

UDC 005

BIG DATA DEVELOPMENT TRENDS, RISKS AND PROSPECTS

STEPAN KUZNETSOV, SERGEI SURTO
 Polotsk State University, Belarus

Management of the structured and unstructured data with the help of new technologies and instruments of their processing – a subject, now relevant both at the level of the concrete enterprises, and at the level of the state. Big Data is even more often discussed at various actions.

Big data is designation of the structured and unstructured these huge volumes and considerable variety which are effectively processed by horizontally scalable program tools which appeared at the end of the 2000th years and alternative to traditional database management systems and solutions of the class Business Intelligence [6].

The IDC company predicts growth of the market of technologies of processing of big data and business analytics by 12% a year till 2020. First of all, growth of the analytical Big Data platforms using machine learning and the systems of artificial intelligence on the basis of neural systems that is demanded by business for the analysis of data on consumers and forecasting of their behavior is expected. The concept of the Internet of Things (IoT) has a great influence on the market. In 2016 the largest volumes of world investments were directed to the Internet of things in the sphere of production, cargo transportation, power supply systems, a consumer segment. The most rapid growth of the Internet of things is expected in insurance area, the sector of consumption, health care and retail trade.

The international analytical agency Gartner predicts increase in demand for experts in the field of data management and increase in number of departments in this area as it is expected that it will lead to maximizing profit. More and more popular is a position of Chief Data Officers (CDO) [2]. Initiatives in the field of analytics, data management, development of strategic tasks and integration of analytics and data in strategy and also ensuring information security at the enterprise have to become its main objective.

Recently many researches are conducted in the field of so-called big social data. A number of experts considers this direction as separate science. Here it is possible to allocate four significant subareas — social computing, Big Data science, data analytics and computational social science(CSS) [5].

The problem of big data covers methods of classification, processing, storage, search, exchange, the analysis and visualization of data. Big data usually have to be processed on crossing of four requirements characteristics known as "requirements of four V":

- the volume of data (Volume), is supposed that and to methods of their processing refer to category of big data only massifs from several terabytes. Quite often monthly accumulation of data array the required volume is characteristic of large enterprises. These data demand the appropriate means of safety relating to the Big Data Security applications;

- speed of data processing (Velocity) if to consider concerning the organization of methods of safety, is meant detection of anomaly and prevention of invasion in real time. The analysis of data, the deep analysis of packages allows to minimize damage of the attack or unauthorized leak from accounts of users;

- variety. Big massifs have to not only correspond on volume to big data, but also turn out from different sources. Concerning cyber security, a variety of data sources and their vulnerabilities is at the same time estimated. Logs can be analyzed, be estimated on the structured and unstructured data, different network streams. For example, going from e-mail, a set of coherent mobile devices of users, by means of capture and the deep analysis of packages, at the same time to be monitored vulnerabilities and to be prevented invasions.

- objectivity (Veracity). Accuracy and reliability of data is meant, from the point of view of the organization of safety – the corporate data belonging to the category big have to be provided with the security aids providing their confidentiality, integrity, availability of sources, such as logs and external channels of data. The relevance of these requirements grows especially when using methods of anonymization of traffic which mean, first of all, destruction of structure of messages and chaotic processing of IP packages.

In connection with the processing of social data, users of various resources have legitimate concerns about the use of their data: where and how they will be stored, who and how can access them, how long they will be stored. Experts confirm the existence of imaginary and real dangers associated with Big Data. To the imaginary can be attributed the effect to the polls. As an example, we can bring the recent scandal with Cambridge Analytica. The company gained access to data analysis of nearly 50 million users of social services. First, a small psychological test was

held by a small group of users, they even paid for it. They gave their consent to the processing of the data, thus, formally did not violate the privacy of the user. But then from their profiles pulled profiles of friends — and on it explicit permission wasn't any more. The data was processed, and the person began to receive commercials designed specifically for him about the candidate. It is difficult to assess the extent to which concerns about big data and its use may have an impact on elections, as many factors play a role. But some research in this area can serve as a theoretical justification for these fears. Research by the famous scientist from Stanford graduate school of business Mikhail Kozinski. One of his studies is devoted to the field of psychometry—the science at the intersection of psychology and sociology, which allows to attribute a person to a particular type of behavior and predict his actions on the basis of sociological data. The scientist M. Kozinski and his colleagues developed a method that allowed analyzing likes on a social network. The method could determine with very high probability the skin color of the user, his sexual orientation, commitment to the party. There are opinions that this technology was applied by Cambridge Analytica to preparation of personal political advertizing during the election of the president of the USA.

The question of the use of certain technologies is associated with the goals of their use and with those in whose hands they fall. Here you can draw an analogy with nuclear development: you can "use the atom" for peaceful purposes - nuclear energy, and you can make an atomic bomb. Similarly, processing and analytics technologies for big data can be used to prevent abnormal situations and transport disasters, to reduce equipment breakdowns, identify customer preferences, and determine the need for medical services. And the same technologies allow filtering people for some purposes, for example, for hiring. Real dangers include the use of Big data for commercial purposes. But what really develops at a rapid pace and has all chances to become one of the main investment targets in the future is the Internet of Things (IoT) [4].

According to experts, there are four important aspects of the use of big data: the data itself, analytics, people, tools [1]. Structured and unstructured data can be distinguished, and human-generated data (texts, office documents, manual input data) and machine-generated data (operational data, satellite images, scientific data, photo and video) can be extracted in both types of data. Now interest has shifted to unstructured data.

There is a number of problems in the field of big data, in particular:

1. Question of quality of unstructured data. It is possible to face a fair share of counterfeited content in the Internet. For example, there are people and programs of artificial intelligence for writing of responses both positive, and negative.

2. Big Data is not only unstructured information. So Banks of the world, except tasks of the analysis of addresses of citizens, identifications of affiliates on the Internet, passes to collecting and processing of financial micro data and operational data from credit institutions that will demand use of Big Data technologies.

The greatest prize from use of Big Data will be received by predictive analytics and operational analytics. But the new direction in business analytics of data discovery continues development. There is an opinion that Data scientist has to have knowledge in many areas, often on a joint of different disciplines. In fact, experts combine on several roles.

Application areas for Big Data technology.

Large and medium-sized companies from the trade and service industries show great interest in using Big Data technologies. These technologies are actively used by banks, mobile operators. In addition, they are used by large manufacturing companies to analyze data on equipment breakdowns and reduce downtime, which allows to reduce costs. For example, in the field of flight control, the analysis of data arrays can increase the reliability of equipment and reduce the number of failures.

But scope of Big Data is much wider.

For example, business purchase can be one more of scopes of Big Data technology. It is about the broadest coverage of various information on the company. It and analysis of ordered data: physical characteristics of objects; financial data. Can be sources of information: file tables, traditional DBMS, accounting systems. It is necessary to analyze also the disordered data: customer reviews; applications for service; IT infrastructure. Here can be sources of information: tables, schemes, social networks, estimates of experts. The so-called passport of an object including data on arrangement, the area, number of storeys, allowing documentation, inventory data, a competitive environment, historical financial data, a seasonal factor of sales can become result of the analysis of such data.

But, as we know, the cost of the company is not equal to the simple sum of cost of its tangible assets. Collecting and the analysis of information on Big Data technology will allow to estimate the cost of intangible assets. It is possible to carry to them: human resources; knowledge and abilities; information resources; organizational and administrative structure; potential of workers; brand; reputation; developments; bases of clients.

Big Data and digital economy:

Abroad already became a norm when at medium-sized and large enterprises the position of Chef Data Officer is entered. Most often he submits directly to the top management. For our enterprises, even large, such

position still remains a big rarity. There is a deficiency of experts in the field of management of big data and the corresponding analytics. The universities in principle practically do not train such experts. Young specialists from the IT sphere should be finished learning, retrained, besides, that they often very overestimated requirements for starting salaries taking into account insufficient qualification.

REFERENCES

- 1 Big Data of danger and prospect [Electronic resource] / Article. – 2018 – Mode of access: <https://www.eg-online.ru/article/372363>. Date of access: 10.11.2018.
- 2 RUSBASE/What is Big Data: collected all the most important about big data/ [Electronic resource]. – 2018 – Mode of access: <https://rb.ru/howto/chto-takoe-big-data>. Date of access: 10.11.2018.
- 3 PROMDEVELOP/ Big Data - What is a big data system? The development of technology Big Data [Electronic resource] / Article. – 2018 – Mode of access: <https://promdevelop.ru/big-data>. Date of access: 10.11.2018.
- 4 TADVISER [Electronic resource] / Article. – 2018 – Mode of access: <http://www.tadviser.ru>. Date of access: 11.11.2018.
- 5 COINNET [Electronic resource] / Article. – 2018 – Mode of access: https://coinnet.ru/big-data-sovremennyj-trend-informatsionnyh-tehnologij/#_Big_Data. Date of access: 11.11.2018.
- 6 Big data [Electronic resource] / Wikipedia, the free encyclopedia. – 2018 – Mode of access: <https://ru.wikipedia.org/wiki/>. Date of access: 11.11.2018.