Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 64 (2017) 369 – 374

www.elsevier.com/locate/procedia

The 9th CIRP IPSS Conference: Circular Perspectives on Product/Service-Systems

A big data analytical architecture for the Asset Management

Jaime Campos[a]*, Pankaj Sharma[b], Unai Gorostegui Gabiria[c], Erkki Jantunen[d], David Baglee[e]

[a]*Linnaeus University, Faculty of Technology, Department of Informatics,Sweden*
[b]*Department of Mechanical Engineering, , IIT Delhi, New Delhi, India*
[c]*Electronics and Computing Department Mondragon University, Mondragon, Spain*
[d] *VTT Technical Research Centre of Finland, P.O.Box 1000, FI-02044 VTT, Finland*
[e]Department of Computing, Engineering and Technology, University of Sunderland, UK

* Corresponding author. Jaime Campos. Tel.: +46-(0) 470-708829 E-mail address:jaime.campos@lnu.se

Abstract

The paper highlights the characteristics of data and big data analytics in manufacturing, more specifically for the industrial asset management. The authors highlight important aspects of the analytical system architecture for purposes of asset management. The authors cover the data and big data technology aspects of the domain of interest. This is followed by application of the big data analytics and technologies, such as machine learning and data mining for asset management. The paper also presents the aspects of visualisation of the results of data analytics. In conclusion, the architecture provides a holistic view of the aspects and requirements of a big data technology application system for purposes of asset management. The issues addressed in the paper, namely equipment health, reliability, effects of unplanned breakdown, etc., are extremely important for today's manufacturing companies. Moreover, the customer's opinion and preferences of the product/services are crucial as it gives an insight into the ways to improve in order to stay competitive in the market. Finally, a successful asset management function plays an important role in the manufacturing industry, which is dependent on the support of proper ICTs for its further success.

Peer-review under responsibility of the scientific committee of the 9th CIRP IPSS Conference: Circular Perspectives on Product/Service-Systems.

Keywords: Asset Management; Big data; Big data analytics; Data mining.

1. Introduction

When preventive maintenance is applied, a large amount of data on operations and maintenance are produced. However, companies and the asset management function are currently storing a huge amount of data and are at the same time not using it [1]. Consequently, there is almost no use of the stored and existing data produced by the equipment that could result in increased efficiency and organisational performance of an asset management function. The analysis of the data produced by a company is extremely important for improved decision making.

The area of Information and Communication Technologies (ICTs) is dynamic, and new technologies emerge rather frequently. Therefore companies need to understand their various characteristics to be able to develop, implement and use these ICTs successfully. The authors provide a big data analytical architecture at a conceptual level where the data scientist and the maintenance staff are part of the system. In addition, it highlights important aspects of a system to be used for the purpose of asset management. There are several ICTs applications and systems suggested and implemented in the industrial domain [2; 3]. However, some of these ICTs do not provide an overall picture of the system where the data

Peer-review under responsibility of the scientific committee of the 9th CIRP IPSS Conference: Circular Perspectives on Product/Service-Systems.
doi:10.1016/j.procir.2017.03.019

scientist and user are part of the system, especially when it is based on the big data approach. In the current work, the authors provide an analytical architecture, based entirely on a big data approach at a conceptual level. A data scientist requires innovative solutions in order to perform different elements of the CRISP Methodology including business and data understating, data preparation, modelling, evaluation and deployment aspects of a big data solution or project. The next part of the paper is ordered in the following way. In section 2, the paper discusses data and big data characteristics. In section 3, big data analytics, machine learning, and data mining features in connection to the domain of interest are discussed. Thereafter, in section 4, important aspects of user interface and visualisation are covered. In section 5, an Asset Management technical and analytical framework is presented and discussed. Finally, conclusions are presented in section 6.

2. Data& big data in the domain of interest

Big data is a term that makes reference to a great quantity of data, which exceeds conventional software's capacity to handle it. The processes that are used to find patterns in them are the so-called predictive analytics and user behaviour analytics. Data gathering is the first step towards an exhaustive analysis for the asset management. Data analysis can use data collected from different sources, which depends on the objective of the analysis. One of the domains where data can be obtained is the web and social media. Search logs clicked web content, or streams are collected from various web services [4]. The information is then used to understand the customer's needs and improve the experience for the user. In the industrial sector, for instance, there is a need to monitor different assets, usually with the help of transducers to keep the equipment in optimal condition and reduce the downtime caused by maintenance. These transducers are typically named sensors. They transform a physical property or phenomena into an electrical signal. The signal then is manipulated by means of filtering to reduce the electrical noise and, if necessary, is converted to a digital signal. After some signal conditioning and manipulation, it is sent to a database where meaningful information can be attained. The data can be obtained in several formats such as structured, unstructured as well as semi-structured [5]. The structured data can be defined as the data that has a well-defined format such as date, numbers or character strings and allows storage and generation of information. Databases or spreadsheets are examples of structured data. Unstructured data refer to data in the same format as were collected, such as emails, PDFs, or documents [6]. The semi-structured data are not limited to determined fields, but contains separators to divide the data. Usually, they cannot be managed in a standardised way, but they contain their metadata that describes the objects and their relationships, i.e. XML or HTML [7].

The storage systems can be classified as DAS (Direct Attached Storage), NAS (Network Attached Storage) and SAN (Storage Area Network) [8]. However, these storage systems have limitations when creating a large-scale distributed storage system [9]. NoSQL (Not-Only SQL) databases provide a more flexible and concurrent storage systems and allow a much faster manipulation and search of information than relational databases. Four different types of NoSQL databases are distinguished: Key-Value, Documental, Graph and Column Oriented [8]. Different data storage architectures have to be tested to find the most appropriate one. Features such as Data Model, Data Storage, Concurrency Control or Consistency have to be taken into account while deciding which architecture suits best the system [8].

The data processing for big data can be done in various ways, depending on the application. Real-time applications such as navigation, finance, Internet of Things (IoT) or intelligent transportation rely heavily on timeliness [9]. The processing power needed for these real-time applications is very high, and the cloud computing and distributed technologies are useful tools to be able to analyse data at such speeds. Also, it is important to understand the characteristics of big data in connection with the 3V's, i.e. volume, velocity, and variety in the domain of interest. Therefore, the maintenance data can be, further on, divided into traditional thinking maintenance data and data connected to the 3Vs for a better understanding of its characteristics in the domain [5]. Traditional data when it comes to volume involves condition monitoring, maintenance plan, work orders, etc. The non-traditional data part of this volume is the data that is indirectly related data, such as purchase contract, production, scheduling, asset depreciation value, etc. The traditional data part of the variety includes semistructured data, such as spreadsheets and data stored in relational databases. Variety data being a part of the big data is semi-structured like emails, XML files and log files with some specific formats. Also, the unstructured data part of the variety consists of pictures, audios, videos, web pages and other documents, such as Word and PDF. Finally, part of the velocity and traditional maintenance data are transaction data and multidimensional data. Whereas, the data considered outside the traditional maintenance data and part of the velocity is the real-time condition monitoring data collected by sensors and instruments. In conclusion, the division of the data is processed in such a way that the data outside the limits of the traditional maintenance data are considered to be part of the big data approach.

3. Big data Analytics & Data Mining

With the emergence of the Big data approach and its technologies, there is a need to use ICTs, such as data mining software algorithms and statistics, for more fruitful information and knowledge as well as possibilities to find hidden patterns from the data. However, to analyse data to discover relationships and non-obvious patterns is not a new concept. Data mining, knowledge extraction, information discovery, information harvesting, data archaeology and data pattern processing are some of the names to the techniques that use data analysis for decision making. For instance, Knowledge Discovery in Databases (KDD) is a concept that involves data mining techniques. It maps low-level data that is

not easy to understand because it is too voluminous, to new forms, like a short report. It can map it into a more abstract form, such as a descriptive estimate or model/example of the activity/process that converts the data to a more useful form, for instance predictive models for estimating the value of future cases [10]. Therefore, the use of KDD and its data mining techniques provides support for the identification of valid, novel and understandable patterns from large and complex datasets [11].

Big data analytics uses, in general, the same techniques as the KDD, however, emergent data mining and machine learning algorithms are used as well. The machine learning techniques consider the characteristics of the big data, i.e. the 3 V's, volume, velocity and variety of the data deluge that it deals with and are, therefore, more challenging compared with traditional data mining methods [9, 11]. These developments are the so-called big data in academia and industry, and most definitions highlight their increasing technological capacity to capture, aggregate and process an increasingly larger volume, velocity, and variety of data, i.e. the 3 Vs [12, 13]. The technologies such as data mining, machine learning and statistical learning, etc., have been there for long, so what differs is the amount and type of data that needs to be processed and analysed. The main objective of data mining and machine learning algorithms is to turn a large collection of data into knowledge and find hidden patterns.

Consequently, companies and the asset management function are currently storing a huge amount of data, which is increasing and at the same time is not used [1]. Thus, there is slightly and/or almost no practical use of the existent data that are produced by the machines or other related data that could increase the efficiency of the asset management process. The data produced by a company is extremely important for improved decision making. It is, therefore, crucial that the asset management makes use of the available and potential data that can be a support for the efficiency and its increased organisational performance. Consequently, there are many benefits of using the big data technologies. The effective use of big data provides advantages, such as increased organisational efficiency, informing strategic direction, better customer service, identifying and developing new products and services [14]. Data mining methods extract information, i.e. hidden patterns from various data. These techniques are, for example, clustering analysis, classification, association, and regression. Thus, these methods involve machine learning algorithms and statistics [9]. There are tangible differences between conventional analytics and big data analytics, for instance, big data involves the gathering, processing, and analytics to unstructured data formats [15]. Its analysis methods are based on machine learning approach, and its primary products are data-based products. In addition, there is a constant flow of data with 100 terabytes to petabytes. The traditional analytics is based on hypothesis-based testing. Its format consists of rows and columns, and its primary purpose is internal decision support and services. In addition, the volume of data is tens of terabytes or less. The visualization of the data part of the decision-making process becomes an important part of the system, which is gone through in the next section.

4. Dashboard/data visualisation

The visualisation part of any decision support system/tool is extremely important since it provides a comprehensive picture of the important issues that a decision maker needs to explore during the decision-making process. The area of visualisation is part of the exploratory data analysis (EDA). It was created during the 1970's by the famous statistician John Tukey. However, there is a distinction between data mining and statistics standpoint when it comes to visualisation. The data mining approach views the descriptive data analysis techniques as an end in themselves. Whereas, the statistics part of EDA interprets it as a hypothesis-based testing and as the final goal [16, 17]. The visualisation of a data mining system is an important process since it facilitates and provides the participation of experts, such as data scientist and domain experts for its interpretation [18, 19]. In conclusion, the data can be visualised in various devices, however, depending on the application. For example, a maintenance technician could have a smartphone where the different machines that need a repair could be mapped or the assembling and disassembling instructions could be shown with a picture [20]. Visualisation could also be done on a desktop computer interface or dashboard where the meaningful information is displayed. Other means of data visualisation could be Virtual and Artificial Reality headsets or tools. In the industry sector, virtualized system or component could be helpful to discover where exactly a fault has happened or what tools are needed for such repair.

5. Asset management technical & analytical framework

The first phase in any Information system (IS) and Information and Communication Technologies (ICTs) is the development of a conceptual model [21]. In the current section, a conceptual analytical framework for asset management is discussed. The framework is presented in Figure 1. The framework consists of three layers which are discussed in the subsequent sections.

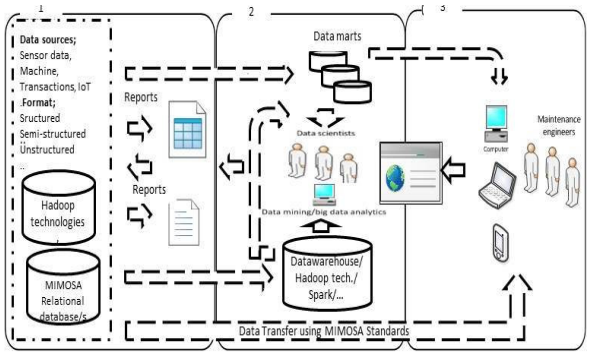

Figure 1.An Asset Management analytical architecture.

5.1. First layer

In this layer, it is important to implement technologies that have the ability to understand and absorb the 3 V's of big data, i.e. volume, velocity, and variety. The latter have helped in revealing hidden patterns which provide an indication towards fault diagnosis and prognosis in assets. The data is collected from a multitude of sources in equally varying formats. This large variety of structured, semi-structured and unstructured data poses unique challenges for the data scientists. Although a large part of asset management data is structured, one cannot rule out a small proportion of unstructured data in the form of, for instance, text and videos. Hadoop is a promising platform since it overcomes many of the constraints with former technologies, such as limitations of storage and capacities of computation of huge volume of data [22]. Also, it supports multiple data formats, i.e. structured, semi-structured and unstructured data. Another advantage is that it is open source software. It is an economical solution with inherent characteristics such as a low learning curve. It also permits multiple cluster nodes at the same time thereby allowing linear scalability. The use of NoSQL solutions, such as Apache Cassandra, help in obviating the problems related to performance, cost, and availability of big data applications (cassandra.apache.org). In addition, it is possible to combine Spark and Hadoop technologies to be able to combine their advantages like economic storage, scalability, and fast data processing capacities. Spark is an open source platform for large scale data process, which is well suited for iterative machine learning tasks and faster real-time analytics [23], (spark.apache.org). Apache Spark includes libraries of SQL and DataFrame as well as such libraries as MLib for machine learning, GraphX and Spark streaming [24], (spark.apache.org). Customer data, namely social network data connected to the customer relation management (CRM) and asset management, should use Hadoop technologies because of its variation when it comes to the data format. To manage this kind of data it is convenient to use the Apache Hive, which is a data warehouse software constructed on Hadoop. It enables reading, writing, and managing large databases stored in distributed storage with the support of SQL. In addition, it involves a command line tool, as well as a database connection performed through the Java Database Connectivity (JDBC), which enables the connection to the Hive software (hive.apache.org).

The MIMOSA relational databases and information exchange standards provide a proper means to handle the structured data. MIMOSA Open System Architecture eradicates the problems of interoperability between disparate hardware and software offered by different vendors. It specifies a standard architecture and framework for which and how to move the information. It has built in metadata to describe the process that is occurring. The Open Systems Architecture for Enterprise Application Integration (OSA-EAI) is one of the MIMOSA standards. OSA-EAI defines data structures for storing and moving collective information about all aspects of equipment, including platform health and future capability, into enterprise applications. This includes the physical configuration of platforms as well as reliability, condition, and maintenance of platforms, systems, and subsystems (www.mimosa.org). While for some applications, the arrival and processing of data can be performed in batch, other analytics applications require continuous and real-time analyses, sometimes requiring immediate action upon processing of incoming data streams [25]. Less critical machines can resort to batch processing of data, whereas more critical equipment used within nuclear and military applications may have to undergo real-time processing of data for possible diagnosis and prognosis. This characteristic of the data is referred to as variability in some literature [26]. Preliminary reports on asset health that require immediate attention by the management crew can be generated at this point. The primary idea is to analyse the streaming data to attend to any high-risk situations that require immediate consideration. Data considered less important can be analysed in detail later on where the situation is not urgent. For example, the data picked up by the sensors which indicate to a big impending failure of the machine needs to be addressed immediately.

5.2. Second layer

It is in this layer where the data analysis takes place, and it is where the data warehouse, Hadoop and Spark technologies are utilised together with the data mining and big data analytics technologies. The data scientist can, as is illustrated in Figure 1, get data from various sources, i.e. data warehouse, specific data from the data marts as well as directly from the source, i.e. unstructured data for further analysis. The people in charge of the data analytics work together with the maintenance expert. If there is a need, for example, to provide service to the equipment then a work order is created based on data provided by the the diagnostic, prognostics and big data analyses.

For the data scientist work and people involved in the process, it is recommended that the industry oriented approach, i.e. CRISP-DM Methodology, is followed. This methodology highlights important aspects in six steps, i.e. the business and data understanding, data preparation, model building, testing and evaluation, and finally the deployment. It reassures best practices as well and provides organisations with the structure needed to realise better and faster results from the data mining and big data analytics projects [27, 28].

Enterprise data gets stored in a data warehouse. Any suitable distributed file system like HDFS can be used for storing a large amount of data into separate clusters. The file system must be fault-tolerant and low cost. Redundancy of data storage to avoid loss of data during failures is required. As it is cheaper to move computation than to move the data, the computational resources are shifted closer to where the data

rests. There are data marts that are established in order to cater for the needs of sub-departments of the enterprise. Any data virtualization software can be used to create virtual data marts by pulling in data from different sources and combining them together to fulfil the requirements of a small section of the enterprise.

The big data being collected by the sensors has two key features; high dimensionality and large sample size. The high dimensionality of data means that firstly, it can be used to develop effective methods that can accurately predict the future observations and secondly, to gain insight into the relationship between the features and response for scientific purposes. In addition, large sample size helps in, firstly, exploring the hidden structures of each subpopulation of the data, which traditionally is not feasible and might even be treated as 'outliers' when the sample size is too small. Secondly, large sample size supports the process of extracting important common features across many subpopulations, which might otherwise be difficult because of the large individual variations [29]. These characteristics of big data can help the data scientists to analyse the data in an effective manner. The immediate analysis of the streaming data is necessary to take corrective actions in order to avoid a failure to take place. However, the data analysts have a considerably larger time in their hands when unearthing hidden patterns from the data sets. The analysts can also use data from the enterprise data warehouse (EDW). In addition, the data being collected from the data sources must be stored in MIMOSA tables. The transfer of data across the enterprise which shown by the dotted arrows should follow MIMOSA standards (www.mimosa.org). However, it is important to be aware that the MIMOSA tables and standards are applicable only to structured sensor data, i.e. and not to unstructured text and video data on Twitter and Facebook.

5.3. Third layer

Visualisation in connection with the big data is about displaying the analytical results in visual or graphic form [15]. The data visualisation provides huge possibilities for the analytics process since it can help to disclose patterns and trends hidden in the data. In the current layer, the maintenance engineer and other related staff gets information about the current state of, for instance, the equipment as well as other important information related to both product and services. Other important aspects of user interfaces are highlighted by [30], where the author covers the Infological equation, which highlights the importance of the personnel/user pre-knowledge. It provides the information systems area with the complex inter and intra-individual aspects of a user. The equation takes into consideration parameters, such as the information (or knowledge) produced from certain data and the interpretation process of each user. Hence, what is information or knowledge from a specific data presented in a user interface depends much on the specific user. The user and the user interface are important parts of a system, which needs to be considered for its optimal implementation. Therefore, the data scientist becomes an important part of the system with all its knowledge about data mining and big data analytics. Thus, data visualisation is an important activity, and it often separates a good analytical system from a bad one. User-friendly visualisation techniques are necessary to narrow the gap between big data system and its users. The visualisation techniques should display the analytic results in an intuitive way so that users can identify the interesting results effectively. To enable a fast response, the back-end system is expected to provide a real-time performance, and the visualisation algorithms need to transform the user's event into a proper and optimised query [31, 32]. The visualisation must convert into actionable instructions for the maintenance crew. One of the major problems when developing user interfaces is to provide the user with queries they can make. If the user is not aware of the querying method, exploration of data will be difficult, and the software system will not be used to its optimal level.

Due to the inherent aspects of the big data technologies ecosystem, the suggested framework provides many advantages for companies that need to take fast and optimal decisions. It is, therefore, valuable to be used for decision support applications like prognostic health management (PHM) that need, for instance, an optimal length of the prognostic distance, which is crucial to diminish the total cost related to maintenance [33]**.** The costs are reduced by relocating the need for unscheduled maintenance into scheduled maintenance within a finite prognostic horizon. In addition, the big data approach provides support for the data gathering for purposes of both product and consumer aspects, which is one of the important requirements to be able to implement a Product/Service (PSS) strategy [34]. Consequently, the big data approach and its ICTs ecosystem provide an optimal way to gather data from various sources resulting in possibilities to be able to take comprehensive decision making as is the case of IPS life cycle ranging from marketing to life cycle data management [35]. Thus, the ICTs used in the big data analytical framework ecosystem simplify aspects related to the relevant data that needs to be gathered, processed and analysed through its various parts of the whole life cycle for purposes of optimisation of various activities, processes as well as decision making.

6. Conclusions

The current work provides an understanding for organisations that are in the process of implementing big data solutions. The paper presents a framework on how to go ahead to make well-organized decisions, implementation of the decisions and analysis of the results, especially for the asset management function. There is a need not only to invest in ICTs, which are crucial to make the big data ecosystem work in the organisation. Moreover, there is a need to invest in human aspects of big data analytics, i.e. a data scientist, because it is necessary to manipulate the big data to be able to make it work and elicit its knowledge for decision-making purposes. Thus, the 3 V's and big data technologies are crucial, however, what is also important for the companies is to

convert the data into information and knowledge that results in increased business value.

Acknowledgements

The research has been conducted as a part of MANTIS Cyber Physical System based Proactive Collaborative Maintenance project. The project has received funding from the Electronic Component Systems for European Leadership Joint Undertaking under grant agreement No 662189. This Joint Undertaking receives support from the European Union's Horizon 2020 research and the national funding organisations Finnish Funding Agency for Innovation Tekes and Ministerio de Industria, Energía y Turismo (Spain).

References

[1] Lee, J., Ni, J., Djurdjanovic, D., Qiu, H., Liao, H. Intelligent prognostics tools and e-maintenance. Computers in Industry, E-maintenance Special, 2006; 57; 476–489. doi:10.1016/j.compind.2006.02.014
[2] Campos, J. Development in the Application of ICT in Condition Monitoring and Maintenance.Computers in Industry, 2009; 60 (1): pp. 1–20. doi:10.1016/j.compind.2008.09.007.
[3] Lee, J., Bagheri, B., & Kao, H. A. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manufacturing Letters, 2015; 3, pp.18-23.
[4] C. Ji, Y. Li, W. Qiu, U. Awada and K. Li. Big Data Processing in Cloud Computing Environments, 2012 12th International Symposium on Pervasive Systems, Algorithms and Networks, San Marcos, TX, 2012; pp. 17-23.
[5] Zhang, L and Karim, R. Big data mining in eMaintenance: An overview, Proceedings of the 3rd International workshop and congress on eMaintenance, 2014, 17-18, Luleå, Sweden
[6] Baltzan, P. Business driven information systems, (3rd ed.). New York: McGraw-Hill, 2012.
[7] Manyika, J., Chui,. M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, 2011.
[8] Min Chen, Shiwen Mao, Yin Zhang, Victor C. M. Leung, Big Data: Related Technologies, Challenges and Future Prospects, 2014; 33-49.
[9] Philip Chen, C. L., and Chun-Yang Zhang. Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data. Information Sciences 275, 2014; pp. 314–47. doi:10.1016/j.ins.2014.01.015.
[10] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. From data mining to knowledge discovery in databases. AI magazine, 1996; 17; 3; 37-54.
[11] Maimon, O. and Rokach, L. Data Mining and Knowledge Discovery Handbook, Springer US, 2010,
[12] Fhom, H. S. Big Data: Opportunities and Privacy Challenges. arXiv preprint arXiv:1502.00823, 2015.
[13] Agnellutti, C. Big Data: An Exploration of Opportunities, Values, and Privacy Issues. Nova Science Publishers, Inc., 2014.
[14] Manyika,J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Hung Byers, A. Big data: The Next Frontier for Innovation, Competition, and Productivity, McKinsey Global Institute, 2012.
[15] Davenport, T.H. Big Data at Work: Dispelling the Myths, Uncovering the Opportunities. Harvard Business Review Press, Boston, Massachisetts, USA, 2016.
[16] Tan, P., Steinbach, M., Kumar, V. Introduction to Data Mining, Addison - Wesley, Reading, MA, 2006.
[17] Friedman, J.H. Data mining and statistics: What's the connection? Proceedings of the 29th Symposium on the Interface Between Computer Science and Statistics, 1997.
[18] Keim, D. (2002). Information visualization and visual data mining,Vis. Comput. Graph. IEEE Trans., vol. 8, no. 1, pp. 1–8, 2002.
[19] De Oliveira, M. and Levkowitz, H. From visual data exploration to visual data mining: A survey,Vis. Comput.Graph. IEEE Trans., 2003; 9; 3; 378–394.
[20] Cavanillas, J.M., Curry, E., Wahlster, W. New Horizons for a Data-Driven Economy – Springer. 2016; 152-154.
[21] Connolly, T and Begg, C.E. Database Solutions: A step by step guide to building databases, 2nd ed. Addison-Wesley, 2004; 191 - 193.
[22] Garg, N., Singla, S., and Jangra, S. Challenges and Techniques for Testing of Big Data, Procedia Computer Science, 2016; 85; 940–948. doi:10.1016/j.procs.2016.05.285.
[23] Meng, Xiangrui, et al. Mllib: Machine learning in apache spark. JMLR 17.34, 2016; 1-7.
[24] Thusoo, A., Sarma, J. S., Jain, N., Shao, Z., Chakka, P., Zhang, N., Antony, S., Liu, H., and Murthy, R. Hive — A petabyte scale data warehouse using Hadoop. In Proceedings of the International Conference on Data Engineering, 2010; 996–1005.
[25] Assuncao MD, Calheiros RN, Bianchi S, Netto MAS, Buyya R. Big Data computing and Clouds: trends and future directions. Special Issue on Scalable Systems for Big Data Management and Analytics. Journal of Parallel and Distributed Computing. 2015; 79 – 80; 3–15.
[26] Kshetri, N. Big data's impact on privacy, security and consumer welfare, Telecommunications Policy, 2014; 38; 1134–1145.
[27] Provost, F., & Fawcett, T. Data Science and its Relationship to Big Data and Data-Driven Decision Making. Big Data, 2013; 1; 1; 51-59.
[28] Mariscal, Gonzalo, Oscar Marban and Covadonga Fernandez. A Survey of Data Mining and Knowledge Discovery Process Models and Methodologies. The Knowledge Engineering Review, 2010; 25; 137–66.
[29] Fan, J., Han, F., and Liu, H. Challenges of Big Data analysis, National Science Review, 2014; 1; 293 – 314.
[30] Dahlbom, B., (Ed.) The infological equation: Essays in honor of Börje Langefors. Gothenburg: Gothenburg University, Dept. of Informatics, 1995.
[31] Chen, G., Wua, S., and Wang, Y. The Evolvement of Big Data Systems: From the Perspective of an Information Security Application". Big Data Research, 2015; 2; 65-73.
[32] Jagadish, H.V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J.M., Ramakrishnan, R., and Shahabi, C.Big Data And Its Technical Challenges, Communications of the ACM, 2014; 57; 7; 86-94.
[33] Fritzsche, R., Gupta, J. N. D., & Lasch, R. Optimal prognostic distance to minimize total maintenance cost: The case of the airline industry. International Journal of Production Economics, 2014; pp.151, 76-88.
[34] Tan, A. R., Matzen, D., McAloone, T. C., & Evans, S. Strategies for designing and developing services for manufacturing firms. CIRP Journal of Manufacturing Science and Technology, 2010; 3(2), pp. 90-97.
[35] Meier, H., Völker, O., & Funke, B. Industrial product-service systems (IPS2). The International Journal of Advanced Manufacturing Technology, 2011; 52(9-12), pp. 1175-1191.