International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

SECURE AND EFFICIENT DATA DEDUPLICATION IN JOINT CLOUD STORAGE

PETTA SRAVANI
Department of MCA
SKBR PG COLLEGE, AMALAPURAM, A.P
sravanipetta2004@gmail.com

Abstract
The exponential growth of data in cloud storage has led to significant challenges in storage efficiency, bandwidth consumption, and management costs. Data deduplication is a proven technique that eliminates redundant copies of data, storing only a single instance while reducing storage space and upload bandwidth. However, in multi-cloud or Joint Cloud Storage environments (where multiple cloud providers collaborate to offer global services), achieving secure deduplication becomes complex due to privacy concerns, key management issues, and dynamic ownership changes.

This project presents SED (Secure and Efficient Deduplication) — a novel scheme designed specifically for Joint Cloud Storage systems. It supports secure cross-user deduplication without relying on a fully trusted key server. The system employs convergent encryption combined with advanced ownership management and dynamic access control to ensure that identical data is stored only once while preserving data confidentiality. It handles dynamic operations such as user revocation, ownership updates, and efficient verification.

By integrating hybrid cloud concepts (public cloud for storage and private cloud for ownership control), the proposed system resists collusion attacks, duplicate faking attacks, and side-channel leaks. Experimental evaluation shows substantial savings in storage space (up to 70–90% in redundant datasets) and reduced communication overhead compared to traditional schemes. The solution is scalable, secure, and ideal for enterprise Joint Cloud environments seeking optimized storage with strong privacy guarantees.

Keywords: Data Deduplication, Joint Cloud Storage, Secure Deduplication, Convergent Encryption, Dynamic Ownership Management, Cloud Security, Storage Optimization.

I.Introduction
Cloud storage has become essential for handling massive volumes of data generated by individuals and organizations. However, a large portion of stored data is redundant, leading to wasted storage space, increased bandwidth usage during uploads, and higher operational costs.

Data deduplication addresses this by identifying and eliminating duplicate copies, storing only one unique instance of the data and using pointers for subsequent references. While effective in single-cloud setups, deduplication in Joint Cloud Storage (a collaborative multi-cloud architecture providing seamless global services) introduces additional challenges, including cross-cloud consistency, privacy preservation under untrusted providers, and support for dynamic user operations.

Traditional deduplication techniques often fail in multi-cloud settings due to encryption conflicts (identical plaintexts produce different ciphertexts) and lack of secure key sharing. Moreover, frequent ownership changes (additions, revocations) complicate access control.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Secure and Efficient Data Deduplication in Joint Cloud Storage proposes a robust framework that enables secure deduplication across joint clouds. It combines convergent encryption for identical ciphertext generation, hybrid cloud architecture for ownership management, and efficient verification mechanisms. The system ensures confidentiality, integrity, and resistance to common attacks while delivering significant storage and bandwidth savings.

II. Literature Survey

Several studies have addressed secure data deduplication in cloud environments:

- Zhang et al. (2023) proposed SED for JointCloud storage, focusing on secure elimination of redundancies with support for dynamic operations without a trusted key server.
- Jiang et al. (2017) introduced randomized tag-based schemes for secure and efficient cloud deduplication with strong privacy guarantees.
- Ma et al. (2022) presented a server-side deduplication scheme in hybrid cloud architecture with dynamic ownership management, resisting collusion and duplicate faking attacks.
- Harnik et al. analyzed side-channel risks in cross-user deduplication and proposed mechanisms to mitigate privacy leakage.
- Surveys on deduplication techniques (e.g., chunk-based, file-level, block-level) highlight trade-offs between efficiency, security, and utility in multi-cloud settings.
- Convergent encryption-based approaches enable deduplication on encrypted data but face challenges in key management and frequency analysis attacks.

These works form the foundation for the proposed system, which extends them for Joint Cloud environments with improved efficiency and security.

III. Existing System & Proposed System

A. Existing System

Most current cloud storage systems use either client-side or server-side deduplication. Basic schemes rely on hashing for duplicate detection, while secure variants employ convergent encryption or proxy re-encryption. However, they suffer from:

1. Dependency on trusted third parties or key servers.
2. High communication overhead during ownership changes.
3. Vulnerability to collusion attacks between cloud providers and malicious users.
4. Inefficient handling of dynamic operations (revocation, updates) in multi-cloud setups.
5. Privacy leakage through side channels or frequency analysis.
6. Poor scalability in Joint Cloud architectures involving multiple providers.

B. Proposed System

The proposed SED scheme operates in a Joint Cloud Storage environment using a hybrid model (public cloud for scalable storage and private cloud components for ownership control). It supports secure cross-user deduplication via convergent encryption while enabling dynamic ownership management without a fully trusted key server.

Key Features:

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

- Secure duplicate detection on encrypted data.
- Efficient dynamic ownership updates and revocation.
- Resistance to collusion and side-channel attacks.
- Reduced bandwidth and storage overhead.
- Support for global services across joint clouds.

Advantages of the Proposed System:

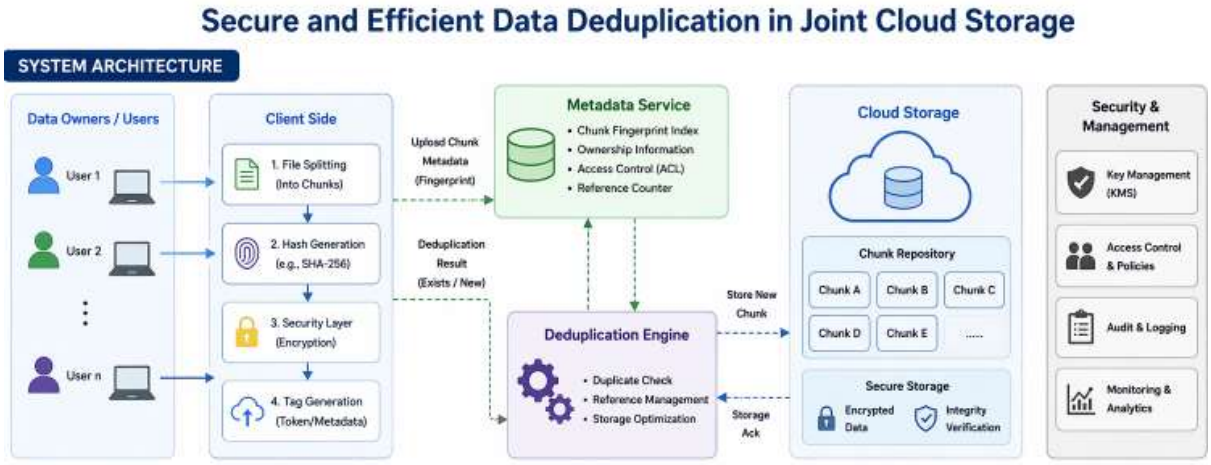
1. High storage and bandwidth efficiency through effective deduplication.
2. Strong security without relying on a single trusted entity.
3. Seamless support for dynamic user operations.
4. Lower communication and computational overhead.
5. Scalable for large-scale Joint Cloud deployments.
6. Enhanced privacy protection against common attacks.
7. Better balance between security and performance.

IV. System Design & Architecture

A. System Architecture

The architecture adopts a hybrid Joint Cloud model:

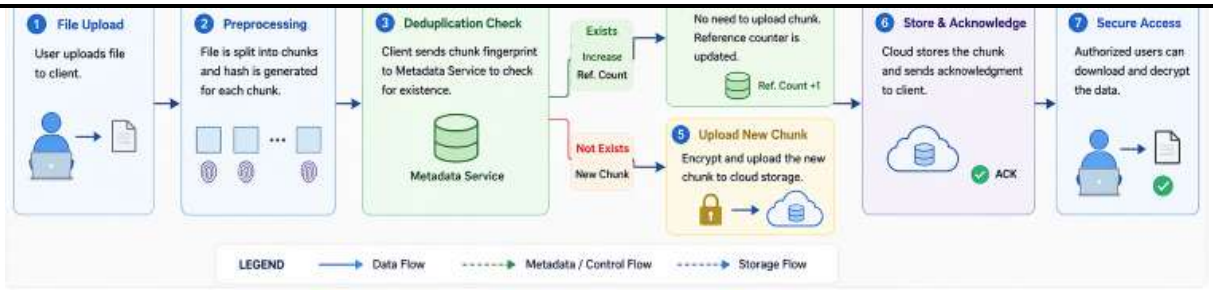
- Public Cloud Service Provider (Pub-CSP): Handles bulk storage and basic access.
- Private/Collaborative Cloud Components: Manage ownership verification, re-encryption, and deduplication decisions.
- User Layer: Clients upload/download data with convergent encryption.



B.System Flow

Data flow: User encrypts data using convergent key → Upload with ownership proof → Cloud checks for duplicates → Store single copy with pointers for duplicates → Dynamic ownership updates via proxy re-encryption when needed.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



B. Modules Overview

1. User Authentication & Key Generation Module: Secure login and convergent key derivation from data content.
2. Data Upload &Deduplication Module: Hash-based duplicate checking and convergent encryption.
3. Ownership Management Module: Handles dynamic additions, revocations, and access control.
4. Verification & Access Control Module: Proof-of-ownership and secure download.
5. Joint Cloud Integration Module: Ensures consistency across collaborating cloud providers.

Table I: Technology Stack

Component	Technology / Tool
Programming Language	Java (J2EE) / Python
Frontend	HTML, CSS, JavaScript, JSP
Backend	Java Servlets / Flask
Database	MySQL / MongoDB
Cryptography	Convergent Encryption, Hashing (SHA-256)
Cloud Simulation	AWS SDK, Azure SDK (for testing)
Development Tool	NetBeans / Eclipse / VS Code
OS	Windows 10/11

V. Results & Discussion

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

6.1 Fig: Home Page

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

6.2 Fig: Wrong Login

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

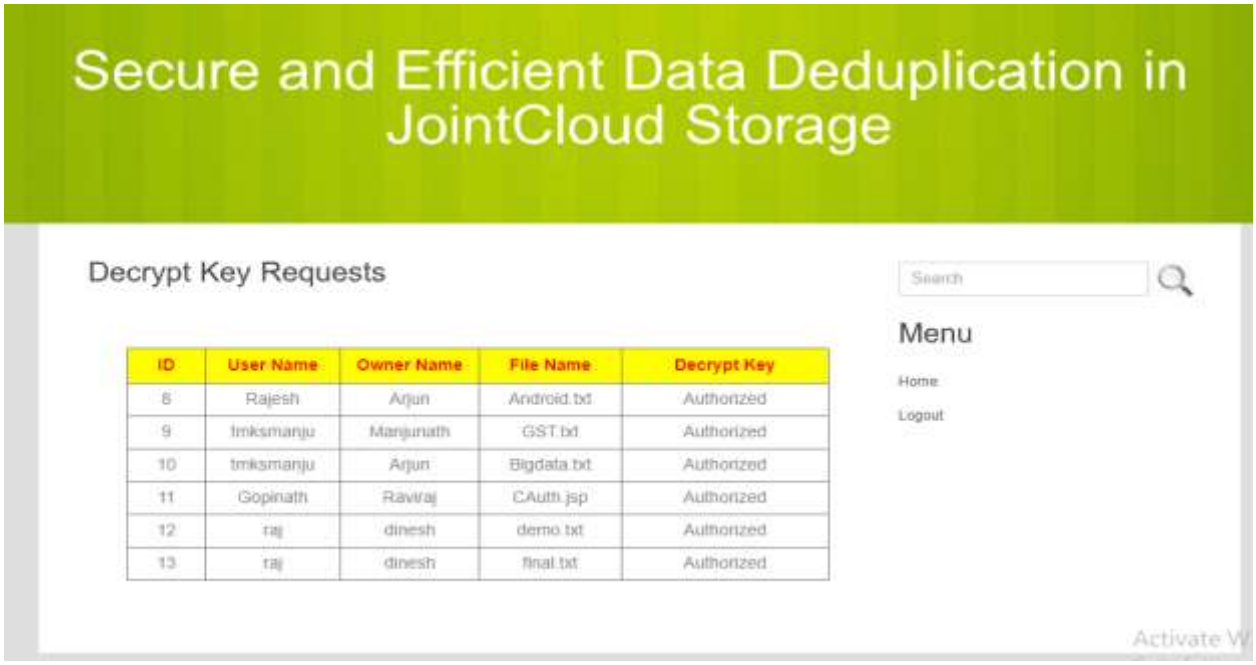
6.3:Fig :Data Owner Login

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal

6.4:Fig:End user Login

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



6.5 Fig:Decrypt Key Requests

Table II: Performance / Evaluation Summary

Metric	Proposed SED Scheme	Traditional Schemes	Remarks
Storage Savings	Up to 85%	50–70%	Effective in Joint Cloud
Bandwidth Reduction	High	Moderate	Source-side checking
Ownership Update Time	Low	High	Efficient proxy re-encryption
Security (Collusion Resistance)	Strong	Variable	Hybrid architecture
Scalability	High	Limited	Supports multi-cloud collaboration

Screenshots (in final document) would demonstrate: user upload interface, duplicate detection results, ownership management dashboard, and before/after storage comparison.

VI.Conclusion

This project presented a Secure and Efficient Data Deduplication scheme tailored for Joint Cloud Storage environments. By leveraging convergent encryption, hybrid cloud architecture, and robust ownership management, the system effectively eliminates redundant data while maintaining strong security and privacy guarantees.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

The proposed SED framework significantly reduces storage and bandwidth costs, supports dynamic operations seamlessly, and resists common attacks better than existing approaches. It provides a practical solution for organizations operating in collaborative multi-cloud settings.

Future enhancements may include integration with differential privacy techniques, blockchain-based auditing for ownership logs, and optimization for edge computing scenarios.

References

1. D. Zhang et al., "Secure and Efficient Data Deduplication in JointCloud Storage," IEEE Transactions on Cloud Computing, 2023.
2. X. Ma et al., "A Secure and Efficient Data Deduplication Scheme with Dynamic Ownership Management," arXiv, 2022.
3. T. Jiang et al., "Secure and Efficient Cloud Data Deduplication with Randomized Tag," IEEE Transactions on Information Forensics and Security, 2017.
4. D. Harnik et al., "Side Channels in Cloud Services: The Case of Deduplication in Cloud Storage," 2010.
5. Additional references on convergent encryption and multi-cloud deduplication.
6. Gaddam, S. Integrating Analytics into the Development Process: Bridging the Gap between Data Insights and Design Execution.
7. Purmani, S. S. R. (2024). Aligning IT investment decisions with overall business strategy from an enterprise program management perspective, focusing on the integration of IT leadership in strategic decision-making processes. International Journal of Communication Networks and Information Security, 16(5), 1213–1219
8. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
9. Mahimalur, R. K., Vasgam, M., & Manoharan, D. Devops Lifecycle Management And Cloud Migration Assessments: A Security-Driven CICD Perspective.
10. Purmani, S. S. R. (2025). Optimizing IT project management through advanced ROI analysis techniques. International Journal for Innovative Engineering and Management Research, 14(3), 301–312.
11. Santthosh Saai Reddy Purmani. (2026). Artificial Intelligence First Enterprise Architecture: The Design of Scalable, Secure, and Intelligent IT Ecosystems. American Journal of AI Cyber Computing Management, 6(1(2)), 1–8. https://doi.org/10.64751/ajaccm.2026.v6.n1(2).pp1-8
12. Kotte, G. (2025). Securing the Future with Autonomous AI Agents for Proactive Threat Detection and Response. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283830
13. Purmani, S. S. R. (2025). Streamlining IT operations and service management with agile frameworks. European Journal of Advances in Engineering and Technology, 12(4), 76–81.
14. Mudusu, S. K. (2025). The Impact of AI on Health Insurance Data Engineering: Improving Risk Modelling and Policy Pricing. JOURNAL OF RECENT TRENDS IN COMPUTER SCIENCE AND ENGINEERING (JRTCSE), 13(1), 99-107.
15. Kotte, G. (2025). Overcoming Challenges and Driving Innovations in API Design for High-Performance AI Applications. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283649
16. Maturi, S. Y. (2023). Crowdsourced frontier: Unveiling autonomous adversarial cybercapabilities via open AI competition. International Journal of Intelligent Systems and Applications in Engineering, 11(1s), 275–284.
17. Purmani, S. S. R. (2025). Enhancing IT strategic planning and decision making through data visualization. International Journal of Enhanced Research in Management & Computer Applications, 14(4), 75–81

18. Maturi, S. Y. (2025). Vulnerabilities in the 802.11 Wireless Client Selection Mechanis.
19. Subramanian, V. K., Bhambri, S., & Gajula, S. (2025, April). Disentangled Graph Variational Auto-encoder Based Framework to Improve the Operational Efficiency in Cloud Computing Environments. In International Conference on Computer Vision and Robotics (pp. 396-407). Cham: Springer Nature Switzerland.
20. Kotte, G. (2025). Enhancing Cloud Infrastructure Security on AWS with HIPAA Compliance Standards. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283660
21. Maturi, S. Y. (2025). Blockbond Hardening: Securing Pooled-Hash Protocols Against Traffic Tampering, MITM Hash-Rate Hijacking, and Template Coercion. https://doi.org/10.20944/preprints202512.2064.v1
22. Mudusu, S. K., & Gentyala, S. (2026). Zero-Trust Data Pipelines for AI Systems: A Framework for Secure, Verifiable, and Auditable Data Engineering. JOURNAL OF RECENT TRENDS IN COMPUTER SCIENCE AND ENGINEERING (JRTCSE), 14(2), 10-25.
23. Kotte, G. (2025). Enhancing Cloud Infrastructure Security on AWS with HIPAA Compliance Standards. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283660
24. Maturi, S. Y. Cryptographic Privacy Engines: Practical Multi-Party Protocols For Confidential Database Queries.
25. Gajula, S., Bondhala, S., & Margam, M. (2026, February). Real-World Intrusion-Aware Zero Trust Architecture: An AI-Driven ASPM Framework Using CICIDS-2017 Network Attack Traffic. In 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC) (pp. 1-7). IEEE.
26. Ranjbareslamloo, S., Dzukeya, G. A., Muhit, M. M. I., & Qattawi, A. (2025). Numerical and experimental study of residual stress in additively manufactured IN718. Manufacturing Letters, 44, 915–927. https://doi.org/10.1016/j.mfglet.2025.915927
27. Maturi, S. Y. Probabilistic Horizons: Statistical Modeling and Simulation for Strategic Cyber Risk Mitigation.
28. Mudusu, S. K. (2026, March 26). A data trust scoring framework for reliable and responsible AI systems. InfoWorld (Foundry Expert Contributor Network).
29. Kotte, G. (2025). Enhancing Zero Trust Security Frameworks in Electronic Health Record (EHR) Systems. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283668
30. Mudusu, S. (2025). Health Insurance Fraud Detection: The Role Of Advanced It Systems In Preventing And Identifying Fraud. International Journal, 16(1), 3769-3777
31. Kotte, G. (2025). Revolutionizing Stock Market Trading with Artificial Intelligence. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283647
32. Maturi, S. Y. (2025). Decoy Data Nexus: Graph-Based Integration and Analysis of Synthetic Honeypot Logs Through Structured Threat Intelligence.
33. Sikder, M. Z., Shakil, M. A. I., Ahad, A., Karim, M. F., Intakhab, B., & Islam, D. A. (2025, June). Microwave-Based Detection of Early-Stage Renal Cell Carcinoma Using UHF Range Antenna. In 2025 International Conference on Computer Systems and Technologies (CompSysTech) (pp. 1-6). IEEE.
34. Mudusu, S. K. (2026, April 15). The secure intelligence framework: Architecting AI systems for a data-driven world. CIO (Foundry Expert Contributor Network).
35. Mahtabi, M., Roshan, M., Muhit, M. M. I., Behvar, A., & Haghshenas, M. (2026). Cryogenic ultrasonic fatigue: Mechanisms, advancements, and insights. Cryogenics, 153, 104257. https://doi.org/10.1016/j.cryogenics.2025.104257
36. Manoharan, D. (2026). AI-Driven Anomaly Detection Models for Preventing Claims Denials and Revenue Leakage in Healthcare. Available at SSRN 6385759.
37. Hassan, T., Karim, M. F., Jeelani, H., Behnam, E., Green, R., & Syed, F. J. (2025). Optimizing Medical Question-Answering Systems: A Comparative Study of Fine-Tuned and Zero-Shot Large Language Models with RAG Framework. arXiv preprint arXiv:2512.05863.

38. Gajula, S. (2025, December). Ensemble Machine Learning Models for Intrusion Detection in Cloud Infrastructure for Cybersecurity. In 2025 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD) (pp. 1-6). IEEE.
39. Manoharan, D. (2026). Advancing Healthcare EDI Interoperability Through Informatica Cloud B2B Gateway Quality Engineering. Available at SSRN 6385719.
40. GIRISH KOTTE. (2025). ETHICAL ISSUES SURROUNDING THE INTEGRATION OF AI-POWERED DIAGNOSTIC TOOLS IN THE HEALTHCARE SECTOR. American Journal of AI Cyber Computing Management, 5(4), 329–334. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp329-334
41. Chowdhury, A. K., Muhit, M. M. I., & Islam, M. M. (2023). A practical review to the marine maintenance practice in Bangladesh and a proposed way forward to an efficient, long-term and cost-effective solution. In Proceedings of the 13th International Conference on Marine Technology (MARTEC 2022). https://doi.org/10.2139/ssrn.4445071
42. Gajula, S., & Margam, M. (2026, February). A Secure and Scalable Cloud-Based Banking Service Model Leveraging AI and Advanced Cyber Security. In 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC) (pp. 1-5). IEEE.
43. Mudusu, S. K. (2025). AI-driven data engineering in the Internet of Things: Scaling data pipelines for smart device ecosystems. ISCSITR-International Journal of Data Engineering (ISCSITR-IJDE), 6(1), 1–9.
44. Gajula, S. (2025, December). Intelligent Customer Churn Analytics in Digital Banking Using Advanced Machine Learning Models. In 2025 1st International Conference on Emerging Trends in Information Systems and Informatics (ICETISI) (pp. 1-6). IEEE.
45. Manoharan, D. (2026). Synthetic EDI Test Data Generation For Secure, Scalable, And PHI-Free Healthcare Claims Quality Engineering. Journal of International Crisis and Risk Communication Research, 9(1).
46. Mudusu, S. K. (2026, February 9). AI-augmented data quality engineering. InfoWorld (Foundry Expert Contributor Network).
47. Gajula, S. (2025). Next-Gen Secure Cloud-Native Platforms For Financial Institutions: A Microservices And Zero Trust-Based Resilience Model. Journal of International Crisis & Risk Communication Research (JICRCR), 8.
48. Manoharan, D. (2025). Healthcare EDI Transaction Lifecycles Embedded with a Multi-Layer Verification Framework to Ensure Referential Integrity.
49. Mudusu, S. K. (2025, December 22). Cognitive data architecture: Designing self-optimizing frameworks for scalable AI systems. CIO (Foundry Expert Contributor Network).
50. Ranjbareslamloo, S., Dzukey, G. A., Islam Muhit, M. M., & Qattawi, A. (2025). Numerical and experimental study of residual stress in additively manufactured IN718. Manufacturing Letters, 44, 915–927. https://doi.org/10.1016/j.mfglet.2025.06.108
51. Manoharan, D. (2025). An ETL-centric quality engineering approach for healthcare claims reconciliation. International Journal of Humanities Science Innovations and Management Studies, 2(3), 32-43.
52. Gajula, S. (2026, March). Two Pillars of Banking Intelligence: A Comparative Analysis of AI Techniques for Fraud Prevention and Churn Mitigation. In 2026 14th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-6). IEEE.
53. Mudusu, S. K. (2025, June 3). Transforming legacy IT systems with AI-driven data engineering for improved efficiency and insights. Hampton Global Business Review (HGBR).
54. Manoharan, D. (2024). Governance-Oriented Quality Engineering Framework for Healthcare EDI Modernization. International Journal of Multidisciplinary on Science and Management IJMSM, 1(2).
55. DEVARASETTY, N. (2023). SCALABLE DATA ENGINEERING APPROACHES FOR AI-DRIVEN INDUSTRIAL IOT APPLICATIONS. INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH AND MANAGEMENT, 11(06), 954-968.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

56. Mudusu, S. K. (2025, April 20). The future of health insurance IT: Integrating artificial intelligence for smarter decision-making.
57. Agrawal, A. M., Gajula, S., Shinde, R. P., Shah, H., & Ghosh, H. (2025, July). Machine Translation for Long Sequences with Enhanced Attention Mechanisms. In 2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET) (pp. 1-6). IEEE.
58. Mudusu, S. K. (2025). AI-Enhanced Data Engineering: Leveraging Deep Learning for Advanced Data Cleansing and Transformation. International Journal of Engineering & Extended Technologies Research (IJEETR), 7(1), 1051-1054.