Big data and cloud computing: roles and relationships, techniques and tools

Jan Jarosz [1]

[1] Faculty of Management, University of Gdansk, Gdansk, Poland; j.jarosz.108@studm.ug.edu.pl;

ARTICLE INFO	ABSTRACT
	In the past years, the increase in data has been accompanied by rapid growth in various fields. It is difficult to analyze large volumes of data using traditional and relational database technology. Therefore, new databases have emerged, and for this reason, big data has become one of the new topics in IT and business today. Also, the cloud environment is increasingly used to store and process big data. Cloud processing refers to processing anything, including Big Data Analytics, on the "cloud". A "cloud" is a collection of high-powered servers from providers that can often view and query large data sets much faster than a regular computer. These two topics differ from each other in various aspects, including definition, collection references, usage method, form and format, and application. In this research, the dimensions and basic concepts, characteristics, tools and techniques, classification, and communication of data are examined. Big has been dealt with cloud computing, and in addition, storage systems, opportunities and challenges, and big data design principles in the cloud environment have been analyzed.

1. Introduction

Society is becoming increasingly instrumented and as a result, organizations are generating and storing large amounts of data. Managing and gaining insight into the generated information is a challenge and the key to competitive advantage. Analytics solutions that mine structured and unstructured data are important to help organizations gain insight into private data and the vast amounts of public data available on the web (Aliahmadi & Nozari, 2022). The ability to correlate private information on consumer preferences and products with the information from tweets, blogs, product evaluations, and data from social networks opens up a wide range of possibilities for organizations to understand customer needs, anticipate their demands, and optimize usage (Nozari & Aliahmadi, 2022). This paradigm is called the popularity of big data.

Despite the popularity of big data and analytics, putting them into action is still a complex and time-consuming endeavor. Big data offers significant value to organizations willing to adopt it, but at the same time, it raises a significant number of challenges to realize the such added value (Nozari & Nahr, 2022). An organization wishing to continuously use analytics technology requires expensive software licenses, hiring large computing infrastructure, and paying for hourly consulting analysts who work with the organization to better understand the business, organize data, and make it suitable for It is analysis (Aliahmadi et al., 2022).

---

[1] Corresponding Author: j.jarosz.108@studm.ug.edu.pl

Cloud computing is one of the important changes in modern communication and information technology and services for organizational programs and has become a powerful architecture for performing large-scale and complex calculations (Nozari et al., 2022). The advantages of using cloud computing include virtual resources, parallel processing, security, and data service integration with scalable data storage. Cloud computing not only minimizes the costs and limitations of automation and computing by individuals and companies, but also provides reduced infrastructure maintenance costs, efficient management, and user access. This is the biggest advantage of cloud computing, which is achieved by eliminating the investment in software or standalone servers. By using cloud capabilities, companies can save on licensing costs and at the same time eliminate additional costs such as the cost of data storage, software updates, management, etc (Aliahmadi et al., 2022).

Public clouds provide services that are available wherever the end user is located. This method provides easy access to information and meets the needs of users in different time zones and geographical locations. As a side benefit, collaboration thrives because it's now easier than ever to access, view, and modify shared documents and files (Aliahmadi et al., 2022). In addition, uptime is guaranteed in most cases, thereby providing continuous access to resources. Different cloud vendors usually use multiple servers for maximum redundancy. In case of system failure, replacements are automatically generated on other machines (Ghahremani-Nahr et al., 2022).

Cloud computing providers typically use a "software as a service" model so that customers can easily process data. Typically, a console that can receive specialized commands and parameters is available, but everything is also done through the site's user interface. Some of the products that are typically part of this package include database management systems, cloud-based virtual machines and containers, identity management systems, machine learning capabilities, and more. In turn, Big Data is often generated by large network-based systems and this can be in a predominantly standard or non-standard format (Tavakkoli-Moghaddam et al., 2021). If the data is in a non-standard format, in addition to machine learning, artificial intelligence providing "cloud computing" may also be used to standardize the data. From there, the data can be harnessed through the cloud computing platform and used in different ways. For example, it can be searched and edited, and used for future insights.

This cloud infrastructure enables real-time processing of big data. It can take huge "bursts" of data from compact systems and interpret them in real-time. Another commonality between big data and cloud computing is that the power of the cloud enables big data analysis in a fraction of the time.

The purpose of this research is to implement a comprehensive survey on the state of big data in cloud computing environments and to present the definition, characteristics, and classification of big data along with some topics of cloud computing. Also, the relationship between big data and cloud computing, big data storage systems, and technology has been investigated.

2.Big data, dimensions, and features

In an ever-changing business world, many companies now face increasing pressure to develop their business intelligence efforts quickly and at a low cost in order to remain competitive. The recent rise of cloud computing is changing the way companies provide IT services and the way businesses and users interact with IT resources. Big data is an evolving term that describes any large volume of structured, semi-structured, and unstructured data that can be mined for useful information. Big data is data that exceeds the processing capacity of traditional databases. The data is too large to be processed by a single machine. The field of big data analytics examines large amounts of data to discover hidden patterns, correlations, and other insights. Big data technology is made possible by the latest advances in computer technology as well as algorithms and approaches developed to manage big data (Nozari et al., 2021).

Storing and analyzing a large amount of data, which is very important for the work of a company, requires an extensive and complex hardware infrastructure. With the continuous growth of data, the data storage device becomes more important, and many cloud companies pursue large storage capacity for competitiveness (Nozari et al., 2022). Accuracy and timely availability of data is very important for decision-making. Big data is only useful when an information management process is implemented to ensure data quality. Figure 1 shows the architectural big data analytic framework.
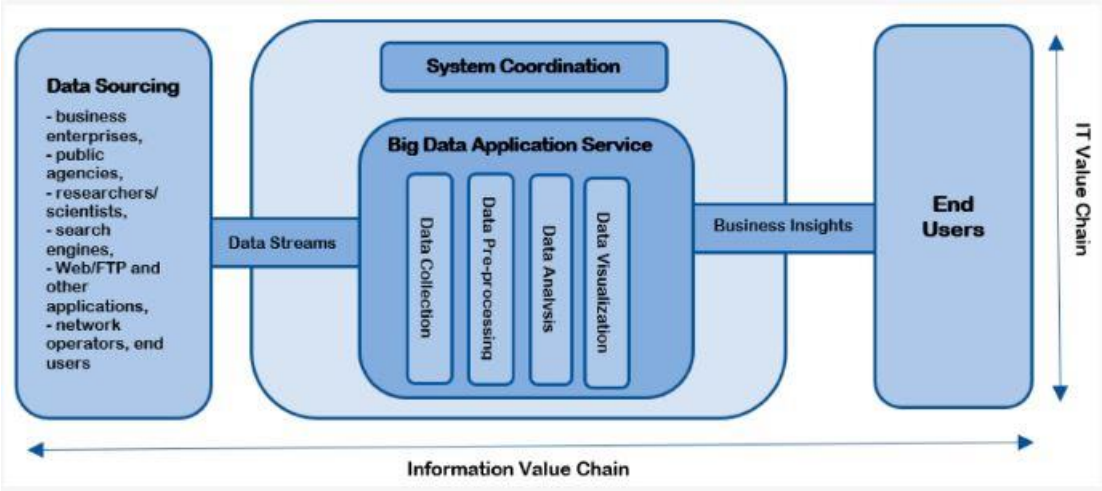

Figure 1: A framework for big data analysis (Mathrani and Lai, 2021)

Security is one of the most important concerns related to big data. To make more sense of big data, organizations must begin integrating parts of their sensitive data into big data. Companies should start creating security policies that are self-adjusting: these policies should take advantage of existing trust relationships and share data and resources across organizations while ensuring that data analysis is optimal. and are not limited due to such policies. Hacking and various attacks on cloud infrastructure affect multiple customers even if only one site is attacked. These risks can be mitigated by using security programs, encrypted file systems, data loss software, and purchasing security hardware to track unusual behavior on servers (Nozari & Sadeghi, 2021).

Social computing includes Social Network Analysis (SNA), Online Communities, Recommender systems, Reputation Systems, and Prediction Markets, and indexing web searches including ISI, IEEE, and Scopus. Considering the advantages of big data (metadata), it should be said that this topic provides new opportunities for researchers in knowledge processing tasks. However, these opportunities often bring challenges. To manage these challenges, it is necessary to know the computational complexity, information security, and computational methods for big data analysis. One of the challenges of big data analysis is related to the diversity of data. With the rapid growth of data sets, data mining tasks have grown significantly. In addition, data reduction, data selection, and feature selection are essential tasks, especially when working with large datasets. This issue reveals an unprecedented challenge for researchers. Because existing algorithms may not be responsive in real-time when working with these high-dimensional data (Nozari et al., 2021).

One of the issues related to Big Data is the exponential growth of raw data. Data centers and databases store huge amounts of data. This data is still growing rapidly. The exponential growth of data often makes it difficult to store it properly. The next challenge is choosing the right big data tool. There are various tools for analyzing and working with big data, but the wrong choice can waste effort, time, and money.

2-1 Application of big data in digital marketing

Big data is one of the practical topics in the field of digital marketing and has various uses. In order to better understand the benefits of using big data in the field of digital marketing, in this section of the article "What is Big Data", first the applications of big data in digital marketing are introduced and then each of them is described specifically (Nozari et al., 2021).

- *Understanding customers and audience segmentation*: big data allows marketers to collect, discover and analyze various aspects related to customer behavioral criteria, such as how they use products and services, as well as social and demographic factors. In this way, due to the information found, the personality of the customers and more precisely, their interests are determined, and after that, the strengthening and optimization of the marketing messages are facilitated.
- *Sentiment analysis*: Marketers can better understand how customers feel about their brand by analyzing social media posts, reviews, and search queries.
- *Predictive and Prescriptive Analysis*: With the cooperation of marketers with the "Supply Chain", it is possible to predict the demand for products by using big data, and in this way, produce more suitable products. and be presented.
- *Targeted Marketing*: Big data analysis is used for activities such as product recommendations, social media advertising, and email drip campaigns. So, with the help of big data, more suitable content is provided for customers.
- *Measurement of results*: with the help of big data, it is possible to measure the performance of campaigns in order to optimize the budget in real-time.
- *Understanding market trends*: By analyzing past data, you can benefit from market trends. This is done using predictive analytics with the goal of predicting demand and performing prescriptive analysis of items.
- *Competitive Analysis*: With the help of big data, you can get a good insight into your competitor's marketing campaigns and by using this information, examine their performance and their dos and don'ts.
- *Sales growth and potential profitability increase*: The above factors can all lead to an increase in sales as well as the profitability of different businesses. In fact, using big data for targeted marketing can help reduce advertising costs, shorten supply chains for on-time delivery, and run more successful marketing campaigns.

Along with the expansion and growth of data, cloud storage providers such as AWS, Microsoft Azure, and Google Cloud will play a prominent and important role in the field of big data storage. Due to this, it is possible to increase scalability and efficiency for companies. In addition, more people will be hired to manage and handle this data, and more job opportunities will be created for "data managers" to manage a company's database. On the other hand, the future of big data also has dark sides. Many tech companies are facing various difficulties due to data and privacy issues. Simply put, laws governing individuals' rights to their data make the process of data collection much more restrictive. Nevertheless, the numerous applications of big data in human life are undeniable, and acquiring skills related to it can lead to growth and prosperity in various fields, especially in the field of digital marketing (Nahr et al., 2021).

3.Cloud computing, dimensions, and features

Cloud Computing provides various services using the Internet. These services include tools and applications such as data storage, servers, databases, networking, and software. In fact, cloud-based

storage allows you to store files in a remote database instead of storing them on a dedicated hard drive or local storage device. With cloud computing, you can access data and software applications at any time and place just by using an Internet connection. Among the important advantages of cloud computing, which have made it a popular option among users and businesses, we can mention cost-effectiveness, increased productivity, speed and efficiency, performance, and security. The reason for naming cloud computing is that it enables access to information through the cloud or virtual space (Nozari et al., 2021).

Companies that provide cloud services allow users to store their files and applications on remote servers and then access their information using the Internet at their desired time and place (Ghahremani-Nahr et al., 2021). This means that the user does not need to be in a specific place for access and can easily control and manage his stored data remotely. Cloud computing does all the heavy work related to data processing and transfers all this work to very distant computers in virtual space; As a result, the Internet becomes a cloud space and you can access your data and files anywhere in the world with any device.

In summary, cloud computing consists of three basic parts:

- Cloud service providers store data and applications on physical machines; That is, places known as data centers.
- Users have access to these data and applications.
- The Internet connects companies, providers, and users quickly, even over long distances.

3-1 Different model of cloud computing

There are different types of clouds, each of which is different from the other. In general, cloud computing is divided into three categories: "public", "private" and "combined", each of which is described below.

- *public cloud*

Public clouds offer their services on servers and storage space on the Internet. These clouds are operated by third-party companies that manage and control all hardware and software and overall infrastructure. Users can access these services using accounts that are available to almost anyone.

- *Private clouds*

Private clouds are provided for specific customers (usually businesses or organizations). A company's data center may host a cloud computing service. Many private cloud computing services are provided on a private network. Companies to universities and organizations can host private clouds for their exclusive use. When they do, they own the underlying cloud infrastructure and host it in a remote location.

- *Combined clouds*

Hybrid clouds, as their name suggests, are a combination of public and private services. This type of model gives the user more flexibility and helps to optimize his infrastructure and security. In general, organizations use private clouds for critical functions and public clouds to accommodate increased computing demand. Data and programs are often switched between them automatically. This gives organizations more flexibility without requiring them to abandon existing infrastructure and security.

3-2 Advantages of cloud computing

Undoubtedly, every new technology has its own advantages and disadvantages. But experience has proven that the benefits of cloud computing are much more than the challenges or perhaps its few disadvantages. Of course, the role of service-receiving companies can also be effective in determining

these benefits. In the following, we state the most important advantages of cloud processing in simple language.

- *Cost savings*

for companies and especially start-up businesses; Cost control is very key to the survival of the organization. In the past, it was the big companies that were able to create, build and install software in physical form at a very high cost. Therefore, small and medium-sized companies that did not have this possibility, would automatically leave the field of competition. But cloud processing has made these companies access the software they need in a shared form easily and remotely and according to their needs and demands. Therefore, they pay according to their consumption and do not need to spend money on infrastructure.

- *Ease of access*

Cloud technology has made access very easy and has completely removed time and place restrictions. Users only need good internet bandwidth to use anywhere in the world. This advantage can be a great opportunity for telecommuting and remote management of activities.

- *Reducing the need for manpower*

In traditional systems, there was a need for specialized and skilled manpower to create and prepare the software infrastructure. But one of the most important advantages of cloud computing is that these actions are performed by the service company. Therefore, service companies can use their energy in more important activities.

- **Automatic Updates**

In traditional methods, a fee should always be spent on updating the software. But cloud software is automatically updated at no cost. For example, the update of the cloud accounting software of cloud solutions is done automatically.

- **High Security**

The infrastructure created by large companies is much more reliable than the infrastructure of other companies whose main expertise is not in this field. At the same time, the act of cloud backup and backup of data and information is done by them so that nothing special happens if the information is lost at one point. Therefore, the security of cloud computing is usually more than traditional methods.

- **Data sharing**

Employees can collaborate in real-time from different geographic locations and remotely, share different data and have access to the same program at the same time.

4.Big data and cloud computing together

When we combine big data and cloud computing, the possibilities are endless! If we only have big data alone, we will have huge, valuable, and high-potential data sets that are unused. Analyzing these data using normal home computers is impossible due to the large volume and the time it takes to process the data. However, cloud computing allows us to use advanced infrastructure and pay only for our time and power consumption. Cloud application development is also boosted by big data. Without big data, there

would be far fewer cloud-based applications because there would be no need for them. Remember that big data is often collected by cloud-based applications as well (Nozari et al., 2022).

In short, cloud computing services are mainly used due to the presence of big data. Thus, the only reason we collect big data is that we have services that are able to receive it and decode it even in seconds. Cloud computing and big data are inextricably linked as neither can exist without the other. Figure 2 shows the simultaneous combination of big data and cloud computing technologies (Ghahreman-Nahr et al., 2021).
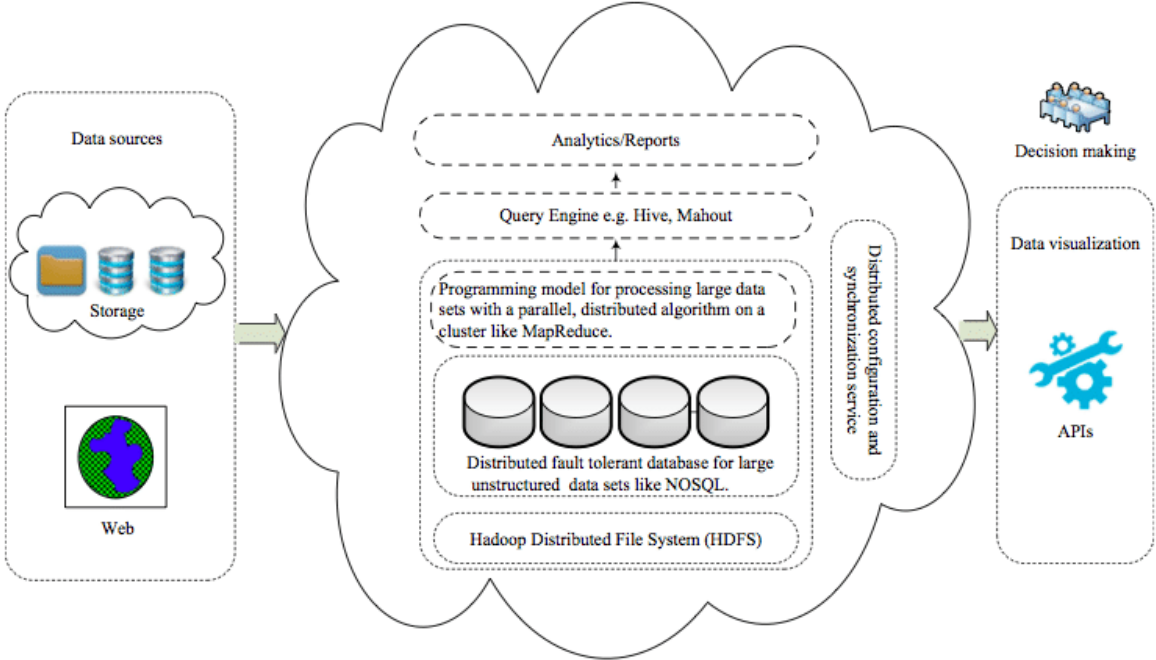

Figure 1: Use of Cloud Computing in Big Data (Khan et al., 2018)

5. Conclusion

In this paper, an overview of big data issues including the concepts of opportunities and challenges, techniques and tools, and the principles of big data design were discussed. There is no doubt that big data analysis is still in its early stages and needs to be developed. Yes. It is also possible to use cloud computing modeling in cloud services to store, process, and analyze data and classify data in big data in order to save on hardware and processing. The most worrying issue of the current era is privacy and information security. Since privacy is essential for individual and all kinds of organizational data, it has become a major challenge for big data. Preventing data leakage during processing and defending against external attacks requires a reliable data-centric security model. This technology should also take care of the security threats that may occur when storing such big data.

In the future, we will see huge advances in the field of big data because rapid advances in cloud computing technologies and data analysis will lead to an increase in the ability to store data and its accessibility. In addition, it is possible to conduct research in the field of future research using quantum computing for big data.

References

Aliahmadi, A., & Nozari, H. (2022, July). Evaluation of security metrics in AIoT and blockchain-based supply chain by Neutrosophic decision-making method. In Supply Chain Forum: An International Journal (pp. 1-12). Taylor & Francis.

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.

Aliahmadi, A., Nozari, H., & Ghahremani-Nahr, J. (2022). Big Data IoT-based Agile-Lean Logistic in Pharmaceutical Industries. International Journal of Innovation in Management, Economics and Social Sciences, 2(3), 70-81.

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.

Ghahremani-Nahr, J., Nozari, H., & Sadeghi, M. E. (2022). Investment modeling to study the performance of dynamic networks of insurance companies in Iran. Modern Research in Performance Evaluation, 1(2), 67-79.

Khan, S., Shakil, K. A., & Alam, M. (2018). Cloud-based big data analytics—a survey of current research and future directions. Big data analytics, 595-604.

Mathrani, S., & Lai, X. (2021). Big data analytic framework for organizational leverage. Applied Sciences, 11(5), 2340.

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.

Nozari, H., & Aliahmadi, A. (2022). Lean supply chain based on IoT and blockchain: Quantitative analysis of critical success factors (CSF). Journal of Industrial and Systems Engineering.

Nozari, H., & Nahr, J. G. (2022). The Impact of Blockchain Technology and The Internet of Things on the Agile and Sustainable Supply Chain. International Journal of Innovation in Engineering, 2(2), 33-41.

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.

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., Ghahremani-Nahr, J., Fallah, M., & Szmelter-Jarosz, A. (2022). Assessment of cyber risks in an IoT-based supply chain using a fuzzy decision-making method. International Journal of Innovation in Management, Economics and Social Sciences, 2(1).

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., Szmelter-Jarosz, A., & Ghahremani-Nahr, J. (2022). Analysis of the Challenges of Artificial Intelligence of Things (AIoT) for the Smart Supply Chain (Case Study: FMCG Industries). Sensors, 22(8), 2931.

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.

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.

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.