

Modern aspects in development of branch applications on the basis of Big Data: possibilities, prospects and limitations

Ramzaev M.V.

International Market Institute, Samara

Abstract. Storage, processing and intellectual analysis technologies of Big Data receives extended popularity throughout the world. The areas of implementation are also expanding, opening up new possibilities for solution to a variety of problems. However, it is important to consider the risks and limitations in implementation of such technologies and conduct detailed studies into revealing such issues. Therefore, it is necessary to form new analytical algorithms for Big Data, considering combination of wide use possibilities and limitations within rapidly changing conditions. One of the prospective areas for the comprehensive application of Big Data technologies is the State government. Having conducted the study into potentials of such technologies it's possible to formulate generalized methodological approach to working with them, which will allow resolving a number of current problems in real time.

Keywords: IT technologies, IT innovation, database, Big Data, Big Data analysis, Big Data risk, data storage, data processing, forecast, competitiveness of the territory, data security, Big Data competence

Citation: Ramzaev M.V. Modern aspects in development of branch applications on the basis of Big Data: possibilities, prospects and limitations. Proceedings of Information Technology and Nanotechnology (ITNT-2015), CEUR Workshop Proceedings, 2015; 1490: 355-363. DOI: 10.18287/1613-0073-2015-1490-355-363

In today's world more and more spheres of human activity are covered by various types of information technologies and their various field applications. Despite the fact that the existing mechanisms of data storage and processing are actively improving, the relevance of development and innovation in the software sector is one of the most important challenges for the economies of countries around the world. The most advanced, discussed and, according to researchers, promising in this field directions are the processing, storage and intelligent analysis of Big Data using cloud technology.

These technologies are already widely used in many fields of science and life. For example, marketing, energy, industry, oil and gas, public and municipal administration, housing and utilities (smart city), banking and insurance activities, bio

- and nanotechnology, GIS technology, the aerospace industry, the safety of various areas of life, health, etc.

The use of Big data technologies in Russia according to the analytical Agency CNews Analytics and Oracle is gradually increasing [1]. However, at the moment there is some skepticism about such trends, which can considerably limit the spread and application of these technologies. According to the international expert community, one of the most problematic issues of implementation and spreading of Big Data Analytics is currently the confidentiality of personal information.

Every year the number of companies which provide access to a variety of databases, as well as provide services for the analysis increases. For example, according to the analytical Agency Mind Commerce market size of Big Data in the U.S. in 2013 amounted to \$ 20 billion, and in 2014 already 29 billion, which means that growth was 45%. This is an important sign of the framing trends and tendencies in the IT sector.

You can also give an example of a study made by Terradata on the subject of practical use of Big Data technologies. The study is based on survey data 316 of managers engaged in information technologies and data processing in leading companies. The results showed that the vast majority of respondents stated a high proportion of their investments in big data analytics, and high rates of return.

About a fifth (21%) of respondents support the idea that technology is big data Analytics are a key activity to identify and create specific competitive advantages, and 38% include these technologies in line of five priorities.

However, the study showed that Big Data is promising and effective from side of innovation capacity in three key areas: the creation of new models of business activity (54%), search and formation of a new product or service (52%), selling data to third-party commercial organizations.

However, according to the results of the study there are a number of obstacles on the implementation of IT technologies in the current activities of the company, behavior models and strategy.

Thus, according to survey, one of the key tasks in the field of corporate culture is encouraging the use of data and experiments in working with them [2].

In this regard, due to the lack of a regulatory framework for activities in the field of storage, transfer to third parties and Big data Analytics, it is logical to assume that there will be a growing risk of wrong use of arrays of information.

However, the law existing in Russia [3] imposes a number of limitations and identifies new opportunities in the BIG DATA analysis.

Although now you can get "personal portraits" of any particular person based on the mappings of IP addresses. And it's not just bank accounts and property register list, but also the possibility to calculate the algorithms of the codes needed to control them.

However, one of the most soft elements of personal data is personal health information. The current trend toward cloud storage also provokes risk increase of information leak.

On the one hand, the organization of the process of collecting and analyzing all available and incoming real-time information about the medical data of millions of

patients around the world can have a significant positive impact on the health sector as a whole. That is true about defining the most effective drugs, schemes and methods of treatment, species diagnostics and other medical interventions. On the other hand, the patient is one of the most weak points in this chain. Patients have no influence on health information processing. He has the right only not to personalize the data and remain anonymous. However, anonymity in this case is a relative term, the guarantee of which is impossible to provide to full extent.

Close to medicine–genetics area recently experiences some difficulties in storing and processing of information, the volume of which is growing rapidly. According to rough estimates of experts, the scale of genetic data is comparable to the scale of the data in astronomy and physics, and in some ways even surpasses them. Therefore, it is logical to assume that further research in this area will soon need a completely new analytical and computational tool is capable of delivering results based on such a large streaming data. Of course, this will require substantial funding, but the scientific and practical relevance of this topic will allow access to all necessary means and resources, both financial and intellectual. However, in this scientific field, there are also certain potential risks from a wider and deeper implementation and application of Big Data technologies, but they are long-lasting and today represent a much smaller threat than potential usefulness.

But experts have other concerns too. They are connected primarily with the fact that almost any analytical agency will be able in real time to segment the society on different sets of criteria that, on one hand, will allow to predict public reactions within each segment, on the other hand simplifies the process of manipulating them for personal gain. Moreover, the prices of all these analytical services in the near future will only become more accessible because of the growing number of proposals. Due to the Economist Intelligence Unit agency survey, which surveys have identified the priorities for the development of Big Data for several years ahead, one of the most dramatically developing areas is human resource management [4]. Thus, it is believed that the mechanisms of active control in society have long been used, for example, in the United States. The expert community believes that the commercial interest of companies to this subject over the next three years will increase by almost 2 times in comparison with 2014.

Another aspect of the complexities of Big data usage, as Professor of Berkeley University Michael Jordan says, is associated with a qualitative segmentation of information and its analysis. The increase in the number of data, the increasing number of data sets and the increasing level of accessibility to them forms specialists trend of inflated expectations, which shows the history of the development of scientific disciplines that often leads to contradictions and unreliability of the results afterwards. In this regard, it is important to pay attention to the study of not only research and development opportunities of Big Data, but also to the research of the risks of negative consequences, constraints, and adapting. For example, if the user for several months studied the information about TV sets and then bought it on-line then he will no longer be an effective consumer for targeted marketing that will haunt him for another few months and more in the Internet and mobile devices. Hardly anyone will buy a TV every time you see an advertising banner in the corner of your screen.

There is also significant potential from the use of Big Data technologies in industrial production. Especially it concerns the production of large objects, parts produced and supplied by many suppliers. We are talking about millions of parts and thousands of manufacturers. So in Russia and in the world exist and work quite effectively standards of quality and technological features, on the other hand, the existing mechanisms do not allow to lead full adaptation of innovative design in real time in different areas of production and objects.

However, the development of Big Data technologies and their attendant widespread adoption of the Internet of things, allows not only to account for all the standards and requirements in the design of any component or construction, but also to optimize the production and bring to a new level control over each part, each node and object in General.

Such approach makes possible to create new mechanisms of standardization of the quality of production, which will run in real time. That is, in case of identifying any inefficient work of a particular piece, which is provided with a sensor, the data will automatically be transferred to the computer not only of the owner of the object itself, but to all project agencies and designer.

Thus, the use and implementation of Big data technologies in industrial production will reveal the rejection and its causes in the shortest possible time, and will also make it possible to significantly improve the predicting risks system associated with quality problems, use and operating characteristics of parts, components and objects in each case.

At the same time, technology in the world of Big Data are applied directly in the production process. Such innovations will allow, for example, to identify the optimal values of components in compositions of various metals, which, in its turn, will impact on the economy, on the accuracy of risk calculation and lifetime of the final products of production.

In modern terms, such global innovation is possible only in case of effective and constructive cooperation between large companies and corporations, bringing together the experience and tools involved in the creation of new technologies that will be useful to all partners. This also applies to the IT sector, and technologies of collection, processing and analysis of Big Data, as to extend the scope of their application is appropriate with the scale of funding and potential profit.

Another important field of application of Big data technologies is the Banking sector. On the one hand, existing in large banking institutions, the mechanisms for collecting, analyzing and processing Big data to solve a wide range of existing problems, on the other hand, current economic conditions and technology development of electronic fraud make Big Data a new vector of innovation in the work of the serious financial structures of the world level. First of all, the Bank needs to know and understand its customer. And it's not just his average annual income and the property which he owns. In particular, it its a detailed expenses structure with regard to the season of year and time of day. The Bank, as a credit institution must not only guess the possible customer need for money but has to some extent to form this need in order to be able to offer its products on the terms that will be acceptable here and now, in real time. These are the conditions of the market today. At the same time,

the Bank should partner with its customer and take care of the security. Widely implement of Big data technologies will be possible to warn the client at the time of any transaction, if the payment recipient for example showed himself earlier as unfair. Or, identifying special features of the movement of any client's funds in his account, to start negotiations with him about possible difficulties when repaying them next payment of the loan, for example. That is, to solve the customer's problem with him, without waiting for the consequences of those difficulties, which could result with significant debts.

Overall, it should be noted that the banking and financial sector nowadays are one of the most interesting directions in the search for practical applications of the use of Big Data technologies.

First of all, the reason for such trend is a direct customer economic benefit. It should be noted that previously the average growth in this sector was about 80% per year. Now competition among banks is already underway for each client, in addition to this is the need for a deeper and wider analysis of the current situation for each existing customer, the decision to change the bank is not a rare thing now.

Any miscalculation in individual work with consumers can lead to the situation when within an hour is not understood by the employees of one Bank, the client will become a client of another Bank. Therefore, the internal Bank study on prediction of customer care becomes a very urgent and necessary from the point of view of efficiency and appropriateness of current activities in general.

To create more definite personal portraits are used social networks, search engines, existing statistical data and data submitted by the clients themselves. Information about the state of the financial and banking sector in general is collected from data provided by the Central Bank and state statistical reports. Although, almost all of the specialists working with these sources note that the quality and the list of such data and the speed of changes are the essential problems in the development of IT applications innovation in the financial sector in Russia.

Competition in the banking sector has led to the situation that market participants are directly interested in the development of new outlets and applications of Big Data technologies in their work because it depends primarily on their profits. In particular, technologies of big data analysis help to rebuild the geographic locations of Bank branches, taking into consideration the passability. This allows to reduce their number and to calculate the necessary and sufficient space and staff for each object separately.

Interesting developments on framing personal practical recommendations for optimizing the client's expenditure, part of which indicates the amount of savings per month. For sure it can be taken or left behind, but the potential for success lies in those elements of expenditure that is least painful for each individual client. This calculation can be particularly effective if the person specifies, for example, the subject for whose benefit saving are held and ongoing savings. This approach largely corresponds to the concept of working with clients based primarily on exactly the formation of a trusting relationship between the Bank and the customer.

However, now relevant are the issues of innovative projects financing and start-up companies, particularly in the area of small business. All this is now starting to be addressed through non-banking technologies on the basis of the Croud and the Croud

Sourcing Funding. Now there are many forms of this kind of financing – from philanthropy to the acquisition of a stake in the project, indicating a wide distribution of these phenomena in the world.

In this area a variety of approaches in handling and use of big data can be used in identifying the most relevant and popular areas of business to create a startup project for a specific target audience and on stage to test the idea of its potential popularity and consumer demand. In this connection it is useful to optimize the search of potential Internet investors in accordance with their personal content. That is, to propose to the users mainly the ideas that are most likely to interest them. Thus, we are talking about target management not only the realization of a particular innovative project or business idea, but also the process of developing this new product, search for optimal forms of financing and search of investors.

However, with the development of computer technology and the widespread use of electronic payments, real time data (database, updated in real time) of any Bank is a threat to all clients, because such information when released into the hands of fraudsters, makes possible to deal with all customer accounts without their knowledge. Of course, the security services of major financial institutions are constantly monitoring these risks and try to prevent them, but without continuous improvement it is the computer protection mechanisms with the use of Big data technologies it will be extremely difficult. Don't forget that fraudsters often use innovative technology and are no less competent than those who protect them.

In the context of the possible application areas of Big Data technologies is increasingly seen public administration, including social services, and housing, security, economics and much more.

As for the security sector one of the most interesting experience is the one of Pakistan, there was created one of the world's largest database storage and processing of data of citizens, including their biometrics NADRA (National Database & Registration Authority), the national Agency for managing databases and registrations of citizens [5]. The main strategic goal of establishing such a system and the organization of the information base is the fight against corruption, combating terrorism, optimization of the process of reforms, as well as improving the efficiency of working with natural disasters. In general, the expert community agrees that this initiative is implemented successfully and today is a unique implementation model of IT innovation and Big Data technologies in public administration. At the same time, the NADRA today – is an independent and financially independent organization, which increased its income by 3 times over the last 5 years.

Considering the peculiarities of application of modern technologies of Big data in public administration, it seems appropriate to adapt a number of existing models and mechanisms in this category for new software and hardware capabilities. Information technology today is one of the prior directions of development in public administration on the Federal, regional and municipal levels. Thus, it seems logical the use of the concept of regional competitiveness, as a base for recording and processing of streaming data. Existing mathematical models for calculating the level of marketability of the territory is able to show results to quite a sufficient degree of

accuracy in the presence of true meanings of the required performance criteria in real time.

Developed additively weighed mathematical model of management of development of competitiveness of the territory has the following form [6]:

$$KS = (\xi_1 * GF + \xi_2 * PRF + \xi_3 * EF + \xi_4 * PPF + \xi_5 * APF + \xi_6 * SF + \xi_7 * FEF + \xi_8 * IFF + \xi_9 * RF + \xi_{10} * IF + \xi_{11} * InF + \xi_{12} * DF) \rightarrow \max,$$

where KS - marketability of municipal area; ξ - coefficient of importance of group of factors (is defined by expert questioning); GF - geography factor; PRF - nature and resource factor; EF - ecological factor; PPF - industry production factor; APF - argo-industrial factor; SF - social factor; FEF - economy and finance; IFF - infrastructure factor; RF - municipal area development factor; IF - innovation factor; InF - investment factor; DF - spiritual factor.

Each of the factors of this model is a separate function that takes into account a number of indicators of regional and municipal competitiveness. Factors of importance in the present formula are constant, which are usually defined by analysis of the views of the expert community. However, with the use of Big data technologies, they can be calculated almost in real time.

To do this we need to create queries criteria in the Internet space to identify the expert community on this or that factor, and then by personal content to determine the importance of a necessary factor in the opinion of the user with the desired characteristics. In this context, it is logical to establish and run the search algorithm of significant correlations excluding the binding to the experts. Thus, it is possible to detect a peak of the correlation values of certain ratios of opinions of experts and Internet content in general. Based on the results of this analysis are generated coefficients of significance and the calculation of the model. At the same time, while analyzing the streaming information in real time you can predict the occurrence of previously unforeseen factors and conditions for change of the model of competitiveness or the occurrence of critical values of significance of a factor (which may be indicative of the likelihood other emergency outbreaks of epidemics, the occurrence of environmental disasters, social disturbances, etc...).

Thus, on the basis of the study of the potential of Big Data technologies in state and municipal management, you can generate a generalized methodological approach in the use and implementation of these IT innovations. This approach consists of four main blocks, which are characterized by a certain kind of task.

The first block is the analysis and identification of certain types of relationships of current indicators of the competitiveness of the territory for now and over past periods. This will help to identify risk trends and growth of potentials of socio-economic modus. The second part is the analysis of factors of competitiveness and identifying those which can be managed in real time. The third block is predicting, which with use modern software and hardware working with stream data becomes more accurate and almost continuous. The fourth block is a simulation of possible situations. In this aspect it is relevant to consider different urgency forecasts and to rank from the point of view of optimism and dark scenario, as a result getting one

real-time integrated programme for the formation of certain decisions that will be the most effective in each particular case.

However, the analysis shows that the widespread adoption and use of Big Data Analytics in public administration (including real time administration) has a number of limitations and problems. First, is the difficulty in framing databases themselves or data sets on socio-economic indicators of particular territories. Second, the problem of the formation of streaming data: collection and analysis of information in real time, which reduces any algorithms of risk identification and optimal decisions in specific situations. Third, an integral part of the use of IT innovation is a competence of the staff, where there is a need for state service staff but also to form a scientific school, specializing in this area that it is extremely difficult for countries with low and medium level of development, as well as having an extensive territory. Number four is that Big data technologies generally do not give accurate guidance that can be adapted in a logical algorithm, they give a set of results, connection of which with the request is not always visible. This means that in reality you can only talk about the calculation of the probability of occurrence of certain situations or critical values of certain indicators, which in its turn means that the prediction based on Big Data Analytics is just about the forecast probabilities.

However, this does not mean that such approaches to public management are not effective. On the contrary, this suggests that there must be a more balanced approach to the means of storing, processing and analyzing Big data in general, taking into consideration data security issues, and the likelihood of their inaccuracies, as well as to approach the analysis from the point of view of logical connection. Fifth, the system of state and municipal management is structurally in legislative frameworks, and the management of it in real time is difficult. Many of key decisions often require certain regulatory changes, which take time. However, the application of Big data technologies in the public sector can increase the accuracy of forecasts in general, so in some cases there is an opportunity to identify in advance the values that need to be adjusted in the future and to prepare appropriate legislative framework.

Now the practitioners and scientists work with the technology for collecting, analyzing, and processing large amounts of data to active discussions on the subject of new opportunities and fields of application. Today it becomes clear that the terminology criterion value of the volume of data goes by the wayside, as almost all streaming data to date are great. Another issue is that the potential of Big Data gets its bloom just by investing respective resources. Thus, the main motor of these technologies today is business that focuses mainly on the profit from the invested funds. This means that the research in the field of technologies of work with streaming large amounts of data will be held and already held in the format of sectional or thematic research aimed at solving specific tasks oriented on obtaining profit. Success cannot be achieved without considering and analyzing all available information and all available data, it can be assumed that this area of it innovation itself can be considered as micro industry, which may find applications in most sectors of the economy and state institutions in general. This means that the need in the educational mechanisms and, consequently, highly qualified specialists will grow. There is a strong likelihood that investments in education on the subject have

practicability and huge potential in terms of payback period as a separate line. However, the demand for such specialists today has no geographical or linguistic boundaries throughout the world, and their significance in the promotion and development of technologies for working with stream data is huge. Obviously, there are a number of tasks of the global level of significance, which nowadays have not been resolved partially because of lack of data, information and mechanisms of work. Environmental issues, and forecasting and prevention of natural disasters, research in genetics and space and global security concerns are in the list. However, as described above, you should carefully consider the possible problems and threats when trying the widespread implementation of Big Data technologies, particularly in the commercialization of search process for applications.

References

1. Research of Oracle и CNews Analytics: Big Data came to Russia: <https://www.oracle.com/ru/corporate/pressrelease/study-of-oracle-and-cnews-analytics-20150226.html/>
2. http://ko.com.ua/teradata_bolshie_dannye_sozdayut_potencial_rosta_112145.
3. Federal Law from 27.07.2006 N 152-FZ "About personal data" (edited 21.07.2014).
4. Analytical survey of Big Data market / <http://www.eiu.com/home.aspx/>
5. National Database & Registration Authority / <https://www.nadra.gov.pk/>
6. **Ramzaev M.** Management of marketability of municipal area development (on the basis of Samara region small towns). Samara: Economic Science, 2009; 3.