

Twenty years survey of Big Data: Definition, concepts and applications in engineering.

Massaki de O. Igarashi¹[0000-0003-4251-8855], Paulo E. Sartorelli¹[0000-0002-3142-0361], Mariana Z. T. de Lima¹[0000-0002-6592-5149]

¹ Mackenzie Presbyterian University, Campinas-SP, Brazil
massaki.igarashi@mackenzie.br

Abstract. In last decade there was an exponential growth in data generation from different sources especially due to advances in information and communication technology. Thus, organizations have seen potential to gain competitive edges from the analyses of this data, changing it in information that, without Big Data tools, could not be obtained. In this context, this work brings a survey about Big Data and explains this concept have changed during the years. Moreover, this paper aims to elucidate about of the last twenty years of Big Data and its applications in different areas of engineering: civil, electrical, manufacturing, mechanical, materials, chemical and software engineering.

Keywords: Big Data, Engineering Areas, Data Generation.

1 Introduction

The last two decades the internet has shown a significant growth, that was especially responsible for the amount and the speed of data generation [1]. Currently, most of actions performed on the Internet generating data, have been identified in order to analyze customer preferences, behavior patterns, evaluate trends and even detect potential crises and fraud [2]. The importance of data generation and its applications increases considering the estimative of data volume growth for 2020 around 40 zettabytes [2][3]. Among the possibilities to extract value from this large amount of data (structured or not); one of the outstanding ways it is to identify existing patterns in databases through the most frequently used information. Another way is that companies can create and store data and get detailed information across a range of areas, such as inventory forecasts, demand prospects over the coming months, and then use that information to make better decisions and improve organizational performance [4].

The term Big Data is used to define a large and complex data set whose traditional database techniques, tools and software are no longer efficient. Therefore, scaling this data, the diversity and complexity requires new techniques, architectures and algorithms for its management and analysis, allowing, an easier extraction of value and knowledge.

Big Data tools and techniques help to extract value and useful information for better decision making in the most diverse areas and possibilities (Table 1).

Table 1 - Different areas of Big Data applications

PRECISION AGRICULTURE	GAMES AND SOCIAL NETWORKS	PRECISION MEDICINE
BANKS AND CREDIT	GOVERNMENT AND PUBLIC MANAGEMENT	HEALTH
INSURANCE	FINANCIAL SECTOR	DECISION-MAKING
EDUCATION	LOGISTICS AND SUPPLY CHAIN	TRANSPORTATION
JUDICIAL	MARKETING	

1.1 Theoretical Framework

According to Hsinchun Chen [05] the use of data as information is not recent. Since the 50's, the administration and information technology areas have already used this concept and information systems. The first research was reported in the 70's; giving rise to data processing methods. In the 90's, softwares have emerged and have been used to analyze the amount of data [06].

Big Data started to be used more frequently in 2000's, with the expansion of new softwares, processing techniques, storage and data transmission [06]. Hsinchun Chen [05], Thomas Davenport [07] and his collaborators emphasized the emergence of new technologies and the spread of the internet and global e-commerce was defined as the beginning of a new era of data generation that is transmitted over this worldwide network. The authors mentioned examples of actions taken from the Internet that began to drive a steady and growing stream of data, such as patterns of web browsing, clicks, online commerce, content that users themselves generate on social media, websites, blogs and platforms. Thus, the term Big Data is not related to a whole new concept, but to the updating of several technologies, tools and techniques that have already existed; this union provides a clearer understanding of big data today [06].

A. The 5 V's of Big Data

There is not a well-established consensus on the definition of the term Big Data, although it is known as a large amount of data that is daily generated daily. The first attempts it is defined by three pillars: volume, velocity and variety [08]. But other authors, when defining Big Data, use five pillars (the 5 V's of Big Data): volume, velocity, veracity, variety and value.

Volume: is related to the large amount of data that is generated daily around the world, coming from different sources [09].

Velocity: the speed which data is generated [10].

Variety: the data diversity by the many sources. Structured or unstructured data [11].

Veracity: the data authenticity is essential for organizations [12]. It is related to the quality of the information available for decision making [09].

Value: value that influences decision making based on data analytics [09].

Trying to systematize these V's that define Big Data, a timeline was created to relate publication dates, definitions and their number of letters V's (Fig. 1).



Fig.1. Big Data Evolution Timeline

Although most authors define only the five pillars (volume, velocity, variety, veracity, and value), there are other authors who add more pillars. For example, M. Ali-ud-din Khan [09] added two other pillars: validity and volatility. Monali Sahoo [18] and collaborators highlight two others: visualization and variability. There are even some current blogging attempts to define the term through 10 V's, but without academic recognition.

2 Methodology

This research has a qualitative approach (which objective is to analyze Big Data definitions in time), with explanatory purpose about the potential of Big Data application in different areas of engineering and to describe its definition along the time. A literature review was conducted from academic publications from national and international literature was searched and used from academic bases Google Scholar, ScienceDirect and ISI Web of Science.

The search was performed based on the keywords: “Big Data” + “Engineering area”, obtaining publication numbers from year to year, from 2000’s to the present date.

3 Analysis and Discussion

3.1 Big Data applications in engineering areas

This study presents some Big Data investigation and applications in 7 engineering areas (civil, electrical, manufacturing, mechanical, materials, chemistry and software).

A. Civil Engineering

Liang Wang [19] reports a new vehicle detection and tracking system based on image data collected by unmanned aerial vehicles (UAV). This system uses consecutively captured frames to generate dynamic vehicle information: position and speed.

This system can be used by “smart cities” in traffic management centers, for the location of accidents, in the real monitoring of a highway, among others, as area of construction, search and rescue applications, structural inspection and health monitoring; and with some modifications it can also be used in jungle animal tracking.

Mazhar Rathore [20] presents an IoT-based system for smart city development and urban planning using Big Data analytics. The current system uses different types of real time sensors (weather sensors, water sensors and intelligent parking), surveillance cameras, emergency buttons on the streets, among other devices. Its implementation is structured in the collection, data filtering, classification, preprocessing, processing and decision making. The sensors generate large amounts of data at high speeds, which are processed by the Hadoop and MapReduce frameworks and guarantee system scalability and efficiency. It can not only benefits citizens, but also authorities by providing information that makes decision making faster and more efficient.

B. Electrical Engineering

Jui-Sheng Chou [21] reports a Big Data analysis framework for smart grids and components of an energy-saving decision support system. The structure analyzes, in real time, the data of electricity consumption, identify consumption patterns and predicts energy consumption; in addition to providing optimal operation schedules for appliances. The measurement infrastructure was installed in a residential building in the experimental simulations.

But, Junghoon Lee [22] reports the development of a smart charging infrastructure for electric cars using Big Data tools. The structure is being implemented in Jeju Province, South Korea, where the authors present the distribution of chargers and the number of chargers in operation. The goal is to combine various statistical and machine learning techniques to identify the charging demand pattern of electric vehicles and integrate renewable energy to power charging networks.

C. Manufacturing Engineering

Yingfeng Zhang [23] presents a proposal for a general Big Data-based analysis architecture for the product lifecycle management (PLM). This architecture integrates Big Data analytics and service-driven standards that aided decision making, coordination and optimization of the Cleaner Production (CP) process. By utilizing the technology and the IoT concept at every stage of the PLM, develop an intelligent manufacturing and maintenance environment.

On the other hand, Jay Lee [24] exposes the possibilities and trends of integrating advanced data analytics into manufacturing, products and services, linking Big Data, the advancement of Information and Communication Technologies (ICT), and the fact that cyber-physical systems (CPS) facilitate the systematic transformation of massive data into information and assist in decision making. In addition, the group also features a systematic architecture called 5C that encompasses the steps required to fully integrate CPS into manufacturing.

D. Mechanical Engineering

Daniel Bumblauskas [25] reports a Smart Maintenance Decision Support System (SMDSS) based on a company's corporate data. The system is able to provide the user with recommendations for improving asset life cycles. The authors illustrate how Big Data analytical tools can be used to prioritize equipment maintenance and define the architecture in which a physical asset owner can enter usage parameters for interconnected equipment and receive a comprehensive proposal for the service to meet the recommendations generated by the analytical model.

Marta Fernandes [26] presents the results of a study in a metallurgy company where data analysis and resource selection methods were employed from continuous equipment monitoring. Machine learning and data mining techniques were used to extract information and assist in decision making. The information gained will assist in the development of adaptive learning models capable of handling complex information that can be applied across a complete line of industrial products and equipment.

E. Materials Engineering

Wencong Lu [27] provides an overview and gives examples of research progress towards the discovery of new materials and classifies this research area as Materials 4.0. The work also highlights the development of machine learning protocols and speaks of quantum material property simulation software, digitized material data, intelligent machine learning algorithms, among others. The author emphasized the considerable reduction in time between concept and commercialization of new products and also speaks of the recovery after conclusion of the product life cycle. The challenge is still aggregating data from multiple sources, as well as managing and analyzing unstructured data to enable 4.0 materials; further recommends that web-based material research platforms need to be developed to explore opportunities and identify gaps.

F. Chemical Engineering

Jody May [28] presents the possibilities of advanced multidimensional separations in mass spectrometry (MS) using Big Data tools; where hybrid analytical instrumentation based on MS has been a powerful technique for meeting the challenges in science and medicine, including helping in drug discovery and synthetic biology. The paper highlights the possibility of large-scale measurements to obtain information and also highlights the challenges of Big Data in chemical analysis: from enumeration of chemical isomers to this field of multidimensional analysis based on MS.

Leo Chiang [29] shows recent advances in some areas such as the chemical and pharmaceutical industry, among others. The objectives of this article are to educate the Chemical Engineering community about the capacity of Big Data and to improve reliability and operational efficiency in various industry sectors, including directing future research. The study also points out that 88% of chemical executives recognize that data analysis will be crucial to maintaining competitive advantage over a 5-year interval.

G. Software Engineering

Timothy Arndt [30] explore how Software Engineering (SE) technology can support the development of big data projects and how big data techniques can be used to develop new processes and evolve SE techniques. The article points out that a lot of research has been done in recent years to improve the production of big data systems and to ensure fast and elastic scaling whenever and wherever needed. One of the main researches focuses on applying SE methods to the production of big data systems is in the area of software architectures development.

According to Giuliano Casale [31] Software engineering as a discipline must be consolidated, because despite presenting impressive achievements in the field, it is one of the latest scientific disciplines. Currently researchers in Software Engineering have been distinguished by the high quality of softwares and the adoption of controlled practices in development and operation.

3.2 Publications with the term Big Data by area

The search for the term “Big Data” + “engineering area” studied at this work in the last 20 years, also held on 27th September 2019 in Google Scholar from the year 2000 until 2019 enabled the drawing of the graph of Fig. 2.

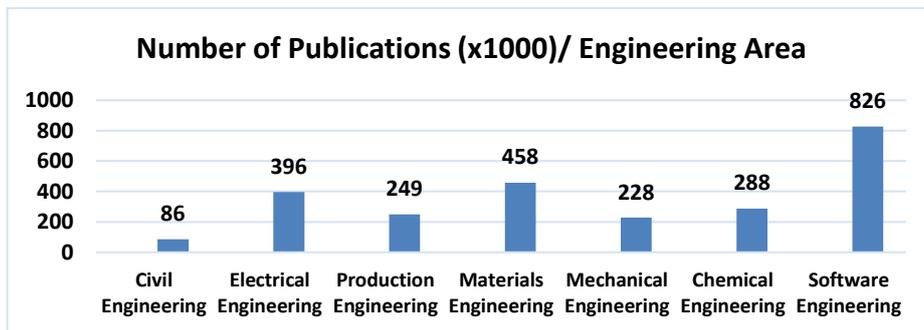


Fig. 2. Number of publications (x 1000) by engineering área

The graph shows the trend of publications in each engineering area studied in this work for the same period. The chart shows a greater number of publications for the terms “Big Data” and “Software Engineering”, with 826,000 publications in the period. The lowest number of publications among those surveyed was “Big Data and” Civil Engineering, “with 86,000 publications.

4 Conclusions

This research presented a brief overview that, in addition to the evolution of publications over a period of 20 years, highlighted some applications in: civil, electrical, manufacturing, mechanical, materials, chemistry and software engineering.

The results of this research demonstrate that some engineering areas, as expected, are of greater interest. But in any case, the results highlight applications that may help in future engineering research.

References

- [1] M. Ge; et al. Big data for internet of things: a survey. Future generation computer systems, 2018.
- [2] J. Gantz, John; D. Reinsel, David. The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the far east. Idc iView: IDC Analyze the Future, 2012. <<https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>> Accessed in 28th Aug. 2019.
- [3] WE ARE SOCIAL. Digital in 2018: world's internet users pass the 4 billion mark. New York: We Are Social Ltd, 30 Jan. 2018. <<https://wearesocial.com/blog/2018/01/global-digital-report-2018>> Accessed in 31st July 2019.
- [4] J. Manyika. Big data: the next frontier for innovation, competition, and productivity, 2011. <http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation.> Accessed in 23 Sept. 2018.
- [05] H. Chen et al. Business intelligence and analytics: from big data to big impact. MIS Quarterly, 2012.
- [06] H. W. Park; L. Leydesdorff. Decomposing social and semantic networks in emerging “big data” research. Journal of Informetrics, 2013.
- [07] T. H. Davenport et al. How 'big data' is different. MIT Sloan Management Review, 2012.
- [08] A. McAfee et al. Big data: the management revolution. Harvard Business Review, 2012.
- [09] Khan, et al. Seven V's of big data understanding big data to extract value. In: Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education. IEEE, 2014.
- [10] D. Barton; D. Court. Get started with big data: tie strategy to performance. Harvard Business Review, Oct. 2012. <<https://hbr.org/2012/10/getting-started-on-a-big-data>> Accessed in 15th Aug 2018.
- [11] J. Tabuena. What every internal auditor should know about big data. Compliance Week, 2012.
- [12] C. Taurion. Você realmente sabe o que é big data. Blog da IBM, 2012. <https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/voce_realmente_sabe_o_que_e_big_data?lang=en> Accessed in 18th July 2018.
- [13] D. Laney. “3D Data Management: Controlling Data Volume, Velocity and Variety” Meta Group Research Note, February 6, 2001.
- [14] P. Russom et al. Big data analytics. TDWI Best Practices Report, Fourth Quarter, v. 19, n. 4, p. 1-34, 2011.

- [15] B. Feldman et al. Big data in healthcare hype and hope. Dr. Bonnie, 2012.
- [16] E. G. Ularu et al. Perspectives on big data and big data analytics. Database systems journal, 2012.
- [17] A. Gandomi; M. Haider. Beyond the hype: big data concepts, methods, and analytics. International journal of information management, 2015.
- [18] M. Sahoo et al. The emergence of big data: a survey. International journal of computer science and mobile applications, india, jun. 2018.
- [19] L. Wang et al. Detecting and tracking vehicles in traffic by unmanned aerial vehicles. Automation in construction, 2016.
- [20] M. Rathore et al. Urban planning and building smart cities based on the internet of things using big data analytics. Computer networks, 2016.
- [21] J. Chou; N. Ngo. Smart grid data analytics framework for increasing energy savings in residential buildings. Automation in construction, 2016.
- [22] J. Lee et al. Big data analysis for an electric vehicle charging infrastructure using open data and software. In: proceedings of the eighth international conference on future energy systems. Acm, 2017.
- [23] Y. Zhang et al. A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products. Journal of cleaner production, 2017.
- [24] J. Lee et al. Industrial big data analytics and cyber-physical systems for future maintenance & service innovation. Procedia cirp, 2015.
- [25] D. Bumblauskas et al. Smart maintenance decision support systems (smdss) based on corporate big data analytics. Expert systems with applications, 2017.
- [26] M. Fernandes et al. Data analysis and feature selection for predictive maintenance: a case-study in the metallurgic industry. International journal of information management, 2019.
- [27] W. Lu et al. Data mining-aided materials discovery and optimization. Journal of materiomics, 2017.
- [28] J. C. May; J. A. Mclean. Advanced multidimensional separations in mass spectrometry: navigating the big data deluge. Annual review of analytical chemistry, 2016.
- [29] L. Chiang et al. Big data analytics in chemical engineering. Annual review of chemical and biomolecular engineering, 2017.
- [30] T. Arndt. Big data and software engineering: prospects for mutual enrichment. Iran journal of computer science, 2018.
- [31] G. Casale et al. Current and future challenges of software engineering for services and applications. Procedia computer science, 2016.