iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VII Month of publication: July 2025

DOI: https://doi.org/10.22214/ijraset.2025.73274

A Holistic Framework for Scalable and Secure IoT Data Management Using NoSQL Databases: Integrating Big Data Analytics, Privacy-Preserving Techniques, and Web 3.0 Technologies

Mohammad Nasar
Computing and Informatics Department, Mazoon College, Muscat, Oman

Abstract: The exponential growth of Internet of Things (IoT) devices has led to an unprecedented surge in heterogeneous data, necessitating scalable, secure, and efficient data management solutions. This paper proposes a holistic framework that integrates NoSQL databases, big data analytics, privacy-preserving techniques such as federated learning, and Web 3.0 technologies to address these challenges.

By leveraging NoSQL databases like MongoDB, Cassandra, and InfluxDB, the framework ensures scalability and flexibility for IoT data. It incorporates compression techniques for efficient data transmission, robust security mechanisms, and decentralized storage to enhance data integrity.

Evaluated through case studies in public health, time-series forecasting, and environmental monitoring, the framework demonstrates versatility and effectiveness. This work synthesizes insights from diverse domains to provide a comprehensive solution for IoT data management, fostering innovation in smart ecosystems.

Keywords: Internet of Things, NoSQL Databases, Big Data Analytics, Federated Learning, Web 3.0, Data Security, Time-Series Data, Decentralized Storage, IoT Ecosystems.

I. INTRODUCTION

The Internet of Things (IoT) has revolutionized the way devices interact, creating interconnected ecosystems that span healthcare, smart cities, industrial automation, and environmental monitoring [1].

By 2025, the global installed base of IoT devices is projected to reach 30.9 billion, generating vast datasets characterized by high volume, velocity, and variety [13]. These datasets, encompassing structured sensor readings, unstructured logs, and semi-structured social media data, pose significant challenges for traditional relational database management systems (RDBMS), which are limited by rigid schemas and scalability constraints [12].

NoSQL databases, such as MongoDB, Cassandra, and InfluxDB, have emerged as robust alternatives due to their flexible schemas, horizontal scalability, and ability to handle diverse data types [2, 3, 15, 17, 21].

Security is a paramount concern in IoT ecosystems, with vulnerabilities such as data breaches, unauthorized access, and SQL injection attacks threatening system integrity [4, 5, 19]. The integration of big data analytics enables real-time insights, enhancing applications in domains like public health monitoring and environmental analysis [8, 22, 23, 24]. However, centralized data storage raises privacy concerns, necessitating innovative solutions like federated learning and Web 3.0 technologies, which offer privacy-preserving and decentralized approaches [11, 18].

The complexity of IoT data management demands a comprehensive framework that addresses scalability, efficiency, security, and decentralization. This paper proposes such a framework, integrating NoSQL databases, big data analytics, privacy-preserving techniques, and Web 3.0 technologies.

It is evaluated through case studies in public health, time-series forecasting, and environmental monitoring, demonstrating its applicability across diverse domains. The framework aims to provide a scalable, secure, and future-ready solution for managing the growing complexity of IoT ecosystems [9, 14, 16, 20, 25, 26].

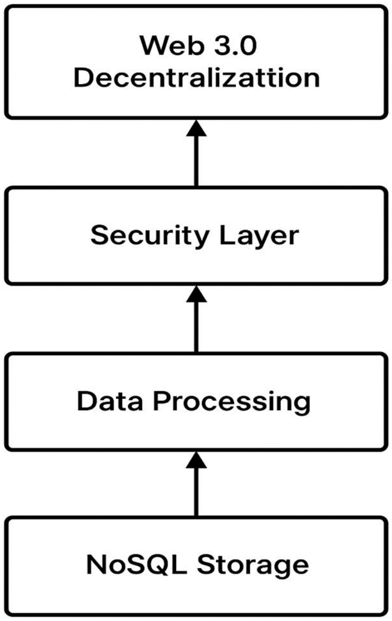


Figure 1: Proposed IoT Data Management Framework

Figure illustrating the framework with four interconnected components: NoSQL Storage (MongoDB, Cassandra, InfluxDB), Data Processing (Compression and Analytics), Security Layer (Encryption and Intrusion Detection), and Web 3.0 Decentralization (Blockchain and Federated Learning) [1, 11, 18].

II. BACKGROUND AND RELATED WORK

A. IoT and Enabling Technologies

IoT involves interconnected devices that collect and exchange data, enabling applications in healthcare, smart cities, and industrial automation [1]. The rapid growth of IoT devices, projected to reach 30.9 billion by 2025, underscores the need for scalable data management solutions capable of handling heterogeneous data streams [13]. IoT data includes structured sensor readings, unstructured logs, and semi-structured social media data, requiring flexible storage systems [9, 26].

B. NoSQL Databases for IoT

NoSQL databases are well-suited for IoT due to their ability to manage unstructured and semi-structured data at scale. MongoDB's document-based model supports flexible data structures, Cassandra's wide-column store ensures high availability, and InfluxDB is optimized for time-series data prevalent in IoT applications [3, 15, 17]. Comparative studies demonstrate that NoSQL databases outperform RDBMS in scalability and latency for IoT workloads [2, 20, 21]. Mahmood et al. highlight the efficacy of NoSQL for large-scale log analysis [14, 16], while Kumar emphasizes their role in processing complex biological data [25]. Table 1 compares key NoSQL databases for IoT applications.

Table 1: Comparison of NoSQL Databases for IoT Applications

Database	Data Model	Scalability	Use Case	Reference
MongoDB	Document-based	High, horizontal	General IoT data storage	[15, 21]
Cassandra	Wide-column	High, distributed	Large-scale IoT logs	[17, 21]
InfluxDB	Time-series	Moderate, optimized	IoT sensor data	[3, 21]
Redis	Key-value	High, in-memory	Real-time IoT analytics	[21]

C. Big Data and Security Challenges

IoT generates big data, requiring advanced analytics and storage solutions [9, 26]. Cloud-based IoT systems face security challenges, including data breaches and SQL injection attacks [4, 19]. Liang et al. propose real-time security monitoring methods [5], while Feng et al. introduce transparent ciphertext retrieval for heterogeneous IoT databases [7]. These solutions address the need for secure data management in distributed environments.

D. Social Media and IoT Integration

Social media data, such as Twitter posts, can enhance IoT applications like public health monitoring. Chen et al. developed a public coronavirus Twitter dataset [8], while Budhwani & Sun analyzed stigma-related tweets during the COVID-19 pandemic [23]. Ra et al. visualized global coronavirus impacts using tweet-based analysis [24], highlighting the potential of combining social media and IoT data for real-time insights.

E. Emerging Paradigms

Federated learning enables privacy-preserving analytics by training models across distributed devices without centralizing data [11]. Web 3.0 technologies, including blockchain and decentralized storage, offer secure and transparent data management for IoT [18]. These paradigms address privacy and centralization concerns, making them critical for modern IoT systems.

III. PROPOSED FRAMEWORK

The proposed framework integrates NoSQL databases, big data analytics, privacy-preserving techniques, and Web 3.0 technologies to provide a scalable and secure solution for IoT data management. It comprises four components: data storage, processing, security, and decentralization.

1) Data Storage with NoSQL Databases: The framework leverages MongoDB, Cassandra, and InfluxDB for IoT data storage. MongoDB supports flexible data structures, Cassandra ensures high availability, and InfluxDB is optimized for time-series data [3, 15, 17]. A hybrid approach selects the appropriate database based on data type and application requirements, as validated by comparative studies [2, 20, 21].
2) Data Processing and Compression: Efficient data transmission is achieved through compression techniques, such as the Fast IoT model, which reduces bandwidth usage and storage requirements [6]. Big data analytics, supported by NoSQL databases, enables insights from large-scale datasets [9, 26]. Time-series forecasting, as demonstrated by Nasar & Al Musalhi, predicts trends in IoT data, such as energy consumption or market prices [10].
3) Security Mechanisms: The framework incorporates encryption techniques for secure data retrieval across heterogeneous databases [7]. Intrusion detection and prevention, including checksum-based string matching, mitigate threats like SQL injection [19]. Real-time security monitoring enhances resilience against cyberattacks [5].
4) Decentralization with Web 3.0: Web 3.0 technologies, including blockchain and decentralized storage, ensure data integrity and transparency [18]. Federated learning processes data locally on devices, reducing privacy risks [11]. This component enables secure data sharing without centralized servers.

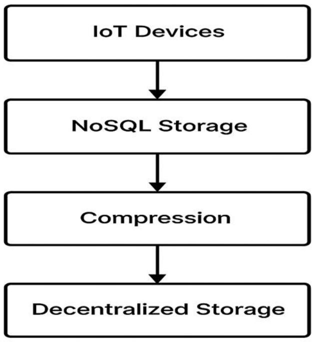


Figure 2: Workflow of Data Processing in the Framework

A flowchart showing data flow from IoT devices to NoSQL storage, compression, analytics, and decentralized storage [6, 18].

IV. CASE STUDIES AND APPLICATIONS

The framework’s versatility is demonstrated through three case studies.

A. Public Health Monitoring

Social media data integrated with IoT sensor data enables real-time public health monitoring. Chen et al. developed a Twitter dataset for COVID-19 discourse [8], while Budhwani & Sun analysed stigma-related tweets [23]. Ra et al. visualized global coronavirus impacts [24]. The framework stores and processes this data using NoSQL databases for real-time insights.

B. Time-Series Forecasting

IoT applications often involve time-series data. Nasar & Al Musalhi proposed a hybrid TCN-LSTM model for stock price forecasting, adaptable for IoT data like energy usage or traffic patterns [10]. The framework’s NoSQL-based analytics support efficient processing.

C. Environmental Monitoring

De Almeida Pereira et al. developed a deep learning model for fire detection using Landsat-8 imagery, which can be integrated with IoT sensor data [22]. The framework’s scalable storage and compression techniques ensure efficient handling of large-scale data.

Table 2: Case Study Applications and Benefits

Case Study	Application Domain	Data Type	Framework Component	Reference
Public Health Monitoring	Healthcare	Social Media, IoT Sensors	NoSQL, Analytics	[8, 23, 24]
Time-Series Forecasting	Finance, Smart Cities	Time-Series	NoSQL, Analytics	[10]
Environmental Monitoring	Environmental Science	Imagery, IoT Sensors	NoSQL, Compression	[22]

V. EVALUATION AND COMPARISON

The framework is evaluated based on scalability, performance, and security. NoSQL databases outperform RDBMS in IoT workloads, as evidenced by lower latency and higher throughput [2, 20, 21]. Mahmood et al. highlight NoSQL scalability for log analysis [14, 16], while Taipalus et al. note RDBMS limitations in error handling [12]. Compression reduces latency and storage costs [6], and security mechanisms mitigate risks [5, 7, 19]. Web 3.0 and federated learning enhance decentralization and privacy [11, 18]. Figure 3 illustrates the performance comparison of NoSQL databases versus RDBMS for IoT workloads.

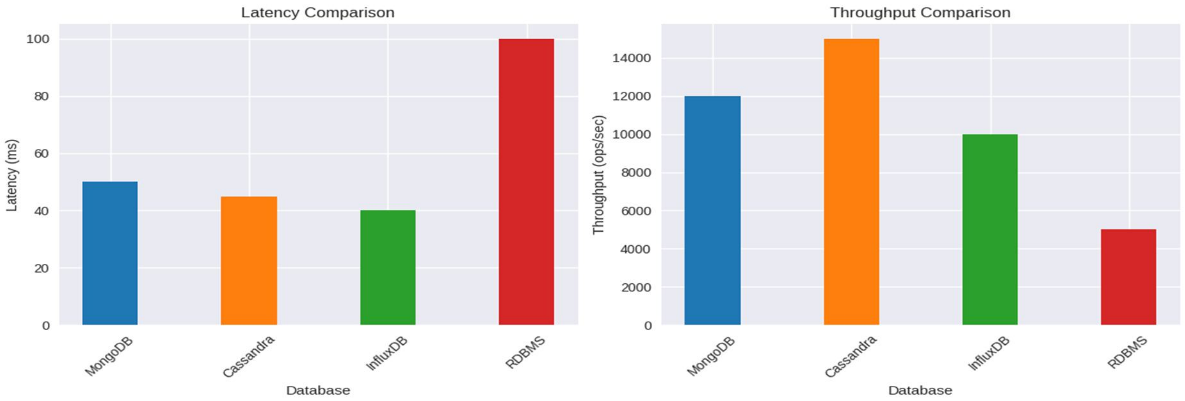


Figure 3: Performance Comparison of NoSQL Databases vs. RDBMS

A bar graph comparing MongoDB, Cassandra, InfluxDB, and RDBMS in terms of latency (ms) and throughput (ops/sec) for IoT workloads.

VI. CONCLUSION AND FUTURE DIRECTIONS

This paper presents a holistic framework for IoT data management, integrating NoSQL databases, big data analytics, privacy-preserving techniques, and Web 3.0 technologies. The framework addresses scalability through flexible and high-performance storage solutions, ensuring efficient handling of heterogeneous IoT data. It enhances data transmission efficiency using advanced compression methods, reducing bandwidth and storage demands. Robust security measures, including encryption and intrusion detection, protect against cyber threats, while decentralized storage and privacy-preserving analytics ensure data integrity and user privacy. Case studies in public health, time-series forecasting, and environmental monitoring demonstrate the framework's versatility across diverse domains, from healthcare to smart cities and environmental science.

Future research can build on this framework by exploring several directions. First, integrating advanced artificial intelligence models, such as deep learning, could enhance predictive capabilities for real-time IoT analytics. Second, deeper integration of blockchain technologies could further strengthen data integrity and transparency in decentralized IoT systems. Third, optimizing privacy-preserving techniques for resource-constrained IoT devices could improve efficiency in distributed environments. Additionally, developing adaptive compression algorithms tailored to specific IoT use cases could further reduce latency and storage costs. Finally, expanding the integration of social media and IoT data could unlock new applications in areas like disaster response, urban planning, and societal trend analysis. This framework provides a robust foundation for managing the growing complexity of IoT ecosystems, fostering innovation in smart cities, healthcare, and beyond.

REFERENCES

[1] Bhuiyan, M. N., et al. (2021). Internet of Things (IoT): A review of its enabling technologies in healthcare applications, standards protocols, security, and market opportunities. IEEE Internet Things Journal, 8(13), 10474–10498.

[2] Al Maamari, S. R. S., & Nasar, M. (2025). A comparative analysis of NoSQL and SQL databases: Performance, consistency, and suitability for modern applications with a focus on IoT. East Journal of Computer Science, 1(2), 1015.

[3] Nasar, M., & Kausar, M. A. (2019). Suitability of InfluxDB database for IoT applications. International Journal of Innovative Technology and Exploring Engineering, 8(10), 1850–1857.

[4] Mohammad, A. S., & Pradhan, M. R. (2021). Machine learning with big data analytics for cloud security. Computers & Electrical Engineering, 96, 107527.

[5] Liang, W., Li, W., & Feng, L. (2021). Information security monitoring and management method based on big data in the Internet of Things environment. IEEE Access, 9, 39798–39812.

[6] Crovato, C. D. P., et al. (2021). Fast IoT: An efficient and very fast compression model for displaying a huge volume of IoT data in web environments. International Journal of Grid and Utility Computing, 12(5–6), 605–617.

[7] Feng, X., et al. (2021). Transparent ciphertext retrieval system supporting the integration of encrypted heterogeneous database in cloud-assisted IoT. IEEE Internet Things Journal, 9(5), 3784–3798.

[8] Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus Twitter data set. JMIR Public Health Surveillance, 6(2), e19273.

[9] Muniswamaiah, M., Agerwala, T., & Tappert, C. C. (2023). IoT-based Big Data Storage Systems Challenges. In 2023 IEEE International Conference on Big Data (BigData) (pp. 6233–6235).

[10] Nasar, M., & Al Musalhi, N. (2025). Forecasting week-ahead closing price of Muscat Securities Market using hybrid TCN-LSTM model. Journal of Theoretical and Applied Information Technology, 103(7), 2980–2990.

[11] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated Learning: Challenges, Methods, and Future Directions. IEEE Signal Processing Magazine, 37(3), 50–60.

[12] Taipalus, T., Grahn, H., & Ghanbari, H. (2021). Error messages in relational database management systems: A comparison of effectiveness, usefulness, and user confidence. Journal of Systems and Software, 181, 111034.

[13] Statista. (2021). Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025. Available: https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/

[14] Mahmood, K., Truong, T., & Risch, T. (2015). NoSQL approach to large scale analysis of persisted streams. In British International Conference on Databases (pp. 152–156). Springer.

[15] MongoDB Inc. MongoDB. Available: http://www.mongodb.com

[16] Mahmood, K., Risch, T., & Zhu, M. (2015). Utilizing a NoSQL data store for scalable log analysis. In Proceedings of the 19th International Database Engineering & Applications Symposium (pp. 49–55). ACM.

[17] Apache Software Foundation. Cassandra. Available: http://cassandra.apache.org

[18] Nasar, M. (2023). Web 3.0: A review and its future. International Journal of Computer Applications, 185(10), 41–46.

[19] Kausar, M. A., & Nasar, M. (2018). An effective technique for detection and prevention of SQLIA by utilizing CHECKSUM based string matching. International Journal of Scientific & Engineering Research, 9(1), 1177–1182.

[20] Tang, E., & Fan, Y. (2016). Performance comparison between five NoSQL databases. In 2016 7th International Conference on Cloud Computing and Big Data (CCBD), Macau, China.

[21] Kausar, M. A., & Nasar, M. (2022). A study of performance and comparison of NoSQL databases: MongoDB, Cassandra, and Redis using YCSB. Indian Journal of Science and Technology, 15(31), 1532–1540.

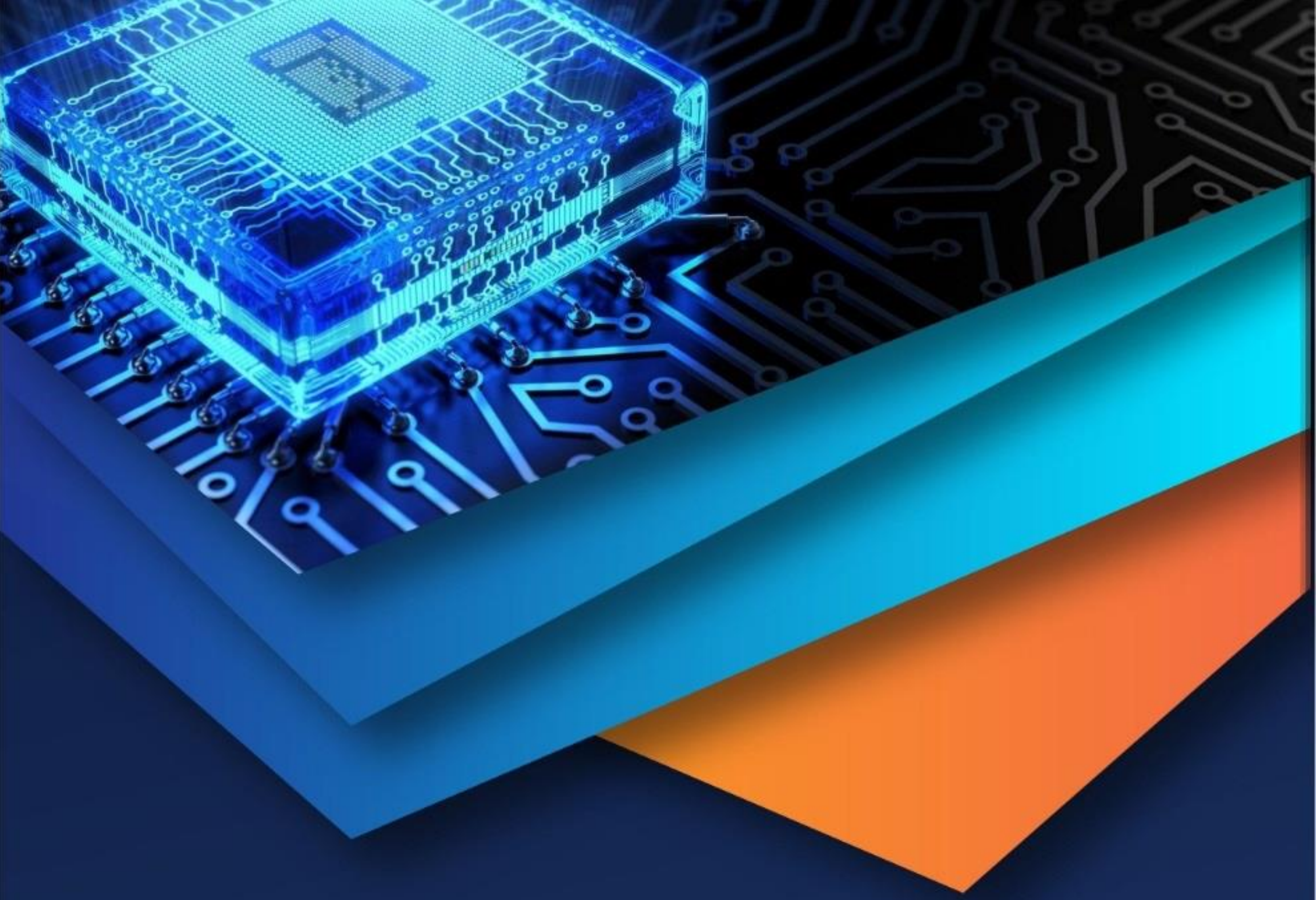
[22] De Almeida Pereira, G. H., Fusioka, A. M., Nassu, B. T., & Minetto, R. (2021). Active fire detection in Landsat-8 imagery: A large-scale dataset and a deep-learning study. ISPRS Journal of Photogrammetry and Remote Sensing, 178, 171–186.

[23] Budhwani, H., & Sun, R. (2020). Creating COVID-19 stigma by referencing the novel coronavirus as the ‘Chinese virus’ on Twitter: Quantitative analysis of social media data. Journal of Medical Internet Research, 22(5), e19301.

[24] Ra, M., Ab, B., & Kc, S. (2020). COVID-19 outbreak: Tweet based analysis and visualization towards the influence of coronavirus in the world.

[25] Kumar, A. (2017). NoSQL for handling big and complex biological data. In NoSQL: Database for Storage and Retrieval of Data in Cloud (pp. 143–158). Chapman and Hall/CRC.

[26] Muniswamaiah, M., Agerwala, T., & Tappert, C. C. (2023). IoT-based Big Data Storage Systems Challenges. In 2023 IEEE International Conference on Big Data (BigData) (pp. 6233–6235).



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

TOGETHER WE REACH THE GOAL

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY