ISSN: 2783-2678

International journal of Innovation in Management Economics and Social Sciences

Journal homepage: www.ijimes.ir

Int. J. Inn. Man. Eco. Soc. Sci. Vol. 1, No 4, 66-72 (2021).

Reviewing the Challenges of Big Data Use in Smart Industries

Ehsan Sabet [1]

[1] *Wolfson School of Mechanical, Electrical and Manufacturing Engineering*
Loughborough University, Leicestershire LE11 3TU, UK

ARTICLE INFO

Received: 14 October 2021

Reviewed: 20 October 2021

Revised:10 November 2021

Accept: 23 November 2021

Keywords: ***Big Data, Smart Industries, Big Data Challenges, Hierarchical Analysis.***

ABSTRACT

Purpose: In recent years, data has been growing on a large scale, and the development of Internet applications, mobile applications, and network-connected sensors has also increased dramatically. These programs and extensive Internet communications continuously generate large volumes of data that are diverse and structurally different, called big data.

Methodology: This article first reviews big data and defines its features, and then discusses the challenges it faces in the smart industry. Finally, using fuzzy hierarchical analysis, the most important challenges of using big data in smart industries have been prioritized.

Findings: Due to the increasing volume of information transfer in the space of industrial generation, big data problem, import and storage of large volume of data information items and its management, preprocessing and post-processing, speed, accuracy and security of information are very important. It has gained a lot of attention and has attracted the attention of many researchers and experts in the field of information technology and active in the industry.

Originality/Value: This article review the challenges of big data use in smart industries.

[1] Corresponding Author: e.sabet@lboro.ac.uk
https://doi.org/10.52547/ijimes.1.4.66

1. Introduction

In recent years, large volumes of data have been stored in various fields including healthcare, general management, retail, biochemistry and other interdisciplinary scientific and research fields. Web-based applications such as social computing, web documents and web search indexing are constantly exposed to big data. This data is stored at a very high cost and is eventually deleted or ignored because there is not enough space to store it. Therefore, the first challenge for metadata analysis is the high-speed input / output storage media. In these circumstances, data accessibility should be a top priority for knowledge discovery and presentation [1]. The reason for this is that this data should be easily and immediately accessible for future analysis. In the past decades, analysts have used hard disk drives to store data, but these devices slow down random input / output performance than sequential inputs / outputs. To overcome this limitation, the concepts of "Solid State Drive" (SSD) and Phase-Change Memory (PCM) were introduced. However, existing storage technologies do not have the efficiency required for large-scale data processing. Big data and artificial intelligence are considered by data scientists or large organizations as the myth of data mechanics [2]. Many organizations hope that artificial intelligence can transform their business using their organizational data. Machine learning is an advanced version of artificial intelligence by which machines can analyze or send and receive data, or learn new concepts. Big data helps companies extract the information they need from them and gain meaningful insights from it. Therefore, it is clear that by integrating artificial intelligence and big data, we come to many new concepts and options, each of which has the ability to transform a large part of different businesses. Artificial intelligence and big data together can help businesses better understand the interests and needs of their customers [3].

Data management is the most challenging issue in the world today. The rapid growth of digital technology has led to an increase in data production rates, which has led to a huge amount of data, so processing this big data with database systems is very difficult. To manage these challenges requires knowledge of computational complexity, information security, and computational methods for big data analysis. For example, many statistical methods that work well for small data sets are not scalable for large volumes of data [4]. Similarly, many computational methods that perform well for low-data data face significant challenges in metadata analysis. Next, the big data challenges in the four groups of "data storage and analysis", discovery of computational knowledge and complexity, "scalability and visualization of data" and "information security" Categorized. All of these are discussed below.

2. Big Data Challenges

With the rapid growth of data sets, data mining tasks have grown significantly. In addition, data reduction, data selection, and feature selection are among the basic tasks, especially when working with large data sets. This reveals an unprecedented challenge for researchers. Because, existing algorithms may not respond at the right time (real time) when working with this high-dimensional data [5].

Therefore, automating the process of analyzing and developing new machine learning algorithms to ensure their robustness is a major challenge. In addition, clustering of large datasets that aid in the analysis of metadata is one of the main concerns in this area (clustering is not what is used in data mining and is an unsupervised learning process; rather, it is the division of data into Parts in order to perform processes faster and more efficiently). Recent technologies such as hadoop and mapReduce make it possible to collect large amounts of "semi-structured" and "unstructured" data in a reasonable amount of time [6]. The types of data that need to be analyzed are shown in Figure (1).

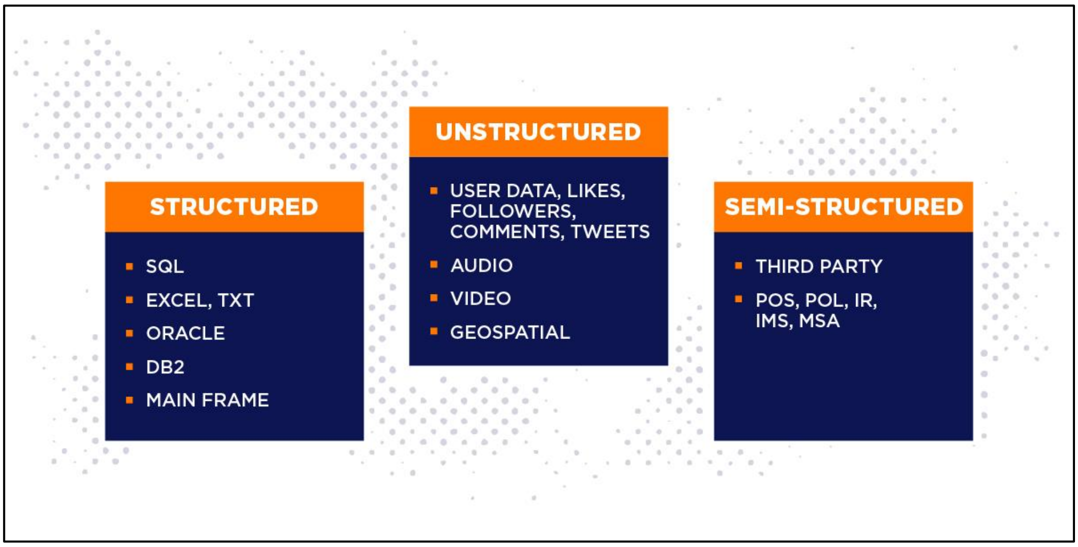

Fig. 1. Data types

A key challenge in this area is how to analyze this data for knowledge. A standard process for this is to convert semi-structured and unstructured data into structured data and then apply data mining algorithms to extract knowledge. A framework for analyzing such data is provided in an article entitled big data analytics: A framework for unstructured data analysis [7].

The details of data analysis for public tweets are also described in an article entitled Opinion mining about a product by analyzing public tweets in twitter by Das et al. The main challenge in this situation is to pay attention to the design of storage systems and increase effective data analysis tools when data comes from various data sources to ensure increased accuracy of outputs. In addition, designing machine learning algorithms to analyze data to improve productivity and scalability is critical [8].

- **Knowledge discovery and computational complexity**

Discovering and presenting knowledge are among the key issues in the metadata debate. These issues have a number of sub-categories such as authentication, archiving, management, protection, retrieval and presentation of information. There are various tools for discovering and presenting knowledge, such as "fuzzy sets", "rough sets", "soft sets", "near sets", "formal concept analysis" (Formal concept analysis), "principal component analysis".

Also, hybrid methods for processing real-world problems have been developed. All of these methods are problem-based. Some of them may not be suitable for large data sets in a "sequential computer". In addition, some also have good scalability characteristics in parallel computers. Because metadata size continues to grow exponentially, existing tools may not be efficient for processing this data to obtain meaningful information. The most popular approaches to managing large datasets are "data warehouses" and "data marts". Data warehousing is primarily responsible for storing data originating from operating systems, while databases are based on data warehouses and facilitate analysis [9].

Big data set analysis requires more computational complexity. The main problem is the management of inconsistent data and the uncertainty that appears in the data set. In general, in principled modeling, the problem of computational complexity is considered. Creating a mathematical system that is comprehensively applicable to metadata is difficult. But domain-based analyzes can be done simply by understanding the specific complexities. A set of such developments can enable big data analysis for a variety of areas [10].

Many researches and studies in this direction have been done using "machine learning" methods (Machine Learning) and using the minimum memory required. The main goal of this research is to minimize the cost and computational complexity. Current metadata analysis tools have poor performance in managing computational complexity, uncertainty, and inconsistencies. This poses greater challenges for the development of methods and technologies that can effectively deal with computational complexity, uncertainty, and incompatibility.

- **Scalability and visualization of data**

One of the most important challenges related to metadata analysis methods is their scalability and security. In recent decades, researchers have focused on accelerating data analytics, which has led to faster processors in accordance with Moore's Law. It is also necessary to develop sampling methods, online analysis and multi-solution analysis methods in order to speed up data analysis. Incremental methods have good scalability in macro data analysis.

Because data size develops much faster than CPU speeds, there is a dramatic change in the technology of processors that have a large number of built-in cores. This change in processors has led to the development of "parallel computing". Real-time application analytics such as social media, finance, and Internet searches are among the things that require parallel processing [11].

The purpose of "Data Visualization" is to present those appropriately using statistical charts, "graph theory" and graphics. Graphic visualization establishes a link between data and its proper interpretation.

- **Information security**

In metadata analyzes, large volumes of data are correlated and analyzed to discover meaningful patterns. Most organizations have different policies in place to protect the security of their sensitive information. Protecting sensitive information is an important issue in metadata analysis because there are so many security risks to metadata. Therefore, information security is a problem for big data analytics. Metadata security can be enhanced using "Authentication", "Authorization" and "Encryption" methods.

The various security metrics faced by large-scale software include network scalability, device diversity, real-time security monitoring, and the lack of "intrusion detection systems | IDS". Security challenges have made "big data" the focus of information security researchers. This focus is on building security policy models and multilevel security systems. Although much research has been done in the field of metadata security, this area still needs further improvement. The main challenge in this regard is to develop a data security model and multi-level privacy for big data [12].

3. Big data technologies

Big data analytics and data science have become the focus of industry and university research. The purpose of data science is to research metadata and extract knowledge from it. Applications of metadata and data science include information science, uncertainty modeling, uncertain data analysis, machine learning, statistical learning, pattern recognition, data warehousing, and signal processing. Effective integration of technologies and analyzes makes it possible to predict future events. The main focus of the following topics is related technologies and topics in need of research in the field of big data.

These technologies fall into four broad categories: Internet of Things (IoT), cloud computing, bio-inspired computing, and quantum computing. However, in reality these issues are not limited to these four categories and are considered much broader [13].

- **Internet of Things (IoT) for big data analytics**

The Internet has reshaped global communications, business, cultural revolutions, and a significant number of human characteristics. Currently, Information Technology activists are trying to control the myriad of automated gadgets on the Internet and build the Internet of Things (IoT).

With the Internet of Things (some call it the Internet of Things, but the Internet of Things is a more general equivalent that better refers to the scope of this topic), devices become Internet users just like humans. The Internet of Things has attracted the attention of researchers and IT companies in recent years due to its numerous opportunities and challenges. It is safe to say that the Internet of Things has economic and social implications for future construction in the field of information technology, networking and communications [14].

- **Cloud computing**

The development of "virtualization technologies" has made "supercomputing" more accessible and cost-effective. The computing infrastructure hidden in virtualization software allows these systems to act like a real computer but with the flexibility to specify details such as number of processors, disk space, memory, and operating system. The use of these virtual computers, known as cloud computing, is one of the most robust macro solutions. Big data and cloud computing technologies are being developed based on the importance of building scalable and accessible data resources. Cloud computing synchronizes large volumes of data with demand-based access to configure computer resources through virtualization methods.

Benefits of using cloud computing include providing resources when there is demand and addressing the resources needed to build and develop a product. At the same time, it improves accessibility and reduces costs. The remaining challenges and research topics in this field have been studied in detail by many researchers, including topics related to data management, data diversity and speed, data storage, data processing, and resource management. Thus, cloud computing helps to develop a business model for a variety of applications with infrastructure and tools [15].

- **Biology-inspired computing for big data analytics**

Computing is inspired by biology, a way of being inspired by nature to deal with complex real-world problems. Big systems are self-organizing without a central control. A nature-inspired cost-cutting mechanism finds the optimal data service solution by searching for data management and service maintenance costs. These methods are developed by biomolecules such as DNA and proteins to guide and perform computational calculations, including data storage, retrieval, and processing.

A notable feature of such computing is that it integrates bio-derived items to perform computational functions and obtain intelligent performance. These systems are more suitable for metadata applications. Because huge amounts of data have been generated from a variety of sources across the web since the advent of digitalization. Analyzing this data and categorizing it into text, photos, video, and more requires a lot of intelligent analysis by data scientists and big data experts [16].

- **Quantum computing for big data analytics**

A quantum computer has a memory that is exponentially larger than its physical size and can manipulate an exponential set of inputs simultaneously. This exponential improvement is possible in computer systems. If a real quantum computer existed, it could solve problems that are difficult for today's computers, and of course macro-data problems. The main technical challenges in the construction of quantum computers are expected to be resolved soon, and humans will witness a revolution in computing. Quantum computing provides a solution for integrating quantum mechanics and information processing [17].

In traditional computers, information is displayed in long strings of bits, so they end up being either zero or one. A quantum computer, on the other hand, uses bits or "qubits". The main difference between qubits and bits is that the quantum system encodes zero and one into two distinct quantum modes and uses mechanical phenomena and laws to decompose and clog. This is because qubits have quantum behavior [18].

4. Conclusion

In recent years, data is growing on a large scale, and the development of Internet applications, mobile applications, and network-connected sensors has also expanded widely. These programs and extensive Internet connections continuously generate large volumes of data that are diverse and structurally different, called big data. Facing such challenges is part of the big data problem. Undoubtedly, there are wide differences between scientists and thinkers, and those who intend to use big data should be aware of the dangers and drawbacks of this work. Otherwise, businesses will suffer serious losses and will continue to face financial crises and heavy fines. For this reason, in this study, the most important challenges facing big data analysis have been evaluated and analyzed. A correct understanding of these challenges can be provided to facilitate the challenges as well as better encounter with big data systems.

References

[1] Fathi, M., Haghi Kashani, M., Jameii, S. M., & Mahdipour, E. (2021). Big data analytics in weather forecasting: A systematic review. *Archives of Computational Methods in Engineering*, 1-29. https://doi.org/10.1007/s11831-021-09616-4

[2] Nozari, H., Fallah, M., Kazemipoor, H., & Najafi, S. E. (2021). Big data analysis of IoT-based supply chain management considering FMCG industries. *Бизнес-информатика*, *15*(1 (eng)). DOI: 10.17323/2587-814X.2021.1.78.96

[3] Nahr, J. G., Nozari, H., & Sadeghi, M. E. (2021). Green supply chain based on artificial intelligence of things (AIoT). *International Journal of Innovation in Management, Economics and Social Sciences*, *1*(2), 56-63. https://doi.org/10.52547/ijimes.1.2.56

[4] Aliahmadi, A., Jafari-Eskandari, M., Mozafari, A., & Nozari, H. (2016). Comparing linear regression and artificial neural networks to forecast total productivity growth in Iran. *International Journal of Information, Business and Management*, *8*(1), 93.

[5] Aliahmadi, A., Jafari-Eskandari, M., Mozafari, M., & Nozari, H. (2013). Comparing artificial neural networks and regression methods for predicting crude oil exports. *International Journal of Information, Business and Management*, *5*(2), 40-58.

[6] Nozari, H., Fallah, M., & Szmelter-Jarosz, A. (2021). A conceptual framework of green smart IoT-based supply chain management. *International Journal of Research in Industrial Engineering*, *10*(1), 22-34. DOL: 10.22105/riej.2021.274859.1189

[7] Das, T. K., & Kumar, P. M. (2013). Big data analytics: A framework for unstructured data analysis. *International Journal of Engineering Science & Technology*, *5*(1), 153.

[8] Das, T. K., Acharjya, D. P., & Patra, M. R. (2014, January). Opinion mining about a product by analyzing public tweets in Twitter. In *2014 International Conference on Computer Communication and Informatics* (pp. 1-4). IEEE. DOI: 10.1109/ICCCI.2014.6921727

[9] Fallah, M., & Nozari, H. (2020). Quantitative Analysis of Cyber Risks in IoT-Based Supply Chain (FMCG Industries). *Journal of Decisions & Operations Research*, *5*(4).

[10] Nozari, H., & Sadeghi, M. E. (2021). Artificial intelligence and Machine Learning for Real-world problems (A survey). *International Journal of Innovation in Engineering*, *1*(3), 38-47.

[11] Mei, H., Guan, H., Xin, C., Wen, X., & Chen, W. (2020). Datav: Data visualization on large high-resolution displays. *Visual Informatics*, *4*(3), 12-23. https://doi.org/10.1016/j.visinf.2020.07.001

[12] Fallah, M., Sadeghi, M. E., & Nozari, H. (2021). Quantitative analysis of the applied parts of Internet of Things technology in Iran: an opportunity for economic leapfrogging through technological development. *Science and Technology Policy Letters*.

[13] Naeem, M., Jamal, T., Diaz-Martinez, J., Butt, S. A., Montesano, N., Tariq, M. I., ... & De-La-Hoz-Valdiris, E. (2022). Trends and Future Perspective Challenges in Big Data. In *Advances in Intelligent Data Analysis and Applications* (pp. 309-325). Springer, Singapore. https://doi.org/10.1007/978-981-16-5036-9_30

[14] Ding, Y., Jin, M., Li, S., & Feng, D. (2021). Smart logistics based on the internet of things technology: an overview. *International Journal of Logistics Research and Applications*, *24*(4), 323-345. https://doi.org/10.1080/13675567.2020.1757053

[15] Sadeeq, M. M., Abdulkareem, N. M., Zeebaree, S. R., Ahmed, D. M., Sami, A. S., & Zebari, R. R. (2021). IoT and Cloud computing issues, challenges and opportunities: A review. *Qubahan Academic Journal*, *1*(2), 1-7. https://doi.org/10.48161/qaj.v1n2a36

[16] Dagdia, Z. C., Avdeyev, P., & Bayzid, M. S. (2021). Biological computation and computational biology: survey, challenges, and discussion. *Artificial Intelligence Review*, 1-67. https://doi.org/10.1007/s10462-020-09951-1

[17] von Burg, V., Low, G. H., Häner, T., Steiger, D. S., Reiher, M., Roetteler, M., & Troyer, M. (2021). Quantum computing enhanced computational catalysis. *Physical Review Research*, *3*(3), 033055.

[18] Piattini, M., Peterssen, G., & Pérez-Castillo, R. (2021). Quantum computing: A new software engineering golden age. *ACM SIGSOFT Software Engineering Notes*, *45*(3), 12-14.