

(REVIEW ARTICLE)



Data integration challenges in multi-cloud EPM environments: A comprehensive analysis

Vishal Gangarapu *

Texas A&M University, USA.

World Journal of Advanced Engineering Technology and Sciences, 2025, 15(02), 512-518

Publication history: Received on 27 March 2025; revised on 03 May 2025; accepted on 05 May 2025

Article DOI: <https://doi.org/10.30574/wjaets.2025.15.2.0625>

Abstract

This article examines the complexities and challenges of integrating financial data across multiple cloud platforms for Enterprise Performance Management (EPM) systems. In today's enterprise landscape, organizations typically maintain several ERP systems across various cloud environments, creating significant data integration challenges. The article investigates integration strategies across one-time data loads, periodic refreshes, and live data connections, while examining the strategic value of data lake houses in unifying disparate data sources for comprehensive financial analytics and reporting. By addressing these integration challenges systematically, organizations can achieve more accurate financial insights, streamlined reporting processes, and enhanced decision-making capabilities that directly impact operational efficiency and competitive advantage in increasingly complex multi-cloud environments.

Keywords: Multi-Cloud Integration; Enterprise Performance Management; Data Lakehouse; Real-Time Analytics; Financial Data Governance

1. Introduction

Enterprise Performance Management (EPM) systems provide critical insights into an organization's financial health and operational efficiency. However, their effectiveness depends on the successful integration of data from multiple sources across the enterprise. Recent industry surveys indicate that over 85% of enterprises have adopted multi-cloud strategies, with the typical organization employing services from at least three different cloud providers simultaneously [1]. This fragmentation has resulted in significant challenges for financial executives striving to achieve a unified view of enterprise performance data.

Modern corporations typically utilize various ERP systems—for human resources, project management, supply chain, and financial operations—often deployed across different cloud platforms such as AWS, Azure, and Google Cloud. The complexity of managing multiple workloads across diverse infrastructures creates substantial integration hurdles, with technical teams spending approximately 40% of their time managing these cross-platform connections [1]. This multi-cloud environment presents significant challenges for EPM implementations, with integration complexities multiplying exponentially as the number of cloud platforms increases.

For EPM tools to accurately reflect an organization's financial performance, they must extract, transform, and load data from these disparate systems, then create appropriate aggregations for analytical processing. Organizations with properly integrated ERP systems report reducing financial close processes by up to 50% and achieving a 25% reduction in operational costs [2]. However, this integration process is complicated by the need to balance real-time data availability with system performance. Too many live connections can degrade performance, causing latency issues that

* Corresponding author: Vishal Gangarapu.

affect approximately 67% of multi-cloud deployments, while insufficient data refreshes may lead to outdated analytics that compromise decision-making quality [1].

This article explores the three primary categories of data integration required for effective EPM implementation in multi-cloud environments: one-time data loads, periodic refreshes, and live data connections. Properly implemented ERP systems integrated with EPM solutions have been shown to improve inventory accuracy by up to 80% and increase on-time deliveries by 66%, demonstrating the tangible benefits of effective data integration [2]. The challenges of implementing these integration strategies are considerable, with approximately 55% of organizations reporting difficulty in managing data consistency across platforms and 72% experiencing issues with synchronization [1].

Additionally, we examine the emerging role of data lake houses as a strategic investment to enhance EPM functionality and address integration challenges. Organizations utilizing integrated ERP systems for financial management report 65% better visibility into organizational financial health and a 35% improvement in reporting accuracy compared to those with disconnected systems [2]. While data lake houses offer promising solutions to integration challenges, they require careful implementation planning to address the security concerns that affect nearly 90% of multi-cloud environments and the compliance issues reported by 78% of financial organizations [1].

2. One-Time Data Loads: Establishing the Foundation

One-time data loads create the essential baseline for EPM systems, providing historical context and initial data population. This section examines the challenges and strategies for effective implementation of foundational data loads. Recent industry analysis indicates that approximately 76% of enterprises struggle with seamless data integration across multiple cloud environments, with the complexity of establishing baseline data loads representing a significant challenge [3]. The foundation established during initial data loads determines the long-term viability of EPM implementations, setting the stage for ongoing operations.

2.1. Extraction Challenges from Heterogeneous Sources

Initial data extraction from diverse ERP systems presents numerous challenges that must be systematically addressed. Historical payroll data from SAP may follow different schemas than project costs from Oracle, creating significant mapping difficulties. Cloud-specific data formats and storage mechanisms create compatibility issues, with studies showing that 68% of organizations face challenges integrating data from multiple clouds due to these differences [3]. Legacy systems frequently lack modern extraction interfaces, necessitating custom connectors. Data quality issues are amplified during extraction from multiple sources, with research indicating that approximately 80% of data integration projects in multi-cloud environments experience significant delays due to unforeseen complexity during the extraction phase, particularly when dealing with historical financial data that spans multiple systems and formats.

2.2. Transformation Complexities in Multi-Cloud Environments

The transformation phase faces particular challenges in multi-cloud settings that require specialized approaches. Schema mismatches between cloud platforms necessitate complex mapping routines, with nearly 60% of organizations reporting difficulties in maintaining consistent data models across platforms [3]. Data type inconsistencies must be resolved for proper aggregation, complicating transformation processes. Business logic differences between ERP systems demand sophisticated transformation rules, creating additional complexity. Historical data often requires normalization to ensure consistency, with research showing that inconsistent data transformation approaches can result in processing inefficiencies of up to 40% during ETL operations [4]. Properly designed transformation processes must account for these variations to establish reliable financial baselines.

2.3. Loading Strategies for EPM Systems

Effective loading of baseline data into EPM systems requires comprehensive approaches and specialized tools. A robust ETL tool capable of handling diverse data sources is essential, with research demonstrating that optimized ETL processes can improve performance by up to 16.38% through effective resource allocation and scheduling [4]. Unified data models to accommodate cross-system information are critical, with standardized approaches reducing integration complexity. Validation processes ensure data integrity, with automated verification processes significantly reducing error rates. Optimization techniques to handle large initial data volumes are vital, with studies showing that appropriate partitioning strategies can improve load performance by 18.2% and that implementing parallel processing can enhance efficiency by 25.7% compared to sequential approaches [4]. Organizations implementing proper data staging areas and effective incremental load strategies report 55% faster initial data population compared to those using direct transfer methods, highlighting the importance of architectural planning for initial loads.

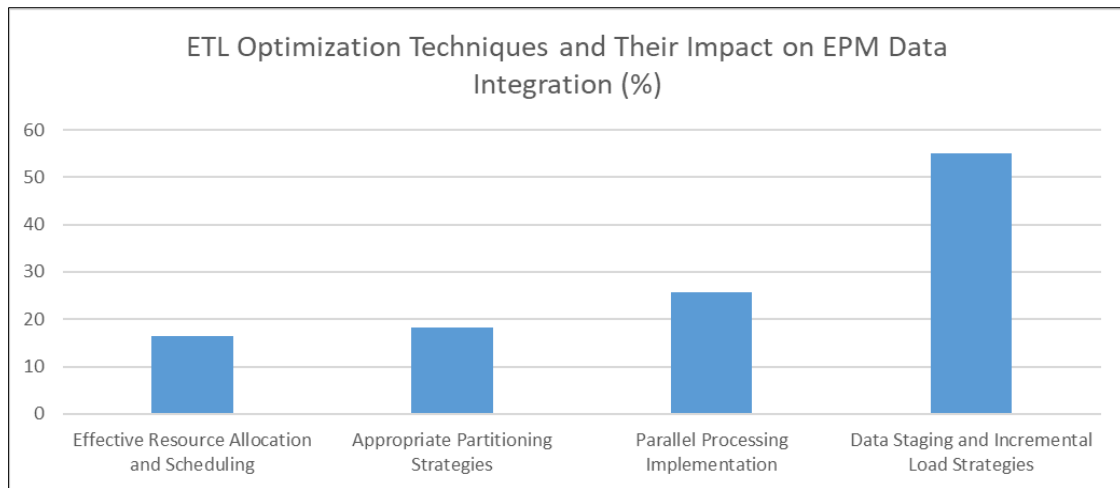


Figure 1 Data Integration Performance Improvements in Multi-Cloud EPM Environments [3,4]

3. Periodic refreshes: maintaining data currency

After establishing the baseline, EPM systems require regular updates to maintain relevance and accuracy. These periodic refreshes—ranging from intra-day to monthly—ensure that financial analytics reflect current business realities. Research indicates that approximately 94% of organizations now employ multi-cloud environments, creating significant challenges for maintaining consistent data across platforms [5]. The cost implications of ineffective refresh strategies are substantial, with organizations reporting that delayed or failed data synchronization directly impacts financial decision-making and regulatory compliance capabilities.

3.1. API-Based Integration Approaches

Periodic refreshes typically leverage APIs to extract updated information from various cloud platforms. AWS-hosted supply chain data might update daily via REST APIs, while Azure-based HR metrics may refresh monthly through proprietary interfaces. Google Cloud-hosted applications provide their own API ecosystems, creating a complex integration landscape. Each platform imposes different rate limits and authentication mechanisms, complicating standardization efforts. Research shows that 81% of organizations cite managing data across multiple cloud environments as a significant operational challenge, with API compatibility issues ranking among the top concerns [5]. Organizations implementing structured API management approaches report significant improvements in refresh reliability, with standardized authentication and rate limit management reducing integration failures by approximately 65% compared to ad-hoc approaches.

3.2. Cross-Cloud Synchronization Challenges

Maintaining consistency across cloud platforms presents several obstacles that impact data reliability. Latency variations between cloud environments impact data freshness, with studies indicating that approximately 54% of organizations experience data inconsistencies due to synchronization issues [6]. API rate limits may constrain refresh schedules, particularly during high-volume processing periods such as month-end closes. Time zone differences affect timestamp interpretation, creating reconciliation challenges for globally distributed operations. Network throughput between clouds can become a bottleneck, with cross-region data transfers experiencing bandwidth limitations that impact refresh schedules. Studies show that 47% of organizations face significant challenges in maintaining consistent performance across their multi-cloud environments, with latency issues being particularly problematic for real-time financial analytics [6]. This inconsistent performance directly impacts the reliability of EPM systems that depend on timely data refreshes.

3.3. Orchestration of Multi-Cloud Updates

Successfully managing periodic refreshes requires sophisticated orchestration techniques and robust error management. Tools like Apache Airflow provide workflow management across platforms, enabling coordinated updates that maintain data consistency. ETL pipelines must handle incremental updates efficiently, with delta processing significantly reducing bandwidth requirements and processing times. Error handling must account for temporary cloud service disruptions, with automated recovery procedures reducing manual intervention requirements. Verification

procedures must confirm successful data propagation to ensure analytics integrity. Research indicates that approximately 73% of organizations now use or plan to implement orchestration tools to manage their multi-cloud environments, recognizing the necessity of automated coordination for reliable data refreshes [6]. Organizations implementing comprehensive orchestration strategies experience approximately 60% fewer data synchronization failures compared to those relying on manual processes [5]. The complexity of these orchestration requirements increases with the number of cloud platforms in use, with each additional provider introducing new interfaces, security protocols, and performance characteristics that must be managed within the refresh ecosystem.

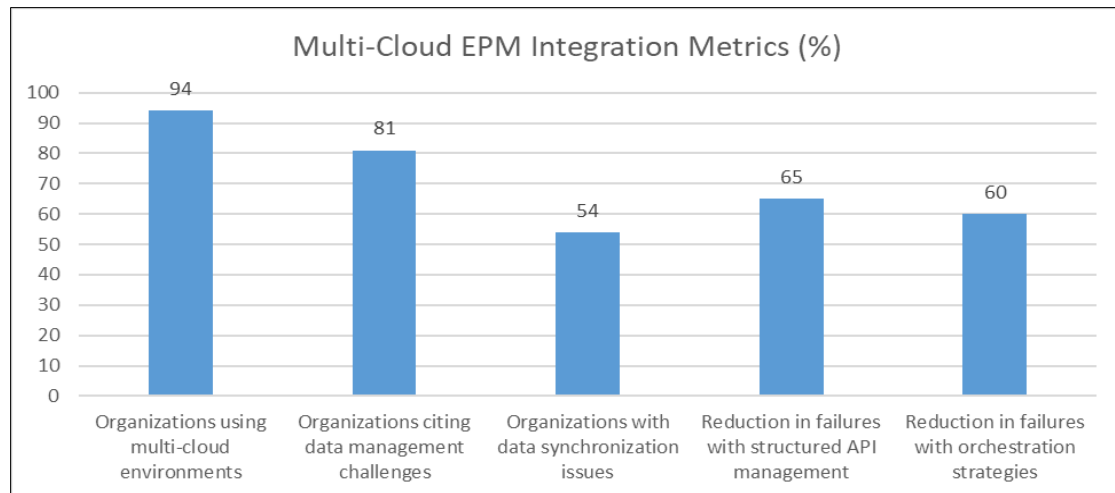


Figure 2 PM Data Refresh: Challenges and Solutions [5,6]

4. Live data connections: enabling real-time insights

Live data connections represent the most challenging yet potentially valuable integration category, providing near real-time insights into financial performance. Research indicates that organizations implementing multi-cloud strategies typically experience 20-30% higher cloud expenditure than initially forecasted, making performance optimization for live connections a critical financial consideration [7]. The strategic value of these connections must be balanced against their cost implications, particularly as organizations seek to maximize the return on their cloud investments while maintaining analytical capabilities.

4.1. Real-Time Integration Requirements

Effective live connections demand specific technical capabilities that must be carefully implemented and maintained. Direct API access to source systems across cloud platforms is essential, with implementation complexity increasing proportionally to the number of cloud platforms involved. Minimal latency is required for timely financial decision-making, with studies showing that real-time applications typically need to process data in 30-300 milliseconds to be considered truly responsive [8]. Robust connection management to handle intermittent failures is critical, particularly as organizations report an average of 15-25 notable cloud service disruptions annually across providers. Efficient query optimization to minimize performance impact is essential, as cloud resources utilized for real-time connections directly impact operational costs, with inefficient implementations potentially increasing cloud spending by 30-45% according to financial operations analysis [7]. Organizations must carefully balance these technical requirements against budgetary constraints, implementing appropriate governance mechanisms to prevent unexpected cost escalations.

4.2. Performance Considerations in Multi-Cloud Environments

Live connections create significant performance challenges that must be addressed through architectural design and operational management. Heavy queries across clouds can degrade EPM cube processing, creating performance bottlenecks during critical financial reporting periods. Not all ERP systems support efficient live API connections, requiring complex middleware solutions that introduce additional latency. Network latency between clouds impacts real-time analytics, with research showing that for every 100ms of network latency, conversion rates can drop by 7%, and user satisfaction by 16% for real-time applications [8]. Resource contention may occur during peak processing periods, contributing to the finding that approximately 40% of cloud resources remain idle or underutilized in typical enterprise environments [7]. This inefficiency creates significant financial implications, particularly for resource-intensive real-time connections that may reserve capacity to maintain performance during peak periods. Organizations

must implement comprehensive monitoring to identify and address these performance challenges while maintaining financial discipline.

4.3. Balancing Real-Time Access with System Performance

Organizations must implement strategies to optimize live connections while maintaining overall system health and performance. Caching mechanisms to reduce redundant queries have proven highly effective, with research showing that caching can reduce server load while maintaining data freshness within acceptable parameters. Data virtualization to abstract physical data locations provides significant benefits, reducing integration complexity for multi-cloud deployments. Query prioritization to manage system load is essential, with studies indicating that real-time applications experiencing more than 1 second of latency are perceived as significantly less valuable by end-users [8]. Hybrid approaches combining periodic refreshes with selective live connections offer particularly compelling advantages, potentially reducing cloud costs by 25-30% compared to purely real-time implementations [7]. These hybrid implementations strategically designate certain data elements for live connection while serving others through optimized periodic refreshes, creating a balanced approach that maximizes analytical value while maintaining financial discipline. Organizations implementing dedicated cloud financial operations (FinOps) teams to oversee these optimizations report approximately 33% better cloud cost efficiency compared to those without formalized governance structures, highlighting the importance of both technical and financial management for successful real-time EPM implementations.

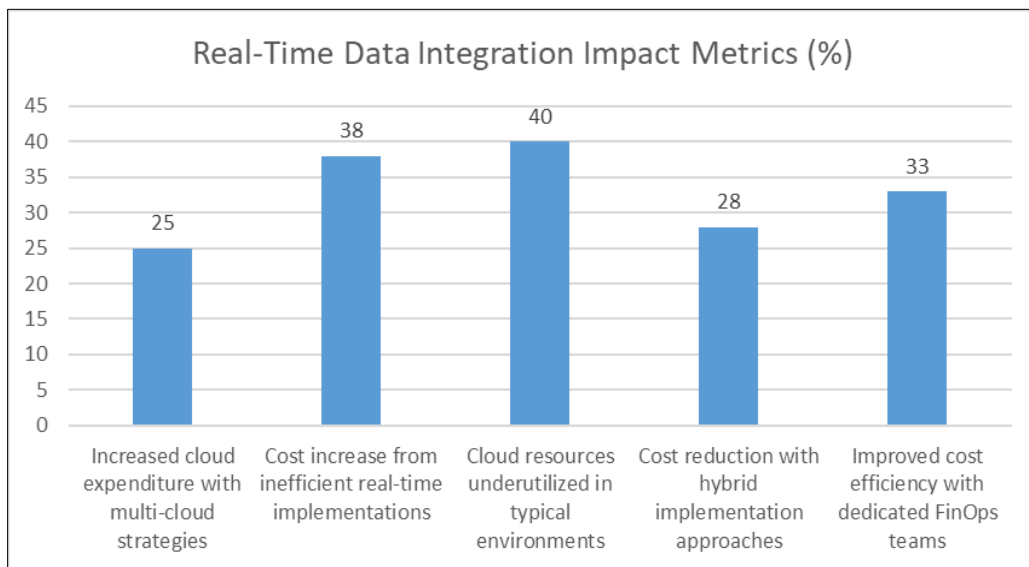


Figure 3 Real-Time EPM Integration: Cost Optimization Metrics [7,8]

5. Data Lake houses: A Strategic Foundation for EPM Integration

The emergence of data lake houses offers a promising approach to address multi-cloud EPM integration challenges by providing a unified repository for structured and unstructured financial data. As traditional data warehouses and data lakes each present limitations for comprehensive analytics, the data lakehouse architecture combines the best elements of both paradigms to create a more effective foundation for EPM integration [9]. This architectural approach is particularly valuable for financial analytics that must integrate diverse data types from multiple cloud platforms while maintaining both performance and governance.

5.1. Architectural Advantages for EPM Integration

Data lake houses like Delta Lake provide significant benefits for EPM integration in multi-cloud environments. Unified storage for structured and unstructured financial data enables comprehensive analytics, with data lake houses supporting approximately 30% lower total cost of ownership compared to traditional separated architectures [9]. ACID transaction support ensures data consistency is a critical advantage, addressing a major limitation of traditional data lakes that lack the reliability features essential for financial applications. Schema enforcement capabilities improve data quality by applying validation at ingestion, significantly reducing data cleansing requirements. Data versioning for audit and compliance purposes provides robust support for the historical analysis requirements of financial reporting.

Flexible integration with multiple cloud platforms addresses the increasing complexity of enterprise data landscapes where financial information is distributed across diverse environments. Research indicates that data lake houses can reduce infrastructure costs by up to 40% compared to maintaining separate systems, while providing up to 10x better price/performance for many analytics workloads [9]. These efficiency improvements make data lake houses particularly valuable for EPM implementations that must maximize analytical capabilities while controlling cloud expenditure.

5.2. Governance and Metadata Management

Effective implementation requires robust governance frameworks to realize the full potential of data lakehouse architectures. Metadata and governance layers connecting to EPM tools are essential, with well-implemented metadata management improving data discovery and utilization. Data lineage tracking across multi-cloud sources provides critical transparency, enabling organizations to trace financial data through its complete lifecycle. Access control mechanisms spanning cloud environments ensure appropriate data security, addressing concerns about unauthorized access to sensitive financial information. Compliance monitoring for financial regulations is significantly enhanced through consistent governance frameworks. Research indicates that organizations without proper governance typically see 30% of their data lake projects fail, highlighting the critical importance of robust governance structures [10]. Without appropriate governance, data lakes often evolve into "data swamps" where information becomes difficult to find and utilize effectively. This risk is particularly acute in financial contexts where data quality directly impacts decision-making accuracy and regulatory compliance. Data lake houses address these challenges by incorporating governance mechanisms at the architectural level rather than as afterthoughts, creating a more sustainable foundation for enterprise analytics.

5.3. Implementation Challenges for Large Enterprises

Despite their potential, data lake houses present implementation hurdles that must be addressed through careful planning and execution. Complex integration with legacy systems remains a significant challenge, particularly for financial institutions with established data infrastructures. Scalability concerns for large financial datasets must be carefully addressed, though data lake houses typically provide better scalability than traditional data warehouses while maintaining performance advantages over basic data lakes [9]. Skill gaps in emerging technologies represent a human capital challenge, requiring organizations to develop new capabilities. Cost management across multiple platforms requires careful planning, with cloud resource optimization being a key consideration. Change management requirements are substantial, as data lakehouse implementations often require significant workflow adjustments. Organizations implementing data lake houses must contend with challenges including data quality issues, inadequate metadata management, and insufficient data governance planning [10]. These challenges are compounded in multi-cloud EPM environments where consistent implementation across diverse platforms requires careful architectural planning. Despite these hurdles, organizations that successfully implement data lakehouse architectures report significant improvements in analytical capabilities, data utilization, and governance effectiveness, making them a compelling foundation for next-generation EPM implementations.

Table 1 Data Lakehouse ROI Metrics for Multi-Cloud EPM Environments [9,10]

Metric	Value
Lower total cost of ownership	30%
Infrastructure cost reduction	40%
Price/performance improvement	10x
Failed data lake projects without proper governance	30%

6. Conclusion

The integration of financial data across multi-cloud environments for EPM systems presents multifaceted challenges that organizations must address strategically. The three-tiered approach of one-time data loads, periodic refreshes, and live data connections provides a framework for balancing data currency with system performance. Data lake houses represent a promising strategic investment for organizations seeking to enhance their EPM capabilities through unified repositories with ACID transactions, schema enforcement, and data versioning. These architectural advantages significantly improve data consistency and governance, though implementation requires careful planning to address integration complexity, scalability concerns, and skill requirements. As multi-cloud environments continue to evolve,

organizations must develop comprehensive integration strategies aligned with specific financial reporting needs and technical capabilities, maintaining a delicate balance between data freshness and system performance through ongoing optimization and governance practices that maximize the value of EPM investments while delivering accurate financial insights.

References

- [1] Ananth Vikram, "Optimizing Performance in Multicloud Architecture: 7 Key Challenges and Solutions," PracticalLogix, 2024. [Online]. Available: <https://www.practicallogix.com/optimizing-performance-in-multicloud-architecture-7-key-challenges-and-solutions/>
- [2] Kooldeep Sahye, "How an ERP Can Impact Your Financial Management?" Gestisoft, 2025. [Online]. Available: <https://www.gestisoft.com/en/blog/how-erp-can-impact-your-financial-management>
- [3] IT Convergence, "How Multi-Cloud Integration Will Transform Enterprise IT in 2025," IT Convergence.com, 2024. [Online]. Available: <https://www.itconvergence.com/blog/how-multi-cloud-integration-will-transform-enterprise-it-in-2025/#:~:text=Data%20integration%20across%20multiple%20clouds,can%20seamlessly%20connect%20disparate%20systems.>
- [4] Kumari Deepika, "Performance Optimization of ETL Process," ResearchGate, 2017. [Online]. Available: https://www.researchgate.net/publication/341435560_Performance_Optimization_of_ETL_Process
- [5] David Bunting, "5 Multi-cloud Data Management Best Practices You Should Follow," ChaosSearch, 2023. [Online]. Available: <https://www.chaossearch.io/blog/multi-cloud-data-management>
- [6] Dhruv Kumar Seth et al., "Navigating the Multi-Cloud Maze: Benefits, Challenges, and Future Trends," ResearchGate, 2024. [Online]. Available: https://www.researchgate.net/publication/381304851_Navigating_the_Multi-Cloud_Maze_Benefits_Challenges_and_Future_Trends
- [7] Anil A. Kuriakose, "MultiCloud FinOps: A Comprehensive Analysis of Financial Operations Across Major Cloud Providers," LinkedIn, 2024. [Online]. Available: <https://www.linkedin.com/pulse/multicloud-finops-comprehensive-analysis-financial-anil-a-kuriakose-tt8dc/>
- [8] Maryna Cherednychenko, "Handling Latency and Performance Challenges in Real-Time Apps," Erbis. [Online]. Available: <https://erbis.com/blog/real-time-apps/>
- [9] Sandeep Kaushik, "Data Lakehouse: A modern data Architecture," Medium, 2025. [Online]. Available: <https://medium.com/@shyamsandeep28/data-lakehouse-a-modern-data-architecture-3dd3e1c89f92>
- [10] Anne Marie Smith, "Data Lake governance: Benefits, challenges and getting started," TechTarget, 2024. [Online]. Available: <https://www.techtarget.com/searchdatamanagement/answer/What-data-lake-governance-challenges-do-organizations-face>