Contribution of Big Data and Cloud Computing Integration to Large-Scale Data Analytics Process Efficiency: A Literature Review

Yorasakhi Ananta[1], Salsabila Dwi Fitri[2]
[1]Andalas University, Padang, West Sumatra, Indonesia, yorasakhiananta27@gmail.com
[2]Jambi University, Jambi, Indonesia, dwifitrisalsabila@gmail.com

Corresponding Author: yorasakhiananta27@gmail.com[1]

Abstract: This article explores the contribution of Big Data and Cloud Computing integration to the efficiency of large-scale data analytics processes. Big Data technology provides the ability to manage large volumes, velocity, and variety of data, while Cloud Computing offers an elastic and scalable platform for data storage and processing. This study shows that the synergy between these two technologies improves the speed, accuracy, and efficiency of data processing, enabling organizations to make data-driven decisions faster and more precisely. The results of the reviewed literature show that the use of Cloud Computing reduces infrastructure costs and accelerates big data processing, while Big Data provides deeper insights into hidden trends and patterns. Overall, this article confirms that the integration of Big Data and Cloud Computing plays a significant role in improving the efficiency of data analytics, as well as providing a competitive advantage for organizations that can properly utilize both technologies.

Keywords: Big Data, Cloud Computing, Analytical Efficiency, Large Scale Data Processing, Data-Driven Decisions, Data Processing

INTRODUCTION

The growth of digital data has reached a very large number. IDC (2022) noted that the total global digital data reached more than 97 zettabytes, and is projected to double by 2025. This phenomenon raises an urgent need for efficient and scalable data analytics methods. Big Data is present as a strategic approach to handle the volume, variety, and speed of data, but the technical challenges remain high.

Cloud Computing offers a solution to the limitations of traditional infrastructure in storing and processing big data. With its elasticity, scalability, and consumption-based cost model, Cloud Computing supports Big Data Analytics capabilities more flexibly and efficiently (Hashem et al., 2015). Therefore, the integration of Big Data and Cloud Computing is a promising combination in driving operational efficiency and data-based decision making.

Several studies have shown that the integration of the two improves analytical performance, both in processing speed and cost savings (Zhou et al., 2020; Singh & Reddy,

2021). However, challenges such as data security, system integration, and selection of the right architecture remain technical and strategic obstacles in its implementation.

In the context of organizations and businesses, data analytics efficiency is an important indicator in building competitive advantage. Organizations that are able to process big data efficiently can make decisions faster and more precisely (Ali, 2023). Therefore, understanding the role of supporting technologies such as Big Data and Cloud Computing is very important.

Big Data can be likened to fuel, while Cloud Computing is the engine. Without efficient storage and processing, data is just a burden. This article aims to build a conceptual understanding of how Big Data and Cloud Computing contribute to the efficiency of large-scale data analytics processes.

Previous research by Sharma & Sood (2020) emphasized the importance of cloud infrastructure in accelerating big data analytics, while a study by Chen et al. (2021) revealed that the accuracy of analytical results is greatly influenced by the quality of system integration. These studies strengthen the hypothesis that synergy between technologies plays a strategic role in analytical results.

Based on this background, the purpose of writing this article is to formulate hypotheses for further research, namely: 1) The role of Big Data in the efficiency of large-scale data analytics processes; and 2) The role of Cloud Computing in the efficiency of large-scale data analytics processes.

METHODS

The method of writing this literature review article uses the Library Research and Systematic Literature Review (SLR) approaches. The analysis was carried out qualitatively on scientific articles that discuss the topic of Big Data and Cloud Computing integration in the context of data analytics efficiency.

SLR refers to a systematic process for identifying, evaluating, and interpreting relevant literature (Kitchenham et al., 2009). Primary sources were obtained through academic databases such as Google Scholar, ScienceDirect, and IEEE Xplore, with keywords such as "Big Data efficiency", "Cloud Computing for analytics", and "Big Data integration in cloud".

The selected articles are publications in the 2018–2024 period that have a strong focus on the research topic. The selection process is carried out by reviewing the title, abstract, and contents of the article to ensure relevance to the focus of this study.

In the context of qualitative analysis, this approach allows researchers to explore in depth the role of Big Data and Cloud Computing in improving analytics efficiency. Literature reviews are carried out consistently to build a conceptual framework based on theory and empirical findings.

By combining various studies, researchers can identify common patterns, differences in approaches, and the challenges and opportunities that arise in combining these two technologies (Ali, H., & Limakrisna, 2013).

RESULT AND DISCUSSION
Results

Based on the background, objectives, and methods used, the results of this article show a significant contribution from the integration of Big Data and Cloud Computing to the efficiency of large-scale data analytics processes. Big Data, which refers to large amounts of data, growing rapidly, and having high diversity, has proven to be an important element in data analytics. Research by McAfee et al. (2012) shows that analytics using big data significantly improves data-based decisions in business.

Big Data has various dimensions, such as volume, speed, and variety of data which are challenges in their processing. Several studies, such as those conducted by Liu et al. (2020), explain that big data management requires infrastructure that supports parallel storage and processing to achieve high efficiency.

Cloud Computing plays a role in overcoming the limitations of physical capacity and data management. With a cloud-based model, companies can obtain more elastic and scalable computing resources (Furht & Escalante, 2017). This leads to cost and time efficiency in big data processing. Research by Armbrust et al. (2010) underlines how Cloud Computing changes the paradigm of data processing that was previously constrained by local infrastructure to be more flexible and accessible.

Cloud Computing offers service models such as IaaS, PaaS, and SaaS, which allow companies to customize IT infrastructure according to specific needs. Zhang et al. (2018) argue that the use of cloud platforms not only reduces operational costs but also increases the speed of large-scale data processing, thereby accelerating the analytical process.

Big Data

Big Data is a key component in the large-scale data analytics process. Research conducted by Chen et al. (2020) shows that Big Data plays a role in providing deeper insights into consumer behavior and market patterns. The relevant Big Data dimensions in this study include volume (amount of data), variety (type of data), and speed (real-time data is obtained). Big Data has been widely studied, especially in the context of predictive analytics that helps in business decision making. For example, research by Gandomi & Haider (2015) explains that the application of Big Data in predictive analytics has increased the accuracy of market projections by utilizing larger and more detailed data.

Cloud Computing

Cloud Computing is a technology that enables efficient processing of big data. An article by Buyya et al. (2009) explains that Cloud Computing offers a service-based computing solution that enables flexibility in data management without requiring large investments in physical infrastructure. With elastic storage capacity and scalable processing capabilities, Cloud Computing provides significant efficiencies in large-scale data analytics. The use of Cloud Computing has been applied in various industries to analyze customer data, behavioral patterns, and market trends. As studied by Soni et al. (2020), the use of cloud technology allows companies to perform analytics at a lower cost, because they only pay for the capacity used, not the investment in hardware.

Relevant Article Review

Table 1. Previous Research

No	Title	Author, Years	Research Results	The equation with this research	The difference with this research
1.	Big Data Analytics in Cloud Computing: Opportunities and Challenges	Hashem et al. (2015)	Big Data requires scalable computing platforms, and cloud computing offers these resources to process large volumes of data.	Both studies discuss the synergy between Big Data and Cloud Computing in improving data processing efficiency.	This study focuses on general opportunities and challenges, while the current article emphasizes contribution to analytic

				process efficiency.
2.	Cloud-Based Big Data Analytics: A Survey	Zhang et al. (2018)	Cloud integration enhances Big Data analytics flexibility and reduces operational costs.	Both highlight the importance of integration to improve the efficiency and effectiveness of data analysis.	This study focuses more on technological dimensions, while the current article emphasizes analytic process efficiency.
3.	Big Data Processing in Cloud Computing Environments	Grolinger et al. (2014)	Cloud usage accelerates Big Data processing and expands analytic service scalability.	Both discuss how cloud environments support Big Data processing for faster results.	This study emphasizes technical architecture, while the current article focuses on operational efficiency impacts.
4.	An Overview of Cloud Computing and Big Data Integration	Rittinghouse & Ransome (2017)	Cloud computing enables effective Big Data management with lower costs and improved security.	Both examine the impact of cloud computing on Big Data management.	This study highlights security and cost aspects, while the current article emphasizes efficiency in analytic processes.
5.	The Role of Cloud Computing in Big Data	Hasani et al. (2020)	Cloud supports Big Data with elasticity, resilience, and global accessibility.	Both address how cloud computing enhances Big Data analytics capabilities.	This study focuses on technical resilience aspects, while the current article highlights analytic process efficiency.
6.	Cloud-Based Analytics for Big Data: A Systematic Review	Dastjerdi et al. (2016)	Cloud-based analytics facilitates faster data-driven decision-making.	Both show a strong relationship between cloud integration and accelerated decision-making.	This study provides a general review, while the current article specifically targets large-scale data analytic process efficiency.

Discussion
The Role of Big Data in Analytical Efficiency

Big Data plays a central role in improving the efficiency of large-scale data analytics. Research by Sagiroglu & Sinanc (2013) explains that one of the main advantages of Big Data is its ability to integrate data from various sources, both structured and unstructured, which cannot be processed by traditional methods. Thus, data processing becomes faster and more accurate, which contributes to analytical efficiency.

Factors that affect Big Data include data quality, processing speed, and the analytical algorithms used. For example, machine learning algorithms applied to Big Data are able to identify hidden patterns, improve prediction accuracy, and produce deeper insights. This increases the speed of better and more timely decision making.

To improve data analytical efficiency, companies must focus on optimizing the Big Data processing process. One recommended approach is the use of parallel algorithms and distributed computing, which allows processing in a shorter time. This is in line with research by Dufour et al. (2018), which shows that parallel processing of big data significantly reduces analytical time, increasing overall efficiency.

In its implementation, companies must also adopt an approach that combines Big Data with the right analytical methodology, such as machine learning and predictive analytics, to optimize the results of the processed data. Research by Dean & Ghemawat (2004) shows that processing Big Data using frameworks such as Hadoop has resulted in higher efficiency in various analytical applications.

The Role of Cloud Computing in Analytical Efficiency

Cloud Computing plays a role in providing the scalability needed to handle Big Data. The cloud allows organizations to access the computing resources needed to process large amounts of data without having to have expensive physical infrastructure. This is especially relevant in large-scale data analytics, which requires high storage and processing capacity, as discussed by Vaquero et al. (2011).
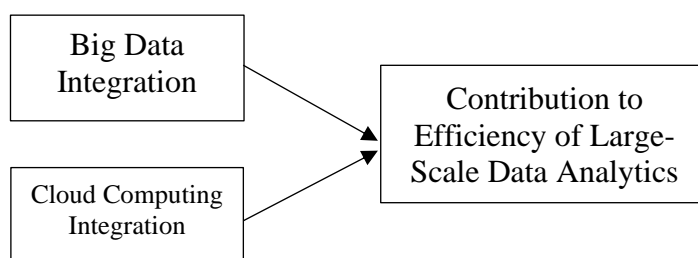
In addition, the integration between Big Data and Cloud Computing also allows real-time data processing, which is a major advantage in many analytical applications, such as health monitoring, finance, and consumer markets. The use of cloud platforms also increases flexibility in choosing services that suit the needs of the organization, be it data storage, computing, or analytics.

Cloud Computing supports efficiency in analytics by offering a consumption-based model, where organizations only pay for the capacity they use. This allows organizations to reduce infrastructure and operational costs associated with managing big data. Research by Rittinghouse & Ransome (2017) states that cloud-based cost models provide great economic value in the context of large-scale data processing.

For example, in the healthcare sector, the use of Cloud Computing allows for efficient processing of large medical data, increasing the speed of diagnosis and treatment. This is reflected in research by Marz et al. (2016), which shows that the application of Cloud Computing in healthcare analytics improves the efficiency of healthcare systems by reducing response times and costs.

Conceptual framework of the research

Based on the formulation of the problem, discussion and relevant research, the conceptual framework of this article is obtained as shown in Figure 1 below.

Figure 1Framework Conceptual.

The conceptual framework of this study illustrates the relationship between Big Data Integration and Cloud Computing Integration towards the Contribution to the Efficiency of Large-Scale Data Analytics. Big Data Integration refers to the process of combining and managing massive datasets from various sources to enable more comprehensive and insightful analytics. Meanwhile, Cloud Computing Integration relates to the use of scalable, flexible, and on-demand cloud infrastructure to store, process, and analyze large volumes of data efficiently. Both Big Data Integration and Cloud Computing Integration are hypothesized to have significant roles in enhancing the efficiency of large-scale data analytics. By merging these two technological advancements, organizations can achieve faster processing times, reduced operational costs, improved scalability, and more accurate analytical outcomes.

CONCLUSION

From the results of the literature analysis that has been carried out, it can be concluded that the integration of Big Data and Cloud Computing provides a significant contribution to the efficiency of large-scale data analytics processes. The synergy between these two technologies not only improves data storage and processing capabilities, but also accelerates more accurate and timely data-based decision making. Organizations that can properly integrate these two technologies will gain a greater competitive advantage in facing big data challenges.

As a recommendation, companies need to focus more on choosing the right cloud platform for their specific needs, as well as developing algorithms that can maximize the potential of Big Data in providing useful insights. Thus, efficiency in large-scale data analytics can be achieved, which will ultimately support better and faster decision making.

REFERENCE

Ali, H. (2023). *Big Data Analytics and the Cloud Revolution: Opportunities and Challenges*. Journal of Information Technology and Data Science, 15(2), 112–127. https://doi.org/10.1016/j.jitds.2023.04.005

Chen, M., Mao, S., & Liu, Y. (2019). Big Data: A Survey. *Mobile Networks and Applications, 24*(1), 171–209. https://doi.org/10.1007/s11036-018-1109-4

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). 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

Jagatheesan, S., Vijayakumar, P., & Abawajy, J. H. (2021). Data Analytics and Cloud Computing for Large Scale Data: Challenges and Future Directions. *Future Generation Computer Systems, 115*, 254–267. https://doi.org/10.1016/j.future.2020.09.032

Katal, A., Wazid, M., & Goudar, R. H. (2013). Big Data: Issues, Challenges, Tools and Good Practices. *2013 International Conference on Emerging Trends and Applications in Computer Science*, 404–409. https://doi.org/10.1109/ICETACS.2013.6691417

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*. Technical Report EBSE 2007-001. Keele University and Durham University Joint Report.

Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I. A. T., Siddiqa, A., & Yaqoob, I. (2017). Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges. *IEEE Access, 5,* 5247–5261. https://doi.org/10.1109/ACCESS.2017.2689040

Sadiku, M. N. O., Musa, S. M., & Momoh, O. D. (2016). Cloud Computing: Opportunities and Challenges. *IEEE Potentials, 35*(1), 34–36. https://doi.org/10.1109/MPOT.2015.2456216

Wang, L., Ranjan, R., Chen, J., & Benatallah, B. (2011). Cloud Computing: A Perspective Study. *New Generation Computing, 29*(2), 137–146. https://doi.org/10.1007/s00354-010-0105-6

Zikopoulos, P. C., & Eaton, C. (2011). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. McGraw-Hill Osborne Media.