International Journal of Cloud Computing and Database Management

Sachin Mallick
UG Scholar, Department of Computer Science and Engineering, KIIT (Deemed to be University), Bhubaneswar, Odisha, India

Shubham
UG Scholar, Department of Computer Science and Engineering, CHRIST (Deemed to be University), Kengeri Campus, Bengaluru, Karnataka, India

Rishabh Raj
UG Scholar, Department of Computer Science and Engineering, Sinhgad Institute of Technology & Science, Narhe Campus, Pune, Maharashtra, India

Corresponding Author:
Sachin Mallick
UG Scholar, Department of Computer Science and Engineering, KIIT (Deemed to be University), Bhubaneswar, Odisha, India

Cloud-based data analytics for scalable and efficient data processing: A review

Sachin Mallick, Shubham and Rishabh Raj

Abstract
Cloud-based analytics has emerged as a transformative solution for organisations seeking to harness large-scale data processing capabilities. The ability to store, process, and analyse vast datasets on demand has revolutionised how businesses, research institutions, and governments operate. This review examines the core components, benefits, challenges, and applications of cloud-based data analytics. We analyze the integration of cloud computing technologies with data processing frameworks, exploring the scalability, cost-effectiveness, and flexibility that cloud solutions provide. Additionally, we highlight the major concerns surrounding security, privacy, and cost management, while also reviewing best practices for optimizing cloud-based analytics. The paper concludes by discussing future directions and the potential for continuous innovation in cloud-based data analytics systems.

Keywords: Cloud computing, big data, data analytics, scalable data processing, real-time data analytics, and cloud platforms

Introduction

The studies searched using keywords like cloud computing, big data, data analytics, scalable data processing, real-time data analytics, and cloud platforms highlighted several key aspects of cloud-based data analytics. One of the most prominent benefits discussed across the studies was the scalability of cloud platforms. Cloud computing enables organizations to dynamically adjust resources to meet the fluctuating demands of data processing tasks, making it ideal for handling large and variable datasets in real-time. This scalability allows businesses to efficiently manage data without the need for substantial upfront infrastructure investments, which is particularly crucial in big data environments where data volumes can change rapidly.

Another important point emphasized in the studies was the cost efficiency and flexibility provided by cloud platforms. Through a pay-as-you-go pricing model, cloud services allow businesses to pay only for the resources they use, which helps optimize costs, especially for organizations with unpredictable workloads. The flexibility of cloud platforms also means businesses can provision resources on-demand, ensuring they are not over-provisioning or underutilizing their infrastructure, thus reducing unnecessary costs and improving overall efficiency.

In addition, real-time data processing capabilities were a significant focus of the studies. Many cloud platforms, with tools like Apache Kafka, AWS Kinesis, and Apache Flink, enable businesses to process and analyze data streams in real-time. This capability is particularly beneficial for industries such as finance, healthcare, and e-commerce, where quick decision-making based on live data is essential. Cloud platforms ensure low-latency services, enabling businesses to gain actionable insights immediately as data is generated.

The integration of artificial intelligence (AI) and machine learning (ML) into cloud-based data analytics was another key trend observed across the studies. Many organizations are leveraging cloud platforms to integrate AI and ML algorithms that automate data analysis and enhance predictive capabilities. These integrations not only streamline data processing but also enable organizations to detect patterns and derive insights more quickly and accurately, which significantly improves decision-making and operational efficiency.

However, data security and compliance challenges remain a concern for cloud-based data analytics. While cloud providers offer robust security measures such as encryption, access controls, and regular audits, the shared nature of cloud environments increases

the risk of unauthorized access and data breaches. Moreover, organizations must navigate complex data protection regulations such as GDPR and CCPA, especially when handling sensitive customer information. Ensuring compliance and securing data in the cloud are ongoing challenges that businesses must address to fully realize the benefits of cloud-based analytics.

These findings from the reviewed studies provide a comprehensive view of the advantages and challenges of cloud-based data analytics, shedding light on both its transformative potential and the obstacles that businesses need to address in order to maximize its effectiveness.

The digital age has brought with it the relentless growth of data across industries. From social media to the Internet of Things (IoT), the volume of data generated is increasing at an unprecedented rate. The need for sophisticated data analytics tools that can scale efficiently to handle such vast amounts of information is critical for organizations aiming to leverage data for strategic advantage. Traditional on-premise data solutions have struggled to keep pace with the challenges presented by this explosion of data, leading to the adoption of cloud computing as a solution. Cloud computing offers the scalability, flexibility, and cost-efficiency required to process big data and perform real-time analytics, making it an invaluable tool in modern data science and analytics.

Cloud-based data analytics refers to the use of cloud platforms to collect, store, process, and analyze large-scale datasets. The integration of cloud computing with data analytics has allowed businesses to handle massive datasets across various industries, enabling them to extract valuable insights for informed decision-making. This paper systematically reviews the benefits, challenges, and techniques related to cloud-based data analytics, with a specific focus on scalability and efficiency in data processing.

The integration of Big Data and Cloud Computing has emerged as a pivotal development in the IT field, with significant implications across various industries. According to Berisha, Mëziu, and Shabani (2022), the exponential growth in data from diverse sources has created a need for advanced processing tools capable of handling vast, rapidly changing, and heterogeneous datasets. Traditional data processing tools are insufficient to address the challenges posed by such large volumes of data. This has led to the rise of Big Data, which encompasses techniques for storing, processing, and analyzing massive datasets. Cloud computing, in turn, provides the necessary infrastructure for efficient Big Data processing, offering scalable and cost-effective solutions for data storage and computation. The combination of these two technologies has proven transformative in sectors such as healthcare, business, and education, where Big Data analytics is utilized for a variety of purposes, including reducing treatment costs, predicting disease outbreaks, and improving decision-making. A key shift in Big Data analytics has been the transition from the traditional ETL (Extract, Transform, Load) model to the ELT (Extract, Load, Transform) paradigm. This change, driven by the demands of Big Data, allows for more efficient data processing by ingesting data in its raw form and transforming it after loading. This shift is particularly suited for Cloud environments, where it enables faster data processing and greater flexibility in handling large-scale datasets [1].

The paper also highlights Google's BigQuery as a case study for demonstrating the capabilities of Cloud-based Big Data analytics. BigQuery, a fully-managed, serverless data warehouse, enables the scalable analysis of petabytes of data using ANSI SQL. The authors conducted experiments to assess performance metrics such as read, compute, and write times across different dataset sizes, showcasing the tool's efficiency and scalability in Cloud-based analytics.

Overall, the literature emphasizes the transformative role of Big Data analytics in Cloud computing, underscoring the practical applications, such as in healthcare, where it drives cost reductions and predictive insights. The shift to the ELT paradigm and the use of advanced tools like BigQuery represent crucial steps in the evolution of Big Data processing, making it more accessible, scalable, and efficient for various industries.

In their 2018 paper, dos Santos *et al.* (2018) introduce *Lemonade*, A cloud-based platform designed to facilitate secure, scalable data analysis and mining workflows for professionals outside the field of Computer Science. The platform simplifies the process by providing a visual interface for constructing data analysis flows, abstracting the complexities of cloud infrastructure, and incorporating authentication, authorization, and accounting (AAA) guarantees to ensure data security. An example of its application is demonstrated through a classification model for detecting fake news, showcasing Lemonade's ability to provide high-level abstractions for data source access and algorithm deployment [4].

Al-Jumaili *et al.* (2023) provide a comprehensive review of the integration of Big Data analytics and cloud computing frameworks in power management systems. The authors address the limitations of traditional parallel computing methods, including execution time, computational complexity, and inefficiencies in power system condition monitoring, particularly in managing consumer power consumption, weather data, and power generation. To overcome these challenges, they propose cloud-based methodologies that enhance data management, enabling real-time monitoring and improving performance across various application scenarios. The paper reviews cloud computing architectures capable of meeting multi-level real-time requirements and discusses emerging parallel programming models such as Hadoop, Spark, and Storm, emphasizing their advancements and constraints in Big Data analytics. Additionally, the authors introduce a new design concept for integrating cloud computing to solve data mining challenges within power management systems, underlining the importance of cloud infrastructure for managing real-time Big Data. Their work contributes significantly to the development of efficient, scalable frameworks for Big Data analytics in the power sector [2].

Understanding Cloud Computing and Big Data
Cloud computing is a model for enabling ubiquitous, on-demand access to shared resources, such as servers, storage, and databases. This service model allows users to avoid investing in physical hardware and provides the ability to scale resources according to demand. There are several key benefits associated with the cloud model that directly impact data analytics processes.

Big Data Challenges: Big data is often characterized by the three Vs — Volume, Velocity, and Variety — which

together make traditional data processing systems inadequate for handling such massive datasets. The rapid growth of data, including structured, semi-structured, and unstructured data, demands new technologies capable of managing this scale. Data sources such as social media, IoT devices, and transactional systems generate data at a rate that requires immediate processing and analysis. The ability to scale up processing power on demand, which cloud computing offers, makes it an attractive solution for big data analytics [8, 9].

The Need for Cloud Computing in Big Data Analytics
Big data analytics involves collecting, organizing, and processing vast quantities of data to identify patterns, correlations, and trends that may otherwise go unnoticed. Cloud computing platforms, with their distributed architecture, provide the infrastructure necessary for handling big data. Services like Amazon Web Services (AWS), Microsoft Azure, and Google Cloud provide the elastic computing resources needed to scale up or down depending on the workload. This flexibility is a key enabler of cloud-based big data analytics, allowing organizations to quickly adapt to changing business requirements [8, 9].

The main Techniques for Cloud-Based Data Analytics
Cloud-based platforms enable a variety of data processing and analytics techniques that help organizations handle big data efficiently. These techniques are crucial for managing the enormous volume and variety of data that organizations need to analyze.

Batch Processing
Batch processing refers to the processing of data in chunks or batches, where data is collected over time and then processed at scheduled intervals. This method has been widely used for tasks such as reporting, data warehousing, and offline analytics. Technologies such as Apache Hadoop use a batch processing model for storing and analyzing large datasets across distributed systems. The cloud's scalable infrastructure is ideal for this model, as it allows for efficient parallel processing of data across multiple cloud instances, thus reducing processing time and increasing throughput [10, 6].

Stream Processing
Stream processing, on the other hand, involves the continuous processing of real-time data streams. In industries such as financial services, telecommunications, and e-commerce, the ability to process real-time data is essential for immediate decision-making. Cloud platforms, with their low-latency capabilities, can provide the infrastructure necessary to support stream processing. Tools like Apache Kafka, Apache Flink, and AWS Kinesis are commonly used for real-time analytics in the cloud. These tools allow businesses to process data as it is generated, offering insights that enable rapid responses to events as they unfold [10, 6].

Machine Learning and AI Integration
Cloud platforms also integrate machine learning (ML) and artificial intelligence (AI) algorithms to automate data processing and decision-making. By leveraging cloud-based AI tools, organizations can enhance their data analytics capabilities, providing faster and more accurate insights.

Cloud computing services, such as AWS SageMaker, Google AI, and Azure Machine Learning, offer scalable resources for training machine learning models. AI can be used for various tasks such as anomaly detection, predictive analytics, and real-time decision-making, further enhancing the effectiveness of cloud-based data analytics [9, 11].

2. Methodology
The methodology for this systematic review paper on Cloud-Based Data Analytics for Scalable and Efficient Data Processing follows a structured approach to evaluate and synthesize relevant studies, frameworks, and technologies that pertain to the integration of cloud computing with data analytics for scalable and efficient data processing. The review was conducted in a manner that ensures all available and recent literature is included, and the findings are accurately represented and analyzed. The methodology is divided into several steps:

The first step in the methodology involved an extensive search of academic and industry-related sources to gather relevant studies on cloud-based data analytics, focusing on scalability and efficiency in data processing. The inclusion criteria were as follows:

Inclusion criteria
Time Frame & Types of documents:
Only studies published within the last six years (2018-2024) were considered to ensure the relevance and timeliness of the data and findings. Peer-reviewed journal articles, conference papers, and reputable industry reports were included. Studies that presented new algorithms, frameworks, or tools for cloud-based data processing were particularly sought after. Articles were gathered from reputable academic databases such as IEEE Xplore, SpringerLink, Scopus, and Google Scholar. Studies from these sources were prioritized due to their established credibility in the academic and technological fields.

Exclusion criteria
a) Studies that were not available in full text or did not meet the required time frame.
b) Articles that focused on traditional, non-cloud-based data processing techniques.
c) Studies with insufficient data or methodology descriptions that made it difficult to assess the validity of the findings.

After applying the inclusion and exclusion criteria, a total of 11 relevant articles were selected for the review, which were cited and analyzed throughout the paper.

2.1. Data Extraction and Analysis
Data extraction was conducted systematically by reviewing the selected articles. The following information was extracted from each study:

Research Focus: The core objectives and topics of each study, such as the exploration of specific cloud platforms (AWS, Azure, Google Cloud), technologies used (e.g., Apache Hadoop, Apache Spark), and the application of these technologies in big data analytics.
Techniques and Frameworks: Techniques for data partitioning, processing, and storage in the cloud, as well as

frameworks used to manage and process big data in a distributed environment.

- **Key Findings:** The main conclusions drawn from each study, including the effectiveness of cloud-based data analytics for scalability, real-time analytics, and cost efficiency.
- Challenges and Limitations: Common challenges highlighted across the studies, including issues related to data security, latency, privacy concerns, and compliance with regulatory standards.
- Technological Advancements: Innovations and emerging trends in cloud-based data analytics, such as serverless computing, hybrid cloud models, AI integration, and real-time data streaming.
- This data was then categorized and grouped according to the themes of benefits, challenges, techniques, and future trends. The data extraction helped to identify gaps in existing research and technological advancements that are relevant to scalable and efficient data processing in cloud environments.

2.2. Synthesis of Findings

1. Once the data was extracted and categorized, a thematic synthesis was conducted. The findings from each study were grouped under the following main themes:
2. Scalability and Elasticity: The ability of cloud platforms to scale computing resources based on demand, enabling efficient processing of big data and the handling of fluctuating workloads.
3. Cost Efficiency: The reduction in infrastructure costs due to cloud-based pay-per-use models, which offer organizations the flexibility to only pay for the resources they use. This also included the cost benefits of not needing to manage physical hardware.
4. Real-Time Data Analytics: The integration of real-time data processing tools and platforms, such as Apache Kafka, Apache Flink, and AWS Kinesis, to provide actionable insights and decision-making in real-time.
5. Cloud Security and Privacy: The challenges associated with storing sensitive data in the cloud and ensuring compliance with data protection regulations such as GDPR and CCPA.
6. Emerging Technologies: The role of AI, machine learning, and serverless computing in enhancing the scalability and efficiency of cloud-based data processing.

2.3 Critical Evaluation

A critical evaluation of the selected articles was performed to assess the strengths and limitations of the methodologies used in the studies. Several factors were considered during the evaluation:

- **Study Design:** Whether the studies used qualitative, quantitative, or mixed-method approaches. The evaluation of data quality was critical, particularly when assessing performance metrics, scalability results, and real-time data analytics capabilities.
- **Cloud Platforms and Tools:** The studies that focused on different cloud platforms, such as AWS, Google Cloud, and Microsoft Azure, were evaluated to understand the differences in their capabilities and effectiveness in big data analytics.

- **Challenges:** A detailed analysis was made of the common and unique challenges highlighted by the studies, including security concerns, latency issues, and compliance with regulatory frameworks.

This evaluation helped to synthesize the benefits and challenges of cloud-based data analytics, providing a balanced view of the current landscape and identifying areas where future research or improvements are needed.

3. Results and Discussion

The results from the systematic review of the selected studies on cloud-based data analytics for scalable and efficient data processing reveal a comprehensive understanding of the various techniques, benefits, and challenges associated with this technology. The findings across the studies offer valuable insights into the ongoing advancements in cloud computing and big data analytics, as well as the emerging trends that are reshaping the landscape of data processing. The following discussion outlines the main results and their implications.

3.1. Scalability and Elasticity in Cloud-Based Data Analytics

One of the primary advantages highlighted in the studies is the scalability offered by cloud platforms. Scalability refers to the ability to dynamically adjust resources based on the workload. Cloud platforms such as Amazon Web Services (AWS), Google Cloud, and Microsoft Azure provide elastic computing resources, allowing organizations to scale up or down based on their processing needs. This elasticity is particularly crucial in the context of big data, where data volumes can fluctuate significantly.

Several studies emphasized the benefits of scalability for improving the efficiency of data processing workflows. For instance, the use of Apache Hadoop and Apache Spark in distributed environments enables parallel processing of large datasets, significantly reducing the time required for processing. By utilizing the cloud's elastic capabilities, these frameworks can efficiently manage workloads, ensuring optimal performance without requiring massive upfront infrastructure investments [8, 5].

Furthermore, scalability plays a critical role in real-time data processing. In industries like e-commerce, telecommunications, and finance, where real-time analytics is necessary, the ability to scale resources quickly ensures that cloud-based platforms can handle high-velocity data streams without compromising performance [6]. This scalability allows organizations to maintain smooth operations even during peak demand periods, thereby improving overall business agility.

3.2. Cost Efficiency of Cloud-Based Data Analytics

The cloud's cost-effectiveness is another major benefit that emerged from the studies. Traditional on-premise data centers require significant upfront investments in hardware, as well as ongoing maintenance and operational costs. In contrast, cloud platforms operate on a pay-as-you-go model, meaning organizations only pay for the resources they actually use. This model offers cost savings, particularly for businesses with fluctuating data processing needs. The pay-per-use pricing model was found to be particularly advantageous for businesses that do not have consistent or predictable workloads. For example, cloud platforms allow

organizations to increase their computing resources when needed for big data processing tasks and scale them back down when the demand subsides. This flexibility ensures that businesses are not locked into costly infrastructure investments that remain underutilized during off-peak periods [9, 7]. Additionally, this cost efficiency extends to storage, where cloud-based storage solutions enable organizations to store vast amounts of data without investing in physical hardware or infrastructure.

Moreover, cloud-based data analytics reduces the need for in-house data management teams, as the cloud service providers offer maintenance, security, and support services. This reduction in operational overhead is particularly beneficial for small and medium-sized enterprises (SMEs) that may not have the resources to manage on-premise infrastructure [11, 8].

3.3. Real-Time Data Analytics Capabilities
Real-time data analytics has emerged as a critical capability in the context of cloud-based data processing. With the rapid growth of data from sources such as IoT devices, social media, and transactional systems, organizations must be able to process data as it is generated to gain immediate insights. Studies on stream processing highlight the cloud's ability to handle real-time data streams efficiently.

The integration of cloud-based platforms like Apache Kafka, Apache Flink, and AWS Kinesis has enabled real-time data processing by allowing businesses to ingest and analyze data continuously. These tools enable businesses to detect anomalies, track performance, and make informed decisions in real time. For example, in e-commerce, real-time analytics can help track customer behavior and adjust inventory or pricing strategies instantaneously. Similarly, in financial services, real-time analytics allows for the detection of fraudulent transactions as they occur [10, 11].

Real-time data analytics is also crucial in the healthcare industry, where the ability to monitor patient data continuously can lead to quicker diagnoses and interventions. By leveraging cloud-based real-time analytics, healthcare providers can manage patient data more efficiently and respond to emergencies faster, ultimately improving patient outcomes (Ponnusamy & Gupta, 2024).

3.4. Integration of AI and Machine Learning in Cloud Data Analytics
The integration of Artificial Intelligence (AI) and Machine Learning (ML) into cloud-based data analytics has further enhanced its capabilities. AI and ML algorithms can automate various stages of data processing, from data cleaning and transformation to predictive analytics and decision-making. The cloud's scalability allows these algorithms to be run on large datasets, providing businesses with more accurate and actionable insights.

Machine learning models, such as those used for predictive analytics, are trained on cloud platforms where computational resources can be allocated dynamically to improve model accuracy. For instance, cloud services like AWS SageMaker and Azure Machine Learning provide the infrastructure needed for businesses to train and deploy ML models without worrying about resource limitations. AI is also used to optimize data workflows by automatically identifying patterns in data, making the analytics process faster and more efficient [9, 6].

The use of AI in cloud-based data analytics also extends to enhancing data quality. Machine learning algorithms can automatically classify and filter raw data, reducing manual intervention and improving the quality of data being processed. This ability to clean and organize data automatically is critical in big data environments, where the volume and variety of data often result in inconsistent or incomplete datasets [11].

3.5. Data Security and Privacy Concerns
While cloud-based data analytics offers numerous advantages, data security and privacy remain significant concerns for organizations. The ability to store sensitive data on cloud platforms introduces risks related to unauthorized access and potential breaches. Several studies highlighted the importance of implementing robust security measures to protect data stored in the cloud.

Encryption, access controls, and regular security audits were identified as essential strategies for mitigating security risks. Moreover, compliance with data protection regulations such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) is critical for organizations that process sensitive information, especially in sectors such as finance and healthcare [7, 8].

Cloud service providers have made strides in addressing security concerns by offering built-in security features such as end-to-end encryption and secure data transfer protocols. However, businesses must also ensure that they configure these security features properly and implement their own additional security measures where necessary [11, 7].

3.6. Latency and Network Issues
Another challenge that emerged from the studies is latency, particularly in real-time data processing. While cloud platforms provide low-latency capabilities, network delays can still affect the speed of data processing, especially when large datasets are being transferred across long distances. This can be particularly problematic for industries that require immediate insights, such as financial trading and emergency response systems.

To address latency issues, some studies suggest the adoption of edge computing, where data is processed closer to the source rather than being sent to centralized cloud servers. This approach reduces the need for long-distance data transmission, thereby improving response times and reducing latency. Edge computing is particularly useful in scenarios where real-time data processing is critical, such as in autonomous vehicles or industrial automation [10].

Overall, the integration of cloud computing with data analytics has brought significant improvements in scalability, cost-efficiency, and real-time data processing. Cloud platforms provide the flexibility to scale computing resources based on demand, which is crucial for processing big data and performing real-time analytics. The pay-as-you-go pricing model allows businesses to optimize costs while ensuring access to powerful computational resources. Moreover, cloud-based data analytics tools such as Apache Hadoop, Apache Spark, and real-time processing frameworks like Apache Kafka and AWS Kinesis enable businesses to derive actionable insights from massive datasets.

However, challenges such as data security, privacy concerns, and latency issues remain, requiring businesses to implement robust security measures and consider new

technologies like edge computing to minimize network delays. Despite these challenges, the cloud remains a transformative technology for enabling scalable and efficient data analytics, with the potential to reshape industries across the globe.

Conclusion

Cloud-based data analytics offers numerous benefits, including scalability, cost-efficiency, real-time data processing, and advanced analytics capabilities powered by AI and machine learning. These advantages enable businesses to handle large datasets efficiently, gain real-time insights, and make data-driven decisions. However, challenges such as data security, latency, regulatory compliance, and cost management must be addressed to ensure that cloud-based data analytics solutions are effective and sustainable. As technology continues to evolve, organizations must stay aware of emerging trends and adopt best practices for securing their data and optimizing their cloud resources. With the right strategies in place, businesses can harness the full potential of cloud-based data analytics to drive innovation and maintain a competitive edge in the data-driven world.

References

1. Berisha B, Mëziu E, Shabani I. Big data analytics in cloud computing: an overview. Journal of Cloud Computing. 2022;11(1):24. https://doi.org/10.1186/s13677-022-00301-w
2. Al-Jumaili AHA, Muniyandi RC, Hasan MK, Paw JKS, Singh MJ. Big data analytics using cloud computing-based frameworks for power management systems: status, constraints, and future recommendations. Sensors. 2023;23(6):2952. https://doi.org/10.3390/s23062952. PMID: 36991663; PMCID: PMC10051254
3. Charles E, Iseal S, Olusegun J, Akintola J. Cloud computing for scalable financial data analytics. Issues in Information Systems. 2024;24(1):416-431. https://doi.org/10.1007/385863527
4. Dos Santos W, Avelar GP, Ribeiro MH, Guedes D, Meira W Jr. Lemonade: A platform for cloud-based data analysis and mining with AAA guarantees. Proceedings of the VLDB Endowment. 2018;11(12):2070-2073. https://doi.org/10.14778/3229863.3236262
5. Gupta U, Sharma R. A study of cloud-based solution for data analytics. In: Springer Nature. 2023. https://doi.org/10.1007/978-3-031-33808-3_9
6. Mishra AK, Srinivas K, Verma H, Chawla S. Cloud-based big data analytics platform. International Journal of Engineering Research in Computer Science and Engineering. 2024;11(11):8-19. https://doi.org/10.48009/1_ijercse_2024
7. Naamane Z. A systematic literature review: benefits and challenges of cloud-based big data analytics. Issues in Information Systems. 2023;24(1):291-304. https://doi.org/10.48009/1_iis_2023_125
8. Ponnusamy S, Gupta P. Scalable data partitioning techniques for distributed data processing in cloud environments: a review. IEEE Access. 2024;12:26735-26746. https://doi.org/10.1109/ACCESS.2024.3365810
9. Raghunath V, Kunkulagunta M, Nadella GS. Scalable data processing pipelines: the role of AI and cloud computing. International Journal of Engineering Research in Computer Science and Engineering. 2020;2(2):44-61.
10. Sharif S. Big data processing in the cloud: scalable and real-time data analytics. International Journal of Computer Science and Engineering. 2024;13(1):1-18.
11. Wondegeorgise YW, Samira Z, Osundare OS, Ekpobimi HO, Kandekere RC. Framework for real-time data analytics using cloud technologies. Issues in Information Systems. 2024;24(1):291-304. https://doi.org/10.48009/1_iis_2023_125