

Industrial application of big data services in digital economy

O L Surnin¹, P V Sitnikov², A A Khorina², A V Ivaschenko³, A A Stolbova⁴ and N Yu Ilyasova^{4,5}

¹SEC "Open Code", Yarmarochnaya Str., 55, Samara, Russia, 443001

²ITMO University, Birzhevaya liniya, 14, lit. A, Saint-Petersburg, Russia, 199034

³Samara State Technical University, Molodogvardejskaya street, 244, Samara, Russia, 443100

⁴Samara National Research University, Moskovskoe Shosse, 34, Samara, Russia, 443086

⁵Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

e-mail: surnin@o-code.ru, sitnikov@o-code.ru, anastasiakhorina@mail.ru, anton.ivashenko@gmail.com, stolbova@o-code.ru

Abstract. Nowadays, the world is moving to automation. Appropriate programs for the implementation of industrial applications are developed by many companies. But is it so easy to implement systems capable of processing large amounts of information in production? 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. We carried out the analysis of existing data processing works. 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. Was formed a basic algorithm for processing data from various sources (sensors, corporate systems, etc.). Examples are given for automobile industry with a reference of Industry 4.0 paradigm implementation in practice. The given examples are illustrated by trends graphs and by subject area ontology of the automotive industry.

1. Introduction

Big data processing remains one of the key technological directions of digitalization of the economy in Russia. According to the National Program for the Development of the Digital Economy, one of the most important areas of application of the information technology (IT) infrastructure is a smart factory, implemented in accordance with the concept of Industry 4.0. This concept is based on the development of cyber-physical systems capable of controlling industrial processes, providing contextual and decentralized decision support. Solving these problems requires analyzing big data in real time.

Therefore, modern industrial enterprises are considered as a source of big data that describe the process of various information exchanges. Information obtained from various sources allows to draw conclusions about the processes of production and manufacturing. The use of methods and software solutions for working with big data in the course of solving this problem allows us to build efficient production at all levels.

Despite the positive results in research and development of big data technologies, their practical application in industrial applications remains a challenge. The challenges are related to the requirement to use unified and integrated software services based on a reliable platform, instead of developing individual algorithms and software solutions. To cover this gap in this article, it is proposed to summarize existing experience in this area [1, 2] and develop an architectural solution for a software development kit that can be used as a reliable platform for building industrial applications.

2. Technology review

Big data technologies include a range of approaches, tools, and methods for processing structured and unstructured data of vast amounts and considerable diversity. These technologies are used to obtain the obtained results that are effective in the conditions of constant growth, spreading information over numerous nodes of a computer network.

Big data processing services are usually implemented as part of research in developing research institutes, but in modern conditions organizations create a large amount of unstructured data, such as text documents, images, video tapes, 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 important conclusions based on them. Big data processing technologies can make a big contribution to the industrial world.

The modern level of automation of modern industrial enterprises allows the introduction of intelligent technologies for analyzing production and business processes, focusing on the concept of a smart factory [3,4]. In accordance with this concept, human resources and robotic production equipment are integrated into a single information space and form a virtual community of participants with autonomous behavior and self-organization. The concept of Industry 4.0 is intensively and massively investigated in [5, 6]. The concept of an integrated solution based on the introduction of modern IT technologies for the development of cyber-physical systems for intelligent factories is described.

The process of user interaction in a single information space in modern manufacturing enterprises and in supply chains generates a sequence of events exchanging documents, messages and other information objects. The number of events is large (large physical volume of data); they vary and require high-speed processing. In this regard, the task of managing the collection and processing of information data in a system for collecting and processing data with a multi-layered architecture can be attributed to the problem of big data [7, 8].

One of the solutions may be close to a subject-oriented approach to business process management (S-BPM), which presents the process as a joint work of several actors organized through structured communication [9]. A model of interaction of subjects (subjects) in a single information space, which can be implemented using multi-agent software, can be proposed. The ideas of indirect and conditional project management, creating a soft influence on highly motivated autonomous actors, are successfully implemented in online communities and social networks [10-14].

During the development, a significant increase in key performance indicators is expected, according to the McKinsey Digital Economy Report [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 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 personnel monitoring and control can be very diverse and include: workflow, work time tracking system, equipment sensors, personal computer (PC) monitoring programs, motion detection sensors in video cameras, information about calls from an employee's work phone, mailbox activity analysis, etc.

When analyzing the workflow systems, conclusions can be drawn about the time spent working with each of the documents and the effectiveness of the tasks in working with documents. Viewing and analyzing the movement of documentation can help identify bottlenecks in working with each of the documents, displaying up-to-date information about the employee, which slows down the chain of work on the document.

The system of accounting of working time and time allows you to analyze the work of the employee, as well as track the periods when the employee is absent from the office, at inappropriate times.

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

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

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

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

Events that will be monitored in the system include the following events: the fact and time of staff in the enterprise and at the workplace, the performance of official duties, the operation of equipment and their changes. An event that we call a change in the state of production facilities or an employee over 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. An event of the same type can relate to different objects. Events are discrete values.

Flow characteristics are a continuous quantity that describes 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. There are positive and negative behaviors in this input array. Positive behavior is the user's behavior without visible deviations, his work patterns, the algorithm works. Negative behavior will be considered as behavior that does not correspond to the standard behavior of the user in a particular situation.

Analysis of data taken from various sources allows you to track the work of staff, track downtime, identify the causes of reduced productivity and production efficiency, as well as personalize personal development plans and recommendations for 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 (Figure 1).

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

Specialized domain services are narrowly focused tools aimed at solving specific problems. The example shows the services of the system for analyzing user behavior 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.

Consider one of the data processing algorithms that allows you to analyze data obtained from various production systems. In this scheme (Figure 2), information is presented in the form of events characterizing a particular user action (sending a message, passing through sensors, entering production, etc.). The data is sorted in accordance with the algorithm in relation to the user, after which the resulting graphs are constructed allowing to draw conclusions about the patterns of user behavior in production, which makes it possible to organize the workflow more efficiently. Not only data inside the production process can be analyzed, but also external data sources, for example, news sources, for analyzing the reputation on the market, the next paragraph considers the possibility of analyzing information from open sources, largely built using the same algorithm

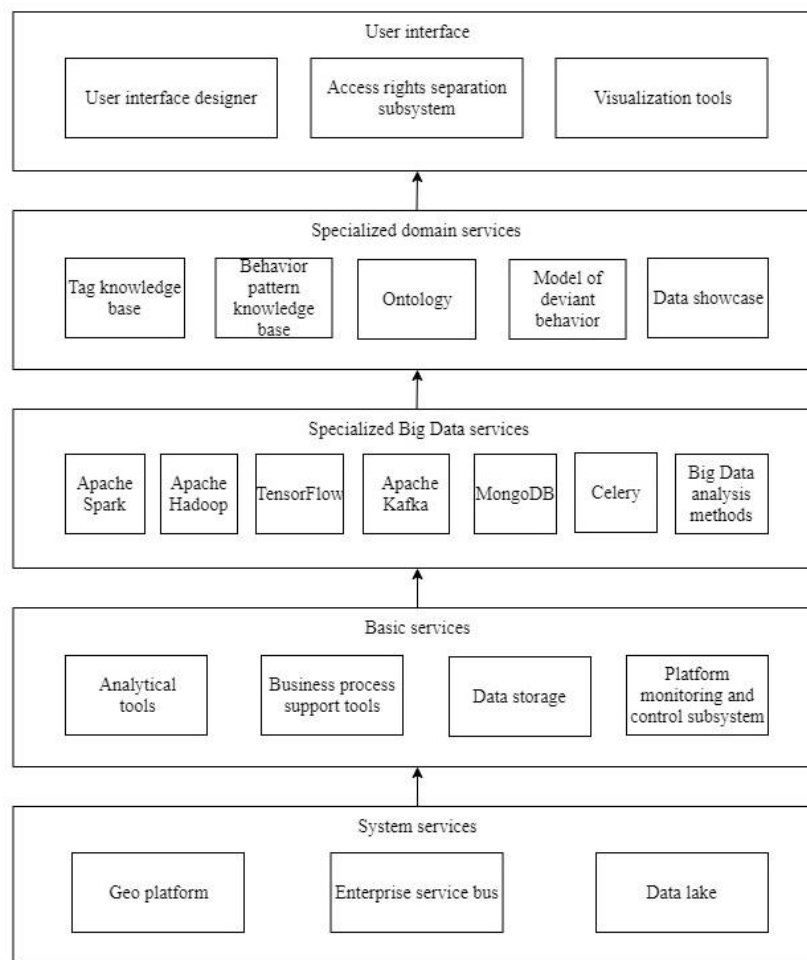


Figure 1. The distribution scheme of services in the system for detecting deviant behaviour.

5. Implementation and practical use

Let's see how information is analyzed for an open source enterprise. Consider an example of improving the marketing effectiveness of a car company.

Take information from open sources and analyze the data on the mention of a specific part of the production to determine the main trends and reviews about the production, to identify the weak points of production or marketing.

Within 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) . Deviations were found in both groups that corresponded to various aspects of the commodity market at a given point in time.

The figure 3 describes the ontology of these groups, tags and limitations of each tag. Below is an analysis of information about several production machines for the most popular tags of a particular production.

As an example (Figure 4-6) 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).

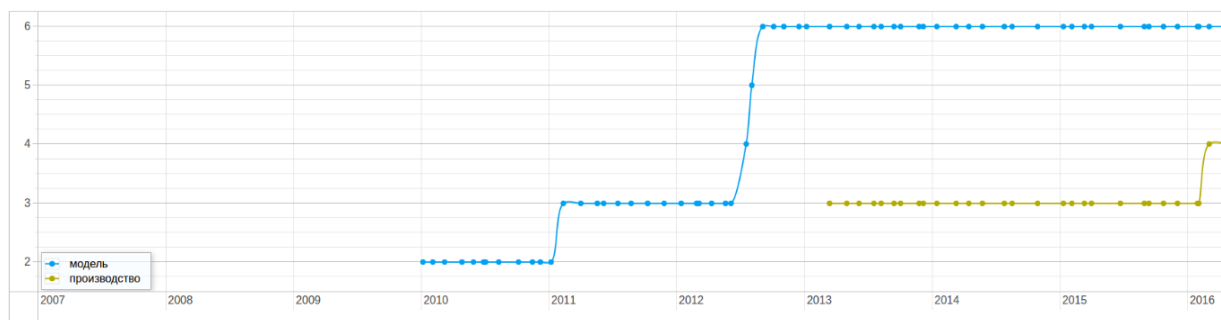


Figure 4. Dynamics of content change for the Lada trademark on the topic “Production”.

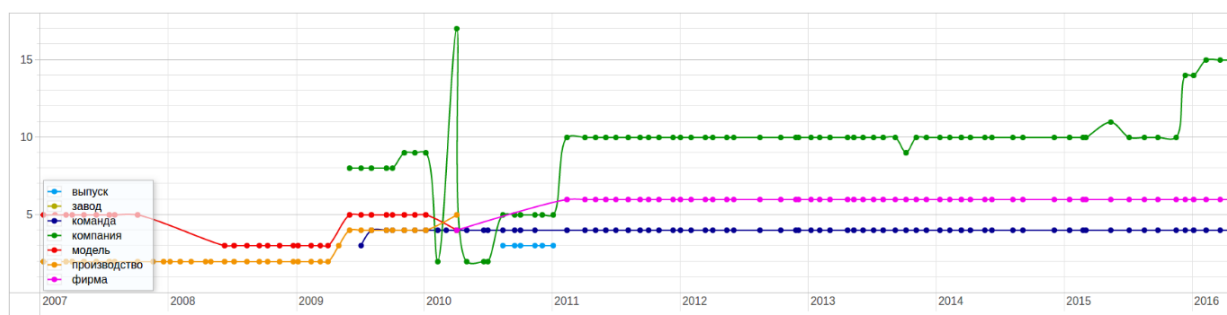


Figure 5. Dynamics of content change for the Audi brand in the theme “Production”.

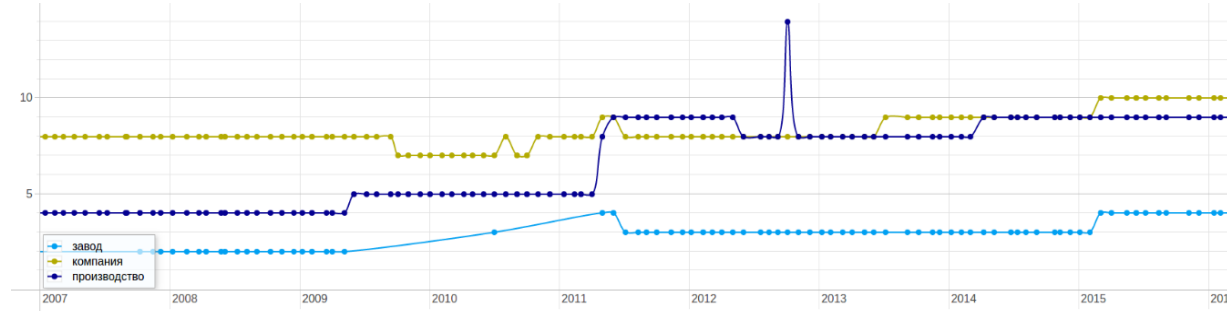


Figure 6. Dynamics of content change for the BMW brand under the theme “Production”.

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, Sitnikov P V, Ivaschenko A V, Ilyasova N Yu and Popov S B 2017 Big Data incorporation based on open services provider for distributed enterprises *CEUR Workshop Proceedings* **1904** 42-47
- [2] Ivaschenko A V, Ilyasova N Yu, Khorina A A, Isayko V A, Krupin D N, Bolotsky V A and Sitnikov P V 2018 Integration issues of Big Data analysis on social networks *CEUR Workshop Proceedings* **2212** 248-254
- [3] Digital Russia 2017 *New Reality Electronic resource* (Digital McKinsey)
- [4] Internet of things - Electronic resource
- [5] Lasi H, Kemper H-G, Fettke P, Feld T and Hoffmann M 2014 Industry 4.0 *Business & Information Systems Engineering* 239-242
- [6] Kagermann H, Wahlster W and Helbig J 2013 Recommendations for implementing the strategic initiative Industrie 4.0 Final report of the Industrie 4.0 *Working Group* p 82
- [7] Baesens B 2014 *Analytics in a Big Data world: The essential guide to data science and its applications* (Hoboken: Wiley) p 232
- [8] One Internet 2018 *Global commission on Internet Governance* (Electronic resource)
- [9] Fleischmann A, Schmidt W and Sary C 2014 *S-BPM in the wild* (Springer) p 282
- [10] Balakrishnan H, Deo N 2006 Discovering communities in complex networks *Proceedings of the 44th annual Southeast regional conference* 280-285
- [11] Bessis N, Dobre C 2014 *Big Data and Internet of Things: A roadmap for smart environments* (Berlin: Springer) p 450
- [12] Gubbi J, Buyya R, Marusic S and Palaniswami M 2013 Internet of Things (IoT): A vision, architectural elements, and future directions *Future generation computer systems* 1645-1660
- [13] Mikhaylov D V, Kozlov A P and Emelyanov G M 2016 Extraction of knowledge and relevant linguistic means with efficiency estimation for the formation of subject-oriented text sets *Computer Optics* **40(4)** 572-582 DOI: 10.18287/2412-6179-2016-40-4-572-582
- [14] Rycarev I A, Kirsh D V and Kupriyanov A V 2018 Clustering of media content from social networks using bigdata technology *Computer Optics* **42(5)** 921-927 DOI: 10.18287/2412-6179-2018-42-5-921-927

Acknowledgments

This work was financially supported by the Russian Foundation for Basic Research under grant # 19-29-01135, # 17-01-00972 and by the Ministry of Science and Higher Education within the State assignment to the FSRC "Crystallography and Photonics" RAS.