

PAPER • OPEN ACCESS

Big Data as the basis for the innovative development strategy of the Industry 4.0

To cite this article: T Latinovic *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **477** 012045

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Big Data as the basis for the innovative development strategy of the Industry 4.0

T Latinovic¹, D Preradović², C R Barz³, A Pop Vadean³ and M Todić¹

¹University of Banja Luka, Faculty of Mechanical Engineering, Banja Luka, Bosnia and Herzegovina

²University of Banja Luka, Faculty of Economics, Banja Luka, Bosnia and Herzegovina

³Technical University of Cluj Napoca, Department of Mechatronics and Machine Dynamics, Cluj Napoca, Romania

E-mail: tihomir.latinovic@mf.unibl.org

Abstract. The aim of the paper is establishing a link between Big Data connected with Industry 4.0. We are developing a modern economy which is based on supply chains and logistics. Within each logistics are IT solutions. Some of the world's solutions are compiled in this paper and guidelines are set in the development of such solutions. Industry development 4.0, and in particular the development of Big data, gives us the possibility to introduce new ways of the industry. In the development of industry, maximizing production and reducing costs is extremely important. Real-time data and advanced algorithms give us the possibility to obtain real-time operational industry. Big data and their analysis are of great use in the industry because they improve the accuracy of the decision-making process. We must organize a structure that can design, manage and develop projects based on IoT and Big Data / Analytics technologies. In the industry, good sensor technology provides access to information. The paper has a focus on the managing the database with Big data in the Industry 4.0.

1. Introduction

Data processing with computers, network and internet give us a big benefit. One of the advantages was that information could easily be generated and easily exchanged between people, from people to machines and machines to machines. All this data that generates people and machines, and when a man manages a machine and when we have a collaborative work of man and machine, generate huge amounts of information called "Large Data". Companies then use these large data, through the so-called. "Data Mining", to achieve a better service to its customers and achieve greater profit. It is much important to know both software and hardware for various ways of generating large data, how to memorize them, and to study applications for commercial use.

Data becomes Big Data depending on the capacity of an organization. Something Big Data is for one organization, for the other it is not. There is some opinion that is data transfer over 1 TB. In a real system in production, service, weather, etc., a different large number from different types of devices sometimes are smart and transmit data. Such data has a high speed, a large amount and variety (unstructured data type).

For a better insight into BD size, we set some examples in Table 1 for some of the events that result in a large amount of data.



Table 1. Table with event and description about the amount of data [1]

Event	Description
Flight of Air Bus	10 TB of data is generated every 30 minutes of flight
Smart meters	In 2009, is about 76 million digital electricity meters, which produced 350 million transactions a year. It is estimated if it goes with the current tempo by 2020, the number of these devices amount to 800 million
Cell phone cameras	In 2013 in the world was about 5 million phones cameras. Most of them have GPS. This number increasing from day to day, and also increasing the number of photos and videos that users create
Web users	At the end of 2014, the number of internet users was nearly 3.08 billion which contributed to the volume of Internet traffic measured in ZB
Blogs	Data from 2013 indicate that there are 200 million registered blogs
Email	About 300 billion emails are sent every day
RFID	The use of RFID tags (chips) is increasing at the global level, 2005 was 1.5 million tags, and for 7 years later, 30 billion tags
Twitter	Twitter generates about 12 TB of data per day, with 200 million users and 230 million tweets per day (97,000 tweets per second)
NYSE	New York Stock Exchange daily generates about 1 TB of data

2. Digital Manufacturing

The digitization process consists of simulation, three-dimensional representation and various collaborative tools for creating products through appropriate production processes. Production has been enhanced by the introduction of improved design such as DFM, Computer Integrated Manufacturing (CIM), flexible production, slim production and other production systems that use an information technology the entire production process [2]. 3D tools that handle production from a point of view are developed (Orientation according to demand, Data-driven and Digitally executed). Various strategies are in place to guide the life cycle of products. Production is becoming global, multidisciplinary, based on innovation and customer-oriented [3].

Consequently, Big Data should be understood not only as data sets but also as a set of techniques for processing and analyzing them. Below are the main attributes of Big Data [4]:

- volume – this feature refers to the quantity and means the high dynamics of data growth requiring advanced technologies for efficient processing,
- velocity – data is streaming delivered to the organization in real time, which requires sufficient computing power to provide fast response
- variety –data, the lack of structuring, in different formats, which is generated by various sources,
- variability – Change in the intensity of data during a time,
- complexity – Is a variety way of complex data and unstructured data (eg. documents, video, photos),
- value – refers to the ability to capture unique information from large data sets that significantly influence organizational efficiency.

2.1. Big data

Big data is recently being arrived and applied into the dataset, big data growth is very large and even they don't work with traditional database management system. According to the IDC Big Data is a new generation technology and architecture designed to efficiently extract value from the huge volume of a wide variety of data. Big data is a term for massive data sets having the large, varied & complex type of structure with the challenges of analyzing, Storing and visualizing for extra processes or results. [5]

2.2. Traditional data

Traditional data Traditional data is basically stored in the database or is in the data-warehousing. Example of the traditional database is RDBMS, DBMS, SQL and more. Data Warehouse is a set of tools and techniques to enable the collection of data from the operational system, integration and management of that data into a central database and then the analysis, visualization and other operation can perform in a dashboard. machine. [6],[7]. In Table 2 we see a comparative study of traditional and Big Data.

Table 2. Comparative Study of Traditional vs. Big Data

	Traditional Data	Big Data
1.	Data generated in Enterprise level or it is including traditional data	Data generated in Outside and Enterprise level.
2.	Traditional data sources are include ERP transaction data, CRM data, Web transaction data, Financial data	Non-traditional data sources are including Social Media, Log data, Device data, Sensor data, Video, Images
3.	Data store in gigabytes or terabytes	Data store in Petabytes, Zettabytes, Exabyte's
4.	Data managed in the centralized form	Data manage in physically distributed form
5.	By default, stable and interrelationship	Unknown Relationship
6.	Specialized high-level software as well as hardware used	Inexpensive commodity boxes in cluster mode

3. Industry 4.0

In industry 4.0 products and production systems such as machines, warehouses and operational resources are enhanced for Cyber-Physical Systems (CPS) and are connected to the world's manufacturing networks (Kagermann et al., 2013) [8]. These intelligent entities in production have the ability to interpret data, encourage actions, and have the ability to autonomy self-control and self-optimization (Lee, 2008) [9]. Intelligent products in the future will be identified, know their history, status and how they are produced. Intelligent manufacturing systems are linked with intelligent business processes. This gives us the ability to control and optimize the value chain, starting from the production to the final delivery of the product (Kagermann et al., 2013) [8]. A gradual approach to the physical world and the digital world with CPS systems enables autonomous and decentralized production (Brettel et al., 2014, Monostori, 2014) [10].

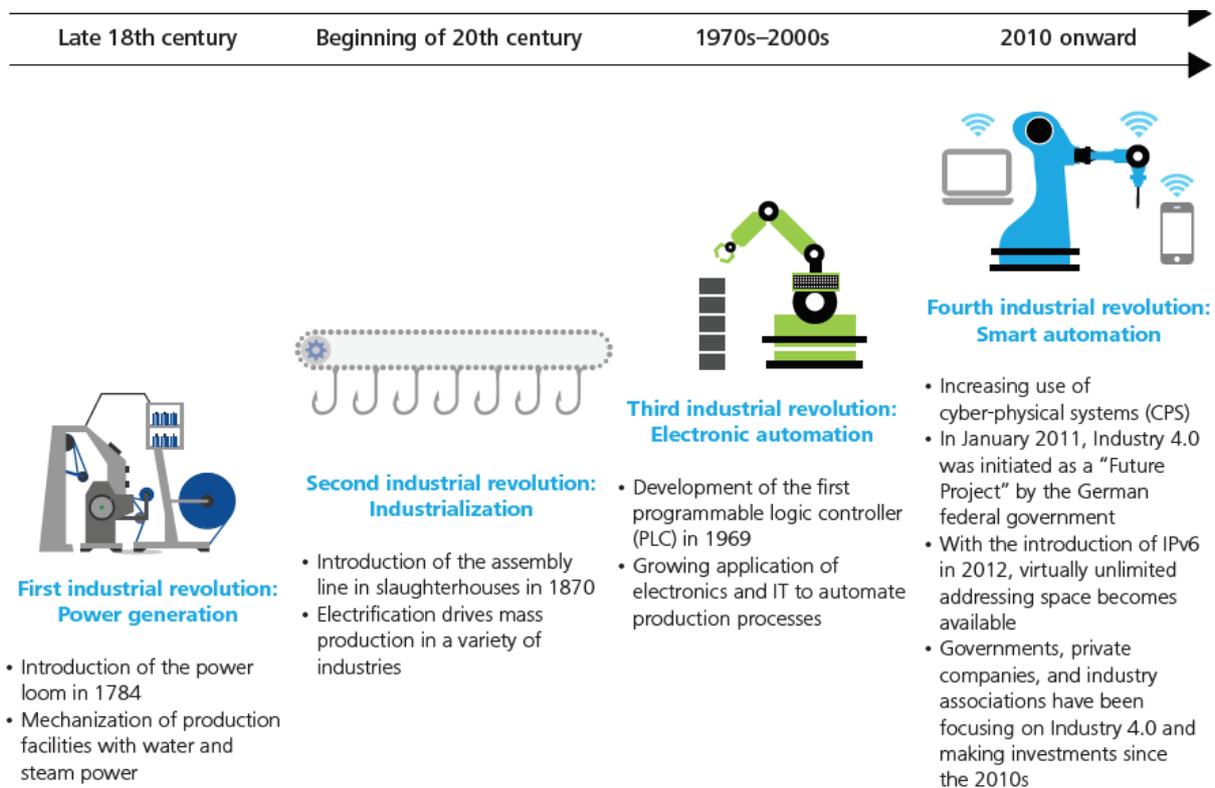


Figure 1. History of Industry 4.0 Graphic: Deloitte University Press | DUPress.com [11]

3.1 Industry 4.0.

Industry 4.0 has six design principles:

- **Interoperability:** With the Internet of things humans communicate with Smart Factories
- **Virtualization:** We working on a virtual copy of the Smart Factory created by 3D software and linked with real sensor data from physical processes.
- **Decentralization:** CPS systems inside Smart Factories have the possibility to make decisions
- **Real-Time Capability:** real-time collecting data
- **Service Orientation:** Service must be oriented on the **Internet of Services**
- **Modularity:** Smart Factories must be flexible adaptation with the module and working islands.

4. Big Data software solutions

There are software solutions that are directly implemented for Big Data in Industry 4.0. There are Hadoop, Cassandra, Scylla, Hypertables, Accumulo, Amazon SimpleDB, Clouds, MonetDV, Apache Flink, Elassandra.

For memory files are: Elastic, MongoDB, Cloud Datastore, Azure DocumentDB, RethinkDB, ToroDBm IBN Cludant etc.

For graphical bases are: Neo4J, Arango DB, Sparksee, GraphBase, Trinity, etc.

NoSQL term used by Carlo Strozzi is the type of open source relational database data that do not have standard SQL interface, but are still relational [12]. There are several approaches to classify the NoSQL database, there are different categories and subcategories, which are subject to type some of them overlapping. Below is provided a basic classification of the database according to the model data, with examples:

1. The key-value: Redis, Riak, OrientDB
2. Document: CouchDB, MongoDB, eXistDB
3. Graph: InfiniteGraph, Neo4J, Stardog
4. Column: Vertica, Druid, HBase
5. Multi-model: CortexDB, MarkLogic, FoundationDB

Semi-structured data are actually a form of structured data, but not in accordance with the formal structure of the data model associated with relational databases. It could be said that the semi-structured data describing themselves (Self-describing), in general, refers to the fact that no scheme. Semi-structured data have appeared all higher use of the Internet where the network is through this type of data, but also appeared the need for the semi-structured data model.

OEM (Object Exchange Model) is one of the standards for presentation of semi-structured data. The key-value databases have the simplest data model. The data is organized so that the associative array notation. A unique key is used to identify a unique record and can be used with three available operations: delete the default key, change the record that is related with that key and enters a new key for the given file. The document database is semi-structured databases, implemented without tables. Each tuple can contain any number of fields and no size limitations of the same. Even if the use in a structured way, there are no limits or limiting scheme database according to a pre-defined structure. Which means that application developers with ease can create dynamic data. This type database can be implemented as a layer through a relational or object database. Another option is to implement directly as semi-structured file formats, such as JSON or XML. Traditional concepts like and the index keys are designed in terms of function as in relational databases data. Using this, we should achieve results of the same performance as in the system implemented by tables.

Table 3. Capabilities and characteristics of representative Big Data software solutions [13]

BIG DATA SOFTWARE SOLUTIONS				
Capabilities /Characteristics	Hadoop HDFS & MapReduce	Cassandra	MongoDB	SimpleDB
Data model	File system	Column	Document	Key-Value
Batch processing / distributed computing	Yes	No	No	No
Real-time queries	No	Yes	Yes	Yes
Random access	No	Yes	Yes	Yes
Horizontal scaling	Yes	Yes	Yes	Yes
Strength	Data processing	Write	Read	Full Indexing
Architecture type	Master-Slave	Peer-to-Peer	Master-Slave	Web Service / Cloud Computing
CAP theorem	Consistency, Partition Tolerance	Availability, Partition Tolerance	Consistency, Partition Tolerance	Availability, Partition Tolerance

5. Conclusion

BD technology has recently been widely used. A large number of scientific papers and projects confirm the necessity of adopting BD tools for everyday work. There are still no projects in Bosnia and Herzegovina for the introduction of Big Data, any forums or blogs on this topic. Also, great data is not mentioned in the curriculum in higher education at the University of Banja Luka.

Many colleagues dealing with computer science do not get involved in this area, because they do not generate large data in the industry in the Western Balkans, but only in the field of Telecommunication.

Big companies almost do not exist, and therefore the need for innovations in the BD and Industries 4.0 industries. Economic areas that have the greatest potential to exploit BD in order to increase profitability, increase efficiency and reduce costs are primarily in the information sector, followed by the financial sector, insurance, management, and retail.

This paper outlines the guidelines for the potential use of Big Data in companies, where they are created, and the software tools that need to work with BD. The paper also lists digital production as the basis of Industry 4.0. When we digitally handle some of it, we send it to digitized production and it produces it, without any worker. Information technologies fully assume the role of service providers and products. These are created by BD, who must process them to overcome digital production. This paper presents the basis of Industry 4.0 as well as most software tools distributed by the type of data they are processing. In further research, we will deal with Bosnia and Herzegovina and where there is no industry development strategy 4.0 because we are still dealing with small and medium-sized enterprises with Industries 3.0 and below. One of the possible applications is the Tax Administration and Telecom Srpske and BiH. The world is slowly shifting from Big to Fast Data and the ability to exploit data generated in real-time and based on their response time. The need for speed is indispensable because it increases profitability and customer satisfaction. The primary goal is to satisfy the client's needs when he wants, where he wants and how he wants, otherwise, he will take over the competition.

Contemporary companies face a number of difficulties when it comes to applying information technology. Business systems are becoming more complex and require significant human resources, so staff in IT departments has a difficult task to maintain and improve the system, both because of technology advancement and difficulties their acceptance by employees, especially when it comes to large companies. Large companies, for as opposed to small and medium enterprises, have a difficult situation because they are less agile and flexible, so they cant quickly respond to change and adapt. Business software and information technology have become integral and the inseparable part of every business entity. Many companies are thanks to innovation in the field of information systems and business applications have succeeded in expanding the current or emerging markets. It can be assumed that modern business is highly dependent on the information systems and software solutions they use. Because of that tracking contemporary trends in the business software market becomes so important that any inertness can lead to the loss of competitive advantage, and in some cases may even endanger survival companies.

Further research can be linked to more specific technical requirements, infrastructure and BD architecture, exploring the possibilities of introducing new exploitation technologies, exploring new methods of accessing fast data, as well as extreme data.

References

- [1] Latinović T S, Preradović D M, Barz C R, Latinović M T, Pop P P and Pop-Vadean A 2016 Big Data in Industry, *IOP Conf. Ser.: Mater. Sci. Eng.* **144** 012006
- [2] ***http://www.plm.automation.siemens.com/en_us/plm/digital-manufacturing.shtml
- [3] ***<https://ieeexplore.ieee.org/abstract/document/6857843/>
- [4] Gierej S 2017 Big data in the industry – overview of selected issues, *Management Systems in production engineering* **25**(4) 251-264
- [5] Suppiah Y, Tariqi Raja R M and Mohamad Amran M F 2016 A Study on Social Data Analytics and Privacy Concern among Social Media Users, *International Journal of Computer Applications* **149**(5) 46-49
- [6] ***<https://gcn.com/blogs/reality-check/2014/01/hadoop-vs-data-warehousing.aspx>
- [7] Tapan P Gondaliya, Hiren D Joshi and Hardik J Joshi 2015 *NewBig Things in Era of DigitalData : 'BigData' & BigDataChallenges with its Solution Using Different Tools*, 10th

- International Conference CALIBER2015, March 12-14, Shimla, India
- [8] Kagermann H, Wahlster W and Helbig J 2013 *Umsetzungsempfehlungen für das Zukunftsprojekt Industry 4.0 – Abschlussbericht des Arbeitskreises Industry 4.0*, Forschungsunion im Stifterverband für die Deutsche Wissenschaft, Berlin, Germany, pp 51-55
 - [9] Lee L C, Su W S and Tang K 2014 *Scheme of handover self-optimization using big data analytics*, The 16th Asia-Pacific Network Operations and Management Symposium, 17-19 September, Hsinchu, Taiwan
 - [10] Monostori L 2014 *Cyber-physical Production Systems: Roots, Expectations and R&D Challenges*, CIRP Conference on Manufacturing Systems, Vol 17, pp 9-13
 - [11] Sniderman B, Mahto M and Cotteleer M J 2014 *Industry 4.0 and manufacturing ecosystems*, July 1, National Academy of Science and Engineering
 - [12] Raghav Mehra, Nirmal Lodhi and Ram Babu 2015 Column Based NoSQL Database, Scope and Future, *International Journal of Research and Analytical Reviews* 2(4)
 - [13] Gölzer P, Cato P and Amberg M 2015 *Data processing requirements of industry 4.0 – use cases for big data applications*, ECIS 2015 Proceedings, Twenty-Third European Conference on Information Systems (ECIS), Münster, Germany, pp 3