Issues in Information Systems
Volume 24, Issue 1, pp. 291-304, 2023

DOI: https://doi.org/10.48009/1_iis_2023_125

A systematic literature review: benefits and challenges of cloud-based big data analytics

Zaineb Naamane, *Middle Georgia State University, zaineb.naamane@mga.edu*

Abstract

Due to the rapid growth of the Internet and digital economy, the amount of data being stored and analyzed continues to rise exponentially. As a result, big data has emerged as a paradigm for improving and enabling research and decision-support applications. However, as this trend continues, organizations face challenges in storing, transporting, processing, sharing, and visualizing Big Data. The use of cloud computing has become an essential component of Big Data advancements by addressing the challenges associated with shared computing resources such as computing, storage, networking, and analytics software. This study conducts a systematic literature review of research studies published in the last five years on big data analytics in the cloud computing environment. First, this article highlights the relationship between the two technologies – big data and cloud computing – by looking at the latest developments in this area. Next, it examines the different implementations in the cloud that are available for storing, processing, and accessing big data. Finally, it identifies the advantages and challenges of using cloud computing to address Big Data in the digital age.

Keywords: cloud computing, cloud security, big data analytics, data storage, data processing, cloud architectures.

Introduction

Data dominates the world we live in today. Due to the tremendous technological advancements over the past few years, various sources produce a large amount of data daily. Since the introduction of mobile devices, inexpensive and numerous sensor devices on the Internet of Things (IoT), social media networks, etc., the amount of information collected have increased significantly (Berisha et al., 2022). It is estimated that about 44 zettabytes of data are generated daily, and approximately 1.7 megabytes of data are generated per person every second (Statista, 2021). Statista Research Department forecasts that worldwide data will increase exponentially between 2020 and 2025, shifting from 44 to 163 zettabytes (Statista, 2021). The size, variety, and speed at which this data is generated and collected make traditional tools unable to handle it.

The concept of big data pertains to the storage, processing, and analysis of large amounts of data. Alternatively, cloud computing focuses on providing the infrastructure to enable such processes efficiently and cost-effectively (Muniswamaiah et al., 2019). Due to the potential competitive advantages that can be gained from leveraging these two technologies, there is great interest in taking advantage of them. In recent years, cloud computing and big data analytics have emerged as among the most critical technologies that disrupted the mainstream of the IT industry (Hashem et al., 2015). These technologies will provide a powerful solution for businesses that will deliver powerful results and benefits (Jagani et al., 2021).

Even though cloud computing environments seem to provide an ideal solution for storing and processing big data workloads, working with big data in a cloud environment is challenging (Berisha et al., 2022).
As part of this research paper, we examine the two frontiers of Big Data and cloud computing. We analyze the benefits and challenges of utilizing cloud computing to tackle Big Data analytics in the digital age. This research is organized as follows. First, a literature review is presented, followed by a proposed methodology that includes a description of the procedure and data analysis. Second, the findings and the analysis of the result are summarized by themes. Third, the limitations of this research are given; Finally, conclusions and future work are discussed.

The main objective of this paper is to provide a review of the opportunities and challenges associated with big data analytics in cloud computing, which entails the efficient processing of large amounts of data. We also discuss how the consolidation of these two dominant technologies can enhance the process of big data analytics enabling businesses to improve decision-making processes. Finally, based on the literature survey results, we provide some best practices for successfully designing and deploying cloud-based big data analytics architectures.

This research aims to survey the literature on cloud-based big data analytics to understand the current state of the art in this area. In addition, we aim to identify the benefits and challenges of deploying big data analytics through cloud computing. The research will answer the following questions:

RQ1: *What are the benefits of big data analytics on cloud computing?*

RQ2: *What challenges involve implementing and deploying big data analytics through cloud computing?*

Literature review

Relationship between cloud computing and big data

The study of the relationship between cloud computing and big data has received considerable attention since digitalization's acceleration and rapid adoption (Ji et al., 2012; Dzulhikam et al., 2022). There is an integrated relationship between cloud computing and big data. This relationship is determined by the product, storage, and data processing as common factors. Big data represents the product, and the cloud represents the container in which the product is stored. The concept of big data relates to cloud computing's capabilities. However, cloud computing is more interested in how big data is gathered and the source from which it is collected. (Zanoon et al., 2017). Therefore, using cloud computing and big data together is mutually beneficial. Big data demands must be accommodated by something other than traditional storage.

Cloud computing provides the environment to store, handle and share data across multiple distributed storage sites (Pedro et al., 2016). The ability to store and process large amounts of data requires scalability, fault tolerance, and availability. By utilizing hardware virtualization, cloud computing can provide all these benefits (Dzulhikam et al., 2022). As a result, big data and cloud computing complement each other since the cloud offers greater accessibility, scalability, and fault tolerance for large data sets (Patel et al., 2021).
Big Data analytics represents the efficient processing of large amounts of data, whereas cloud computing refers to the virtualization of hardware resources.

Based on a future development trend, cloud computing will provide the lower layer of computing resources that allow fast and efficient processing of big data at the upper layer. As a result, big data will focus primarily on enhancing data analysis capabilities and improving the efficiency of interactive queries in real time (Tang, 2017).

Challenges of big data in the cloud

As noted above, there is a strong correlation between cloud computing and big data technologies. While these two technologies have been partnered, both still present specific challenges (Pedro et al., 2016). Although big data has the potential to solve many current problems related to massive amounts of data, it is a constantly evolving field that still deals with several issues. Due to the rapid growth of data, maintaining a balance between cost and efficiency is a growing concern (He et al., 2017); data governance is also a topic raised as a result of big data and the need to define retention and aging policies for data (Pearce, 2022). The use of cloud-based data has also sparked concerns regarding privacy and security (El-Seoud et al., 2017). Using outsourced databases in the cloud creates significant security concerns because query processing procedures may easily allow unauthorized access to sensitive data sets (Cuzzocrea, 2014).

Furthermore, big data usually involves structured and unstructured data from different sources and formats. Consequently, processing, transforming, and cleaning this heterogeneous data present a significant challenge for big data analytics solutions (Sandhu, 2022). Additionally, the visualization of such large dynamic data requires massive parallelization, representing a challenge for visualizing and summarizing data (Muniswamaiah et al.,2019). Moreover, disaster recovery, data uploading to the cloud, and Exaflop computing are also challenges being investigated at present (Pedro et al., 2016).

Benefits of big data in the cloud

There are many benefits associated with integrating cloud computing with big data. Multiple servers are required to process big data due to the size of the data and the high velocity and variability needed to process it. These servers operate in parallel to meet the high demands of big data. A cloud computing service provides multiple servers and automatically manages resource allocation. Thus, building big data on these cloud multi-servers and using their resource allocation capabilities is a great fit, resulting in more efficient big data analysis (Islam et al., 2019).

A cloud computing solution provides an on-demand pool of resources, making it possible to implement Big Data technology with a minimal financial commitment (Rashed, 2018). Moving to cloud computing may reduce the cost of managing and maintaining IT systems (Balachandran et al., 2017). To accommodate the massive amounts of data generated by big data, clusters of servers and volumes are necessary. By utilizing a cloud computing system instead of adding new servers and increasing storage for big data, greater flexibility and scalability can be achieved while eliminating the need to invest in additional big data servers and computers (El-Seoud et al., 2017).

Furthermore, cloud computing enables faster provisioning of big data servers since provisioning in the cloud is simple and cost-effective (Balachandran et al., 2017). Providing a scalable, convenient, on-demand, and shared computing environment with minimal management effort are some benefits of using cloud computing in conjunction with big data (Amitkumar et al., 2017). Further, using a cloud system as storage for big data would improve performance and reliability and unlock advanced analytics, as cloud systems are primarily based on remote multi-server architecture, which allows massive amounts of data to be handled simultaneously (El-Seoud et al., 2017).

Further, compared to traditional methods of storing and processing data, cloud computing offers fast, easy, and almost unlimited resources, allowing organizations to overcome the volume challenges associated with big data (Rashed,2018). Additionally, it facilitates multi-tenancy and improves the robustness and automation of the environment (Goyal et al., 2019). A further advantage of integrating the two is the

enhanced control, monitoring, and reporting of big data resources (Meng et al., 2013). Further, this integration makes it possible to reduce complexity and increase productivity (El-Seoud et al., 2017). As a result of these advantages, cloud-based approaches have become increasingly popular for deploying big data analytic solutions.

Cloud-based data analytics implementations

Data analytics service models

A comprehensive data analytics service, scalable storage capabilities, and reliable programming tools are essential to identifying valuable information from vast digital datasets. Several fields, including medical, retail, sales, the internet of things, biosciences, and high energy physics, can benefit from advanced data mining techniques and associated tools, which allow them to extract information from large, complex data sets (Talia, 2013). Such information can then be used as a basis for informed business decisions. In addition, a combination of cloud-based computing systems and big data analytics techniques can provide new insights into the problem within a short period through the application of cloud-based computing systems (Sandhu, 2022).

The cloud is becoming an increasingly popular place for developers and researchers to implement big data analytics solutions, using the software as a service (SaaS), Platform as a service (PaaS), and infrastructure as a service (IaaS) model to achieve this (Talia, 2013). As opposed to the on-premises model, the SaaS model provides end users with a comprehensive suite of big data analytics applications that can be accessed remotely on any device, so end users can benefit from the cloud's scalability in storage and processing power to analyze large or complex datasets (Chauhan & Kumar, 2013).

As part of the Platform as a service (PaaS) model, developers of data analytics programs can create scalable analytics services and applications by leveraging developers' programming suites and environments (Purcell, 2014). In addition, infrastructure as a model (IaaS) enables researchers to build a virtualized set of hardware and software resources that can be deployed to run analysis frameworks or applications to analyze data as a result of using virtualized hardware and software resources (Arostegi, 2018; Zanoon et al., 2017).

Orchestrating big data analysis workflows in the cloud

The concept of workflow refers to the systematic breakdown of tasks that must be completed to solve a problem. Streamlining and improving the performance of underlying processes tailored to achieve a desired outcome is ideal for this concept in scientific and business applications (Khan et al., 2017).

As big data analytical problems become increasingly complex, scientific workflows are increasingly used to perform tasks within specific domains. For example, accessing, processing, and visualizing data are all part of data analytics. Consequently, workflow tasks must consider these system components effectively and efficiently. (Barika et al, 2019).

Integrating scientific workflows with cloud-based big data analytics is possible with Scientific Workflow Management Systems. Several advantages are associated with this approach, including better scalability, greater flexibility, and easier deployment. Many frameworks, architectures, and scheduling algorithms have been proposed to take advantage of these benefits applications (Khan et al., 2017; Li et al., 2015; Zhao et al., 2014; Talia, 2013; Vöckler et al., 2011; Lin et al., 2009). Many systems have been designed and implemented over the years; however, it is essential to note that security issues still need to be addressed.

Future research should also focus on optimizing scheduling algorithms and multi-site cloud execution (Khan et al., 2017).

Methodology

This study conducted a systematic literature review to study the current state-of-the-art cloud-based big data analytics. The following four steps are followed (Chen et al., 2022; Kitchenham, 2004; Pereira & Serrano, 2020; Webster & Watson, 2002): 1)- Research outline: Identifying the objectives and defining criteria of the study, 2)- Data collection: applying search criteria and conducting filtration, 4)- Data analysis, 3)- Reporting: reporting the findings, drawing conclusions, providing recommendations and best practices for big data analytics implementations in the cloud.

Research outline

This research explored the benefits and challenges associated with deploying big data analytics through cloud computing. Two distinct keywords are used to explore multiple databases: cloud computing and big data analytics.

Data collection

Data collection was conducted to collect research papers and concept papers on big data analytics in cloud computing. In the discovery phase, as shown in the figure below, several databases were selected to search for articles related to the topic, such as ACM Digital Library, IEEE Explore, and Google Scholar.

The literature review covered articles published between the years 2012 and October 2022. The following keywords were used to identify these articles: "cloud computing" AND ("big data" OR "big data analytics"). The selection of the articles to be included in this literature followed the workflow below to determine the most relevant research papers that are related to the research subjects.

During the filtration and evaluation phase, an exclusion criterion has been applied that includes the following:

a)- The studies cannot be accessed in full text.
b)- non-English studies.
c)- studies other than Peer-reviewed articles and conference proceedings.
d)- Studies that are not relevant to the research question.

Once evaluated, the abstract of the research papers is then examined to determine qualified studies for this research. In the preview and selection phase, full-text articles are briefly accessed and previewed to ensure usability for the current study. These articles are then thoroughly reviewed during the data analysis phase following the preview, question, read, summarize (PQRS) system.

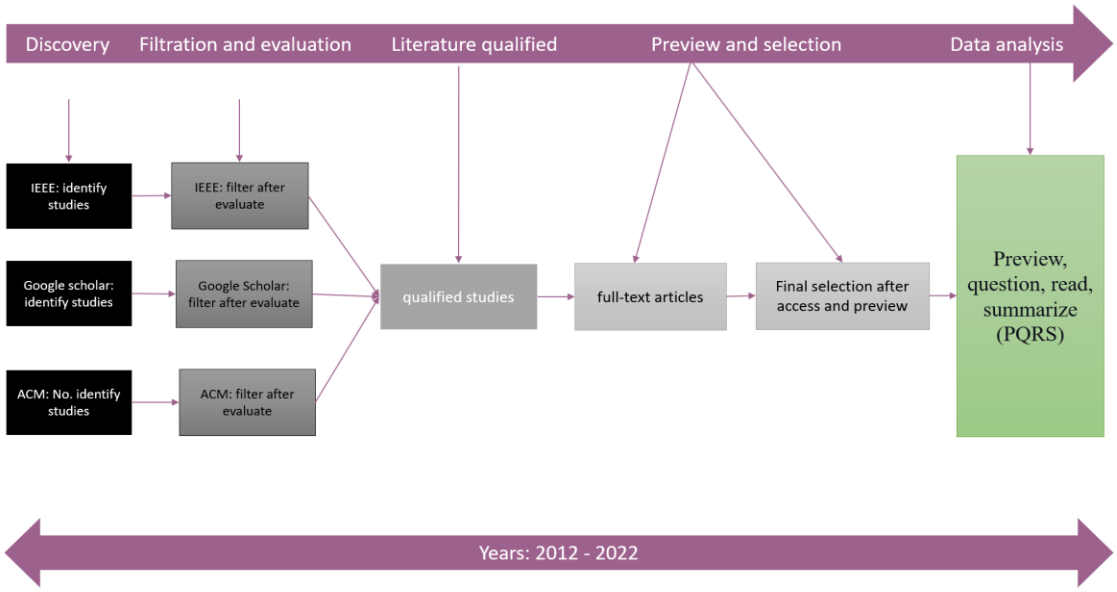

Figure 1: Workflow of systematic literature review of cloud-based big data analytics

Data analysis

As part of the filtration process, the keyword "cloud computing" in the research title is used for an initial search, followed by the keyword "big data" or "big data analytics" for a systematic search. After each iteration, an initial review of the article is conducted to determine whether it is worthy of further reading or inclusion.

A simple methodology proposed by (Cohen, 1990) is followed to conduct a systematic and critical review of the article's contents. This simple method is referred to as the preview, question, read, summarize (PQRS) system that allows researchers to stay organized and focused and facilitates easy identification and retrieval of information, particularly if many publications are being reviewed at once.

During the preview phase, 32 articles deemed relevant to the review have been accumulated. During the question stage, questions are asked about each publication. An indexing and summary system assisted the process (Burns & Grove, 2007), which includes the article's title, author, purpose, methodology used in a research study, findings and outcomes, and relevance to research questions (Timmins, 2005).

The source and complete reference are also included in this system for proper record-keeping and organization. During the reading phase, selected articles are read, and research questions are carefully examined. Finally, in the summary stage, a summary of each article and key thoughts, comments, strengths, and weaknesses of the publication are written (Cronin & Coughlan, 2008). These provided the basis for reporting the review and findings. The findings are summarized into tables describing the benefits and challenges of big data analytics in cloud computing.

Results

The results of this study indicated that Cloud computing and Big Data play an increasingly significant role in a digital society. When combined, these two technologies can affect meaningful change in how organizational objectives are achieved. To answer RQ1, substantial benefits of big data in the cloud were

identified by critically analyzing the identified research papers. Table 1 includes the advantages of cloud-based big data analytics:

Scalability: Managing big data requires clusters of servers and volumes. The cloud serves as an infrastructure for all these systems, allowing greater flexibility and scalability while removing the need for significant investments in big data servers and machines. In addition, cloud computing makes provisioning servers in the cloud significantly easier and faster. As a result, the cloud environment used can be scaled depending on the big data processing.

Agility: Big data projects have different characteristics. Because of this, data analytics projects have varying infrastructure requirements. Cloud technology allows users to use as many resources as they need to accomplish a particular task and then release those resources when the job has been completed.

Cost reduction: Business data centers are costly in terms of capital expenditures. Besides hardware, businesses must also purchase facilities and power and maintain equipment over the long term. However, those costs are absorbed by the cloud into a flexible rental model based on pay-per-use, allowing resources and services to be accessed as needed.

Reduced complexity: Several components and integrations are required to implement a big data solution. Cloud computing can allow Big Data analysis teams to automate several of these components, reducing the process's complexity and enhancing the team's productivity.

Resilience: A big data project's value is derived from its data. As a result, the main advantage of cloud resilience is that it provides reliable data storage. Cloud services are designed to replicate data to ensure the high availability of data across all resources.

Table 1: Advantages of cloud-based big data analytics

Advantages	Reference
Scalability	Li, 2022; P N, 2022; Jagani et al., 2021; Alam et al., 2020; Mukhdoomi et al., 2020; Talia, 2019; Yeotikar, 2019; Rashed, 2018; bt Yusof Ali et al., 2018; Rashed et al., 2018; Dasoriya et. el, 2018; Sohail et al., 2017; El-Seoud et al., 2017; Nachiappan et al., 2017; Neves et.al, 2016; Yang et al., 2016; Hashem et al., 2015; Litchfield et al., 2014; Talia, 2013; Ji et al., 2012
Agility	Jagani et al., 2021; Khan et. al ,2021; Alam et al., 2020; Islam et al., 2019; Rashed et al., 2018; Sohail et al., 2017; Khan et al., 2017; El-Seoud et al., 2017; Yang et al., 2016; Hashem et al., 2015; Litchfield et al., 2014
Cost reduction	Jagani et al., 2021; Mukhdoomi et al., 2020; Islam et al., 2019; Yeotikar, 2019; Rashed et al., 2018; Zala et.al, 2018; bt Yusof Ali et al., 2018; Sohail et al., 2017; Khan et al., 2017; El-Seoud et al., 2017; Mathur et .al, 2017; Nachiappan et al., 2017; Hashem et al., 2015
Reduced complexity	Jagani et al., 2021; Mukhdoomi et al., 2020; Islam et al., 2019; Yeotikar, 2019; Zala et.al, 2018; Rashed et al., 2018; Sohail et al., 2017; Nachiappan et al., 2017; Khan et al., 2017; El-Seoud et al., 2017; Yang et al., 2016; Hashem et al., 2015
Resilience	Mukhdoomi et al., 2020; Islam et al., 2019; Yeotikar, 2019; Dasoriya et al., 2018; Tsagkaropoulos et al., 2018; Rashed et al., 2018; Mathur et al., 2017; El-Seoud et al., 2017; Khan et al., 2017; Neves et al., 2016; Venkatesh et al., 2015; Litchfield et al., 2014

As an emerging technology, cloud computing and big data analytics are constantly developing and face challenges and obstacles. To answer RQ2, key challenges of big data in the cloud were identified by

critically analyzing the identified research papers. Table 2 includes the challenges of cloud-based big data analytics:

Security and privacy: out of 32 research articles, 28 discussed data security and privacy issues related to deploying big data analytics in the cloud. There can be a wide variety of data types involved with big data projects, from proprietary data to personally identifiable information (PII) subject to data protection regulations and other industry, governmental, or social regulations. Additionally, data security becomes particularly critical in the cloud computing environment as the data is in different places all around the globe. Concerns about cloud technology are primarily related to the security of data and the protection of privacy. Even though many techniques in cloud computing have been explored in academic circles and industry, data security and privacy protection are becoming increasingly crucial for developing cloud computing technologies in government agencies, industries, and businesses in the future.

Network dependence: An organization's ability to utilize the cloud highly depends on the ability to connect across the local area network, the internet, and the cloud provider's network. The network path can be outage-prone, increasing latency and rendering the cloud inaccessible. Even though an outage might not affect a big data project in the same manner as a mission-critical workload, it is still essential to consider the impact of outages when using the cloud for big data applications.

Heterogeneity: Variety, one of the significant aspects of big data characterization, is the result of the growth of virtually unlimited different sources of data. This growth leads to the heterogeneous nature of big data. Data from multiple sources generally have different types and representation forms and are significantly interconnected; they have incompatible formats and are inconsistently represented (Sohail et al., 2017). In a cloud environment, users can store data in a structured, semi-structured, or unstructured format. Structured data formats are appropriate for today's database systems, whereas semi-structured data formats are appropriate only to some extent. Unstructured data are inappropriate (Che et al., 2013) because they have a complex structure that is difficult to represent in rows and columns. The challenge is handling multiple data sources and types (Hashem et al., 2015).

Visualization: The purpose of data visualization is to arrange large amounts of complex data clearly and understandably by making it visual to the end user. It is easy to represent structured data graphically. However, creating real-time visualizations of unstructured data can be challenging (Sandhu, 2022). Moreover, massive parallelization is required to visualize and summarize large volumes of dynamic data, which represents a challenge for data visualization and summarization in the cloud (Muniswamaiah et al., 2019).

Data governance: An organization is responsible for creating and enforcing the rules of law and transparency related to data and holding people and information systems accountable to achieve their business objectives. However, several policies, principles, and frameworks can create a significant number of challenges when it comes to striking a delicate balance between risk and value as data size increases and in the context of delivering better and faster data management technology.

Table 2: Challenges of cloud-based big data analytics

Challenges	Reference
Security and privacy	Sandhu, 2022; Khan et al., 2021; Su et al., 2021; Jagani et al., 2021; Alam et al., 2020; Mukhdoomi et al., 2020; Yeotikar, 2019; Islam et al., 2019; Tsagkaropoulos et al., 2018; Rashed et al., 2018; bt Yusof et al., 2018; Maohong et al., 2018; Dasoriya et al., 2018; Kibona et al., 2018; Zala et al., 2018; Chaoui et al., 2017; Mathur et al., 2017; Khan et al., 2017; Sohail et al., 2017; El-Seoud et al., 2017; Khan et al., 2017; Neves et al., 2016; Hashem et al., 2015; Venkatesh et al., 2015; Litchfield et al., 2014; Talia, 2013
Network dependence	Sandhu, 2022; Jagani et al., 2021; Khan et al., 2021; Su et al., 2021; Yeotikar, 2019; Tsagkaropoulos et al., 2018; bt Yusof et al., 2018; Nachiappan et al., 2017; Khan et al., 2017; Sohail et al., 2017; Mathur et al., 2017; Hashem et al., 2015; Litchfield et al., 2014
Heterogeneity	Sandhu, 2022; Jagani et al., 2021; Islam et al., 2019; Rashed et al., 2018; Mathur et al., 2017; Sohail et al., 2017; El-Seoud et al., 2017; Neves et al, 2016; Venkatesh et al., 2015; Hashem et al., 2015; Litchfield et al., 2014; Che et al., 2013
Visualization	Sandhu et al., 2022; Jagani et al., 2021; Islam et al., 2019; bt Yusof et al., 2018; Rashed et al., 2018; Khan et al., 2017
Data governance	Su et al., 2021; Yeotikar, 2019; Rashed et al., 2018; Sohail et al., 2017; El-Seoud et al., 2017; Mathur et al., 2017; Neves et al., 2016; Venkatesh et al., 2015; Hashem et al., 2015

Conclusion

Using data analytics has long been a way for businesses to maximize profits and guide strategic decision-making. However, cloud computing and big data technologies are becoming critical technologies that will soon shape the business world. In addressing the most significant challenges associated with big data, cloud computing is poised to offer a cost-effective, high-quality method of transforming big data into valuable information and to enable companies to increase profitability (Mathur et .al, 2017). Undoubtedly, cloud computing has become an essential service that many modern organizations rely on today, not only in the technology industry – it has spread to every sector. By using big data technologies and cloud computing, businesses can make proactive, knowledge-driven decisions that will allow them to anticipate future trends and behavioral trends, enabling them to make informed decisions at the right time. In addition, companies can now store their data remotely and gain access to their data and services at any time and location.

Furthermore, cloud-based data analytics allows companies to leverage an easily deployable, time- and cost-effective infrastructure. In an environment where data are abundant, simplification is essential for progress to be made and to extract value from large amounts of data.

This study aimed to review other literature on cloud-based data analytics implementations and identify challenges and advantages related to integrating cloud computing and big data. The systematic literature review results showed that cloud-based data analytics offer significant advantages in protecting business competitiveness, such as cost savings, scalability, agility, reduced complexity, and resilience.

Although there are many advantages to using cloud computing for deploying big data analytics (Jagani et al., 2021; Alam et al., 2020; Mukhdoomi et al., 2020; Talia, 2019; Yeotikar, 2019) there are also some challenges and disadvantages, particularly regarding privacy and security, data governance, dependency on the network for resource availability, data heterogeneity, and visualization (Sandhu, 2022; Khan et al., 2021; Su et al., 2021; Jagani et al., 2021; Alam et al., 2020; Mukhdoomi et al., 2020; Yeotikar, 2019; Islam et al., 2019; Tsagkaropoulos et al., 2018; Rashed et al., 2018). The success of any big data analytics

initiative based on cloud computing depends heavily on the organization's understanding of what is involved in the project before it is initiated. Although cloud analytics offers several advantages, it is essential that proper planning is undertaken to ensure that all analytics process phases are considered (Balachandran, 2017).

Cloud computing and big data analytics implementation research are gaining momentum. However, despite the amount of literature available in this field, the study's primary objective is not building new theories that address big data challenges, such as security, privacy, and data governance, but rather the practical implementation of cloud-based analytics solutions (Berisha et al. , 2022; Alam et al., 2020; Barika et al., 2019; Goyal et al., 2019; Arostegi et al., 2018; Dasoriya et al., 2018; He et al., 2017; Khan et al., 2017; Tang et al., 2017; Yang et al., 2016; Chauhan et al., 2013; Talia, 2013; Ji et al., 2012).

However, it is anticipated that new studies regarding the security and privacy issues associated with big data and cloud computing implementations will appear more frequently due to the increased interest in research integrating big data with cloud computing research and its applications and the emphasis on cybersecurity awareness. Therefore, researchers and cloud developers should continue paying attention to these issues. Furthermore, many companies have taken part in developing big data analytics applications in the cloud, which will contribute to a higher level of privacy and security concerns.

References

Alam, A., Ullah, I., & Lee, Y. (2020). Video Big Data Analytics in the Cloud: A Reference Architecture, Survey, Opportunities, and Open Research Issues. IEEE Access, 8, 152377-152422.

Amitkumar Manekar, S., & Pradeepini, G. (2017). Opportunity and Challenges for Migrating Big Data Analytics in Cloud. *IOP Conference Series: Materials Science and Engineering*, *225*, 012148. https://doi.org/10.1088/1757-899x/225/1/012148

Arostegi, M., Torre-Bastida, A., Bilbao, M. N., & Del Ser, J. (2018). A heuristic approach to the multicriteria design of IaaS cloud infrastructures for Big Data applications. Expert Systems, 35(5), e12259.

Balachandran, B. M., & Prasad, S. (2017). Challenges and Benefits of Deploying Big Data Analytics in the Cloud for Business Intelligence. *Procedia Computer Science*, *112*, 1112–1122. https://doi.org/10.1016/j.procs.2017.08.138

Barika, M., Garg, S., Zomaya, A. Y., Wang, L., Moorsel, A. V., & Ranjan, R. (2019). Orchestrating Big Data Analysis Workflows in the Cloud. ACM Computing Surveys, 52(5), 1–41. https://doi.org/10.1145/3332301

Berisha, B., Mëziu, E., & Shabani, I. (2022, August 6). Big data analytics in Cloud computing: an overview. *Journal of Cloud Computing*, *11*(1). https://doi.org/10.1186/s13677-022-00301-w.

bt Yusof Ali, H. B., bt Abdullah, L. M., Kartiwi, M., & Nordin, A. (2018). Risk Assessment for Big Data in Cloud. Proceedings of the 2018 Artificial Intelligence and Cloud Computing Conference on ZZZ - AICCC '18. https://doi.org/10.1145/3299819.3299841

Chauhan, R., & Kumar, A. (2013, November). Cloud computing for improved healthcare: Techniques, potential and challenges. In *2013 E-health and bioengineering conference (EHB)* (pp. 1-4). IEEE.

Che, D., Safran, M., Peng, Z. (2013). From big data to big data mining: challenges, issues, and opportunities, *Database Systems for Advanced Applications, Springer, Berlin Heidelberg*, pp. 1–15

Chen, W., He, W., Shen, J., Wang, X., & Chen, H. (2022). A Systematic Literature Review: Benefits and Challenges of Blockchain-Based Digital Transformation. *Issues in Information Systems*, *23*(3), 36–44. https://doi.org/10.48009/3_iis_2022_104.

Cohen G (1990) Memory. In: Roth I, ed. The Open University's Introduction to Psychology. Volume 2. Milton Keynes, Lawrence Eribaum: 570–620 Colling J (2003) Demystifying the clinical nursing research process: the literature review. Urol Nurs 23(4): 297–9

Cronin, P., Ryan, F., & Coughlan, M. (2008). Undertaking a literature review: a step-by-step approach. *British Journal of Nursing*, *17*(1), 38–43. https://doi.org/10.12968/bjon.2008.17.1.28059

Cuzzocrea, A. (2014). Privacy and Security of Big Data. Proceedings of the First International Workshop on Privacy and Secuirty of Big Data - PSBD '14. https://doi.org/10.1145/2663715.2669614

Dasoriya, R., & Samdani, K. (2018). Advancements in Data Analytics using Big Data and Cloud Computing. International Journal of Applied Information Systems, 12(10), 1–5. https://doi.org/10.5120/ijais2018451735

Dzulhikam, D., & Rana, M.E. (2022). A Critical Review of Cloud Computing Environment for Big Data Analytics. *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, 76-81.

El-Seoud, S. A., El-Sofany, H. F., Abdelfattah, M. A. F., & Mohamed, R. (2017). Big Data and Cloud Computing: Trends and Challenges. *International Journal of Interactive Mobile Technologies (IJIM)*, *11*(2), 34. https://doi.org/10.3991/ijim.v11i2.6561

Goyal, N., Pandey, A. K., Gupta, S. K., & Pandey, R. (2019, February 24). *Suppleness of Multi-Tenancy in Cloud Computing: Advantages, Privacy Issues and Risk Factors*. Papers.ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3358249

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015, January). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, *47*, 98–115. https://doi.org/10.1016/j.is.2014.07.006.

He, Q., Zhu, X., Li, D., Wang, S., Shen, J., & Yang, Y. (2017, June 1). *Cost-Effective Big Data Mining in the Cloud: A Case Study with K-means*. IEEE Xplore. https://doi.org/10.1109/CLOUD.2017.124

Islam, M., & Reza, S. (2019). The Rise of Big Data and Cloud Computing. *Internet of Things and Cloud Computing*, *7*(2), 45. https://doi.org/10.11648/j.iotcc.20190702.12

Jagani, N., Jagani, P., & Shah, S. (2021, March 1). Big Data In Cloud Computing: A Literature Reviw. *International Journal of Engineering Applied Sciences and Technology*, *5*(11). https://doi.org/10.33564/ijeast.2021.v05i11.029.

Ji, C., Li, Y., Qiu, W., Awada, U., & Li, K. (2012). Big Data Processing in Cloud Computing Environments. *2012 12th International Symposium on Pervasive Systems, Algorithms and Networks*, 17-23.

Khan, S., Shakil, K.A., & Alam, M. (2017). Workflow-Based Big Data Analytics in The Cloud Environment Present Research Status and Future Prospects. ArXiv, abs/1711.02087.

Kitchenham, B. (2004). Procedures for performing systematic reviews. Keele, UK, Keele University, 33(2004), 1-26.

Li, B. 2022. Research Review of Cloud Computing Technology Based on Big Data. In 2022 3rd Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC'22), April 14–16, 2022, Dalian, China. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3544109.3544145

Litchfield, A.T., & Althouse, J. (2014). A Systematic Review of Cloud Computing, Big Data and Databases on the Cloud. Americas Conference on Information Systems.

Li, X., Song, J. and Huang, B. (2015). Scientific workflow management system architecture and its scheduling based on cloud service platform for manufacturing big data analytics. Int J Adv Manuf Technol DOI 10.1007/s00170-015-7804-9

Lin C., Lu S., Fei X., Chebotko A., Pai D., Lai Z., Hua J. (2009). Reference architecture for scientific workflow management systems and the VIEW SOA solution. IEEE Trans Serv Comput 2(1): 79–92

Meng, S., & Liu, L. (2013). Enhanced Monitoring-as-a-Service for Effective Cloud Management. *IEEE Transactions on Computers*, *62*(9), 1705–1720. https://doi.org/10.1109/TC.2012.165

Mukhdoomi, M. A., Oberoi, A., & Gupta, A. (2020). Cloud and Big Data Electronic Age: A Review. International Journal of Computer Applications, 175(37), 31–42. https://doi.org/10.5120/ijca2020920941

Mukhdoomi, M. A., Oberoi, A., & Gupta, A. (2020). Coming Together of Big Data and Cloud Computing: A Review. International Journal of Scientific Research in Computer Science, Engineering, and Information Technology, 118–137. https://doi.org/10.32628/cseit206613

Muniswamaiah, M., Agerwala, T., & Tappert, C. (2019, August 31). Big Data in Cloud Computing Review and Opportunities. *International Journal of Computer Science and Information Technology*, *11*(4), 43–57. https://doi.org/10.5121/ijcsit.2019.11404.

Nachiappan, R., Javadi, B., Calheiros, R. N., & Matawie, K. M. (2017). Cloud storage reliability for Big Data applications: A state of the art survey. Journal of Network and Computer Applications, 97, 35–47. https://doi.org/10.1016/j.jnca.2017.08.011

Neves, P. C., Schmerl, B., Cámara, J., & Bernardino, J. (2016). Big Data in Cloud Computing: Features and Issues. *Proceedings of the International Conference on Internet of Things and Big Data*. https://doi.org/10.5220/0005846303070314.

P N, S. (2022). Impact and Implications of Big Data Analytics in Cloud Computing Platforms. *International Journal for Research in Applied Science and Engineering Technology*, *10*(5), 4661–4666. https://doi.org/10.22214/ijraset.2022.43407

Patel, P., & Patel, R. (2021). The Correlation of Cloud Computing and Big data: Review & Benefits.

Pearce, G. (2022, May 25). Real-World Data Resilience Demands an Integrated Approach to AI, Data Governance and the Cloud. ISACA. https://www.isaca.org/resources/isaca-journal/issues/2022/volume-3/real-world-data-resilience-demands-an-integrated-approach-to-ai-data-governance-and-the-cloud

Pereira, R., & Serrano, J. (2020). A review of methods used on IT maturity models development: A systematic literature review and a critical analysis. Journal of information technology, 35(2), 161-178.

Purcell, B. M. (2014). Big data using cloud computing. Journal of Technology Research, 5, 1.

Rashed, A. H., Karakaya, Z., & Yazici, A. (2018). Big data on cloud for government agencies. Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age. https://doi.org/10.1145/3209281.3209360

Sandhu, A. K. (2022). Big data with cloud computing: Discussions and challenges. *Big Data Mining and Analytics*, *5*(1), 32–40. https://doi.org/10.26599/bdma.2021.9020016

Sohail, H., Zameer, Z., Ahmed, H. F., Iqbal, U., & Ali Shah, P. A. (2017). Challenges and Opportunities in Big Data and Cloud Computing. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 175–181. https://doi.org/10.1007/978-3-319-51207-5_17

Statista. (2021, June 7). *Data Created Worldwide 2010-2025 | Statista*. Statista; Statista. https://www.statista.com/statistics/871513/worldwide-data-created/.

Talia, D. (2019). A view of programming scalable data analysis: from clouds to exascale. Journal of Cloud Computing, 8(1). https://doi.org/10.1186/s13677-019-0127-x

Talia, D. (2013). Clouds for Scalable Big Data Analytics. Computer, 46(5), 98–101. https://doi.org/10.1109/mc.2013.162

Tang, Z. (2017). On Study of Application of Big Data and Cloud Computing Technology in Smart Campus. IOP Conference Series: Earth and Environmental Science, 100, 012026. https://doi.org/10.1088/1755-1315/100/1/012026

Timmins F, McCabe C (2005) How to conduct an effective literature search. Nursing Standard. 20, 11, 41-47. Date of acceptance: May 10 2005.

Vöckler, J. S., Juve, G., Deelman, E., Rynge, M. and Berriman, G. B. (2011). Experiences Using Cloud Computing for A Scientific Workflow Application, Invited Paper, ACM Workshop on Scientific Cloud Computing (ScienceCloud). pp. 15-24

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. MIS quarterly, xiii-xxiii.

Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2016, November 3). Big Data and cloud computing: innovation opportunities and challenges. *International Journal of Digital Earth*, *10*(1), 13–53. https://doi.org/10.1080/17538947.2016.1239771.

Yeotikar, R. (2019, February). Cloud Computing With Big Data: Challenges & Issues [Review of Cloud Computing With Big Data: Challenges & Issues]. "RESEARCH JOURNEY" International E-Research Journal.

Zala, C., & Patel, P. (2018). Introduction to Big Data Analytics using Cloud Computing. International Journal of Research in Electronics and Computer Engineering, 6(2), 1415–1418.

Zanoon, N., Al-Haj, A.J., & Khwaldeh, S.M. (2017). Cloud Computing and Big Data is there a Relation between the Two: A Study.

Zhao, Y., Li, Y., Lu, S. and Lin, C. (2014). Devising a cloud scientific workflow platform for big data. IEEE World Congress on Services IEEE: 393-401