettabytes (1021 bytes) of data through the Internet [1]. Mention should be made about the fact, that about 90% of generated data in the world was produced during the last few years. Every person, who uses smartphones or PCs, stores and generates data, which includes photos, videos, audio, documents, etc.

The era of data storing started with punch cards. This is a piece of stiff paper that holds digital data represented by holes in predefined positions. Semyon Korsakov is considered to be the first who proposed using punch cards in informatics to store data. In the late 1920s IBM introduced the punch card. The capacity of such a punch card was about 80 bytes. Storing data should be effective because at the beginning of the computer era the storage capacity was small. And as a result, such data storage was focused on the needs of a single application.

The next type of storing data was a magnetic drum. The magnetic drum storage unit contained a large metal cylinder filmed with recording material. The capacity of such storage was about 48 kilobytes (103 bytes). It was invented by Gustav Tauschek in 1932. Magnetic drums weren't widely spread and were used only by US Navy. After World War II they became more popular in computational systems. Later a magnetic tape drive was invented. It was the first widely spread type of storage that could store music, videos, audio. The first drives could store about 231 kilobytes of data. The evolution of such type was VHS

drives and they could hold up to 2 gigabytes (10^9 bytes) of data, they were in great demand for storing videos. The manufacture of VHS was running until 2016.

The next step in the development of data storage was the hard disk drives. HDD is a data storage device that stores and retrieves digital data using magnets to write data on a spinning disk. The first disks could store about 4 megabytes. And these devices are being used at present; their capacity is up to 20 terabytes (1012 bytes). Hard disk drives are used in all PCs, servers because of their reasonable cost, small size and speed, in comparison with other types of storage.

Floppy disks were worked out after hard disk drives. They presented disk storage composed of a thin and flexible disk of a magnetic storage medium. It can store from 80 kilobytes up to 150/200 megabytes. This storage also became popular and the main reason was the size of floppy disks. A lot of PCs of that time could read and write such disks. Now we can see the floppy disk every day in the form of the save-icon button. A zip drive is a floppy disk, and it can store much more data. Its capacity increased from 100 megabytes to 750 megabytes. Zip drives present the final form of floppy-like devices.

The compact disk (CD) was designed in 1982. It could hold up to 700 megabytes of data. CD is a small plastic disc that stores and retrieves computer data using light. CD was used to store music. Even now CDs are used for listening to music. DVDs are the evolution of CDs, they could store 1.5 gigabytes of data and were used to store photos, videos, audio. DVDs replaced VHSs, which were popular at that time. The next step in disk drives is Blu-ray optical disks. Such disks are used to hold high-resolution videos. Such videos consume a lot of data, so one needs a device, which can hold that amount of data. The maximum capacity of Blu-rays is 50 gigabytes. They are still popular, but cost a lot and aren't produced today, because more convenient and affordable storages are available.

In 1991 SSD (solid-state drive) was designed. These drives are sort of upgraded HDDs, but they have another principle of work, so this is not the evolution of HDD. Solid-state storage is a device that uses integrated circuits to store data constantly. The capacity of SSD was 20 megabytes and increased to around 31 terabytes at the present. They provide greater speeds, especially for random-access operations. This fact is the main reason, why they are so widely spread nowadays because we want our PC to perform better (modern operational systems consist of a big number of small files).

The next step in data storing was in 1999. That year SD cards and USB flash drives were invented. They are portable, small drives, but at the beginning, they could hold only 64 kilobytes and 8 kilobytes of data respectively. The USB and SD operate closely to SSD and differ only in size. But nowadays USB Flash drives can store up to 1 terabyte of data and an SD card is capable of 512 gigabytes of data. This type of storage is still very popular today due to its size, performance and cost.

Today some of the types are being improved to store more data, be more performant and reliable. The main types of storage nowadays are USB flash drives, SD cards, HDDs, SSDs. But our world is constantly changing and requires a lot of changes in the IT sphere. Today we are moving from storing our data on some drives at home to cloud storage. When we are speaking about cloud storage, we should begin with the history of cloud storage. The first web-based data storage was developed by AT&T. Cloud storage consists of servers, which are connected to the World Wide Web. HDDs and SSDs are widely used in cloud storage. Cloud storage must be reliable in different aspects. Big companies maintain cloud storage and create backups to keep data safe. All these backups are stored on magnetic tape storage, which holds up to 400 terabytes of data [2]. The main advantage of magnetic tape storage is that there is no need in electricity to store data without any loss. For example, HDDs and SSDs need the energy to ensure that data isn't corrupted, so they are not acceptable as storage for backups.

Now cloud storage is widely spread: people use smartphones, which are always connected to the Internet, store photos, videos, documents, music in the "Cloud" [3]. Nowadays it is very convenient because you do not need to think about data safety in terms of physical safety, you can get your information at any time and the only thing you need is the Internet. The fact about the amount of produced data is not so surprising because people upload and download their information from the cloud, which leads to generating data. Even when we share some news on social media or messengers, we produce data.

Today, a lot of researches are targeted at discovering the ability to store data in DNA. Scientists are continually looking for more space-efficient alternatives to store a great amount of data. The first attempts were done in the previous century, but there is no usable storage based on this technology. The main advantage of such storage is reliability and efficiency. Theoretically, each gram of DNA can store up to 455 exabytes of data [4]. The main problems at present are reading and writing data in DNA because it costs a lot and takes too much time to perform such operations. Modern systems can achieve the time of reading/writing at the level of a few days and the cost of such operation is about 100 USD per GB. Also, mention should be made about the fact that the capacity and error rate of such storage depends on which type of encoding was selected. The time and cost of operation on DNA-based storage depends on how big is the data to be encoded. At present DNA-based storage can significantly benefit from the breakthroughs in DNA synthesis.

As a result of such fast progress, at present, we use the most performant and convenient storage types, such as cloud storage, USB flash drives, HDDs, SSDs. During the last hundred years, storage capacities increased more than 1012 times and became more performant. Due to this progress now, we

are storing and processing more information and this allows us to access new technologies, such as AI, AR, etc., which require a great amount of data. Of course, progress didn't stop at current achievements, but each step is more and more complex and expensive to discover. It's interesting to look over this progress and how such devices work and how they were improved.

References:

1. Data Amount – [Electronic resource]. – Access mode: https://www.statista.com/statistics/871513/worldwide-data-created – Date of access: 22.03.2022.

2. Tape Storage Usage – [Electronic resource]. – Access mode: https://www.datacenterknowledge.com/industry-perspectives/how-tape-storage-changing-game-data-centers – Date of access: 22.03.2022.

3. Brief History of Data Storage – [Electronic resource]. – Access mode: https://www.dataversity.net/brief-history-data-storage – Date of access: 22.03.2022.

4. Carbon-based Archiving – [Electronic resource]. – Access mode: https://academic.oup.com/gigascience/article/8/6/giz075/5521158 – Date of access: 24.03.2022.