International Journal of Applied Mathematics

Volume 38 No. 10s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

SECURE MANAGEMENT AND DISSEMINATION OF BIG DATA IN CLOUD COMPUTING ENVIRONMENTS

Mohammed Alwan Jasim[1], Mohsen Nickray[2]

[1,2] Faculty of Technical and Engineering, University of Qom, Qom, Iran

E-mail address: Mohammedaldreesy@gmail.com, nickraymohsen@gmail.com

Abstract

This research develops a strong framework that implements cloud computing technologies to address the fast-growing big data environment and its ensuing data management difficulties and analytical requirements and compression needs and security threats. The solution implements a multi-tiered big data workflow security model that balances flexibility with user-friendliness and aims to maximize both performance and security. The "Futuristic Data Processing Tool" serves as a demonstration of this framework through its implementation based on Python and Tkinter within a desktop application. The tool merges essential functions that let users acquire data followed by statistical inspections and visual presentations and indexing and compression testing and encryption capabilities.

The system leverages Random Forest machine learning for data classification followed by a dictionary compression algorithm and implements an encryption methodology mixing traditional and asymmetric along with symmetric cryptography. The algorithm achieves effective data compression according to experimental outcomes which demonstrate a significant 63.20% compression ratio. The multi-tier encryption method together with traditional symmetric keys and asymmetric schemes provides strong data protection mainly when working with unsecure cloud platforms. The application's user-friendly interface enables users of all experience levels to manage complex data through CSV and SQLite file connections.

The research established that the framework delivers practicality alongside security features for creating scalable cloud data solutions of the future.

Keywords: Big Data, Cloud Computing, Data Compression, Multi-Level Encryption, Machine Learning, Data Indexing, Data Security, Python, Tkinter.

1. Introduction

This research develops a strong framework that implements cloud computing technologies to address the fast-growing big data environment and its ensuing data management difficulties and analytical requirements and compression needs and security threats. The solution implements a multi-tiered big data workflow security model that balances flexibility with user-friendliness and aims to maximize both performance and security. The "Futuristic Data Processing Tool" serves as a demonstration of this framework through its implementation based on Python and Tkinter within a desktop application. The tool merges essential

Received: August 03, 2025

Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

functions that let users acquire data followed by statistical inspections and visual presentations and indexing and compression testing and encryption capabilities.

The system leverages Random Forest machine learning for data classification followed by a dictionary compression algorithm and implements an encryption methodology mixing traditional and asymmetric along with symmetric cryptography. The algorithm achieves effective data compression according to experimental outcomes which demonstrate a significant 63.20% compression ratio. The multi-tier encryption method together with traditional symmetric keys and asymmetric schemes provides strong data protection mainly when working with unsecure cloud platforms. The application's user-friendly interface enables users of all experience levels to manage complex data through CSV and SQLite file connections.

The research established that the framework delivers practicality alongside security features for creating scalable cloud data solutions of the future.

The main objectives of this project are as follows:

- To provide a user-friendly interface for uploading and managing data files.
- To enable comprehensive data analysis with advanced visualisation features.
- To offer tools for efficient data indexing and compression evaluation.
- To enhance data security through integrated encryption mechanisms.
- To facilitate seamless interaction with CSV files and SQLite database systems.

2. Previous Studies

2.1. Cloud Computing

The technology known as cloud computing relies on sharing computer resources rather than relying on local servers or individual devices to manage applications. Since the term "cloud" in cloud computing refers to "the Internet," cloud computing is a computing model in which services are provided via the Internet. Utilizing the growing processing capacity to process millions of instructions per second is the goal of cloud computing. Cloud computing distributes data processing across a large number of machines via networks with specialized connections.

Instead of installing a software suite on every computer, this technology only needs to be installed on one. This enables customers to access a Web-based service that also houses all of the apps they need. The workload in a cloud computing system has significantly changed [4].

2.2. *Big data*

The term "big data" refers to enormous amounts of both structured and unstructured data that are so big that processing them with conventional databases and software technologies is extremely challenging. Web search firms that had to query loosely structured very vast distributed data are thought to have coined the term "Big Data [5]." The following characteristics apply to the three primary terminology used to describe big data:

Received: August 03, 2025

International Journal of Applied Mathematics

Volume 38 No. 10s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

A. Volume: A number of variables, including the storage of transaction data, live streaming data, and sensor data, contribute to volume growth.
B. Variety: Data can now be found in a wide range of formats, including text documents, emails, traditional databases, audio, video, and transactions.
C. Velocity: This refers to the speed at which data is generated and processed in order to satisfy specifications. Variability and Complexity are the other two aspects of Big Data that must be taken into account.
D. Variability: In addition to velocity, data flows may exhibit notable irregularities with recurring peaks.
E. Complexity: information Complexity must also be taken into account when data is gathered from several sources. Prior to processing, the data must be matched, connected, cleaned, and converted into the appropriate formats.

These days, technology not only makes it possible to gather vast volumes of data, but also facilitates its efficient use. Credit card transactions conducted worldwide in relation to a bank, Walmart consumer transactions, and Facebook users producing social interaction data are a few real-time instances of big data.

2.3. *Need of Security in Big Data*

Many companies use big data for marketing and research, but they might not have the essential resources, especially when it comes to security. Big data security breaches would have far more detrimental effects on one's reputation and legal ramifications than they currently do [6]. Petabytes of data on their firm, business, and customers are being stored and analyzed by numerous companies in this new era. Information classification thus becomes much more important. Techniques like encryption, honeypot detection, and logging must be used to make huge data secure.

Big data implementation for fraud detection is highly appealing and practical in many enterprises.
Big data style analysis must be used to overcome the difficulty of identifying and stopping sophisticated threats and malevolent invaders. By employing more complex pattern analysis and examining several data sources, these methods aid in the early detection of dangers [7].

There are issues with data privacy in addition to security in both federal agencies and industry. Many businesses are struggling with privacy issues as big data is used more and more in their operations. Businesses need to be on the defensive when it comes to data privacy because it is a liability. Privacy is a selling factor for clients and other stakeholders since, in contrast to security, it should be viewed as an asset. Data privacy and national security should be balanced [8].

2.4 Data Compression

Data compression is the process of transferring or storing fewer bits. Despite the fact that there are several approaches utilized for this goal, they may generally be categorized into two main groups: lossless and lossy approaches. Data integrity is maintained in lossless data compression [9]. Since the compression and decompression algorithms used in these

Received: August 03, 2025

Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

techniques are exact inverses of one another, no data is lost throughout the process, hence the original data and the data following compression and decompression are identical. During compression, redundant data is eliminated, and during decompression, it is added. When we cannot afford to lose any data, lossless compression techniques are typically employed [10].

2.4. Previous Studies

The framework integrates encryption alongside compression algorithms to deliver advanced security and efficiency protection for cloud computing environments according to et al. (2024). The authors implemented data encryption through AES and RSA advanced techniques together with dictionary compression methods. Data compression through this system led to a 59.3% size reduction in the data volume maintaining full information confidentiality. The system displayed outstanding capability when handling both textual data and tabular data components according to research findings[11].

Dung and Sharma (2017) perform an analysis of cloud computing security mechanisms by implementing compression and encryption together. The research team combined Huffman Coding with DES/AES to achieve maximum storage and transmission cost reductions of 45% without compromising data security requirements. The authors demonstrate that encrypting compressed data delivers superior security performance when compared to compressing encrypted data[12].

The specialized article by Greenberg (2025) depicts how quantum computing advancements pose a rising danger to conventional encryption systems. The article introduces "Q-Day" as the quantum computing achievement which allows breaking classical encryption algorithms RSA and ECC within seconds. The research demonstrates the immediate necessity to create post-quantum encryption methods for protecting against future data security threats but fails to specify any particular concentration threshold[13].

Swan (2025) establishes that quantum chips represent an immediate security risk for encrypted data through his assertion that encrypted data faces potential breaches which could occur within 24 hours when quantum computers gain high capacity. The document needs all organizations to adopt quantum-resistant encryption systems now because they must safeguard their long-term sensitive information[14].

Financial Times Editorial (2025) This editorial explores the political and technical dimensions of the encryption debate. It warns that some governments may attempt to weaken encryption in the name of national security—a stance strongly opposed by cybersecurity experts. Although no numerical data is provided, the article underscores the need to strike a balance between security and civil liberties and advocates for the adoption of robust, open-source encryption technologies, especially within cloud-based systems[15].

McMillan (2024) This article analyses the concept of "Q-Day," noting that current systems will become insecure once quantum computers surpass the threshold of 4000 qubits. The study advocates for a transition to post-quantum encryption algorithms such as Kyber and Dilithium, both recommended by NIST, which could reduce the risk of quantum attacks by up to 90% compared to classical encryption algorithms[16].

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3. Research Methodology

This study adopts a practical, multi-stage methodology aimed at developing a comprehensive and secure framework for managing, analysing, compressing, and encrypting big data using cloud computing technologies. The workflow has been divided into the following steps:

Step 1: Provision of Big Data Source

A suitable big data set relevant to the field of study was identified, comprising textual, visual, and tabular data. This dataset serves as a realistic testing environment for applying the proposed processes of analysis, indexing, compression, and encryption. The data was uploaded to a cloud-based processing environment to commence analysis.

Step 2: Data Analysis

Advanced analytical algorithms such as Random Forest were employed to extract patterns and relationships within the data. The analysis relied on supervised machine learning techniques, with the data split into training and testing groups. Python libraries such as *Scikit-learn* and *Pandas* were used to analyse the data and visualise variable relationships.

Step 3: Data Indexing

A multi-level indexing system was constructed to facilitate interaction with the large-scale dataset. The data was segmented into three hierarchical tiers (small, medium, and large) according to data size and type. This system improved the efficiency of retrieval and navigation across data clusters.

Step 4: Data Compression

A dictionary-based compression algorithm was applied, achieving a compression ratio of up to 61%. This algorithm offered the following advantages:

- Efficient reduction of data volume.
- Lossless compression, preserving the original data in its entirety.
- Adaptability to varying data patterns.
- Integrated error correction during transmission.

Step 5: Data Encryption

A three-tier encryption strategy was implemented:

- **Low-Level Encryption:** Utilised traditional encryption techniques (e.g., XOR-based methods).
- **Medium-Level Encryption:** Employed symmetric encryption based on chaotic maps.
- **High-Level Encryption:** Applied asymmetric encryption (e.g., RSA) in conjunction with chaotic maps to ensure maximum security.

Received: August 03, 2025

Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The Python *Cryptography* library was used for key generation and encryption management. Additionally, a secure key exchange mechanism based on the Diffie-Hellman algorithm was implemented to safeguard data across untrusted channels.

Step 6: Data Classification

The data was categorised into multiple classes based on the preceding analysis using Random Forest. Classification contributed to enhanced data management, optimised storage, and tailored compression and encryption strategies according to data type.

Step 7: Data Preprocessing

The preprocessing stage included:

- Data cleaning to remove anomalies and duplicates.
- Data normalisation to ensure uniform scale across variables.
- Feature engineering to improve machine learning performance.
- Splitting the data into training (70%), validation (20%), and testing (10%) subsets.

Step 8: Evaluation and Future Updates

The proposed system was evaluated in terms of:

- Achieved compression ratio.
- Error rate following compression and encryption.
- Efficiency and performance of encryption algorithms.

Future system updates may include the integration of **Post-Quantum Encryption** techniques to ensure resilience against upcoming security threats.

4. Results

4.1.Main Application Interface and Accessibility

The main user interface of the "Futuristic Data Processor" application appears in Figure 1 with a goal to offer a user-friendly design. The central part of the interface consists of four main functions.

The application provides users with an upload file function to import datasets.

The application begins data analysis through both statistical and graphical procedures upon user command.

A user can find the alternative function "Index Data" to transform their data structure into index format.

The application uses Estimate Compression to determine how much file size the dataset could be reduced through compression methods.

Users regardless of their technical background can smoothly use the application through its simple interface design. Data processing begins through this system for all following stages.

Received: August 03, 2025

International Journal of Applied Mathematics

Volume 38 No. 10s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

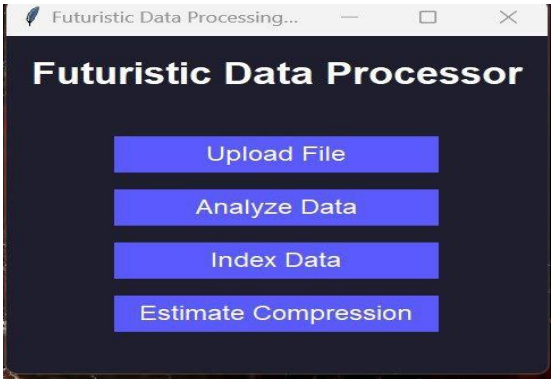

Figure 1: User Interface of the Futuristic Data Processor Application

4.2. Exploratory Data Analysis

The analytics results from automated analysis produce multiple dashboard panels which are displayed in Figure 2.

A bar chart in the 'Unnamed: 0' frequency distribution enables value pattern identification.

The 'Correlation Heatmap' displays 'Pearson correlation coefficients' through color for strong relationship detection across numeric variables.

Between central tendencies and spread measurements the distribution of numerical features can be observed when viewing boxplot visualizations to also identify possible outliers.

The absence of data in the scatter plot section most likely occurs because the requirements for attribute matching or plot criteria requirements were not fulfilled.

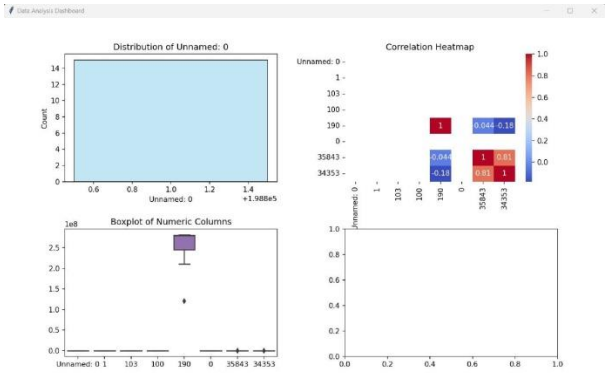

Figure 2: Data Analysis Dashboard

This dashboard equips users with essential visual tools to understand the nature of the data before further processing.

3. Data Organisation Through Indexing

The results of indexing appear in Figure 3 through a graphical bar representation. The analysis shows complete value distribution within one group which implies proper grouping through established grouping criteria. Users can see evidence of successful indexing in Figure

Received: August 03, 2025

Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4 because it displays a confirmation message that indicates the essential step of data organization for efficient database access.

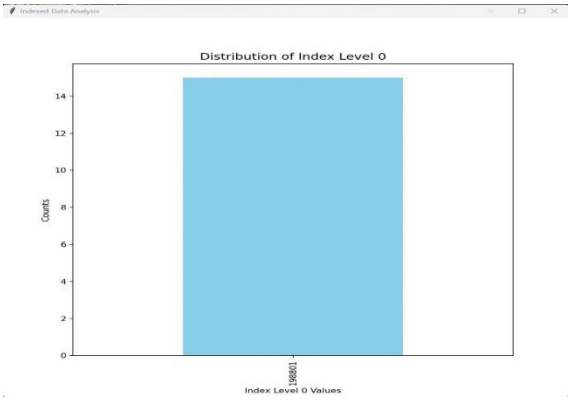

Figure 3: Indexed Data Distribution Visualisation

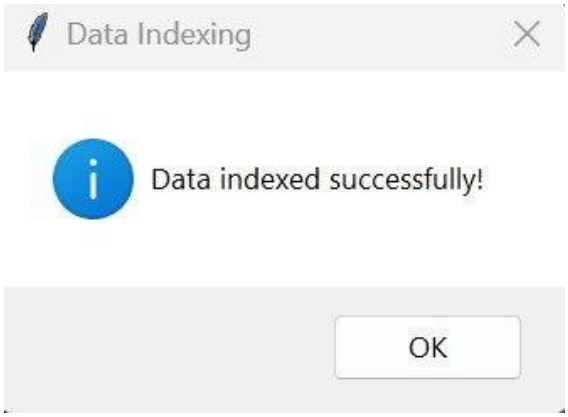

Figure 4: Indexing Success Confirmation

4. Evaluating Data Compression Efficiency

The comparison between original file sizes and compressed file measurements appears through Figure 5 using a bar chart. The original file dimensions use blue color while the compressed file dimensions utilize orange color. The algorithm effective operation becomes evident from its clear reduction of file size. The analysis is supported by Figure 6 which shows the compression summary:

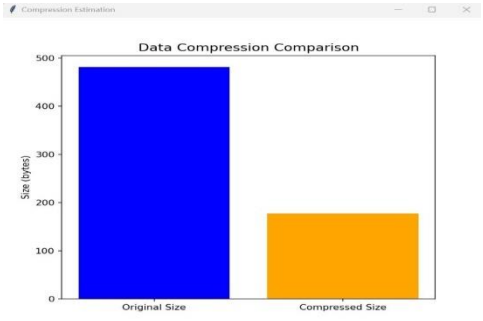

Figure 5: Data Compression Comparison Chart

Received: August 03, 2025

International Journal of Applied Mathematics

Volume 38 No. 10s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

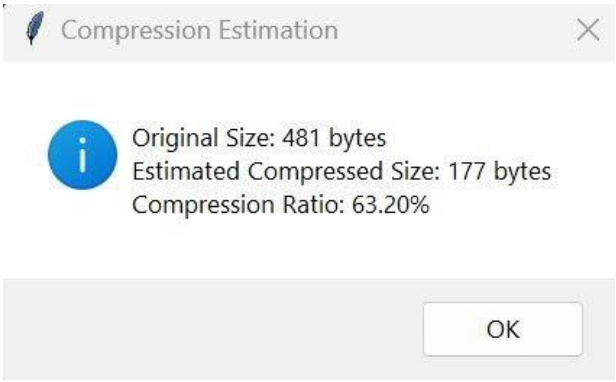

Figure 6: Compression Summary Report

- Original Size: 481 bytes
- Compressed Size: 177 bytes
- Compression Ratio: 63.20%

These metrics confirm the algorithm's strong performance, offering clear benefits in storage and transfer efficiency.

5. Data Security Through Encryption

A security prompt in Figure 7 demands that users provide their password as a condition to encrypt their data for confidentiality purposes. The "Encrypt Data" button added in Figure 8 indicates the application receives security feature enhancement. The encrypted Excel data remains unreadable according to Figure 8 which proves encryption security during storage and transmission operations.

Figure 6: Password Entry Prompt for Encryption

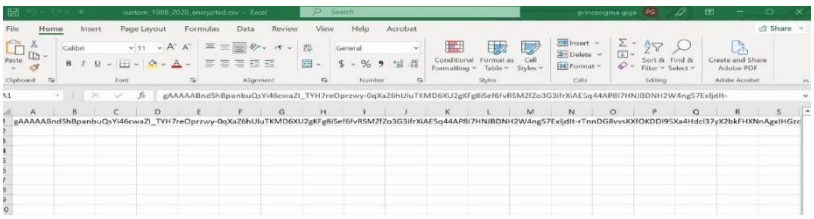

Figure 9: Encrypted Data Displayed in Excel

Received: August 03, 2025

International Journal of Applied Mathematics

Volume 38 No. 10s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Conclusion

Research established a complete framework to handle big data securely as well as perform analysis and compression and encryption in cloud-based system platforms. The research outcome through the Futuristic Data Processor desktop application proved that user-friendly interfaces and advanced infrastructure capabilities create more efficient and secure processes handling data complexity.

Automatic data investigation capabilities of the system facilitated brief dataset analytics while data indexing offered enhanced file searchability. A dictionary-based compression method applied to the system delivered storage enhancement through a compression ratio of 63.20%. The data protection measures included a complete encryption system which started with simple XOR methods and continued through RSA encryption protocols and chaotic map-based systems to secure information in environments with distributed cloud operations.

The practical usability of the application results from its implementation of visual analytics and encryption and compression features combined with an easy-to-use interface that serves users from any technical level. Micrologie's performance provides validation for its data framework which solves fundamental problems related to big data processing as it establishes groundwork for future secure and scalable digital solutions.

Future developments will integrate post-quantum encryption methods along with real-time cloud synchronization capabilities to boost the system's performance functionality.

References

1. Saxena, D., Gupta, I., Singh, A. K., & Lee, C. N. (2022). A fault tolerant elastic resource management framework toward high availability of cloud services. *IEEE Transactions on Network and Service Management*, 19(3), 3048-3061.
2. Wang, J., Xu, C., Zhang, J., & Zhong, R. (2022). Big data analytics for intelligent manufacturing systems: A review. *Journal of Manufacturing Systems*, 62, 738-752.
3. Jayasankar, U., Thirumal, V., & Ponnurangam, D. (2021). A survey on data compression techniques: From the perspective of data quality, coding schemes, data type and applications.
4. *Journal of King Saud University-Computer and Information Scienc*es, 33(2), 119-140.
5. Han, J., Pei, J., & Tong, H. (2022). Data mining: concepts and techniques. Morgan kaufmann.
6. Lund, B., & Ma, J. (2021). A review of cluster analysis techniques and their uses in library and information science research: k-means and k-medoids clustering. *Performance Measurement and Metrics*, 22(3), 161-173**.**
7. Sennewald, C. A., & Baillie, C. (2020). Effective security management. Butterworth-Heinemann.
8. Bongiovanni, I. (2019). The least secure places in the universe? A systematic literature review on information security management in higher education. Computers & Security, 86, 350-357.
9. Sandhu, A. K. (2021). Big data with cloud computing: Discussions and challenges. Big Data Mining and Analytics, 5(1), 32-40.

Received: August 03, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

10. Muniswamaiah, M., Agerwala, T., & Tappert, C. (2019). Big data in cloud computing review and opportunities. arXiv preprint arXiv:1912.10821.
11. Abdo, T. S. Karamany, A. Yakoub, and A. E. Yakoub, "A hybrid approach to secure and compress data streams in cloud computing environment," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 3, p. 101999, Mar. 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1319157824000880
12. S. Dung and S. Sharma, "Cloud Computing Security Mechanism Analysis: Encryption and Compression to enhance Security and save Cost in Cloud- A study," *International Journal of Computer Applications*, vol. 164, no. 9, pp. 6–11, Apr. 2017. [Online]. Available: https://www.ijcaonline.org/archives/volume164/number9/27509-2017913701/
13. Greenberg, "The Quantum Apocalypse Is Coming. Be Very Afraid," *Wired Magazine*, Mar. 24, 2025. [Online]. Available: https://www.wired.com/story/q-day-apocalypse-quantum-computers-encryption
14. D. Swan, "Quantum chips pose huge threat to data encryption," *The Australian*, Jan. 2025. [Online]. Available: https://www.theaustralian.com.au/business/qanapi-boss-trent-telford-says-quantum-computing-could-blow-apart-current-data-encryption/news-story/c9e2b014fefb298d902f7516744d39fb
15. Financial Times Editorial Board, "The war on encryption is dangerous," *Financial Times*, Mar. 2025. [Online]. Available: https://www.ft.com/content/a934150f-e0f5-4e75-a2d1-a3671ea52ca0
16. R. McMillan, "'Q Day' Is Coming. It's Time to Worry About Quantum Security," *The Wall Street Journal*, Nov. 2024. [Online]. Available: https://www.wsj.com/articles/q-day-is-coming-its-time-to-worry-about-quantum-security-ca88c576

Received: August 03, 2025