

Problems, opportunities and prospects of "big data"

Jamalova Jala

Abstract— The article shows how to manage Big Data sets. Areas of application of Big Data and efficiency are considered. The issues of data security and the possibility of using the "openness" mechanism for data management were discussed.

Index Terms— Big Data, analytics, data protection, global agreements

I. INTRODUCTION

For effective data management, accessibility, continuity, usability, scalability, the ability to manage at various levels of detail, privacy and security solutions, and data quality control are important. If the main requirements for data processing by traditional methods are: Reliability, Stability, Security, Visibility, then processing of large amounts of data and unstructured information is added to these requirements: Physically and Logically. It should be noted that "Big Data" is so capacious that it is possible to characterize them much more widely.

II. MANAGEMENT OF BIG DATA MASSES

Analyzing large data sets is a culture of thinking. The ability to build the logic of reasoning and statements based on the interpretation of data, integrated their different areas of science and technology, make judgments based on incomplete data, the ability to integrate data from different sources and integrate them in the context of the task at hand, and to draw conclusions based on data analysis (2). Big Data is a data set that is too large in volume and complex to be processed. Using traditional methods of data processing is expensive or even in some cases impossible. For the maintenance of Big Data, computing clusters that realize parallel processing and a different style of system management compared to traditional transaction systems and data stores began to be used. Analysis of the literature suggests that in the mid-90s, scientists in laboratories decided to use clusters from a large number of small computers connected by conventional technology with the same programming paradigm as supercomputers, for tasks with large volumes of data instead of expensive supercomputers. Files are not stored on each computer. Files are broken into small pieces, the pieces themselves are scattered across different computers with duplication. These technologies are made open source and so the programmers were able to put more energy there and popularize Big Data.

In connection with the exponential growth of data, the inability of analysts within a time, close to the real one, the special technologies are required - this is the use of data directly located in the database for the purposes of analytics. Hadoop is used by analytic and parallel DBMS. This is an open-source framework designed to create and run distributed applications. The Hadoop cluster consists of many machines that store and process large data sets in parallel. Client

computers send jobs to this computing cloud and get results [3]. The key of Hadoop to success is the efficiency of network usage and high reliability. Hadoop consists of two main components: the distributed cluster system of the Hadoop Distributed File System (HDFS) and the Mar / Reduce programming interface. A number of products have been developed based on HDFS and Mar / Reduce. The Big Data technology distinguishes with high flexibility, high scalability and application of cloud-based technologies.

III. ANALYSIS OF OPPORTUNITIES AND PERSPECTIVES OF BIG DATA

Advanced foreign journals describe modern technologies and effective management solutions in a wide range of large data processing areas. They are «Journal Database», "International Journal of Big Data", "International Journal of Internet Science", "Journal of policy analysis and management (2014) which in its article "Big Data in Research on Social Policy", focuses heavily on predictability of many policy decisions with the use of Big Data. The monthly journal "Nature Biotechnology" and many medical journals publish the use of large data in medicine. "Journal of Computational Science" (2012) in the article *Advanced Computing Solutions for Health Care and Medicine* provides an analysis of bioinformatics of "Big Data". In the article *"Toward a Big Data Healthcare Analytics System: A mathematical Modelling Perspective"* (IEEE Computer 2014) a decrease in the mortality and disability of newborns in the neonatal period with the help of large data is predicted. Journal of Big Data covers how to conduct a medical examination of the population in the article "A review of the data mining using the high-quality data in health informatics". The information collected together and analyzed with the help of Big Data theology gives efficiency in many fields of application of Big Data. Using the Big Data, the Medicare Chronic Condition Summary file (4) helps to examine in detail the factors that affect chronic diseases, health prevention, morbidity and mortality.

Informatization of health care with a view to improving the efficiency and quality of medical services is given great attention. In order to improve the health of the population of each patient and generally, the population, the author [5] defined a 4-level survey that includes: bioinformatics, neuroinformatics, clinical informatics and public health informatics. Today, "Big Data" technologies are used to process video streams, and there are solutions for distributed detection of objects and events, for example, in the distributed processing of data streams coming from video surveillance cameras to provide automatic response to irregular phenomena. Video stream processing systems should be distributed, and the rules for them should be universal for all protected objects, so the use of "Big Data" technology is of high demand here [6].

For higher accuracy of notification and general threat analysis, the author [7] considered Data Fusion, heterogeneous intrusion detection architecture and security event (SIEM) systems management. Intrusion detection enhancements can be achieved by encompassing a wider range of heterogeneous sources. Technology of Big Data for intrusion detection can help solve these problems.

The basis for the analysis of corporate databases is the risk management system - the system collects data from various sources and calculates the risks for each individual employee, and then calculates the overall risk. For risk management to be truly effective, it is necessary to implement it on a continuous basis throughout the life cycle of the project [8]. Geo-databases can be useful in the work of environmentalists, disaster response teams and other professionals. A cartographic platform is being developed, where specialists visualize data on the number of populations of rare animals to protect them from extinction. For this project, called Earth Insights, cameras and climate sensors were located in the tropical forests of 16 countries around the world, collecting data on animals, vegetation, temperature, precipitation, humidity, etc. Manage three terabytes of information, including more than 1.4 million photographs and more than three million climate indicators. [9]. The purpose of the Columbia University Medical Centre is to provide timely information to health professionals. If one monitor patients' status in real time, then advanced analyst diagnoses serious complications of brain aneurysm damage for 48 hours before [10].

"Not everything that is technically possible is ethically acceptable," emphasized Benedict XVI, referring to the world community. Working with Big Data calls into question the boundaries of legal human rights (11). Information from us and about us probably responds not so much to social, ethical and political logic as to the logic of technical capabilities: to what the Internet allows." This raises the following questions: information from social networks – is property or as independent information? The right to title can be claimed by unborn relatives whose personal life may be at risk, in the case of disclosure of information about blood relatives. Definitions of location, disclosure of information about surveys. Example: If an accident occurs in New York, for an emergency doctor it will be very important to have information about chronic diseases or allergies to medicines. Quick access to such facilities may help to provide a qualified aid. There is a regional health infrastructure program in the USA. InterSystems participated in the implementation of this program in New York, the system created contain a data for 25-30 million patients [12]. It is necessary to stimulate people who have shared their symptoms of disease. In response to targeted recommendations of the doctor, the patient receives free treatment, procedures. It is necessary to study moral problems. All this can be called - the problem of managing data collection. When making important decisions in the state administration, it is important to compare data from different, unrelated sources, processing credit cards, fingerprints, and transport operations. Big Data are usually aggregated from different sources, for different time bases and collected in some data sets. It seems to us necessary, to link records among themselves, to transform them, to create new variables, while documenting all data transformations.

In countries such as Germany, Australia, Spain, Japan, Brazil and Pakistan, Big Data technologies are used to solve

national issues. These technologies help public authorities more effectively provide services to the population, provide targeted social support. Taking into account these forecasts, the market volume in 2020 will be equal to 68.7 billion US dollars [13].

IV. CONCLUSION

IDC experts identified 3 transforming factors that affect the state of the Big Data market [13]:

- Mass absorption of the client base of companies offering mobile applications and other data platforms;
- Development of cloud infrastructure;
- Changes in data privacy laws

According to the Accenture poll [13,14] questions of data security issues are now the main barrier to the introduction of Big Data technologies. More than 51% of respondents confirmed that they are worried about data protection and their confidentiality. 47% of companies reported on the impossibility of introducing Big Data due to a limited budget, 41% of companies indicated a shortage of qualified personnel as a problem.

In the field of implementation of Big Data technologies, it is necessary to organize stakeholders to achieve global treaty-related issues - on matters of network surveillance, personal data protection, network neutrality and other key issues. It is necessary to adopt an international convention on the possibility of applying an "openness" mechanism for data management, which will include the shared principles of management, integrity, network neutrality.

Organization of effective data management is a priority.

REFERENCES

- [1] «How data and knowledge grew»- online magazine on economics and politics/ <http://slon.ru/specials/data-economics/other/timeline>.
- [2] http://koi.nsu.ru/new/website/koi/var/custom/File/090401_M_TRPS_Ana_litika%20bol60shik//
- [3] Ch. Lam. "Hadoop in action", M. DMK Press, 2012, pp. 23-25
- [4] D.Parker. National Health Interview Survey, NHIS, 2013
- [5] M.Herland, T.Khoshgoftaar, R.Wald, A Review of data mining using Big Data in health. Informatics, Journal of Big Data 2014, 1
- [6] V.Korzhev, Big Data in service of security./Open systems, 2013, No 9, pp 4-49.
- [7] R.Zueh, T.Khoshgoftaar, R.Wald, Intrusion detection and Big Heterogeneous Data: a Survey, Journal of Big Data, 2015, 2
- [8] A.Zakis, How to apply the risk management/Intelligent Enterprise, 2003, No 13
- [9] HP Health Insights. A solution God might choose to monitor real life. www.enterprisecioforum.com/en/blogs/enadhan/hp-earth-insights-solution-god-might-cho
- [10] W.Ranghupati, V.Ranghupati, Big Data analytics in healthcare; promise and potential/Health Information Science and Systems, 2014, v 2
- [11] <http://media.ls.urfu.ru/182/1063/2120>
- [12] Big Data in medicine and healthcare, <http://www.sop.ru/medit/2014/04/13040834.html>
- [13] Analytical survey of Big Data market, <http://habrahabr.ru/company/moex/blog/256747>
- [14] Accenture Big Success with Big Data Survey 2014. <http://Accenture.com/SiteCollectionDocuments/PDF/Accenture-Big-data-POV.PDF>

I, **Jamalova Jala Siyavush** was born in Baku in 1965. In 1995 I graduated from the Azerbaijan State Technical University on mechanical engineering specialty and I work as an assistant on my specialty.

The direction of her scientific work - systems and networks.

Scientific direction area - computer complexes, systems and networks