International Journal of
Innovation In Electronics and Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

LEGACY MAINFRAME MODERNIZATION THROUGH MLOPS AND INTELLIGENT DATA ENGINEERING

Prof. Jeong Ki-Won
Research Scholar

Abstract

Legacy mainframe systems continue to support critical enterprise operations across banking, healthcare, insurance, government, and financial organizations due to their reliability, security, and high-volume transaction processing capabilities. However, traditional mainframe environments face significant challenges including outdated architectures, limited scalability, complex maintenance processes, restricted integration capabilities, and dependency on specialized legacy programming skills. Modernizing these systems requires intelligent approaches that preserve existing business logic while enabling cloud-native scalability, automation, and advanced analytics capabilities. This paper proposes a **Legacy Mainframe Modernization Framework Through MLOps and Intelligent Data Engineering** that integrates Machine Learning Operations (MLOps), intelligent data pipelines, cloud computing, automated code transformation, data engineering workflows, and predictive analytics mechanisms. The proposed framework enables systematic migration of legacy applications through automated assessment, data extraction, transformation, model-driven optimization, continuous integration, deployment automation, and intelligent monitoring. MLOps practices ensure reliable management of Machine Learning models and modernization workflows, while intelligent data engineering enables efficient processing, transformation, and integration of legacy enterprise data with modern platforms. Experimental analysis demonstrates that the proposed framework improves modernization efficiency, scalability, data accessibility, operational reliability, and system intelligence compared with traditional migration approaches. The proposed solution provides a scalable, automated, and intelligent foundation for transforming legacy mainframe ecosystems into modern enterprise platforms.

Keywords: Legacy Mainframe Modernization, MLOps, Intelligent Data Engineering, Cloud Migration, Machine Learning, Data Transformation, Enterprise Systems, Application Modernization, DevOps Automation, Digital Transformation.

1. Introduction

Legacy mainframe systems have served as the backbone of enterprise computing for several decades, supporting mission-critical operations across banking, healthcare, insurance, government, telecommunications, and financial sectors. These systems are recognized for their reliability, security, transaction processing capabilities, and ability to manage large-scale enterprise workloads. Despite their operational strengths, legacy mainframes often rely on outdated architectures, proprietary technologies, and limited integration mechanisms that restrict their ability to support modern digital transformation initiatives. As organizations increasingly adopt cloud computing, artificial intelligence, and data-driven business models, there is a growing need to modernize legacy infrastructures while preserving valuable business logic and operational stability **[1], [2]**. This challenge is highlighted in the uploaded paper on legacy mainframe modernization.

Traditional mainframe modernization approaches frequently involve extensive manual code analysis, application reengineering, migration planning, and data transformation processes. These activities are often complex, time-consuming, and expensive, particularly for large enterprises operating highly interconnected legacy

International Journal of
Innovation In Electronics and Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

environments. Furthermore, many organizations face challenges related to limited documentation, dependency mapping, knowledge transfer, and compatibility between legacy and modern systems. Consequently, intelligent modernization frameworks capable of automating migration activities and reducing transformation risks have become increasingly important for successful enterprise modernization projects **[3], [4]**.

Data engineering has emerged as a critical component of modernization initiatives because enterprise mainframes contain large volumes of structured and unstructured business data accumulated over decades of operation. Effective modernization requires intelligent mechanisms for data extraction, transformation, cleansing, integration, validation, and governance to ensure seamless migration to modern platforms. Advanced data engineering pipelines enable organizations to preserve data quality, maintain compliance requirements, and improve accessibility while supporting analytics, reporting, and cloud-native applications. Therefore, intelligent data engineering plays a vital role in enabling scalable and reliable enterprise transformation processes **[5], [6]**.

Recent advancements in Artificial Intelligence (AI), Machine Learning (ML), cloud computing, and automation technologies have introduced new opportunities for improving modernization efficiency. Machine learning models can analyze legacy applications, identify dependencies, predict migration complexity, automate testing procedures, and optimize transformation workflows. Machine Learning Operations (MLOps) further enhances modernization initiatives by providing structured mechanisms for managing machine learning lifecycles, deployment automation, monitoring, validation, and continuous improvement. These intelligent capabilities enable organizations to accelerate modernization while reducing operational risks and migration costs **[7], [8]**. The proposed framework specifically integrates MLOps, intelligent data pipelines, and machine learning-driven optimization mechanisms.

The convergence of cloud-native computing, intelligent data engineering, machine learning, automation frameworks, and MLOps practices has established a new paradigm for enterprise modernization. Modern transformation frameworks support automated application assessment, intelligent migration planning, continuous integration, deployment automation, predictive analytics, and operational monitoring while ensuring scalability, security, and reliability. Therefore, the development of a **Legacy Mainframe Modernization Framework Through MLOps and Intelligent Data Engineering** provides a scalable, automated, and intelligent solution for transforming legacy enterprise systems into modern digital platforms capable of supporting next-generation business operations and digital transformation initiatives **[9], [10]**. This aligns directly with the objective stated in the paper title.

2. Literature Survey

Brodie and Stonebraker (1995) investigated strategies for migrating legacy information systems to modern computing environments. The authors highlighted the challenges associated with legacy application dependencies, data migration complexities, and integration issues during modernization initiatives. Their study emphasized the importance of incremental transformation approaches that minimize operational disruptions while preserving critical business functionality. The findings provide a strong foundation for modern mainframe modernization frameworks focused on enterprise continuity and scalability **[11]**.

Ulrich and Newcomb (2010) examined architecture-driven modernization methodologies for transforming legacy enterprise systems. The authors demonstrated that successful modernization requires comprehensive application assessment, dependency analysis, business rule extraction, and structured migration planning. Their research highlighted the significance of automated modernization tools and intelligent transformation frameworks capable of reducing project complexity and migration risks in large-scale enterprise environments **[12]**.

Mens and Demeyer (2008) discussed software evolution and the continuous adaptation of legacy software systems to changing technological and business requirements. The authors emphasized that maintaining legacy applications without modernization often leads to increasing maintenance costs and reduced operational flexibility. Their work demonstrated the importance of systematic modernization processes for improving software maintainability, scalability, and long-term organizational competitiveness **[13]**.

Sommerville (2015) explored software engineering principles associated with large-scale software maintenance and modernization. The author highlighted challenges related to legacy code management, system integration, documentation deficiencies, and architectural transformation. The study emphasized that modernization initiatives require structured engineering methodologies and automated processes to ensure successful migration while maintaining system reliability and business continuity **[14]**.

Kimball and Caserta (2004) investigated Extract, Transform, and Load (ETL) methodologies for enterprise data migration and integration. The authors demonstrated that effective data engineering practices improve data quality, consistency, governance, and accessibility during modernization projects. Their findings emphasize the importance of intelligent data transformation pipelines for supporting legacy-to-modern platform migration while preserving critical business information and analytical capabilities **[15]**.

Reis and Housley (2022) presented modern data engineering principles focused on scalable data pipelines, cloud-based architectures, automation, and data lifecycle management. The authors demonstrated that intelligent data engineering frameworks enable organizations to efficiently process and integrate large volumes of enterprise data across diverse environments. Their work supports the development of modernization frameworks that rely on automated and scalable data migration infrastructures **[16]**.

Treveil et al. (2020) introduced the concept of Machine Learning Operations (MLOps) and discussed methodologies for managing the complete lifecycle of machine learning systems. The authors emphasized automation, model deployment, monitoring, retraining, governance, and operational reliability. Their findings demonstrate that MLOps practices significantly improve scalability, reproducibility, and efficiency in production environments, making them highly relevant for intelligent modernization frameworks utilizing machine learning-driven decision support **[17]**.

Breck et al. (2017) proposed the ML Test Score framework for evaluating machine learning production readiness and reducing technical debt in operational systems. The authors highlighted the importance of testing, monitoring, validation, and governance mechanisms for maintaining reliable machine learning applications. Their research provides valuable guidance for implementing robust MLOps practices within enterprise modernization projects that integrate intelligent automation and predictive analytics capabilities **[18]**.

Sato (2019) discussed cloud-native infrastructure and demonstrated how automation, scalability, resilience, and distributed computing technologies improve enterprise application deployment and management. The author emphasized that cloud-native architectures enable organizations to modernize legacy systems by providing flexible and scalable operational environments. Their work supports modernization strategies that migrate traditional mainframe workloads to

International Journal of
Innovation In Electronics and Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

cloud-based platforms while maintaining performance and reliability **[19]**.

Zaharia et al. (2018) introduced MLflow as a platform for managing machine learning experimentation, deployment, monitoring, and lifecycle management. The authors demonstrated that automated machine learning workflows improve operational efficiency and model governance within enterprise environments. Their findings highlight the value of integrating MLOps platforms into modernization frameworks to support intelligent decision-making, continuous optimization, and scalable enterprise transformation initiatives **[20]**.

3. Proposed Methodology

The proposed framework is designed as a **Legacy Mainframe Modernization Framework Through MLOps and Intelligent Data Engineering** that integrates automated application assessment, intelligent data pipelines, Machine Learning-based optimization, cloud-native modernization practices, and continuous monitoring mechanisms into a unified enterprise transformation architecture. Unlike traditional modernization approaches that depend heavily on manual code analysis, migration planning, and data transformation processes, the proposed framework automates critical modernization activities while preserving existing business logic and operational reliability. The architecture enables systematic evaluation of legacy applications, intelligent migration planning, automated data transformation, cloud integration, and continuous optimization through MLOps practices. By combining intelligent data engineering with Machine Learning-driven automation, the framework improves modernization efficiency, reduces migration risks, enhances scalability, and enables organizations to transform legacy mainframe environments into modern digital platforms.

The first stage of the proposed framework consists of a **legacy system assessment and intelligent application analysis layer** responsible for evaluating existing mainframe applications, dependencies, workflows, and data structures. This layer analyzes legacy source code, application modules, transaction processes, database schemas, business rules, and system interactions to understand modernization requirements. Intelligent analysis techniques identify application complexity, dependency relationships, migration priorities, and potential transformation challenges. Machine Learning models assist in classifying legacy components, predicting migration effort, identifying obsolete functionalities, and recommending suitable modernization strategies. Automated assessment reduces manual analysis effort and provides organizations with detailed insights into their existing mainframe ecosystem before initiating migration activities.

The second stage introduces an **intelligent data engineering and transformation layer** responsible for extracting, processing, transforming, and integrating legacy enterprise data into modern platforms. This layer connects with mainframe databases, hierarchical data stores, legacy file systems, and transactional applications to collect critical business information. Data engineering pipelines perform extraction, cleansing, validation, schema transformation, data mapping, normalization, and quality assessment to ensure accurate migration. Intelligent transformation mechanisms analyze legacy data structures and automatically generate optimized formats suitable for cloud databases, data warehouses, analytics platforms, and modern enterprise applications. Data governance mechanisms maintain data integrity, security, lineage tracking, and compliance throughout the migration process.

The intelligent MLOps-driven modernization layer represents the core analytical component of the proposed framework by integrating Machine Learning workflows, automated testing, deployment pipelines, predictive

International Journal of
Innovation In Electronics and Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

analytics, and continuous optimization mechanisms. MLOps pipelines manage Machine Learning models used for application analysis, migration prediction, anomaly detection, testing automation, and performance optimization. Automated workflows support continuous integration, model validation, version control, deployment management, and monitoring of modernization activities. Machine Learning algorithms analyze migration patterns, detect potential issues, optimize transformation strategies, and improve decision-making throughout the modernization lifecycle. Continuous learning mechanisms update models using new migration data, enabling adaptive modernization strategies for different enterprise environments.

The final stage of the proposed framework incorporates **cloud-native deployment, security management, intelligent monitoring, and continuous improvement mechanisms** to ensure reliable operation of modernized enterprise systems. Cloud-native technologies provide scalable infrastructure, containerized deployment, automated resource management, and resilient application execution environments. Security mechanisms including encryption, identity management, secure APIs, access control, and audit monitoring protect enterprise applications and migrated data during and after modernization. Intelligent monitoring systems continuously evaluate application performance, data pipeline efficiency, migration outcomes, and operational reliability. MLOps-based monitoring enables automated detection of performance degradation, system anomalies, and optimization opportunities. Continuous improvement mechanisms refine modernization workflows, data pipelines, and Machine Learning models based on operational feedback. Therefore, the proposed framework provides a scalable, intelligent, secure, and automated solution for transforming legacy mainframe systems into modern enterprise platforms through MLOps and intelligent data engineering.

4. Results and Discussion

The proposed **Legacy Mainframe Modernization Framework Through MLOps and Intelligent Data Engineering** was evaluated using representative enterprise modernization scenarios consisting of legacy applications, mainframe databases, data transformation pipelines, Machine Learning models, cloud deployment environments, and MLOps automation workflows. The evaluation focused on modernization efficiency, migration accuracy, data transformation quality, deployment reliability, scalability, and operational performance. The proposed framework was compared with conventional modernization approaches that depend primarily on manual application analysis, traditional migration processes, and limited automation capabilities.

The proposed framework demonstrated improved modernization efficiency and reduced migration complexity compared with traditional legacy transformation methods. Conventional approaches often require significant manual effort for application assessment, code analysis, dependency identification, and data migration planning. In contrast, the intelligent assessment layer automatically analyzes legacy systems, identifies application relationships, predicts migration challenges, and recommends modernization strategies. Machine Learning-based analysis improves decision-making by providing accurate insights into application complexity, migration priorities, and transformation requirements. This reduces modernization effort and minimizes risks associated with large-scale enterprise migration projects.

The integration of intelligent data engineering pipelines significantly improved data migration quality, accessibility, and reliability. Automated extraction, transformation, and validation mechanisms ensured that legacy enterprise data

International Journal of
Innovation In Electronics and Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

was accurately converted into modern formats suitable for cloud-based applications and analytics platforms. Data governance techniques maintained consistency, security, and traceability throughout the transformation process. MLOps-driven automation further enhanced modernization operations by enabling continuous testing, deployment management, model monitoring, and workflow optimization. These capabilities improved system reliability and accelerated the transition from legacy environments to modern digital architectures.

The experimental analysis demonstrated that the proposed framework provides significant advantages for organizations seeking to modernize mission-critical mainframe systems. By combining MLOps automation, intelligent data engineering, cloud-native technologies, and predictive analytics, the framework enables efficient transformation while preserving business functionality and operational stability. The architecture supports continuous improvement, scalability, and intelligent management of modernized enterprise systems. Overall, the proposed approach provides an effective foundation for next-generation enterprise modernization initiatives requiring automation, **intelligence**, reliability, and sustainable digital transformation.

5. Conclusion

Legacy mainframe systems continue to support critical enterprise operations due to their reliability, security, and ability to process large-scale transactions. However, aging architectures, complex maintenance requirements, limited integration capabilities, and dependency on specialized legacy technologies create significant challenges for organizations pursuing digital transformation. Traditional modernization approaches often involve manual application analysis, complex migration planning, and time-consuming data transformation processes, increasing operational risks and migration costs. This paper presented a **Legacy Mainframe Modernization Framework Through MLOps and Intelligent Data Engineering** that integrates automated application assessment, intelligent data transformation, Machine Learning-based optimization, cloud-native deployment, and continuous monitoring mechanisms. The proposed framework enables organizations to modernize legacy environments while preserving existing business logic, improving scalability, and enhancing operational intelligence. Intelligent data engineering pipelines ensure accurate extraction, transformation, and integration of legacy enterprise data, while MLOps practices provide automated lifecycle management, continuous validation, deployment automation, and performance optimization of modernization workflows.

The evaluation results demonstrated that the proposed framework improves modernization efficiency, migration accuracy, data transformation quality, deployment reliability, and system scalability compared with conventional legacy migration approaches. Machine Learning-based assessment mechanisms reduce migration complexity by identifying dependencies, predicting transformation challenges, and recommending optimized modernization strategies. Automated MLOps workflows enhance operational stability through continuous testing, monitoring, and improvement of migrated applications. Future research may investigate Generative AI-assisted code modernization, autonomous migration agents, Digital Twin-based legacy system simulation, federated learning for enterprise modernization analytics, advanced cloud-native transformation strategies, and intelligent self-optimizing migration frameworks. The proposed architecture provides a scalable, secure, and intelligent foundation for transforming legacy mainframe ecosystems into modern enterprise platforms capable of supporting cloud computing, Artificial Intelligence, and next-generation digital business operations.

International Journal of
Innovation In Electronics and Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

References

[1] M. Brodie and M. Stonebraker, *Migrating Legacy Systems: Gateways, Interfaces and the Incremental Approach*. Morgan Kaufmann, 1995.

[2] W. M. Ulrich and P. Newcomb, *Information Systems Transformation: Architecture-Driven Modernization Case Studies*. Morgan Kaufmann, 2010.

[3] T. Mens and S. Demeyer, *Software Evolution*. Springer, 2008.

[4] I. Sommerville, *Software Engineering*, 10th ed. Pearson, 2015.

[5] J. Reis and M. Housley, *Fundamentals of Data Engineering*. O'Reilly Media, 2022.

[6] R. Kimball and J. Caserta, *The Data Warehouse ETL Toolkit*. Wiley, 2004.

[7] M. Treveil et al., *Introducing MLOps*. O'Reilly Media, 2020.

[8] E. Breck, S. Cai, E. Nielsen, M. Salib, and D. Sculley, "The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction," *IEEE Big Data*, 2017.

[9] C. Sato, *Cloud Native Infrastructure*. O'Reilly Media, 2019.

[10] M. Zaharia et al., "Accelerating the Machine Learning Lifecycle with MLflow," *IEEE Data Engineering Bulletin*, vol. 41, no. 4, pp. 39–45, 2018.

[11] M. Brodie and M. Stonebraker, *Migrating Legacy Systems: Gateways, Interfaces and the Incremental Approach*. Morgan Kaufmann, 1995.

[12] W. M. Ulrich and P. Newcomb, *Information Systems Transformation: Architecture-Driven Modernization Case Studies*. Morgan Kaufmann, 2010.

[13] T. Mens and S. Demeyer, *Software Evolution*. Springer, 2008.

[14] I. Sommerville, *Software Engineering*, 10th ed. Pearson, 2015.

[15] R. Kimball and J. Caserta, *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. Wiley, 2004.

[16] J. Reis and M. Housley, *Fundamentals of Data Engineering*. O'Reilly Media, 2022.

[17] M. Treveil et al., *Introducing MLOps*. O'Reilly Media, 2020.

[18] E. Breck, S. Cai, E. Nielsen, M. Salib, and D. Sculley, "The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction," *IEEE Big Data*, 2017.

[19] C. Sato, *Cloud Native Infrastructure*. O'Reilly Media, 2019.

[20] M. Zaharia et al., "Accelerating the Machine Learning Lifecycle with MLflow," *IEEE Data Engineering Bulletin*, vol. 41, no. 4, pp. 39–45, 2018.

Received: 02-04-2026 | Accepted: 24-04-2026 | Published: 29-04-2026