Industrial application of Big Data services in digital economy

P.V. Sitnikov[1], A.A. Khorina[1], A.V. Ivaschenko[1], A.A. Stolbova[2], N.Yu. Ilyasova[3,4]

[1]ITMO University, Birzhevaya liniya 14A, Saint-Petersburg, Russia, 199034
[2]SEC “Open Code”, Yarmarochnaya str. 55, Samara, Russia, 443001
[3]Samara National Research University, Moskovskoe Shosse 34, Samara, Russia, 443086
[4]Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

Abstract. Despite multiple positive results in research and development of Big Data technologies, their practical implementation and use remain challenging. At the same time most prominent trends of digital economy require Big Data analysis in various problem domains. Based on generalization of theoretical research and a number of real economy projects in this area there is proposed in this paper an architecture of a software development kit that can be used as a solid platform to build industrial applications. Examples are given for automobile industry with a reference of Industry 4.0 paradigm implementation in practice.

1. Introduction

Big Data processing remains one of the key technological trends of economy digitalization in Russia. According to the National program of digital economy development one of the critical areas of IT infrastructure application is a smart factory, introduced according to the concept of Industry 4.0. This concept is based on developing cyber-physical systems capable of monitoring industrial processes, providing contextual and decentralized decision making support. Solving these problems require Big Data analysis in real time.

Therefore modern industrial enterprises are treated as a source of Big Data that describes the process of various information exchanges. Information obtained from different sources allows drawing conclusions about the processes of production and manufacturing. The use of methods and program solutions of working with Big Data, in the course of solving this problem, allows building efficient production at all levels.

Despite positive results in research and development of Big Data technologies, their practical implementation in industrial applications remains challenging. The problems are concerned with a requirement to use unified and integrated program services built on a solid platform instead of developing customized algorithms and program solutions. To cover this gap in this paper there is proposed to generalize existing experience in this area [1, 2] and develop an architectural solution for a software development kit that can be used as a solid platform to build industrial applications.

2. Technology Review

Big Data technologies incorporate series of approaches, tools and methods for processing structured and unstructured data of huge volumes and considerable diversity. These technologies are used to obtain the obtained results, effective in conditions of continuous growth, distribution of information across numerous nodes of the computer network.

Big Data processing services are usually implemented as part of scientific research in the development of research institutes, but in modern conditions organizations create a large amount of unstructured data, such as text documents, images, videotapes, computer codes, tables, etc. All this information is stored in many repositories, sometimes even outside the organization. Companies may have access to a vast array of their own data and may not have the necessary tools that could establish relationships between these data and draw significant conclusions based on them. Big Data processing technologies can make a big contribution to the industrial world.

Modern level of automation of modern production enterprises allows implementing intelligent technologies of production and business processes analysis targeting the concept of smart factory [3]. According to this concept human resources and robotic production equipment are integrated into a solid information space and form a virtual community of actors with autonomous behavior and self-organization. The concept of Industry 4.0 is intensively and massively explored in [4, 5]. It describes a composite solution vision based on implementation of modern IT technologies to develop cyber-physical systems for smart factories.

The process of interaction of users in integrated information space at modern production enterprises and supply chains generates a sequence of events of the exchange of documents, messages and other information objects. The number of the events is big (large physical data volume); they vary and require high-speed processing. In this regard, the task of managing the collection and processing of information data in the system of acquisition and processing system with a stratified architecture may be referred to the Big Data problem [6, 7].

One of the solutions can be close to subject-oriented approach for business processes management (S-BPM), which conceives a process as a collaboration of multiple subjects organized via structured communication [8]. There can be proposed a model for the interaction of actors (subjects) in integrated information space, which can be implemented using the multi-agent software. The ideas of indirect and conditional project management generating soft influence over highly motivated autonomous actors are being successfully implemented in Internet communities and social networks [9, 10].

In the course of development, a significant increase in key production indicators is expected, according to the McKinsey report on the Digital Economy [1]:

- Optimization of production and logistics operations
- Reduced equipment downtime and repair costs, due to increased equipment loading and equipment performance
- Rapid prototyping and quality control, due to the analysis of large data arrays in the development and improvement of products
- Decrease in an n expense of the electric power and fuel, due to reduction of production losses of raw materials

All these characteristics, one way or another, are related to data and the use of Big Data technologies in processing information about processes in production contributes to the efficiency of operations.

3. Big Data industrial sources

Data sources for monitoring and controlling personnel can be very diverse and include: workflow, time and attendance system, equipment sensors, PC’s monitoring programs, motion detection sensors in video cameras, call information from employee’s work phone, mailbox activity analysis, etc.

During the analysis of workflow systems, it is possible to draw conclusions about the time of work with each of the documents and the effectiveness of performing tasks on working with documents. View and analyze the movement of documentation can help identify the bottlenecks in the work with each of the documents, with the display of actual information about the employee, which slows down the chain of work on the document.

Time and attendance system allow you to analyze the work of the employee, as well as track the periods when the employee is out of the office, at the wrong time.

Equipment sensors will assess the effectiveness of the use of equipment, with the ability to determine periods of downtime. This information allows you to build recommendations for the maintenance of equipment.

Analysis of information from PC's monitoring program and identification of user activity working periods on a specific computer will allow evaluating the efficiency of using a PC and the effectiveness of an employee's work on a PC. Data from a PC will help to analyze the statistics of the location of an employee outside the workplace, when analyzing the activity of a PC in combination with motion detection sensors.

Motion detection sensors in video cameras will allow you to evaluate the performance of certain algorithms thanks to tracking the movement of a person on production and analyze statistics of the employee's location outside the workplace.

Information from all possible data sources forms a single information space. A single information space contains information about events occurring at the enterprise, and some streaming characteristics.

The events that will be monitored in the system include the following events: the fact and time of the personnel at the enterprise and at the workplace, the performance of official duties, the operation of equipment and their changes. An event we call a change in the state of production facilities or an employee by the time. Events are characterized by the objects to which they relate, the time when the event occurred, the type of event and the type of data source. The same type of event can refer to different objects. Events are discrete values.

Streaming characteristics are a continuous value describing the trajectory of a person's movement. Critical changes in the trajectory, namely a sharp change in the trajectory, will be considered an event.

Events and characteristics form an array of source data. In this array of input data there is positive behavior and negative behavior. Positive behavior is the behavior of the user without visible deviations he patterns of work, the algorithm works. Negative behavior will be considered behavior that does not match the standard behavior of the user in a particular situation.

Analysis of data taken from various sources allows monitoring the work of staff, monitor equipment downtime, identify the causes of reduced productivity and production efficiency and carry out personalized development of personal development plans and recommendations to a specific employee.

4. Solution architecture

Let us consider what properties the universal information system for processing Big Data should have:

- Flexibility. A system that has been subjected to a specific regulatory or adaptive effect can change its state and behaviour within the limits determined by the critical values of its parameters.
- Scalability. The system should be expanded depending on the purpose of the customer.
- Configurability. The system should be able to customize the features of a particular enterprise and the class of tasks to be solved.
- Interpretability of results. The system should be able to apply measurement results for various purposes.

To respect all conditions, the system should provide the following basic levels of services. System services are located on the lower level of the service distribution scheme and include:

- Geo-platform for working with maps and location data;
- Enterprise Service Bus for interfacing with external systems;
- Data Lake, which is a data warehouse from various external data providers, on the basis of which the necessary databases, knowledge bases and directories are formed.

Basic services are mandatory services and components that define the main functionality of the system, including:

- Analytical tools for data processing;
- Business Process Support Tools;
- Data storage facilities;
- Platform management subsystem that provides logging of system processes.

Specialized services for working with Big Data, including services such as:

- Apache Spark;

- Apache Hadoop;
- TesorFlow;
- Apache Kafka;
- MongoDB;
- Celery;
- Block of Big Data analysis methods (simulation modelling, machine learning, evolutionary algorithms, neural networks, etc.).

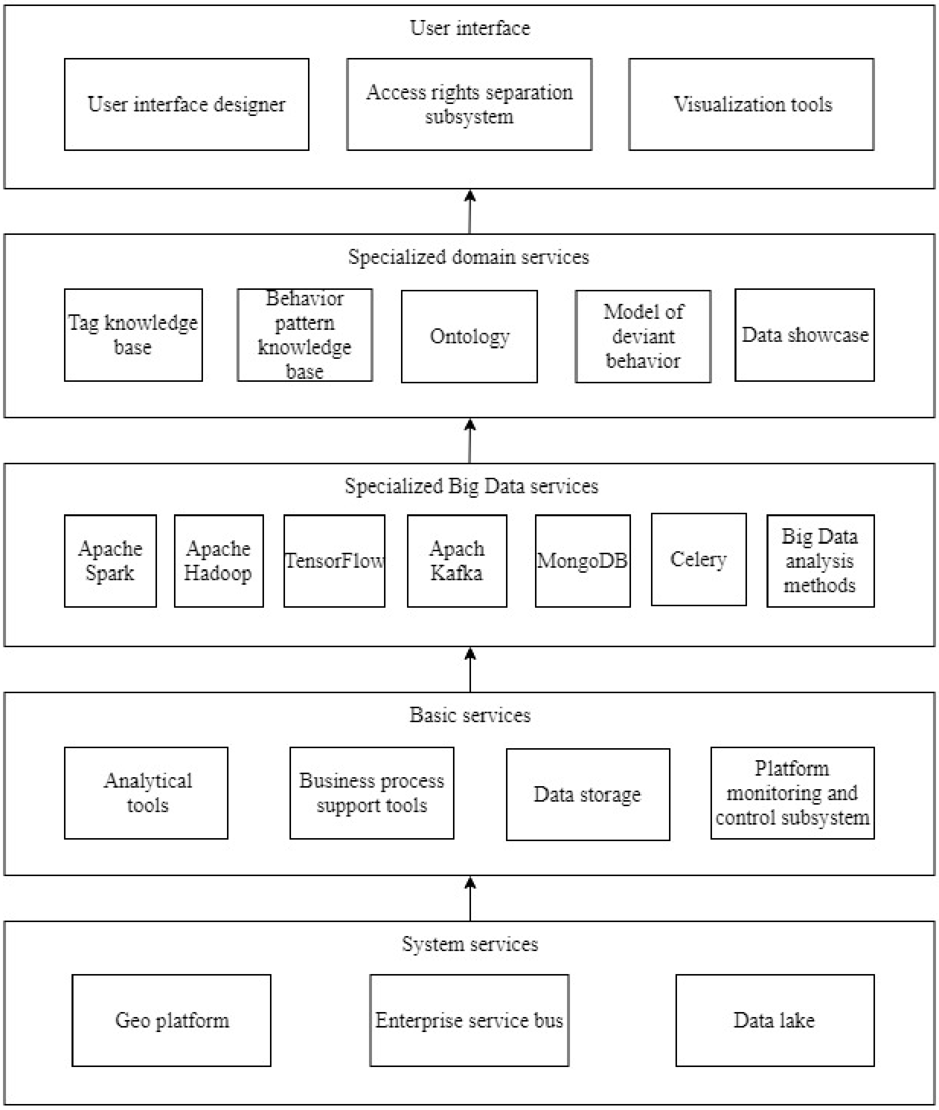

Figure 1. The distribution scheme of services in the system for detecting deviant behaviour.

Specialized domain services are narrowly focused tools aimed at solving specific problems. The example shows the services of the system for analyzing user behaviour in social networks:

- Tag knowledge base containing a list of tags that classify deviant users;
- Behaviour patterns knowledge base, which reflects scenarios of possible user behaviour, both deviant and reference;
- Description of ontology for a specific subject area;
- Model of deviant behaviour, which is the construction of deviant behaviour.

User Interface is the top level of the service distribution system and contains the following components:

- User interface designer with the ability to select the available functionality depending on the goals of the system;
- Visualization tools for graphical presentation of information;
- Access rights separation subsystem.

5. Implementation and practical use

Let's see how information is analyzed for an enterprise from open sources. Let us consider an example of improving the marketing effectiveness of a car manufacturing company. Take information from open sources and analyze data on the mention of a particular part of the production, to identify the main trends and feedback on production, to identify weak points of production or marketing. Within the framework of this example, two groups of consumer interests were identified: “Construction” (which includes tags related to the details of the car’s construction: doors, engines, etc.) and “Production” (including tags related to corporate culture and team management). In both groups, deviations were found that corresponded to various aspects of the commodity market at a given point in time. The figure describes the ontology of these groups, tags and limitations of each tag.

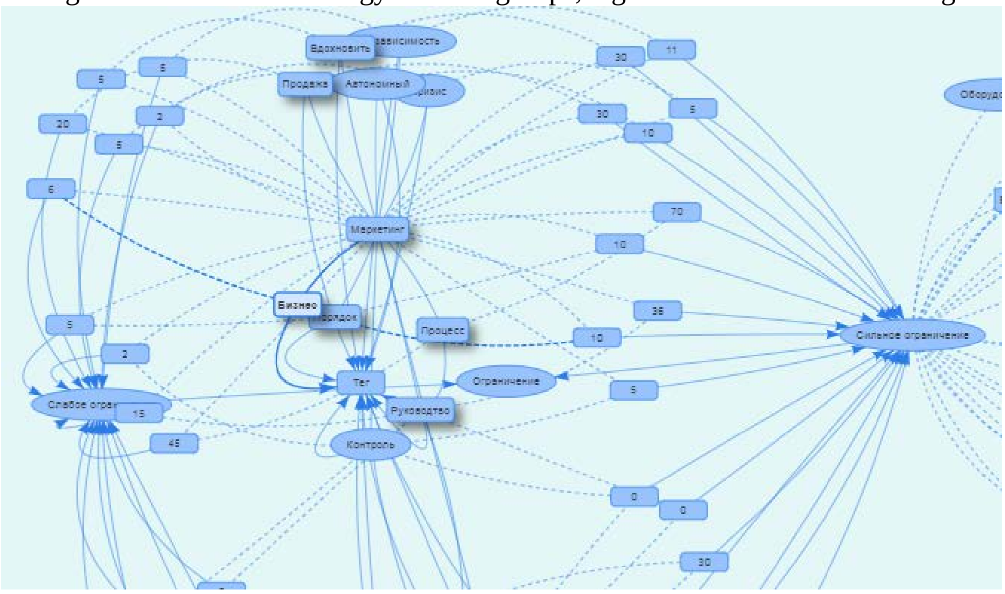

Figure 2. Ontology of groups “Construction” and “Production”.

Below is an analysis of information about several production machines for the most popular tags of a particular production.

As an example (Figure 3-5) illustrated strategic description of industrial companies for the automotive industry. The main Wikipedia articles on popular car trademarks and their changes have been analysed.. The dynamics of such updates turned out to be surprisingly high, despite the fact that these articles contain high-level descriptions that also include historical aspects of this area, additional information about trademarks that are updated, added and edited in real time by various types of users. In the framework of this example, the consumer group “Production” was identified (including tags of corporate culture and team management).

Figure 3. Dynamics of content change for the Lada trademark on the topic “Production”.

Figure 4. Dynamics of content change for the Audi brand in the theme“Production”.

Figure 5. Dynamics of content change for the BMW brand under the theme “Production”.

So, Lada, a popular Russian car manufacturer, is not paying more attention to the production segment due to the emergence of new trademarks. Other popular car manufacturers such as: Audi, BMW were also considered. In the graphs above for other manufacturers, you can also see statistics based on a study showing a surge in interest in a particular market segment.
Analysis of data from several open sources will allow to monitor the work of the marketing department, the work of production, to monitor the trends of bursts of interest among potential customers.

6. Conclusion

Based on the foregoing, it can be concluded that it is advisable to use systems based on Big Data technology in production. The capabilities of these systems allow you to optimize production processes.

7. References

[1] Surnin, O.L. Big Data incorporation based on open services provider for distributed enterprises / O.L. Surnin, P.V. Sitnikov, A.V. Ivaschenko, N.Yu. Ilyasova, S.B. Popov // CEUR Workshop Proceedings. – 2017. – Vol. 1904. – P. 42-47.
[2] Ivaschenko, A.V. Integration issues of Big Data analysis on social networks / A.V. Ivaschenko, N.Yu. Ilyasova, A.A. Khorina, V.A. Isayko, D.N. Krupin, V.A. Bolotsky, P.V. Sitnikov // CEUR Workshop Proceedings. – 2018. – Vol. 2212. – P. 248-254.
[3] Digital Russia. New Reality Digital McKinsey [Electronic resource]. – Access mode: https://www.mckinsey.com/ru/our-work/mckinsey-digital (20.11.2018).
[4] Lasi, H. Industry 4.0 / H. Lasi, H.-G. Kemper, P. Fettke, T. Feld, M. Hoffmann // Business & Information Systems Engineering. – 2014. – Vol. 4(6). – P. 239-242.
[5] Kagermann, H. Recommendations for implementing the strategic initiative Industrie 4.0 / H. Kagermann, W. Wahlster, J. Helbig // Final report of the Industrie 4.0 Working Group. – 2013. – 82 p.
[6] Baesens, B. Analytics in a Big Data world: The essential guide to data science and its applications // Hoboken: Wiley, 2014. – 232 p.
[7] One Internet. Global commission on Internet Governance. [Electronic resource]. – Access mode: https://www.cigionline.org/initiatives/global-commission-internet-governance (20.11. 2018).
[8] Fleischmann, A. S-BPM in the wild / A. Fleischmann, W. Schmidt, C. Stary // Springer, 2014. – 282 p.

[9] Balakrishnan, H. Discovering communities in complex networks / H. Balakrishnan, N. Deo // Proceedings of the 44th annual Southeast regional conference, 2006. – P. 280-285.

[10] Bessis, N. Big Data and Internet of Things: A roadmap for smart environments / N. Bessis, C. Dobre. – Berlin: Springer, 2014. – 450 p.