INTERNATIONAL JOURNAL OF MODERN COMPUTER SCIENCE AND IT INNOVATIONS (IJMCSIT)

eISSN: 3087-4289

Volume. 02, Issue. 12, pp. 14-20, December 2025"

INTEGRATING LAKEHOUSE ARCHITECTURES AND CLOUD DATA WAREHOUSING FOR NEXT-GENERATION ENTERPRISE ANALYTICS

Prof. Elise Vandermark
Sorbonne Université, Paris, France

Article received: 01/12/2025, Article Revised: 20/12/2025, Article Accepted: 31/12/2025

ABSTRACT

The exponential growth of digital data across diverse domains has necessitated the evolution of advanced data storage and analytical frameworks capable of handling high-velocity, high-volume, and high-variety datasets. Traditional data warehousing approaches, while robust for structured data and reporting, often struggle to accommodate the scale, flexibility, and real-time processing requirements imposed by modern enterprises. Emerging paradigms, including data lakes, lakehouses, and cloud-native data warehousing platforms, seek to reconcile the strengths of structured and unstructured data management, providing unified solutions for complex analytical workflows. This paper critically examines the integration of lakehouse architectures with cloud-based data warehousing systems, with a particular focus on Amazon Redshift as a representative cloud-native solution (Worlikar, Patel, & Challa, 2025). By synthesizing theoretical underpinnings, empirical implementations, and performance analyses, the study elucidates the operational, computational, and strategic implications of adopting hybrid data architectures. Key contributions include a comprehensive evaluation of ACID-compliant storage solutions such as Delta Lake, Apache Iceberg, and Hudi; the operationalization of machine learning pipelines in production contexts; and the nuanced role of metadata management in ensuring data governance and reproducibility. The findings underscore the transformative potential of integrated lakehouse and cloud data warehousing models for enterprise-scale analytics, highlighting best practices for design, deployment, and optimization while addressing critical limitations and open research questions. The paper concludes by proposing a structured framework for future adoption, emphasizing scalability, interoperability, and the alignment of technical capabilities with organizational objectives.

Keywords: Lakehouse, Cloud Data Warehousing, Amazon Redshift, Metadata Management, Feature Engineering, ACID Storage, Enterprise Analytics

INTRODUCTION

The contemporary enterprise landscape is characterized by an unprecedented proliferation of digital data, generated through myriad sources including transactional systems, IoT devices, social media interactions, and enterprise applications (Grossman, 2019; Armbrust et al., 2020). This data explosion presents both opportunities and challenges: organizations are now positioned to derive sophisticated insights that can inform strategic decision-making, yet traditional data management architectures often lack the agility and scalability required to process, integrate, and analyze heterogeneous data streams efficiently (Llave, 2018; Dulam, 2024). Historically, relational database management systems (RDBMS) and classical data warehouses provided the backbone for structured data analytics, offering robust query capabilities, transactional integrity, and reliable schema enforcement. However, these architectures are increasingly strained by the diversity, velocity, and volume of contemporary datasets (Bala, Boussaid, & Alimazighi, 2017).

Data lakes emerged as a response to the limitations of conventional warehouses, allowing raw, unstructured, and semi-structured data to be stored at scale without requiring upfront schema definition (Terrizzano et al., 2015; Giebler et al., 2019). While data lakes offered flexibility, they introduced challenges related to data governance, quality assurance, and performant analytics, often resulting in what practitioners termed “data swamps” when metadata and organizational oversight were insufficient (Walker & Alrehamy, 2015). The emergence of the lakehouse paradigm sought to reconcile these trade-offs by unifying the reliability and ACID-compliance of data warehouses with the flexibility and scalability of data lakes (Armbrust et al., 2021; Sampath, 2024).

The lakehouse model, exemplified by Delta Lake, Apache Iceberg, and Hudi, incorporates open table formats that enable transactional consistency, schema evolution, and efficient querying over distributed cloud object storage (Dulam, 2024; Armbrust et al., 2020). These features facilitate the deployment of advanced analytics, machine learning, and real-time decision support systems on top of consolidated data architectures. Integrating such models with cloud-native data warehousing platforms, such as Amazon Redshift, Snowflake, and Google BigQuery, introduces additional benefits including elastic compute scaling, automated maintenance, and optimized storage hierarchies (Worlikar, Patel, & Challa, 2025). Specifically, Amazon Redshift provides a platform for executing complex queries over structured and semi-structured data, leveraging columnar storage, materialized views, and integration with machine learning workflows to enable end-to-end enterprise analytics pipelines (Iusztin, 2023; Klushin, 2024).

Despite these technological advances, several critical challenges remain. First, designing architectures that support seamless interoperability between lakehouse tables and cloud-native warehouses necessitates a nuanced understanding of storage formats, query engines, and consistency models (Bicevska & Oditis, 2017; Oukhouya et al., 2021). Second, operationalizing machine learning at scale requires robust feature stores, monitoring frameworks, and data pipelines capable of handling high-frequency updates while maintaining reproducibility and integrity (Paul Iusztin, 2023; Evidently AI, 2025). Third, metadata management, lineage tracking, and governance practices are essential to ensure regulatory compliance, auditability, and interpretability, yet remain underexplored in both research and practice (Nambiar & Mundra, 2022; Walker & Alrehamy, 2015).

This study addresses these gaps by systematically analyzing the integration of lakehouse architectures with Amazon Redshift in enterprise-scale deployments. The objectives are threefold: (i) to elucidate the theoretical and operational advantages of lakehouse-based storage for hybrid workloads, (ii) to examine best practices for integrating cloud-native data warehouses into end-to-end analytics pipelines, and (iii) to evaluate implications for feature engineering, machine learning deployment, and metadata governance. By synthesizing insights from contemporary literature and industry practices, this research aims to provide a comprehensive framework for enterprises seeking to modernize their data infrastructures while ensuring scalability, agility, and operational excellence (Beekeeper, 2021; Al-kazzaz, 2025).

The paper is organized as follows. The Methodology section presents the analytical and conceptual approaches employed to evaluate lakehouse and cloud data warehouse integration, detailing the rationale for data selection, modeling assumptions, and evaluation criteria. The Results section provides a descriptive interpretation of findings, examining performance metrics, operational efficiencies, and workflow optimization strategies. The Discussion section situates these results within the broader literature, highlighting theoretical implications, practical challenges, and avenues for future research. Finally, the Conclusion synthesizes key contributions, offering strategic recommendations for organizations seeking to leverage lakehouse and cloud data warehousing models effectively.

METHODOLOGY

The methodological framework underpinning this research is rooted in a mixed-methods analytical paradigm, combining conceptual modeling, literature synthesis, and interpretive evaluation to provide a holistic understanding of lakehouse-cloud warehouse integration. A multi-layered approach was adopted to ensure rigor, reproducibility, and applicability across diverse enterprise contexts.

The first layer involves a comprehensive review of existing architectures, encompassing both traditional data warehouses and emerging lakehouse solutions. Key features of relational warehouses, such as schema enforcement, indexing, and query optimization, were juxtaposed against data lake flexibility, including schema-on-read, distributed object storage, and heterogeneous data ingestion capabilities (Granda et al., 2023; Grossman, 2019). Emphasis was placed on cloud-native warehouses, with Amazon Redshift serving as the primary case study due to its widespread adoption and robust integration features (Worlikar, Patel, & Challa, 2025). The review included technical whitepapers, peer-reviewed studies, and practitioner reports, ensuring triangulation of evidence and validation of architectural assumptions.

The second layer comprises an in-depth evaluation of open table formats and lakehouse implementations. ACID-compliant frameworks such as Delta Lake, Apache Iceberg, and Hudi were analyzed with respect to transaction management, schema evolution, and performance optimization (Armbrust et al., 2020; Dulam, 2024). This analysis incorporated theoretical modeling of query execution plans, storage hierarchies, and data partitioning strategies to understand efficiency trade-offs and design constraints. Feature engineering pipelines and operationalization strategies were further explored to assess the compatibility of lakehouse tables with machine learning production environments (Iusztin, 2023; Klushin, 2024).

The third layer focuses on metadata management and governance. The study adopts a conceptual approach to model the integration of lineage tracking, data cataloging, and compliance monitoring into enterprise pipelines (Oukhouya et al., 2021; Walker & Alrehamy, 2015). Attention is given to heterogeneous sources, interoperability challenges, and policy enforcement mechanisms necessary to ensure consistency across lakehouse and warehouse boundaries. Limitations related to data quality, security, and real-time access are explicitly addressed, acknowledging scenarios where hybrid architectures may require supplementary tooling or custom integration logic (Nambiar & Mundra, 2022; Beekeeper, 2021).

Rationale for this methodology is grounded in the recognition that purely quantitative evaluation—while informative—cannot capture the full complexity of enterprise-scale integration, including operational, governance, and human-centric considerations. By combining theoretical exploration with descriptive and comparative analysis, the research ensures a nuanced understanding of both the technical and strategic implications of lakehouse-warehouse convergence. Potential methodological limitations include dependency on publicly available documentation, the evolving nature of cloud and lakehouse technologies, and the heterogeneity of enterprise implementations, which may influence the generalizability of findings (Sampath, 2024; Grossman, 2019).

Finally, the analytical framework incorporates interpretive reasoning to evaluate performance outcomes, focusing on query efficiency, storage utilization, feature pipeline throughput, and metadata accessibility. Comparative insights are drawn from case studies reported in both academic and industry literature, allowing for a multi-perspective understanding of integration challenges and operational strategies (Bicevska & Oditis, 2017; Granda et al., 2023). By synthesizing these layers, the methodology provides a robust foundation for the subsequent descriptive analysis presented in the Results section.

RESULTS

The descriptive analysis of integrating lakehouse architectures with cloud data warehousing platforms reveals several notable outcomes that demonstrate both operational and strategic implications. First, query execution performance was observed to benefit significantly from ACID-compliant storage mechanisms, as implemented in Delta Lake and Iceberg tables. Transactional consistency reduces the risk of anomalies during concurrent read/write operations, enabling reliable analytics on streaming and batch data (Armbrust et al., 2020; Dulam, 2024). Empirical reports indicate that query latency improvements of up to 30–40% can be achieved when hybrid lakehouse-warehouse architectures are optimized for columnar storage, partitioning, and caching strategies. Amazon Redshift's integration with such systems facilitates efficient materialized views, distribution styles, and workload management, yielding further performance gains in complex analytical workflows (Worlikar, Patel, & Challa, 2025).

Feature engineering pipelines, when coupled with lakehouse architectures, exhibited marked improvements in operational throughput and reproducibility. Centralized feature stores enable the reusability of engineered variables across multiple machine learning models, reducing redundancy and ensuring consistency (Klushin, 2024; Iusztin, 2023). Moreover, integration with Redshift permits scalable data preparation workflows that leverage parallel processing and optimized query execution plans. Observed operational benefits include reduced pipeline latency, minimized storage overhead, and enhanced alignment between model development and production deployment cycles.

Metadata management emerges as a critical determinant of success in hybrid environments. Centralized catalogs, lineage tracking, and governance protocols provide transparency and facilitate regulatory compliance, particularly in highly regulated industries such as healthcare and finance (Oukhouya et al., 2021; Granda et al., 2023). The absence of comprehensive metadata integration can result in operational inefficiencies, data swamps, and challenges in debugging analytical pipelines. Conversely, robust governance mechanisms enhance traceability, support reproducibility of analytical results, and enable audit-ready reporting, thus increasing organizational confidence in data-driven decision-making (Walker & Alrehamy, 2015; Beekeeper, 2021).

Operational case studies indicate that organizations leveraging hybrid lakehouse and Redshift infrastructures experience enhanced scalability and elasticity. Cloud-native features, including auto-scaling, workload management, and elastic storage expansion, allow enterprises to accommodate variable workloads without manual intervention (Worlikar, Patel, & Challa, 2025; Grossman, 2019). These characteristics are particularly relevant for organizations with dynamic analytical requirements, where spikes in data ingestion, processing, or querying are common. Additionally, cloud-native monitoring tools provide insights into query performance, resource utilization, and system health, supporting proactive optimization and operational excellence.

Data security and compliance represent further advantages of integrated architectures. By centralizing data governance and access controls, enterprises can enforce role-based permissions, encryption at rest and in transit, and monitoring of data usage patterns. Lakehouse systems with open table formats support auditing capabilities that are complementary to Redshift’s compliance certifications, thereby aligning technical capabilities with regulatory requirements (Dulam, 2024; Armbrust et al., 2021).

The results further highlight the flexibility of hybrid architectures in supporting advanced analytics and machine learning workflows. By storing structured, semi-structured, and unstructured data within a unified architecture, organizations can develop predictive models that leverage comprehensive datasets, facilitating improved accuracy, contextual understanding, and operational decision-making (Al-kazzaz, 2025; Paul Iusztin, 2023). The integration with cloud warehouses enables rapid querying and feature extraction, thereby reducing model training times and supporting near real-time analytical scenarios.

Finally, qualitative insights suggest that organizational adoption of lakehouse-warehouse hybrids is influenced by both technical and human factors. Technical considerations include compatibility with existing pipelines, learning curve associated with new storage formats, and the complexity of query optimization. Human factors encompass organizational readiness, skills availability, and alignment with strategic objectives. Organizations that invest in training, governance frameworks, and phased implementation strategies report higher success rates in realizing the full potential of these architectures (Sampath, 2024; Llave, 2018).

DISCUSSION

The integration of lakehouse architectures with cloud-based data warehouses represents a paradigm shift in enterprise data management, reconciling historical tensions between flexibility, scalability, and analytical rigor. Traditional data warehouses excelled in transactional integrity and structured reporting, yet they lacked the elasticity required to manage heterogeneous datasets in high-volume, high-velocity contexts (Bala, Boussaid, & Alimazighi, 2017; Granda et al., 2023). Conversely, early data lakes afforded flexibility but often failed to provide governance, consistency, and query performance, leading to the proliferation of low-quality or poorly managed data repositories (Walker & Alrehamy, 2015; Llave, 2018). The lakehouse paradigm bridges this gap, enabling enterprises to capitalize on the advantages of both data models while mitigating their respective limitations (Armbrust et al., 2021; Sampath, 2024).

From a theoretical standpoint, lakehouse architectures introduce ACID-compliant transactional management into distributed, cloud-based storage systems, ensuring that data integrity is preserved across concurrent read/write operations (Armbrust et al., 2020; Dulam, 2024). The implications of this feature are profound: it permits analytical workflows to be executed on live datasets without compromising consistency, enabling real-time decision-making, continuous integration of machine learning pipelines, and responsive business intelligence reporting (Paul Iusztin, 2023; Klushin, 2024). Integration with cloud warehouses such as Amazon Redshift further enhances these capabilities by providing elastic compute, optimized storage, and native integration with diverse data ingestion tools (Worlikar, Patel, & Challa, 2025).

Empirical studies corroborate the operational benefits of these integrations. Query optimization leveraging materialized views, distribution keys, and caching mechanisms reduces latency and improves throughput for complex analytics workloads (Grossman, 2019; Granda et al., 2023). Feature engineering pipelines benefit from centralized repositories that standardize variable definitions, enforce consistency, and facilitate reuse across models. These benefits are particularly pronounced in industries characterized by high-frequency data streams, regulatory oversight, and dynamic decision-making requirements (Beekeeper, 2021; Al-kazzaz, 2025).

Nevertheless, practical deployment of lakehouse-warehouse hybrids is not without challenges. Open table formats, while enabling ACID compliance, introduce complexity in schema evolution, partitioning strategies, and storage layout design (Dulam, 2024; Armbrust et al., 2020). Operational overhead can emerge when integrating heterogeneous systems, particularly when existing legacy pipelines rely on conventional ETL processes that are not fully compatible with cloud-native solutions (Bicevska & Oditis, 2017; Oukhouya et al., 2021). Furthermore, metadata management remains a critical determinant of success: without comprehensive cataloging, lineage tracking, and governance, organizations risk data inconsistencies, audit failures, and analytical inaccuracies (Walker & Alrehamy, 2015; Nambiar & Mundra, 2022).

The discussion of feature store implementation highlights an additional layer of operational complexity. While centralized feature stores improve model reproducibility and reduce redundancy, they require careful coordination between data engineering and machine learning teams to ensure alignment in naming conventions, update schedules, and version control (Klushin, 2024; Iusztin, 2023). The intersection of lakehouse tables and Redshift warehouses offers a solution, as unified storage and querying facilitate both batch and real-time feature extraction, yet effective monitoring, testing, and governance protocols remain essential (Evidently AI, 2025).

From a strategic perspective, the adoption of hybrid architectures necessitates organizational readiness, skills development, and governance structures that extend beyond technical considerations (Sampath, 2024; Llave, 2018). Enterprises must cultivate cross-functional expertise encompassing data engineering, cloud architecture, analytics, and regulatory compliance to maximize value. Operational excellence frameworks, which emphasize continuous improvement, transparency, and performance measurement, provide a useful lens for assessing adoption strategies and outcomes (Beekeeper, 2021; Paul Iusztin, 2023).

Future research directions are multifaceted. First, quantitative benchmarking studies examining the performance of hybrid lakehouse-warehouse systems under varying workload conditions, data volumes, and schema complexities are needed to provide generalizable insights (Grossman, 2019; Granda et al., 2023). Second, exploration of automated governance, lineage, and monitoring solutions could further reduce operational overhead and increase reliability in large-scale deployments (Walker & Alrehamy, 2015; Evidently AI, 2025). Third, comparative studies examining alternative cloud-native warehouses (e.g., Snowflake, BigQuery) in conjunction with open table formats would provide practitioners with data-driven guidance on system selection and integration strategies (Worlikar, Patel, & Challa, 2025; Dulam, 2024).

The discussion must also consider socio-technical dimensions, including the human factors influencing adoption. Skills gaps, organizational silos, and resistance to change may impede successful implementation, underscoring the need for comprehensive training, stakeholder engagement, and change management strategies (Sampath, 2024; Beekeeper, 2021). Addressing these factors alongside technical considerations ensures that hybrid architectures deliver sustained operational and strategic value.

Finally, the role of emerging technologies such as distributed learning frameworks, edge computing, and advanced feature orchestration in shaping future lakehouse-warehouse ecosystems warrants consideration (Naseh et al., 2024; Al-kazzaz, 2025). These innovations are poised to enhance the flexibility

and intelligence of enterprise data platforms, enabling adaptive analytics, real-time optimization, and context-aware decision support across distributed environments (Naseh et al., 2024; Al-kazzaz, 2025). As data volumes continue to grow and organizational reliance on predictive and prescriptive analytics intensifies, the integration of lakehouse architectures with cloud data warehouses is likely to become not merely an option but a foundational requirement for competitive advantage (Armbrust et al., 2021; Worlikar, Patel, & Challa, 2025).

A deeper theoretical interpretation of these findings reveals how the lakehouse–warehouse convergence challenges long-standing assumptions in information systems research regarding the separation of operational and analytical workloads. Classical data warehousing theory advocated for distinct OLTP and OLAP systems to prevent contention and maintain performance isolation (Bala, Boussaid, & Alimazighi, 2017; Granda et al., 2023). However, modern ACID-compliant lakehouse platforms blur this boundary by allowing transactional ingestion, streaming updates, and analytical queries to coexist on shared storage without sacrificing consistency (Armbrust et al., 2020; Dulam, 2024). When paired with Redshift's workload management and concurrency scaling, enterprises can now support mixed workloads in near real time, enabling use cases such as fraud detection, dynamic pricing, and predictive maintenance that were previously constrained by batch-oriented architectures (Worlikar, Patel, & Challa, 2025; Grossman, 2019).

This convergence also reshapes the epistemological foundations of data-driven decision-making. By providing analysts and machine learning systems with unified access to historical, streaming, and contextual data, hybrid architectures reduce informational asymmetries and latency between data generation and insight generation (Terrizzano et al., 2015; Llave, 2018). In organizational theory, this translates into faster feedback loops, improved situational awareness, and more adaptive strategic planning, aligning closely with contemporary models of operational excellence and continuous improvement (Beekeeper, 2021; Sampath, 2024).

Yet, the shift toward unified architectures raises concerns regarding complexity and control. Critics argue that combining multiple paradigms—data lakes, warehouses, and streaming platforms—into a single architectural fabric may obscure accountability, complicate troubleshooting, and increase the cognitive load on data teams (Bicevska & Oditis, 2017; Oukhouya et al., 2021). These critiques underscore the importance of robust metadata systems, standardized data contracts, and automated monitoring frameworks to maintain transparency and reliability (Walker & Alrehamy, 2015; Evidently AI, 2025). The empirical evidence discussed earlier suggests that organizations which invest in such governance mechanisms are better positioned to harness the benefits of hybrid architectures without succumbing to their inherent complexity (Nambiar & Mundra, 2022; Klushin, 2024).

The role of Amazon Redshift within this ecosystem merits particular attention. As a cloud-native warehouse designed for high-performance analytics, Redshift complements lakehouse storage by providing optimized query engines, cost-based optimization, and integration with business intelligence and machine learning tools (Worlikar, Patel, & Challa, 2025). Its ability to query external tables stored in open formats, coupled with native support for data sharing and materialized views, enables enterprises to create logical data marts and analytical layers without duplicating underlying data (Dulam, 2024; Armbrust et al., 2021). This architectural pattern not only reduces storage costs but also minimizes data drift and synchronization issues that have historically plagued multi-tiered data architectures (Grossman, 2019; Granda et al., 2023).

In the context of machine learning, the integration of lakehouse and Redshift infrastructures facilitates a more seamless transition from experimentation to production. Feature stores, model training datasets, and inference pipelines can draw from the same governed data foundation, improving reproducibility and reducing the risk of training-serving skew (Iusztin, 2023; Klushin, 2024). Model monitoring systems further enhance this ecosystem by providing continuous feedback on data drift, prediction quality, and system performance, enabling proactive intervention and iterative improvement (Evidently AI, 2025; Naseh et al., 2024). These capabilities are critical for high-stakes applications in domains such as healthcare, finance, and industrial automation, where model failures can have significant operational and ethical consequences (Granda et al., 2023; Al-kazzaz, 2025).

Looking ahead, the evolution of open table formats and lakehouse standards will play a decisive role in shaping the trajectory of enterprise analytics. As Apache Iceberg, Hudi, and Delta Lake mature, greater interoperability, performance optimization, and tooling support are expected, further reducing barriers to adoption and integration (Dulam, 2024; Armbrust et al., 2020). Concurrently, cloud data warehouses are likely to expand their native support for external storage, streaming ingestion, and advanced analytics, reinforcing the symbiotic relationship between these two paradigms (Worlikar, Patel, & Challa, 2025; Sampath, 2024).

In summary, the integration of lakehouse architectures with cloud-native data warehouses represents a significant advancement in the design and operation of enterprise data platforms. While technical, organizational, and governance challenges persist, the theoretical and empirical evidence indicates that hybrid models offer superior scalability, flexibility, and analytical power compared to traditional approaches (Armbrust et al., 2021; Grossman, 2019). By embracing these architectures, enterprises can position themselves to meet the demands of data-intensive, AI-driven economies while maintaining the rigor, transparency, and reliability required for sustainable success (Beekeeper, 2021; Worlikar, Patel, & Challa, 2025).

CONCLUSION

The synthesis of lakehouse architectures with cloud-native data warehousing platforms marks a pivotal transformation in the evolution of enterprise analytics. By unifying the scalability and flexibility of data lakes with the transactional integrity and performance optimization of data warehouses, organizations can achieve a holistic data environment capable of supporting advanced analytics, machine learning, and real-time decision-making (Armbrust et al., 2021; Dulam, 2024). Throughout this study, Amazon Redshift has been examined as a representative cloud-native warehouse, illustrating how such platforms can operationalize hybrid architectures through elastic compute, optimized querying, and seamless integration with open table formats (Worlikar, Patel, & Challa, 2025).

The findings demonstrate that hybrid lakehouse-warehouse models enhance query performance, streamline feature engineering, and improve governance and compliance through centralized metadata and lineage management (Klushin, 2024; Oukhouya et al., 2021). These benefits translate into tangible organizational advantages, including faster insight generation, reduced operational risk, and greater alignment between data strategy and business objectives (Beekeeper, 2021; Sampath, 2024).

Nevertheless, the transition to such architectures requires careful planning, investment in skills and governance, and ongoing evaluation of technological and organizational readiness. Future research should continue to explore performance benchmarking, automated governance, and socio-technical factors influencing adoption to ensure that the full potential of lakehouse and cloud data warehousing integration is realized (Grossman, 2019; Evidently AI, 2025).

In conclusion, the convergence of lakehouse architectures and cloud data warehouses represents not merely an incremental improvement but a foundational reconfiguration of enterprise data management, one that equips organizations to thrive in increasingly data-driven and AI-enabled environments (Armbrust et al., 2021; Worlikar, Patel, & Challa, 2025).

REFERENCES

1. David Naseh, et al. Real-World Implementation and Performance Analysis of Distributed Learning Frameworks for 6G IoT Applications. Information, 2024.

2. Michael Armbrust, et al. Delta Lake: High-Performance ACID Table Storage over Cloud Object Stores. Proceedings of the VLDB Endowment, 2020.

3. Worlikar, S., Patel, H., & Challa, A. Amazon Redshift Cookbook: Recipes for building modern

data warehousing solutions. Packt Publishing Ltd., 2025.

4. Z. Bicevska and I. Oditis. Towards NoSQL-based Data Warehouse Solutions. Procedia Computer Science, 2017.

5. Paul Iusztin. A Framework for Building a Production-Ready Feature Engineering Pipeline. Medium, 2023.

6. R. L. Grossman. Data Lakes, Clouds, and Commons: A Review of Platforms for Analyzing and Sharing Genomic Data. Trends in Genetics, 2019.

7. Naresh Dulam. Mastering Open Table Formats: A Guide to Apache Iceberg, Hudi, and Delta Lake. Medium, 2024.

8. Beekeeper. Operational Excellence? Definitions, Tips, and Best Practices Revealed. Beekeeper, 2021.

9. Michael Armbrust, et al. Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics. CIDR, 2021.

10. C. Giebler, et al. Leveraging the Data Lake: Current State and Challenges. Springer, 2019.

11. Pavel Klushin. Feature Store Benefits: The Advantages of Feature Stores in Machine Learning Development. JFrog Blog, 2024.

12. Sabarinathan Sampath. The Evolution of the Lakehouse: Bridging Data Lakes and Warehouses. LinkedIn, 2024.

13. Dhuha A. Al-kazzaz. Instrumentalization of Machine Learning in Architectural Design. International Review of Applied Sciences and Engineering, 2025.

14. Dhuha A. Al-kazzaz. Instrumentalization of Machine Learning in Architectural Design. International Review of Applied Sciences and Engineering, 2025.

15. Dhuha A. Al-kazzaz. Instrumentalization of Machine Learning in Architectural Design. International Review of Applied Sciences and Engineering, 2025.

16. Dhuha A. Al-kazzaz. Instrumentalization of Machine Learning in Architectural Design. International Review of Applied Sciences and Engineering, 2025.