THE IMPORTANCE OF USING THE BIG DATA SYSTEM AND ITS PROSPECTS

[1]Rakhmonov Nodirjon, [2]Tolibjonov Khurshidbek, [3]Ibrohimova Nargizaxon
Teacher of the Business Department of Kokand University, Kokand, Uzbekistan[1], Student of the Kokand University, Kokand, Uzbekistan[2], Student of the Kokand University, Kokand, Uzbekistan[3]
tolibjonovxurshidbek@mail.com

ANNOTATION

This article details the emergence, development stages, and current role of Big Data. Provides accurate information on the history of Big Data and its applications. The growth of digital data in the world over the years, as well as the performance of the Hadoop project are analyzed. The problems of the Big Data and Hadoop projects have been studied and the necessary instructions have been given.

Keywords: Big date, hadoop, digital data, tools.

MAIN PART

There is so much information in our lives that we need for our lifestyle, our education, our work. To store them, we need modern devices in addition to human memory. We also need a mobile phone or, in addition, memory cards to record the memorable moments of ordinary people from birth. A person or a group of people has enough mobile phone memory to store family information, but the above mentioned is too small to store the data of large organizations, enterprises and companies. In this case, you need to use the Big Data service.

Big data is a very large amount of non-homogeneous and rapidly declining digital data that cannot be processed in the usual way. In some cases, along with the concept of big data, it is also understood to process that data. Basically, the object of analysis is called big data.

According to experts, the term "Big Data" refers to streams that receive more than 100 GB of data per day. Later, with the rapid increase in information, this concept became more widespread. The term is commonly used to refer to large amounts of data (at the terabyte, exabyte, and petabyte levels).

The term weighty first appeared in the press in 2008, when Clifford Lynch, editor-in-chief of Nature magazine, published an article in his journal on the development of the future of science using large-scale data processing technologies. Until 2009, the term was approached only from the point of view of scientific analysis, but after the publication of several more articles on the subject in the press, the concept of "Big Data" began to be widely used. In 2010, the first attempts were made to address the growing problem of heavy data. Software products have been developed to reduce the risks of using large data streams. Much of the “big data” is not data collection. Now “Big Data” is real. Google has big data and CERN also has big data. Most people probably do not. Data only starts when the store needs 1,000 computers. Big information technologies like Hadoop are also real. They are not always used logically (do not bother using more than 100 nodes in hadoop clusters, as non-clustered machines that do not work well for you can work well). But of course, people write such software. But most of the work that is being done is not data collection. This is Extract, Transform, Load (ETL), so replace the database. Instead of using a database with structure, indexes, and accelerated queries, the data is just hacked, and when you figure out what to do, you re-read all the data and extract the data you need, transit, and upload it to an Excel spreadsheet. Because after selection, extraction, and conversion, it is usually not “big”.

The volume of digitalized data in the world is growing exponentially. According to IBS, in 2003, 5 exabytes (1 exabyte - 1 billion gigabytes) of data were collected. In 2008 it reached 0.18 zettabytes (1 zettabyte = 1024 exabytes), in 2011 it reached 1.76 zettabytes, and in 2013 it reached 4.4 zettabytes. In May 2015, the

amount of data collected in the world exceeded 6.5 zettabytes. By 2020, humanity will be producing 40-44 zettabytes of digital data.

According to IBS experts, only 1.5% of the data collected in 2013 was of any information value. Unfortunately, there are now big data processing technologies in the world that can be used to extract interesting, useful information that people need from a very large array of data.

Many marketing promises of big data don't materialize. Twitter produces far less insight for most companies than your ad (unless you're a teenie rockstar); and Twitter's user base is very rich. Such a contradiction is difficult to resolve and requires very experienced statisticians.

Suitable for initial study of big data processing technologies - it will introduce you to the course easily and clearly. Gives an idea of how the abundance of information affects everyday life and all its areas: science, business, medicine and more. It contains a lot of pictures, so it can be taken without much effort.

The Big Data book, which explains how to work with big data on a "simple to complex" basis, is also useful for newcomers. The initial stage covers many important aspects: processing, visualization, OLAP, as well as some methods of data analysis and classification.

A practical guide to using and working with large data using the Python programming language. Suitable for both engineering students and professionals who want to deepen their knowledge.

Hadoop is a project designed specifically to work with distributed applications that perform actions on thousands of nodes at once. Getting to know him will help you understand in more detail the practical application of big data.

Big data, "Big data" has been a topic of discussion in the city's information and marketing press for years. And it is clear: digital technology has entered modern human life and "everything is written". The volume of data on various areas of life is growing, and at the same time, data storage capacity is also growing.

Most experts agree that accelerating data growth is an objective reality. Social networks, mobile devices, metering device data, business data are just a few of the types of sources that generate huge amounts of data. According to research, IDCR Digital World was published in 2012, and in the next 8 years the amount of data in the world will reach 40 Zb (zettabytes), which is equal to 5200 Gb for each inhabitant of the planet.

Problems: The problem is that Hadoop is good for linear problems, but most of the data is not linear. And nonlinear algorithms don't have big data; study linear time approximations carefully and you should lose with more accurate data loss, which should be smaller than you lose by processing smaller data. A good example of this trading problem is the k-tool. K is actually (mostly) a linear problem; so it is possible to move a little in Hadoop. The only iteration is linear, and if you have a good program, it will fit big data. However, the number of approximations also increases with the size of the data set, and therefore it is not linear. However, since this is a statistical method for finding "tools," the results are not actually large with the size of the data set. So if you can use k-tools in big data, it does not make much sense - you can just sample the data, use a very efficient single-node version of k-tools, and the results will be just as good. Additional information is explained by the accuracy of the value, which should not be obvious to you. Since this applies to so many problems, it does not seem to open up in a real data mine in Hadoop. Everyone tries to do it, and many companies sell these things. But it does not work much better than the smaller version. However, if customers want to buy it, companies sell this feature. And if you get a grant, researchers will write articles about it. Whether it works or not. It is life: There are a few situations where all of this works. Google search engine - Cern. But also image recognition (but without the use of Hadoop, GPU clusters seem like ways to get there) has recently taken

advantage of the increase in data size. But in any of these cases, you have clean data. Google indexes everything; Cern removes any interesting data and only analyzes interesting metrics - there are no spammers who send spam to Cern ... and when analyzing an image you don't use an internet camera or random images, if any, they you evaluate them as random images, not as competent data).

The problem with weighted data is that the various pieces of information that have been collected over the decades are still very important and open to any system. Another major problem is the cost of processing them. This can include the cost of expensive equipment and the salaries of qualified information professionals. Obviously, the equipment needs to be updated regularly so that it does not lose its efficiency as the data volume increases. The third problem is the large amount of information that needs to be processed. For example, studies yield results in a large number of studies rather than 2-3 times, because it is very difficult to separate data from the general flow and make an objective assessment in order to have a real impact on an event.

Data loss problem. Precautions should not be limited to simple one-time backups, but at least 2-3 backups. However, the increase in volume further complicates the backup - IT professionals are trying to find the optimal solution to this problem.

The bottom line is that we can not hide from technology. Big Data is changing the world, slowly penetrating our city, our homes and our gadgets. How fast technology is taking over the planet is hard to say. But one thing is clear - as Bob Kelso said in the series "Clinic", keep the tradition or go back to the old way.

REFERENCES

1. Hilbert and Lopez, "Technological Capabilities of the World for Data Storage, Communication, and Computing," Science, 2011 Global.
2. N.Rahmonov. X.Tolibjonov. Advanced countries in the digital economy and Industry 4.0. Conference "Hopeful youth of Uzbekistan". March 2021.2.
3. Andreas Vaygend. Big Data. All technology in one book.
4. Alexey Blagirev. Big data. Simple languages.