IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

ISSN 2347–3657

International Journal of Information Technology & Computer Engineering

Volume 6, Issue 2, 2018

FINANCIAL DATA PROCESSING USING CLOUD-BASED DATA ANALYTICS & DATA MANAGEMENT

[1] **Rajya Lakshmi Gudivaka**
Jawaharlal Nehru Technological University Kakinada,Andhra Pradesh,India
rlakshmigudivaka@gmail.com

[2] **R. Mekala**
Sri Ranganathar Institute of Engineering and Technology
Coimbatore, India.
mail2dr.mekala@gmail.com

ABSTRACT

In the modern financial landscape, the exponential growth of data generated by digital transactions, IoT devices, and online banking has created significant challenges for traditional data processing methods. This paper proposes a cloud-based solution for financial data processing, utilizing advanced data compression, exploratory data analysis (EDA), and cloud data management. The adoption of Optimized Row Columnar (ORC) storage format reduces storage requirements by up to 60%, while cloud platforms like AWS, GCP, and Azure provide scalable and cost-effective resources. Real-time analytics powered by AI-driven tools enable better decision-making for financial institutions. Despite the benefits, challenges related to data privacy, latency, and system integration persist, which are mitigated through hybrid cloud models and enhanced encryption techniques.

Keywords: Cloud Computing, Data Analytics and Data Management

1 INTRODUCTION

In the modern financial landscape, the volume, velocity, and variety of data have grown exponentially, driven by digital transactions, IoT devices, and online banking Archenaa & Anita, 2015 [1]. Financial institutions rely heavily on data to make informed decisions, manage risks, and enhance customer experiences Jain & Kumar, 2015[2]. Traditional data processing methods often struggle to handle this "Big Data," necessitating advanced solutions like cloud-based analytics and optimized storage formats such as Optimized Row Columnar (ORC) Diamantoulakis et al., 2015 [3]. These technologies enable efficient data compression, exploratory data analysis (EDA), and seamless cloud data management, transforming raw data into actionable insights.

The shift toward cloud-based financial data processing is fueled by several factors Manogaran et al., 2017 [4]. The increasing complexity of financial markets demands real-time analytics and scalability, which on-premise systems often lack Agrawal et al., 2011[5]. Additionally, regulatory requirements and the need for data security push organizations to adopt robust cloud platforms. The rise of distributed computing and cost-effective cloud storage solutions further accelerates this transition, allowing businesses to process vast datasets without heavy infrastructure investments.

Despite its advantages, cloud-based financial data processing faces significant challenges Yang et al., 2017[6]. Data privacy and compliance with regulations like GDPR and PCI-DSS remain critical concerns Balachandran & Prasad, 2017 [7]. Latency issues can arise when processing large datasets in the cloud, impacting real-time decision-making. Moreover, integrating legacy systems with modern cloud platforms often leads to compatibility issues, while the risk of data breaches necessitates stringent security measures.

To address these issues, financial institutions are adopting hybrid cloud models, combining the flexibility of public clouds with the security of private infrastructure. Advanced encryption techniques and zero-trust architectures enhance data protection. Optimized frameworks like ORC improve processing efficiency, while AI-driven analytics tools streamline EDA. By partnering with trusted cloud providers and investing in employee training, organizations can mitigate risks and fully leverage the potential of cloud-based financial data processing.

1.1 PROBLEM STATEMENT

The proposed work addresses challenges in cloud-based financial data processing through key solutionsGupta et al., 2012 [8]. It incorporates hybrid cloud models to enhance security and compliance, combining the flexibility of public clouds with the security of private infrastructure Sezer et al., 2017[9]. Optimized data compression using ORC format reduces storage requirements by up to 60%, ensuring efficient data management. Real-time analytics through Exploratory Data Analysis (EDA) and AI-driven tools enables timely insights for decision-making, while cloud data management leverages platforms like AWS, GCP, and Azure for scalability and cost-effectiveness, reducing the need for heavy infrastructure investments Mansouri et al., 2018[10]. This comprehensive approach ensures efficient, secure, and scalable financial data processing.

1.2 OBJECTIVES

- Analyze the challenges in cloud-based financial data processing, particularly focusing on security, compliance, and efficient data management.
- Implement hybrid cloud models combining public cloud flexibility with private cloud security to enhance data privacy and compliance.
- Evaluate the impact of optimized data compression using the ORC format to reduce storage requirements by up to 60%.
- Apply real-time analytics through Exploratory Data Analysis (EDA) and AI-driven tools to provide timely insights for decision-making.
- Leverage cloud platforms like AWS, GCP, and Azure for scalability and cost-effective data management, reducing infrastructure investments.

2 LITERATURE SURVEY

The continuous growth of information systems and the rise of mobile computing have made it essential for users to access enterprise applications across various devices like tablets, smartphones, and desktops. To enable effective decision-making, businesses are leveraging Business Intelligence (BI) tools, such as data mining, OLAP, querying, and reporting, to analyze raw data and provide actionable insights. Cloud computing has emerged as a popular solution for storing and processing business data, especially with the challenges associated with local storage. However, organizations face obstacles in fully transitioning to the cloud, which hampers its adoption. Tangirala, 2016 [11] discusses these challenges and proposes solutions for overcoming them, making cloud computing a promising architecture for better information management. Additionally, a new data model for Big Data is presented to address the unique needs of modern business analytics.

Sun et al., 2016 [12] introduces the concept of Smart and Connected Communities (SCC), an evolution of smart cities aimed at addressing the needs of preservation, livability, and future planning. The vision of SCC is to enhance a community's livability, preservation, revitalization, and attainability, focusing on living in the present, planning for the future, and remembering the past. The paper argues that the Internet of Things (IoT) can provide a connected network of devices and sensors, while big data analytics enables real-time control for SCC. Mobile crowdsensing and cyber-physical cloud computing are identified as key IoT technologies in advancing SCC. As a case study, TreSight is presented, integrating IoT and big data analytics for smart tourism and sustainable cultural heritage in Trento, Italy.

Cai et al., 2016 [13] highlights the significance of IoT-related applications in addressing data-related challenges within contemporary business organizations, particularly in the context of cloud computing. It introduces a functional framework that outlines the acquisition, management, processing, and mining of IoT big data, with key technical modules and their capabilities defined. The paper also analyzes current IoT application research, identifies challenges and opportunities in IoT big data research, and reviews critical academic and industry publications on the topic. Finally, open issues and examples are discussed within the proposed IoT research framework.

Sharma & Wang, 2017 [14] discusses the growing importance of big data analytics in various domains, with a focus on the Internet of Things (IoT), which connects people, processes, data, and objects to improve daily life. Key challenges include extracting useful features from the vast amount of heterogeneous data generated by IoT devices and using this data to enhance wireless IoT network performance. While cloud and edge computing offer solutions, each has its benefits and limitations. The paper proposes a novel framework for coordinated processing between edge and cloud computing, leveraging the strengths of both platforms to enable real-time data analytics in wireless IoT networks. The framework integrates network-wide knowledge and historical data from the cloud to guide edge computing units, addressing performance requirements and identifying key enablers, challenges, and future research directions.

Van Oosterom et al., 2015 [15] presents a comprehensive inventory of point cloud data management user requirements, gathered through structured interviews with users from government, industry, and academia. Based on these requirements, a benchmark has been developed to evaluate various point cloud data management solutions, including database models in Oracle and PostgreSQL, compared with the flat table model in Oracle, PostgreSQL, and MonetDB. The study uses the second national height map of the Netherlands (AHN2) as the main dataset, containing 640 billion points. The benchmark results are presented for different stages with increasing data volume and complexity, from mini (20 million points) to full benchmark (complete AHN2). Several improvements in point cloud data management, such as data ordering algorithms and a unique vario-scale LoD organization, are proposed and partially tested for better performance.

Manikyam & Kumar, 2017 [16] provides an analysis of Big Data applications in Cloud Computing Architecture, focusing on four key areas: data management and supporting architectures, model development and scoring, visualization and user interaction, and business models. Through a detailed examination, the paper identifies technological gaps and offers guidance for Big Data analysis in Cloud Computing. It also highlights challenges in Big Data management, such as data variety, storage, integration, processing, and resource management. The paper discusses three major areas of analytics—predictive, descriptive, and prescriptive data—enabled by visualization and user interaction techniques.

Chaudhary et al., 2017 [17] addresses the challenges arising from the exponential increase in Internet-enabled devices, leading to the growing popularity of fog and cloud computing. With high data rates and secure access expected by end users, the bidirectional data flow between users and devices can cause congestion at cloud data centers, especially considering the high mobility of devices like vehicles. To address these challenges, we propose an integrated cloud and fog computing architecture in the 5G environment, leveraging advanced technologies such as SDN and NFV with the NSC model. This architecture automates virtual resources for fast computing and supports data analytics, while addressing issues of device mobility. The paper also compares core and edge computing, focusing on node heterogeneity, security, and potential attacks on data in the 5G environment.

Kune et al., 2016 [18] explores the evolution of Big Data computing, which arises from the explosion of information in various fields such as business, engineering, medicine, and science. As the volume of data grows, knowledge discovery and decision-making become increasingly complex, requiring new data-intensive techniques and mathematical models. Big Data computing demands substantial storage and computing resources, which can be provided through on-premise or cloud infrastructures. The paper contrasts traditional data warehousing with Big Data, presents a taxonomy of Big Data computing and its underlying technologies, and introduces the integrated platform of Big Data and Clouds, known as Big Data Clouds. It also discusses the layered architecture, components of Big Data Cloud, and identifies open technical challenges and future research directions.

Ahmed et al., 2017 [19] examines the growing intersection of Big Data and the Internet of Things (IoT), driven by the rapid expansion of connected devices and increasing data consumption. Managing Big Data within the continuously growing IoT network presents challenges related to data collection, processing, analytics, and security. The paper explores recent advancements in Big Data analytics for IoT systems, identifying key requirements for managing Big Data and enabling analytics in IoT environments. It also provides a taxonomy of relevant literature, highlighting the opportunities arising from the convergence of Big Data, analytics, and IoT. Finally, the paper discusses the role of Big Data analytics in IoT applications and outlines several open challenges as future research directions.

Khan et al., 2015 [20] explores the management and analysis of big data in smart cities, focusing on land-use, environment, socio-economic, energy, and transport data to support decision-making in science, policy, planning, governance, and business. It proposes a cloud-based analytics service for big data analysis, demonstrated through a prototype built with Hadoop and Spark. The service analyzes the Bristol Open Data, identifying correlations between urban environment indicators such as crime, safety, economy, and employment. Experiments comparing Hadoop and Spark results are presented, with an assessment of trends over the years based on quality-of-life indicators to identify positive and negative patterns.

3 METHODOLOGY

This flow illustrates the end-to-end process of financial data management using modern technologies. It begins with Finance Data Collection, where large volumes of structured and unstructured data (Big Data) are gathered from various sources. Next, Data Compression techniques like Optimized Row Columnar (ORC) are applied to reduce storage costs and improve processing efficiency. The compressed data then undergoes Data Analytics, specifically Exploratory Data Analysis (EDA), to uncover patterns, trends, and insights. Finally, the analyzed data is stored and managed in Cloud Data Management systems, ensuring scalability, accessibility, and security. This streamlined approach enables financial institutions to handle massive datasets effectively while supporting real-time decision-making.

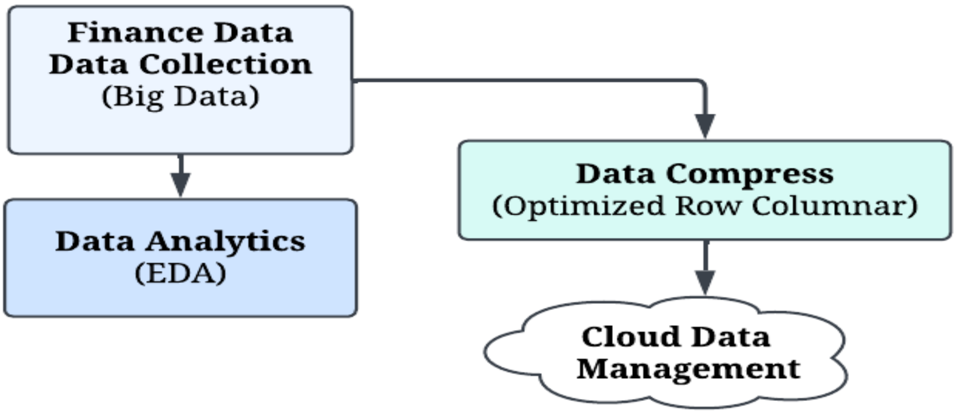

Figure 1: Optimized Financial Data Management and Analytics

3.1 DATA COLLECTION

Financial data collection involves gathering large volumes of transactional and market data from diverse sources such as transaction records, market APIs (e.g., stock prices, forex), IoT and mobile banking logs, and regulatory filings. The challenge lies in the real-time ingestion of these high-velocity data streams, as financial data constantly evolves. Additionally, schema variability across different data sources makes it difficult to standardize the information. To address these challenges, distributed frameworks like Apache Kafka and Spark Streaming are employed to manage and process this large-scale, dynamic data in real time. These technologies ensure smooth data flow, even during peak volumes, enabling efficient storage and analytics.

3.2 DATA COMPRESSION USING OPTIMIZED ROW COLUMNAR

The Optimized Row Columnar (ORC) format significantly improves data storage efficiency in big data environments. By using columnar storage, ORC allows queries to only access relevant columns, reducing the amount of data read during data retrieval. Lightweight compression techniques, such as run-length encoding and dictionary encoding, are applied to minimize the data footprint without compromising performance. The use of predicate pushdown further optimizes query execution by filtering data early in the process, thus minimizing I/O operations. Benchmarks have demonstrated that ORC can reduce storage requirements by up to 60% when compared to traditional row-based formats, making it a preferred choice for handling large datasets in financial applications.

3.3 DATA ANALYTICS USING EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis (EDA) is an essential step in understanding the structure and patterns within financial data. Techniques such as statistical summaries (mean, variance) are used to grasp the central tendencies and spread of the data. Visualizations, including time-series trends and risk heatmaps, help identify patterns and anomalies over time, aiding in risk management and forecasting. Anomaly detection, especially for fraud patterns, is another crucial aspect of EDA in financial datasets. Tools like Python (with libraries such as Pandas and Seaborn) and Spark ML automate the process of generating insights, making it easier for analysts to detect significant patterns and anomalies in vast datasets, thus driving better decision-making.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i, \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2 \quad (1)$$

ISSN 2347–3657

International Journal of
Information Technology & Computer Engineering

Volume 6, Issue 2, 2018

Were, μ is the **mean** of the dataset $x1, x2, \ldots, xn$, σ^2 is the **variance**, a measure of the spread or dispersion of the data, n is the number of data points in the dataset, xi represents individual data points.

3.4 CLOUD DATA MANAGEMENT

Cloud data management enables financial organizations to scale their data processing capabilities and ensure secure and efficient storage. Platforms like AWS, Google Cloud Platform (GCP), and Microsoft Azure offer scalable resources that can automatically adjust during peak demand, ensuring continuous availability and performance. Security is a critical concern in financial data management, and these cloud providers offer robust encryption methods (such as AES-256) and compliance with industry standards like GDPR and PCI-DSS, safeguarding sensitive data. Additionally, cloud-based solutions are more cost-effective, offering pay-per-use pricing models, which are significantly cheaper compared to traditional on-premise setups with high capital expenditures. This approach provides flexibility, cost savings, and security, making cloud data management ideal for the finance industry.

4 RESULT AND DISCUSSION

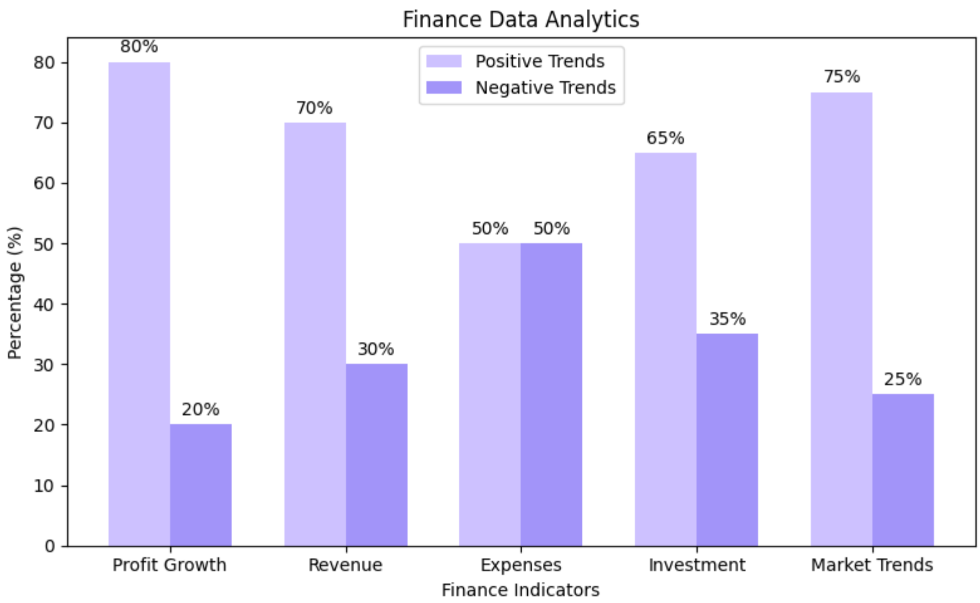

Figure 2: Data Analytics

The Figure 2 represents the results of finance data analytics, comparing the positive and negative trends for five key finance indicators: Profit Growth, Revenue, Expenses, Investment, and Market Trends. The percentage of positive trends is shown in a lighter shade, while negative trends are displayed in a darker shade of purple. Profit Growth has the highest positive trend at 80%, with a smaller negative trend of 20%. Revenue shows a positive trend of 70%, but the negative trend is 30%. Expenses are split evenly, with 50% positive and 50% negative trends. Investment shows a higher positive trend of 65%, while Market Trends have 75% positive and 25% negative trends. This analysis highlights how different financial indicators are performing in terms of positive and negative trends.

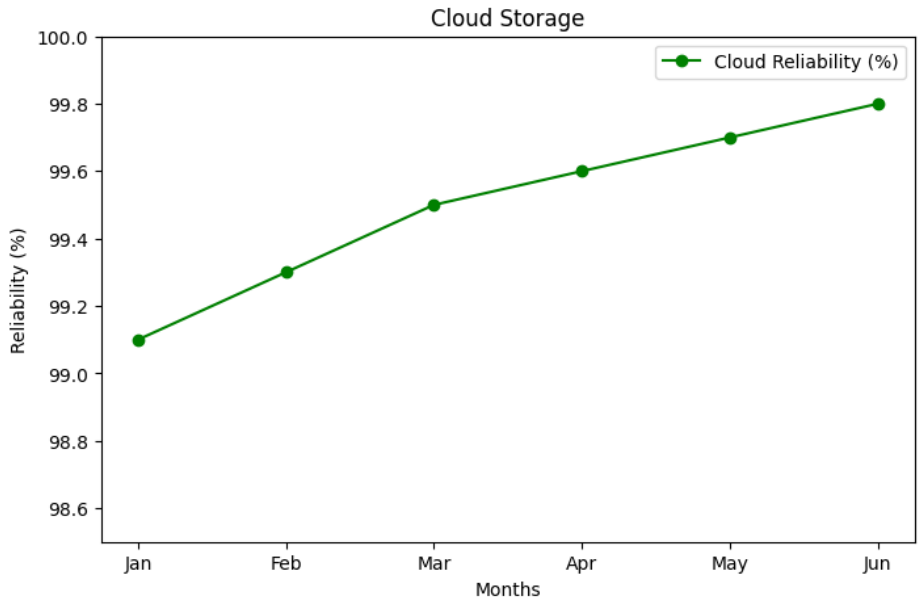

Figure 3: Cloud storage

This Figure 3 depicts the cloud storage reliability over a six-month period, from January to June. The reliability percentage steadily increases each month, starting at 99.2% in January and rising to 99.8% in June. This demonstrates an improvement in cloud storage reliability, with the trend showing consistent growth in performance. The green line represents the reliability percentage, and each data point corresponds to the monthly reliability value. The graph indicates that the cloud storage system is becoming more reliable over time.

5 CONCLUSIONS

This paper presents an effective framework for managing and processing financial data through cloud-based platforms, enabling organizations to overcome traditional data management challenges. By leveraging the ORC format for optimized storage and integrating Exploratory Data Analysis (EDA) for actionable insights, financial institutions can make more informed decisions while ensuring scalability and security. The use of cloud data management platforms further enhances efficiency, providing real-time analytics to stay competitive in the rapidly changing financial landscape. However, the integration of legacy systems and compliance with data privacy regulations remain crucial considerations for fully adopting these solutions. The proposed hybrid cloud approach, combined with robust encryption and AI tools, ensures the secure and efficient management of vast datasets, paving the way for improved financial services.

REFERENCE

[1] J. Archenaa and E. M. Anita, "A survey of big data analytics in healthcare and government," *Procedia Comput. Sci.*, vol. 50, pp. 408–413, 2015.

[2] V. K. Jain and S. Kumar, "Big data analytic using cloud computing," in *2015 Second International Conference on Advances in Computing and Communication Engineering*, IEEE, 2015, pp. 667–672. Accessed: Apr. 01, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7306766/

[3] P. D. Diamantoulakis, V. M. Kapinas, and G. K. Karagiannidis, "Big data analytics for dynamic energy management in smart grids," *Big Data Res.*, vol. 2, no. 3, pp. 94–101, 2015.

[4] G. Manogaran, D. Lopez, C. Thota, K. M. Abbas, S. Pyne, and R. Sundarasekar, "Big Data Analytics in Healthcare Internet of Things," in *Innovative Healthcare Systems for the 21st Century*, H. Qudrat-Ullah and P. Tsasis, Eds., in Understanding Complex Systems. , Cham: Springer International Publishing, 2017, pp. 263–284. doi: 10.1007/978-3-319-55774-8_10.

[5] D. Agrawal, S. Das, and A. El Abbadi, "Big data and cloud computing: current state and future opportunities," in *Proceedings of the 14th International Conference on Extending Database Technology*, Uppsala Sweden: ACM, Mar. 2011, pp. 530–533. doi: 10.1145/1951365.1951432.

[6] C. Yang, Q. Huang, Z. Li, K. Liu, and F. Hu, "Big Data and cloud computing: innovation opportunities and challenges," *Int. J. Digit. Earth*, vol. 10, no. 1, pp. 13–53, Jan. 2017, doi: 10.1080/17538947.2016.1239771.

[7] B. M. Balachandran and S. Prasad, "Challenges and benefits of deploying big data analytics in the cloud for business intelligence," *Procedia Comput. Sci.*, vol. 112, pp. 1112–1122, 2017.

[8] R. Gupta, H. Gupta, and M. Mohania, "Cloud Computing and Big Data Analytics: What Is New from Databases Perspective?," in *Big Data Analytics*, vol. 7678, S. Srinivasa and V. Bhatnagar, Eds., in Lecture Notes in Computer Science, vol. 7678. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 42–61. doi: 10.1007/978-3-642-35542-4_5.

[9] O. B. Sezer, E. Dogdu, and A. M. Ozbayoglu, "Context-aware computing, learning, and big data in internet of things: a survey," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 1–27, 2017.

[10] Mansouri, Y., Toosi, A. N., & Buyya, R. (2017). Data storage management in cloud environments: Taxonomy, survey, and future directions. *ACM Computing Surveys (CSUR)*, *50*(6), 1-51.

[11] S. Tangirala, "Efficient big data analytics and management through the usage of cloud architecture," *J. Adv. Inf. Technol. Vol*, vol. 7, no. 4, pp. 302–307, 2016.

[12] Y. Sun, H. Song, A. J. Jara, and R. Bie, "Internet of things and big data analytics for smart and connected communities," *IEEE Access*, vol. 4, pp. 766–773, 2016.

[13] H. Cai, B. Xu, L. Jiang, and A. V. Vasilakos, "IoT-based big data storage systems in cloud computing: perspectives and challenges," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 75–87, 2016.

[14] S. K. Sharma and X. Wang, "Live data analytics with collaborative edge and cloud processing in wireless IoT networks," *IEEE Access*, vol. 5, pp. 4621–4635, 2017.

[15] P. Van Oosterom *et al.*, "Massive point cloud data management: Design, implementation and execution of a point cloud benchmark," *Comput. Graph.*, vol. 49, pp. 92–125, 2015.

[16] N. R. H. Manikyam and S. M. Kumar, "Methods and techniques to deal with big data analytics and challenges in cloud computing environment," *Int. J. Civ. Eng. Technol.*, vol. 8, no. 4, pp. 669–678, 2017.

[17] R. Chaudhary, N. Kumar, and S. Zeadally, "Network service chaining in fog and cloud computing for the 5G environment: Data management and security challenges," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 114–122, 2017.

[18] R. Kune, P. K. Konugurthi, A. Agarwal, R. R. Chillarige, and R. Buyya, "The anatomy of big data computing," *Softw. Pract. Exp.*, vol. 46, no. 1, pp. 79–105, 2016.

[19] Arulkumaran, G., & Gnanamurthy, R. K. (2014). Improving Reliability against Security Attacks by Identifying Reliance Node in MANET. Journal of Advances in Computer Networks, 2(2).

[20] Z. Khan, A. Anjum, K. Soomro, and M. A. Tahir, "Towards cloud based big data analytics for smart future cities," *J. Cloud Comput.*, vol. 4, pp. 1–11, 2015.