NOVATEUR PUBLICATIONS
JournalNX- A Multidisciplinary Peer Reviewed Journal
ISSN No: 2581 - 4230
VOLUME 6, ISSUE 12, Dec. -2020

BASICS AND IMPLEMENTATION OF BIG DATA

A.V. KHALILOV,
BUKHARA ENGINEERING-TECHNOLOGICAL INSTITUTE

A.K. KHOJIEV,
BUKHARA ENGINEERING-TECHNOLOGICAL INSTITUTE

Z.R. KHUJANAZAROV
BUKHARA ENGINEERING-TECHNOLOGICAL INSTITUTE

ABSTRACT:

This article presents the key concepts for Big Data technology: the main characteristics, methods, stages of transition, the scope of technology. The analysis of world achievements in this area is made, examples of use in works of foreign authors are given. The modern market for the use of Big Data technology is analyzed.

KEYWORDS: Big Data, Business Intelligence, technology, NewSQL.

INTRODUCTION:

Big data is a designation of structured and unstructured data of huge amount and a significant variety, which are efficiently processed by horizontally scalable software tools that appeared in the late 2000`sand alternative to traditional database management systems and Business Intelligence class solutions.

Currently, the amount of information is growing exponentially. In order to quickly respond to market changes, gain competitive advantages, increase production efficiency, there is a need to obtain, process and analyze a huge amount of data. To work with such volumes of information, engineers were forced to upgrade the tools to work on the analysis of all data. The concept of Big Data was formed, which was interesting only to a narrow circle of specialists. Now this word is heard by anyone who is interested in the field of information technology. In addition, this definition, or rather the direction of IT development, has become extremely popular and strategically important recently [1].

Big Data technologies allow processing a large amount of unstructured data, systematizing them, analyzing and identifying patterns where the human brain would never have noticed them. This opens up completely new possibilities for using data. The very concept of Big Data means more than just big data. These are huge stored and processed arrays of hundreds of gigabytes, and even petabytes of data. Data that can be processed and extracted from it some useful information. In short, Big Data can be defined as a collection of information processing technologies for obtaining information.

BIG DATA TECHNOLOGIES:

The technologies used to collect and process Big Data can be divided into 3 groups:
- Software;
- Equipment;
- Utility services.

The most common data processing (SW) approaches include:

SQL is a structured query language that allows to work with databases. Using SQL allows creating and modifying data, and the corresponding database management system manages the data array.

NoSQL - the termis identified as Not Only SQL. It includes a number of approaches

aimed at implementing a database that differ from models used in traditional, relational DBMSs. They are useful for constantly changing data structures. For example, to collect and store information on social networks

Map Reduce is the model of calculation distribution. Used for parallel calculations of very large data sets (petabytes* or more). Thus, the query is a separate program. The principle of operation is the sequential processing of data using two Map and Reduce methods. Map selects preliminary data, and Reduce aggregates it.

Hadoopis used to implement search and context mechanisms of highly loaded sites, e.g. Facebook, eBay, Amazon, etc. A distinctive feature is that the system is protected from failure of any of the cluster nodes, since each unit has at least one copy of data on the other node.

SAP HANA is a high-performance New SQL platform for data storage and processing. Provides fast processing of requests. Another distinguishing feature is that SAP HANA simplifies the system landscape by reducing the cost of supporting analytical systems [4].

The first Big Data technologies had been used by those industries whose activities are tied to the processing of large flows of information daily - banks, mobile operators, retail networks. Work with data in these areas is aimed at creating a portrait of the client in order to offer the most suitable services for him [2].

STORAGE AND MANAGEMENT:

This is just the case when we have to admit that there are problems in BigData. The more data accumulated, the more demanding it is to store and manage it. You will have to buy expensive hardware or accept the disadvantages of storing data in the cloud. You will need specialists who can foresee possible problems when analyzing large amounts of data that can organize all the nuances in such a way that you really use the data effectively.

BIAS:

Bias is another major problem in Big Data. It is quite easy to draw a specific conclusion if the results of one or two studies are at your disposal, but if results become significantly larger, there is a rather large space that allows changing the overall meaning of the results by changing the presentation of the data. It is therefore very important to ensure that research results are not influenced by the views of any of the stakeholders.

The more data you have, the more difficult it is to highlight exactly what you need now. Of course, the nature of this problem is directly related to the specifics of Big Data and Data Mining in general, but it should not be overlooked.

“Each time you use Google or Yandex search systems, you work with big data”.

For example, the full potential of big data in medicine has not yet been revealed. Machine learning algorithms are already actively used in the diagnosis of cancer, but this approach is not used in other areas, for example, in the treatment of fluor personalized diet advice.

It would be interesting to look at the bundle of big data and complemented reality. City and museum guides, instructions for everything that gets into the lens of your mobile camera, first aid tips – now there is simply not enough fantasy to imagine the synergy effect of these two technologies in the future [3].

In the future, as a full-fledged partner of the ELBA project (Establishment of training and research centers and Courses development on Intelligent BigData Analysis in Central Asia) in frame of the Erasmus+ program, we will increase our knowledge and apply these technologies in various areas of our Republic.

PERCEIVING AND RETRIEVING BIG DATA FOR SSAM:

As shown in the upper left corner of Fig. 1, this is the main layer of frame. Devices of IoT (Internet of Things), such as smart devices, RFID readers, RFID tags, intelligent sensors, etc. [5], are aimed for the entire product production cycle (i.e., intelligent design, intelligent manufacturing, intelligent maintenance and services, intelligent delivery) of the AM (Automatic Management) intelligent environment.

Then, huge heterogeneous data from several sources of AM, such as product planning and design, materials and procurement, AM systems, AM production control and status, product qualification, energy consumption, product delivery and customer feedback, maintenance and services, product recovery, etc., are monitored and recorded for further evaluation. Intelligent sensors and calibration devices are used to control and monitor the operation of the AM system and produce quality products in accordance with the requirements of the customer. Before the assembly, product quality data is also tracked and collected at each stage of production. Then standard communication procedures (e.g. Modbus, Intranet, Internet RS-485/323, 5G, etc.) [5] are used to transfer large captured data for further processing at the next working level of the structure.

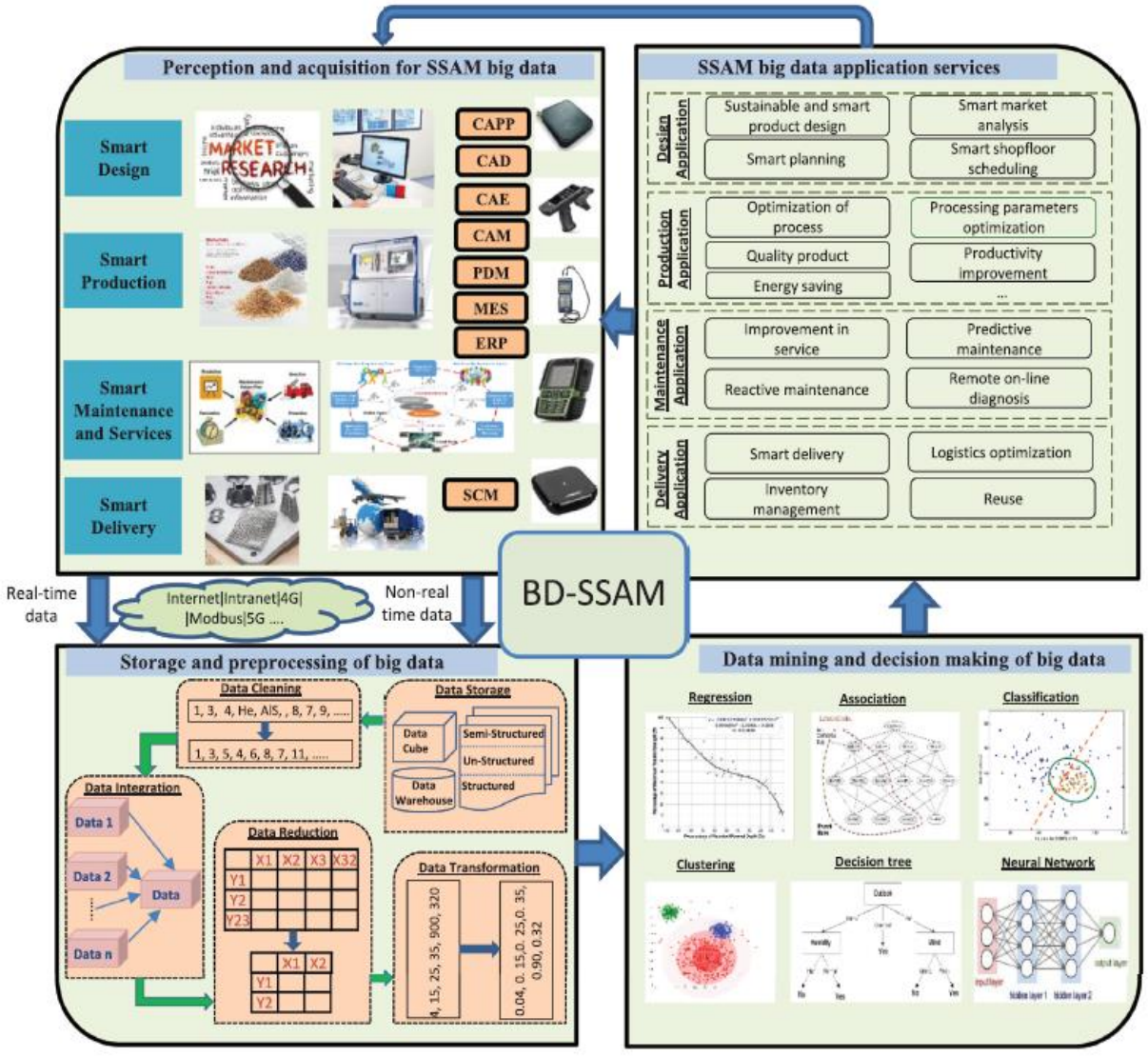

Fig.1. BigData Analyses for the SSAM.

SSAM BIG DATA STORAGE, PREPROCESSING, INTEGRATION, AND MANAGEMENT:

Real-time and non-real-time data are collected for the entire production cycle of AM plants, which consist of a large amount of unstructured, structured and semi-structured data [6]. Conventional data storage and management tools and technologies are not sufficient to further process and process too large and complex data. Therefore, this layer supports further processing of disordered and huge datasets or data cubes using extensible markup language (XML), not only structured query language (NoSQL) and distributed database system (DDBS) [7]. The Hadoop computing environment is used to process non-real-time production cycle data. The Storm real-time computing environment is used for real-time intensive processing of product AM production cycle data [8], [9]. At this level, the data flow for preprocessing and control is followed by the sequence below and shown in Figure 4 (bottom left):

- I Storage Method
- II Data cleansing method
- III Method of data integration
- IV Data processing method
- V Data conversion method

Initially, the production run data of the AM product is selected to form a data cube, and build a production run data store so that large SSAM data can be integrated using predefined logic [10].

Created data stores are used to store and manage the production cycle data cube and to describe the complex logical relationship between massive production cycle data cubes. Further, the generated production cycle data cubes have a large number of redundancies, which are reduced by the data purge operation. Common models that are used to purge data are the information structure model [11] and the RFID-Cuboid model [12]. During a cleanup operation, the production run data cube set from the production run data store is an input that provides output as an organized production run data cube set that transmits important information about the production run state.

In addition, metamodelshave been developed for the integration of huge data, which includes design, market, production, maintenance, size quality, product delivery, logistics, etc. [13]. Obviously, integrated datasets are still huge, which may not be possible for data analysis.

Therefore, the data reduction process is performed to obtain a reduced illustration of the data set, which is much smaller in volume, but it must maintain the validity of the original data. In addition, the short data of the production cycle is converted, so that the subsequent mining method can be more efficient, and the detected patterns are easier to recognize [13]. Finally, the pre-processed SSAM big data is stored in various aforementioned data management systems, such as DDBS, for further use in decision-making.

INTELLIGENT DATA ANALYSIS AND DECISION MAKING ON LARGE SSAM DATA:

Throughout the production cycle, AM plants generate a huge amount of data in the form of 4V (volume, speed, diversity and cost), which makes it difficult to study them using traditional analysis methods. This framework uses impressive technology that has the potential to identify hidden knowledge from large data sets about AM products, namely mining models (e.g. clustering, classification, association, neural network, prediction, etc.) [5]. By combining AM product analysis approaches and big data mining techniques, valuable information and knowledge can be derived from these AM product production cycle datasets. Based on these results, AM enterprise managers will provide intelligent

and sustainable AM product solutions for application services [5].

SSAM BIG DATA APPLICATION SERVICES:

This layer of application services can be seen in the upper right corner of Fig.1. Big Data Application Services are used to deliver significant real-time and non-real-time applications based on acquired information and knowledge to end users [5], [13]. Multiple forms of application services are planned for the different stages of the production cycle at this structure level. In particular, environmentally friendly and intelligent product design, intelligent workshop planning, AM process optimization, AM processing parameter optimization, energy consumption reduction, logistics optimization, sustainable and intelligent maintenance, service forecasting, intelligent delivery, etc. are intended for SSAM, which contributes to the advancement of CP strategy in a sustainable and intelligent AM environment. Thanks to the introduction of intelligent decision-making and real-time feedback, the above-mentioned services are used in AM enterprises for efficient and sustainable intelligent production.

CONCLUSION:

After reading this article, it may seem that there are more problems than benefits in analyzing data, but do not forget that when used skillfully, it is a powerful and effective tool that can help make effective decisions. In particular, for competent work with Big Data, you need to understand the specifics of a particular market and business, so many analysts advise creating data analysts inside the company, and not hiring them from the outside.

Big Data opens up new horizons for us in production planning, education, healthcare and other industries. If their development continues, Big Data technologies can raise information, as a factor in production, to a completely new high-quality level. Information will become not only equivalent to labor and capital, but also, perhaps, will become the most important resource of the modern economy.

REFERENCES:

1) Andreas Wygend BIG DATA. All the technology in one book "Exmo" 2017
2) Victor Mayer-Schoenberger, Kenneth Kukier “Big data. A revolution that changes how we live, work and think ”Moscow; 2014
3) Blagirev A. “Big data in simple language. Blagirev - "AST Publishing House", 2019 - (Business book)
4) Bill Franks A revolution in analytics. How to improve your business in the era of Big Data with the help of operational analytics, Moscow; 2016.
5) Y. Zhang, S. Ma, H. Yang, J. Lv, Y. Liu, A big data driven analytical framework for energy-intensive manufacturing industries, J. Clean. Prod. 197 (2018) 57–72.
6) A. Majeed, M. Muzamil, M.Z. Awan, A. Siddiqui, Developing of a manufacturing cycle architecture for Fused deposition modeling technique, Int. J. Lightweight Mater. Manuf. (2019).
7) K. Shvachko, H. Kuang, S. Radia, R. Chansler, The Hadoop distributed file system, Proceedings of the IEEE 26th Symposium on Mass Storage Systems and Technologies, MSST 2010, May 6, 2010 - May 7, 2010, IEEE Computer Society, Lake Tahoe, NV, United states, 2010 IEEE.
8) W. Yang, X. Liu, L. Zhang, L.T. Yang, Big data real-time processing based on storm,Proceedings of the 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2013, July 16, 2013 – July 18, 2013, IEEE Computer Society, Melbourne, VIC, Australia, 2013, pp. 1784–1787.

9) M.H. Iqbal, T.R. Soomro, Big data analysis: apache storm perspective, Int. J. Comput. Trends Technol. 19 (2015) 9–14.
10) R.Y. Zhong, C. Xu, C. Chen, G.Q. Huang, Big data analytics for physical internetbased intelligent manufacturing shop floors, Int. J. Prod. Res. 55 (2017) 2610–2621.
11) D.F. Xu, Q. Li, H.B. Jun, J. Browne, Y.L. Chen, D. Kiritsis, Modelling for product information tracking and feedback via wireless technology in closed-loop supply chains, Int. J. Comput. Integr. Manuf. 22 (2009) 648–670.
12) R.Y. Zhong, G.Q. Huang, S. Lan, Q.Y. Dai, X. Chen, T. Zhang, A big data approach for logistics trajectory discovery from RFID-enabled production data, Int. J. Prod. Econo. 165 (2015) 260–272.
13) Y. Zhang, S. Ren, Y. Liu, T. Sakao, D. Huisingh, A framework for Big Data driven product lifecycle management, J. Clean. Prod. 159 (2017) 229–240.
14) Khalilov FV, Kutliev NB PROCESSING THE FLOW OF CHAINS // Modern materials, equipment and technology. - 2014 .-- S. 457-460.
15) Khalilov FV, Khozhiev AK, Rakhmatov SS Modeling and forecasting of social processes at the regional level using information and communication systems // Modern instrumental systems, information technologies and innovations. - 2014 .-- S. 259-263.