Journal of Data Analytics, Vol 1, No 1, (2022), 16-24

Review the challenges of using big data in the supply chain

Hadeel S. Obaid [1]

[1] *PhD student, Tampere University, Finland*

ARTICLE INFO

Received: 12 March 2022
Reviewed: 10 May 2022
Revised: 18 May 2022
Accepted: 25 May 2022

Keywords:

Big Data, Big Data-Based Supply Chain, Big Data Analysis, Big Data Challenges, Data Based Supply Chain Challenges.

ABSTRACT

The increasing growth of computer networks and Internet-based technologies, followed by the growth of data and information required by their users and consumers, has led to the emergence of new concepts in this field. Big data is one of these concepts that has been considered by researchers in various fields of business in recent years. When looking at it from the outside, it is fair to assume that the more data a company or organization has, the better, because the company in question will have a larger amount of data for mining, and as a result their data will be more accurate. However, this is not always the case, because learning how to effectively manage Big Data has become a very challenging task for many businesses around the world. Working with big data involves collecting data from information sources, exploring and analyzing them, modeling them based on the desired features, and providing data security measures. For this reason, this paper examines the challenges of working with big data and the big data revolution in general and big data mining in the business supply chain as fundamental business processes.

[1] Corresponding Author: hadeel.obaid@tuni.fi
https://doi.org/10.52547/jda.1.1.16

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. However, it is important to note that the term "big data" is relatively new, as it was first coined in 2005. Big data is so expensive and complex that it is virtually impossible to analyze and process this data using traditional methods. Big data analytics helps organizations harness their data and use it to identify new opportunities (Nozari, Sadeghi et al., 2022). This, in turn, leads to smarter business moves, more efficient operations, more profits and better customers. Over time, the problems with big data have become more challenging, as more data is generated than ever before, and this data is growing exponentially (Nozari, Fallah et al., 2021). One of the most important challenges that businesses face is how to cope with the huge amount of data they are bombarded with every day. It is not easy to get through the data quickly. If an organization wants to get some value from its data, it must find a way to process and organize it quickly, which is much easier said than done (Ghahremani-Nahr et al., 2022).

These days, companies have data that is transferred to them very quickly and in bulk, and they cannot process it faster than they receive it. This leads to inefficient organization and can even be detrimental to the company's productivity. To manage these challenges requires knowledge of computational complexity, information security, and computational methods for macro data analysis (Aliahmadi et al., 2022). Companies must not only learn to cope with the growing volume of data, but also learn how to turn data into insights. Because if the data is not analyzed enough, they alone cannot provide intrinsic value to an organization. This is what every company does with data and it is very important (Aliahmadi et al., 2022).

Using Big Data is a big change for any company, it has to be accepted first by the senior management and then by the lower-level employees. To accept more big data in the organization, the implementation and use of new big data solutions requires monitoring and control (Najafi et al., 2022). However, excessive control by senior managers may have a negative impact. With the arrival of new technologies and tools needed for big data solutions, having the skills to work with these tools is one of the concerns of managers of organizations involved in big data discussions. Also, a higher and more professional level is required to work with this Systems and their management are needed because the existing systems in big data still do not have very simple user interfaces. Therefore, as it is known, the use of big data can always face major challenges. Therefore, due to the importance of the supply chain and its processes, it seems that a correct understanding of the challenges facing it in the face of this huge amount of data can be essential. In this research, the dimensions and components of the supply chain based on the use of big data are described and its most important challenges are investigated using a decision-making method based on hierarchical analysis.

The structure of this article is as follows. First, in the second part, the literature review about big data and its use in the supply chain is examined, and also the most important challenges are raised, and the research method is presented in the third part. Analytical results are presented in the fourth part and finally, the conclusion is presented in the fifth part.

2. Literature Review

In recent years, the volume of data through various tools - data generation - such as mobile devices, sensor technologies, "remote sensing" and "radio frequency identification readers" has grown exponentially. The concept of big data has been around for years. Most organizations now realize that if they get all the data in their business, they can use their analysis and get significant value from it (data-driven decision making) (Nozari et al., 2021). Big data contains an abstract concept; That is, data sets that are not visible, acquired, managed, and processed by traditional IT and software and hardware tools. This data is stored at a very high cost and is eventually deleted or ignored because there is not enough space to store it. This is because this data should be easily accessible and real-time for future analysis (Lotfi et al., 2016).

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. Various tools for discovering and presenting knowledge such as "fuzzy sets", "rough sets", "soft sets", "near sets", "formal concept analysis" (There is formal concept analysis, principal component analysis, and more (Nahr et al., 2021). 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 (Nozari et al., 2022).

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 macro data analysis for a variety of areas (Ghahremani-Nahr et al., 2021).

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 complexity of computation. 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 (Aliahmadi et al., 2013).

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. Based on what has been said, it is clear that metadata has created many challenges for the development of hardware and software, and this has led to the creation of "parallel computing", "cloud computing", and "distributed computing" (Mohammadi et al., 2015).

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 (Nozari et al., 2016).

It uses big data analytics and quantitative supply chain methods to improve decision-making for all activities throughout the supply chain. Specifically, this analysis does two innovative things. First, it expands the dataset for analysis beyond the traditional internal data held in enterprise resource planning (ERP) and supply chain management (SCM) systems (Szmelter-Jarosz et al., 2021). Second, it applies powerful statistical methods to new and existing data sources. This creates new insights that help improve supply chain decision-making, from improving front-line operations to strategic choices, such as selecting appropriate supply chain operating models. Typically, planning is now the most data-driven process in the supply chain, using a wide variety of inputs from enterprise resource planning (ERP) and SCM planning tools(Nozari & Szmelter et al., 2018). There is now significant potential to truly redefine the planning process, however, by leveraging new internal and external data sources to transform demand and supply in real-time. Figure 1 shows the impact points of big data on the supply chain.

Fig. 1. The role of big data on the supply chain

The use of big data analytics has a wide range of benefits for business networks, but it goes without saying that it also brings challenges. In the following, we will describe the challenges we face when implementing and analyzing big data in the supply chain.

- ***Amount of data collected (C1)***: With today's data-driven organizations and the introduction of big data, managers and other employees are often overwhelmed by the amount of data collected. A supply chain may receive information about every incident and interaction that takes place on a daily basis, leaving analysts with thousands of interconnected data sets (Nozari, zmelter-Jarosz and Ghahremani-Nahr, 2021).
- ***Collect critical data in real-time (C2)***: With so much information, it can be difficult to drill down and get the insight you need most. When employees are overworked, they may not fully analyze data or focus only on metrics that are easiest to collect, rather than other measures that are truly more valuable(Nozari et al., 2021).
- ***Visual presentation of data (C3)***: To be understood and effective, data often needs to be presented visually in the form of charts and photographs. While these tools are incredibly useful, they are difficult to build by hand. Taking the time to pull information from multiple fields and put it into a reporting tool is frustrating and time-consuming (Nahr et al., 2021).
- ***Inaccessible data (C4)***: Moving data into a centralized system has little effect if it is not easily accessible to the people who need it (Tavakkoli-Moghaddam et al., 2021).

- ***Poor quality data (C5)***: Nothing is more harmful to data analysis than inaccurate data. Without proper input, you will have unreliable output. One of the main causes of incorrect data is manual errors made during data entry.
- ***Big Data Security Vulnerabilities (C6)***: Big Data security challenges are a broad topic. Oftentimes, security is forgotten in the adoption of big data projects. Big data technologies evolve, but their security features are still neglected (Fallah and Nozari, 2020).
- ***Increasing users, interactions and generated content (C7)***: This is a challenge for all parts of data management. Both the storage layer and the exploration layer must be very fast and scalable.
- ***Energy management (C8)***: The energy consumption of a processor's computing system is very important from an economic point of view. With the increase in the volume of data and the demands of analysis, processing, storage and transmission of big data, a lot of electrical energy is inevitably consumed. Hence, system power consumption control should be done for big data to ensure scalability and availability (Fallah et al., 2021).
- ***Data privacy (C9)***: Big data that is produced in today's world contains a large amount of personal life information of people that should be protected in privacy (Nozari et al., 2021).

In general, the challenges that the supply chain faces in the face of big data are shown in Figure 2. Correctly understanding the challenges and how to face them helps to implement an intelligent supply chain with high sustainability.

For this purpose, in this research, we have evaluated this basic challenge using a decision-making method based on hierarchical analysis. This evaluation shows the importance of each of the key sections in dealing with big data.

Fig. 2. Big data challenges in the supply chain

3. Research Methods

In this research, a fuzzy nonlinear hierarchical analysis method is used, which is represented by the Mikhailov method (Nozari et al., 2019). In this method, fuzzy even comparisons are assumed to be triangular fuzzy numbers. The definite weight vector (priority) $w = (w_1, w_2, \ldots, w_n)$is extracted in such a way that the priority rate is approximately within the range of the initial fuzzy judgments. In other words, the weights are determined so that relation (1) is established.

$$l_{ij} \le \frac{w_i}{w_j} \le u_{ij} \tag{1}$$

Each definite weight vector (w) holds with a degree in the above fuzzy inequalities which can be measured by the linear membership function of Equation (2) (in terms of unknown rate):

$$\mu_{ij}\left(\frac{w_i}{w_j}\right) = \begin{cases} \dfrac{(w_i / w_j) - l_{ij}}{m_{ij} - l_{ij}} & \dfrac{w_i}{w_j} \le m_{ij} \\ \dfrac{u_{ij} - (w_i / w_j)}{u_{ij} - m_{ij}} & \dfrac{w_i}{w_j} \le m_{ij} \end{cases} \tag{2}$$

Given the specific form of membership functions, the fuzzy prioritization problem becomes a nonlinear optimization problem in the form of Equation (3).

$$\begin{aligned} &\max \lambda \\ &Subject\ to: \\ &(m_{ij} - l_{ij})\lambda w_j - w_i + l_{ij} w_j \le 0 \\ &(u_{ij} - m_{ij})\lambda w_j + w_i - u_{ij} w_j \le 0 \\ &i = 1,2,...,n-1\ ,\ \ j = 2,3,...,n\ ,\ \ j > i, \\ &\sum_{k=1}^{n} w_k = 1 \qquad w_k > 0, \qquad k = 1,2,...,n \end{aligned} \tag{3}$$

5. Research findings

The stages of evaluation and ranking are divided into two main parts, namely determining the matrix of pairwise comparisons and using mathematical modeling for ranking. This pairwise comparison table is shown in Table (1). These tables are used for calculations by the Mikhailov method.

Table 1. Pair comparison matrix for big data challenges in supply

	C1			C2			C3			…	C7			C8			C9		
C1	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-
C2	2.2	3.5	4.8	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-
C3	2.2	2.8	3.5	3.2	4.5	5.1	-	-	-		-	-	-	-	-	-	-	-	-
C4	1.2	2.5	4.1	2.2	2.8	3.5	2.1	3.5	4.7		-	-	-	-	-	-	-	-	-
C5	2.5	3.2	4.2	1.2	2.5	3.7	1.2	2.5	4.8	…	-	-	-	-	-	-	-	-	-
C6	3.1	4.1	4.8	2.2	3.4	4.1	1.8	2.1	2.8		-	-	-	-	-	-	-	-	-
C7	2.1	2.7	3.7	3.1	4.1	5.2	3.1	3.7	4.1		-	-	-	-	-	-	-	-	-
C8	1.8	2.7	3.5	1.7	2.2	3.5	2.2	2.8	4.1		2.1	2.8	3.1	-	-	-	-	-	-
C9	1.5	3.2	3.8	2.4	3.2	3.7	2.4	3.2	3.8		1.8	2.5	3.5	2.2	3.1	4.5	-	-	-

By placing the data from Table (1) in the nonlinear model (3) and solving the model using *LINGO* software, the weight and rank of each challenge can be obtained. This is shown in Table (2).

Table 2. Weight of IoT Big data Challenges in supply chain

Challenge	Code	Weight	λ
Amount of data collected	C1	0.149852	0.452147
Collect critical data in real-time	C2	0.108581	
Visual presentation of data	C3	0.068741	
Inaccessible data	C4	0.064751	
Poor quality data	C5	0.107413	
Big Data Security Vulnerabilities	C6	0.188114	
Increasing users, interactions and generated content	C7	0.090125	
Energy management	C8	0.067485	
Data privacy	C9	0.155241	

Figure 3 shows the priority of big data challenges in the supply chain.

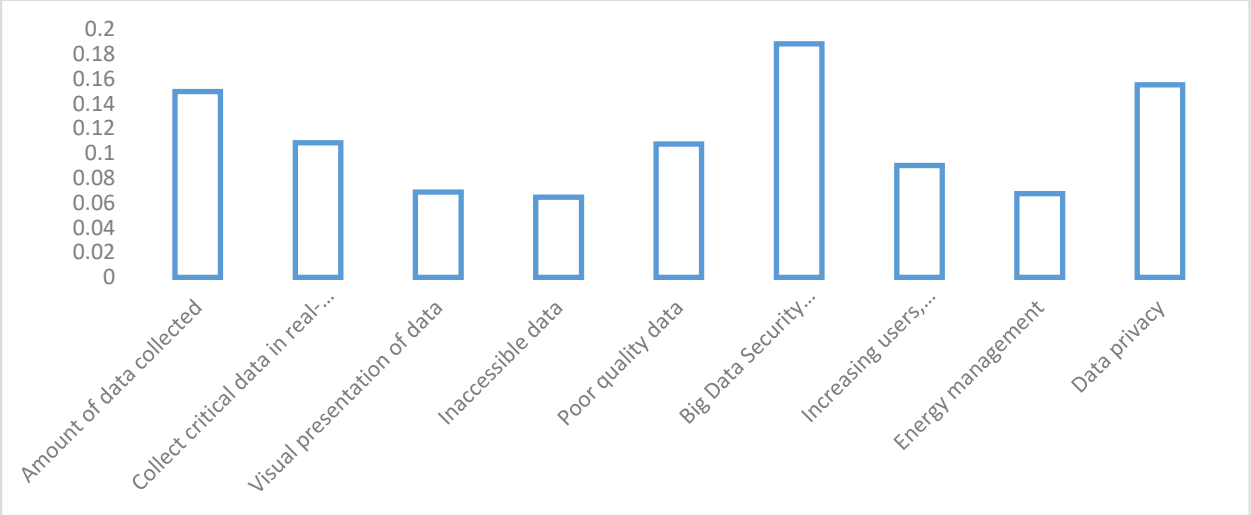

Fig. 3. Priority of big data challenges in supply chain

5. Conclusion

The rapidly growing interest from academics and practitioners in the use of big data analytics (BDA) in supply chain management (SCM) has necessitated a review of up-to-date research development in order to develop a new agenda. In the era of big data, due to the large volume of information, although organizations have been facing many problems, by using big data analysis, they have been able to improve their efficiency to a great extent, and by integrating information in the supply chain by The topic of cloud computing and big data achieve coordination between components and improve communication.

In addition to all the benefits that the presence of big data brings for businesses and especially supply chain processes, the use and presence of this huge amount of data also have challenges. Correct understanding of these challenges can help to implement an intelligent supply chain based on big data

in addition to reducing their amount. For this reason, in this research, the need to recognize and understand the challenges of big data in business processes and the use of integrated information in the era of technology has been discussed, so that in this direction, it can be used in various parts of the supply chain, including customer relations, marketing. , inventory optimization and supplier control and demand forecasting using big data analysis applications reached satisfactory results. The results show that privacy and security challenges are among the most important challenges that big data managers and experts in supply chain processes should pay special attention to, and therefore attention to these challenges can increase the ability of the supply chain.

References

- 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.
- Aliahmadi, A., Nozari, H., & Ghahremani-Nahr, J. (2022). A framework for IoT and Blockchain Based on Marketing Systems with an Emphasis on Big Data Analysis. International journal of Innovation in Marketing Elements, 2(1), 25-34.
- Aliahmadi, A., Nozari, H., & Ghahremani-Nahr, J. (2022). AIoT-based Sustainable Smart Supply Chain Framework. International Journal of Innovation in Management, Economics and Social Sciences, 2(2), 28-38.
- Fallah, M., & Nozari, H. (2020). Quantitative analysis of cyber risks in IoT-based supply chain (FMCG industries). Journal of Decisions & Operations Research, 5(4).
- 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, 11(4).
- Ghahremani-Nahr, J., & Nozari, H. (2021). A Survey for Investigating Key Performance Indicators in Digital Marketing. International journal of Innovation in Marketing Elements, 1(1), 1-6.
- Ghahremani-Nahr, J., Aliahmadi, A., & Nozari, H. (2022). An IoT-based sustainable supply chain framework and blockchain. International Journal of Innovation in Engineering, 2(1), 12-21.
- Lotfi, F. H. Z., Najafi, S. E., & Nozari, H. (Eds.). (2016). Data envelopment analysis and effective performance assessment. IGI Global.
- Mohammadi, H., Ghazanfari, M., Nozari, H., & Shafiezad, O. (2015). Combining the theory of constraints with system dynamics: A general model (case study of the subsidized milk industry). International Journal of Management Science and Engineering Management, 10(2), 102-108.
- Nahr, J. G., Bathaee, M., Mazloumzadeh, A., & Nozari, H. (2021). Cell production system design: a literature review. International Journal of Innovation in Management, Economics and Social Sciences, 1(1), 16-44.
- 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.
- Najafi, S. E., Nozari, H., & Edalatpanah, S. A. (2022). Investigating the Key Parameters Affecting Sustainable IoT-Based Marketing. In Computational Intelligence Methodologies Applied to Sustainable Development Goals (pp. 51-61). Springer, Cham.
- Nozari, H., & Szmelter, A. (Eds.). (2018). Global supply chains in the pharmaceutical industry. IGI Global.
- 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.
- 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)), 78-96.
- Nozari, H., Fallah, M., Szmelter-Jarosz, A., & Krzemiński, M. (2021). Analysis of security criteria for IoT-based supply chain: a case study of FMCG industries. Central European Management Journal, 29(4), 149-171.
- Nozari, H., Najafi, E., Fallah, M., & Hosseinzadeh Lotfi, F. (2019). Quantitative analysis of key performance indicators of green supply chain in FMCG industries using non-linear fuzzy method. Mathematics, 7(11), 1020.
- Nozari, H., Najafi, S. E., Jafari-Eskandari, M., & Aliahmadi, A. (2016). Providing a model for virtual project management with an emphasis on IT projects. In Project Management: Concepts, Methodologies, Tools, and Applications (pp. 476-496). IGI Global.
- Nozari, H., Sadeghi, M. E., & Najafi, S. E. (2022). Quantitative Analysis of Implementation Challenges of IoT-Based Digital Supply Chain (Supply Chain 0/4). arXiv preprint arXiv:2206.12277.

- Nozari, H., Szmelter-Jarosz, A., & Ghahremani-Nahr, J. (2021). The Ideas of Sustainable and Green Marketing Based on the Internet of Everything—The Case of the Dairy Industry. Future Internet, 13(10), 266.
- Nozari, H., Tavakkoli-Moghaddam, R., & Gharemani-Nahr, J. (2022). A neutrosophic fuzzy programming method to solve a multi-depot vehicle routing model under uncertainty during the covid-19 pandemic. International Journal of Engineering, 35(2), 360-371.
- Szmelter-Jarosz, A., Ghahremani-Nahr, J., & Nozari, H. (2021). A neutrosophic fuzzy optimisation model for optimal sustainable closed-loop supply chain network during COVID-19. Journal of Risk and Financial Management, 14(11), 519.
- Tavakkoli-Moghaddam, R., Ghahremani-Nahr, J., Samadi Parviznejad, P., Nozari, H., & Najafi, E. (2021). Applications of Internet of Things in the Food Supply Chain: A Literature Review. Journal of Applied Research on Industrial Engineering,(),-. doi, 10.