Open Access   **Review Article**

Strategic Optimization of ETL Architectures for Financial Data Warehouse Systems: A Multi-Objective Analysis

Rakesh Mittapally*

Business Intelligence Architect /AI and ML Engineer, USA

Abstract

In the ever-changing landscape of financial technology, organizations struggle with the dual imperative of handling massive data volumes while adhering to strict regulatory frameworks and maintaining market dominance. This study refines the extraction, transformation, loading (ETL) algorithms designed for financial data warehousing in the fintech industry. As financial transactions grow in complexity, the ability to seamlessly integrate, process, and analyze data from disparate sources emerges as a key factor in organizational performance. This research has particular relevance in mitigating systemic inefficiencies that often plague large-scale financial data warehouses, such as computational lags, resource allocation challenges, and scalability barriers. By evaluating state-of-the-art ETL methods, this study provides financial institutions with empirically based strategies for improving data processing efficiency, while preserving data integrity and complying with evolving regulatory mandates. Using a systematic approach, this research uses the MOORA (Multi-Objective Optimization Based on Ratio Analysis) technique to rigorously evaluate six state-of-the-art ETL frameworks: PySpark-Optimized ETL Framework, AWS Glue-Pipelineed S Dabrita-Bas Redshift, Athena-Driven Server less Analytics, Data Lakehouse with Delta Lake, and Hadoop-based Batch Processing. The evaluation is structured around seven

Key Metrics: processing speed, data integrity, query efficiency, cost-effectiveness, execution time, data propagation latency, and computational resource consumption. The comparative investigation identifies the AWS Glue-based data pipeline as the most effective framework, securing the highest evaluation score (0.07604) over the PySpark-Optimized ETL Framework (0.07496) and Athena-Driven Server less Analytics (0.077). The results underscore AWS Glue's exceptional ability to preserve data consistency while simultaneously improving query execution and overall processing efficiency. This study highlights how cloud-based ETL solutions are revolutionizing financial data processing by delivering better scalability and cost-effectiveness. Furthermore, embedding artificial intelligence within ETL workflows strengthens data integrity through intelligent anomaly detection and dynamic transformation algorithms. Beyond technological advancements, research underscores the need for financial institutions to rigorously implement data governance to mitigate the persistent issues of redundancy and inconsistency. By providing a well-defined framework, these findings equip financial institutions with the tools to evaluate and deploy ETL systems according to their operational landscapes, ultimately refining analytical accuracy and securing a strategic edge in an industry where data is at its peak.

Keywords: MOORA method, multi-objective optimization, AWS Glue, PySpark, Data consistency, Query performance, Cloud-based ETL, Big Data processing and Artificial intelligence in ETL.

Introduction

Efficient data management is a cornerstone for maintaining a competitive edge while navigating fintech's regulatory challenges. At the heart of this effort is the Extract, Transform, Load (ETL) framework, a structured approach that encompasses data extraction, cleansing, and systematic integration into data warehouses. Accordingly, this study aims to explore the advancements in ETL methodologies designed for financial data warehousing in the fintech landscape. Given the continuous technological evolution in this domain, fintech organizations are increasingly embracing industry-wide advancements, refining data processing algorithms, and enhancing the accuracy of business analytics and strategic decision-making [1]. Extensive data warehouses are often plagued with bottlenecks that prevent the rapid manipulation and interpretation of vast datasets. Primary barriers include processing inefficiencies, resource limitations, and scalability constraints, all of which undermine the effectiveness of an organization's data management architecture. Substandard ETL workflows exacerbate integration challenges, reduce analytical capabilities, and reduce the agility needed for informed decision-making and dynamic business operations [2]. Financial data warehousing is a highly specialized domain dedicated to the structured integration, retrieval, and analytical processing of financial information obtained from trading platforms, market intelligence providers, and regulatory agencies. Given the complexity and heterogeneity of financial datasets, an effective ETL (extract, transform, and load) framework is essential for seamless integration. The extraction phase integrates structured and unstructured datasets from various sources, such as banking infrastructures and stock market feeds.

Transformation ensures standardization and refinement, consistency, and regulatory compliance. Finally, the loading

phase organizes the deposition of massive datasets in the warehouse, enabling instant analysis, pattern recognition, and data-driven strategic financial management [3]. With the digital transformation driving the fintech sector, data warehousing and ETL methodologies are playing a key role in organizing the influx of vast, heterogeneous financial data. As transaction complexity increases, the seamless integration of data from traditional banking institutions and decentralized blockchain ecosystems necessitates sophisticated ETL strategies. These processes meticulously extract, cleanse, and integrate structured and unstructured data, ensuring accuracy and instant access for analytical applications. Sophisticated ETL frameworks improve strategic decision-making by improving data integrity and operational efficiency. This research explores the evolutionary trajectory of ETL, its instrumental role in FinTech performance, and its ability to mitigate industry-specific barriers, thereby underpinning data-driven advancements in financial technology [4].

The advent of Big Data has revolutionized the way organizations extract actionable insights from vast datasets, yet its management presents formidable challenges. Complexity stems from four fundamental dimensions: volume, type, velocity, and truthfulness. The massive influx of data generated daily requires sophisticated storage paradigms and highly scalable architectures. Furthermore, the heterogeneity of data demands adaptive architectures that can seamlessly handle both structured and unstructured formats. The relentless pace of data influx necessitates real-time analytics capabilities, especially in domains such as fraud detection. As digital interactions proliferate, organizations must adopt avant-garde technologies to effectively process, interpret, and harness the power of Big Data [5]. Cloud computing has emerged as an essential catalyst for Big Data ecosystems, providing elastic, cost-effective computational resources. However, it presents a wide spectrum of challenges in architectural design, cost-effectiveness, system performance, operational reliability, cybersecurity, and data consistency.

The inherently distributed nature of Big Data necessitates complex mechanisms for synchronization, data replication, and workload orchestration to maintain fluid operations. Security and confidentiality concerns are further magnified by the sheer volume of sensitive information that traverses multiple cloud environments. Mitigating these risks mandates strong encryption methods, strict access control policies, and rigorous compliance frameworks to maintain data integrity. It is essential to refine cloud-centric Big Data architectures to achieve an optimal balance between performance, security, and scalability [6]. Integrating artificial intelligence into ETL workflows is reshaping data management by leveraging machine learning, natural language processing, and advanced analytics to refine data quality and operational efficiency. By automating data cleansing, anomaly detection, and optimization of transformation procedures, machine learning algorithms ensure data integrity before analysis. AI-driven transformations dynamically adjust to changing data structures, fine-tuning ETL processes for real-time adaptability. Furthermore, natural language processing facilitates intuitive engagement with data, giving business users access to self-service analytics. This paradigm shift is reducing reliance on IT departments while fostering a culture focused on data-driven insights. In essence, AI-augmented ETL improves operational efficiency, increases decision-making accuracy, and redefines financial data warehousing practices [7]. The introduction of artificial intelligence and machine learning is redefining financial

data management by increasing automation, accuracy, scalability, speed, and insight-driven capabilities across ingest, extraction, transformation, and load operations. AI-powered automation reduces manual intervention, enabling data engineers to prioritize high-level strategic analysis while ensuring seamless and error-free workflows. Machine learning strengthens data accuracy by flagging anomalies, thereby mitigating financial uncertainties.

The scalability of these systems is enhanced by adaptive models that seamlessly handle expanding data volumes. Real-time processing accelerates data access, which is a critical factor in fraud detection and risk mitigation. In addition, AI refines data integration processes and drives predictive analytics, empowering financial institutions to anticipate market trends and improve risk optimization strategies [8]. Risk management serves as the cornerstone of banking, safeguarding financial stability while ensuring compliance with regulatory mandates. With the advent of the new Basel Capital Accord, the paradigm is shifting from conventional credit risk assessment to comprehensive models that encompass credit, market, and operational risks. Traditional frameworks rely primarily on reactive strategies, limiting their effectiveness in mitigating unexpected threats. The integration of big data technology is revolutionizing risk oversight by facilitating instant data aggregation, predictive modeling, and automated decision-making.

By leveraging cutting-edge data mining techniques, social media analytics, and transaction monitoring, banks are strengthening fraud detection, improving liquidity risk strategies, and strengthening regulatory compliance, fostering a data-centric approach to risk governance [9]. Financial reporting under IFRS fosters financial transparency across the European Union, Asia, and South America, while the United States adheres to GAAP. IFRS prescribes essential financial statements, such as the statement of financial position, comprehensive income statement, and statement of cash flows, structuring the representation of assets, liabilities, and profit and loss in a consistent manner. At the heart of corporate finance systems is the general ledger, a centralized repository that integrates transactions with sub-ledgers for granular financial monitoring. However, ensuring seamless financial reporting is hampered by challenges in data accessibility, making it necessary to implement advanced data warehousing solutions. The integration of real-time reporting mechanisms improves strategic decision-making, promotes automation, and streamlines operational workflows [10]. Financial institutions, coupled with national and international compliance mandates, are navigating the increasingly demanding regulatory landscape by deploying sophisticated data management and engineering frameworks. Protecting data integrity is critical, as discrepancies arising from redundant or outdated customer records introduce financial vulnerabilities and hinder AI-driven applications. Effective data governance requires implementing accurate customer identification protocols, ensuring compliance and mitigating risks. The IT infrastructure underlying financial institutions uses virtualization and automation, yet discrepancies in data quality persist, with expert analyses indicating an error range of 1% to 5% in financial datasets. Redundancies often emerge from corporate mergers, unique product system requirements and software inefficiencies. Achieving seamless data integration requires the deployment of robust architectures, mainly relying on data warehousing solutions [11-12].

Materials and Methods

The MOORA (Multi-Objective Optimization Based on Ratio Analysis) technique serves as a key tool for evaluating and ranking ETL architectures across multiple performance dimensions. This assessment consists of four benefit-driven parameters—processing speed, data consistency, query efficiency, and cost-effectiveness—where higher values indicate better performance. At the same time, three unproductive factors—execution time, data latency, and resource consumption—are reduced to increase overall performance. The MOORA methodology involves normalizing these criteria, deriving a comprehensive performance score, and ranking the available alternatives.

Processing Speed (min):

Processing speed refers to the time required to execute data processing operations, where higher values correspond to improved performance. Accelerated processing facilitates seamless management of vast datasets, fostering rapid decision-making, and real-time analytics capabilities. Within large-scale ETL (Extract, Transform, Load) systems, improving processing speed is critical to maintaining seamless data workflows. By mitigating system bottlenecks and increasing operational efficiency, high-speed processing emerges as a fundamental criterion in the evaluation of data-intensive architectures.

Data Consistency (%):

Data consistency reflects the accuracy and reliability of processed information, ensuring consistency across different platforms and processing stages. A higher percentage indicates fewer inconsistencies, thereby reducing analytical errors and flawed decision-making. This aspect is especially important in financial and transactional domains, where accuracy is non-negotiable. Robust ETL frameworks use rigorous validation protocols and reconciliation mechanisms to enforce consistency, while protecting data integrity while protecting against corruption, loss, or deviation from standardized formats.

Query Performance (QPS - Queries Per Second):

Query performance, measured in queries per second (QPS), encompasses the responsiveness and efficiency of an analytics system in retrieving and processing data. High QPS values indicate a system's ability to handle concurrent queries with minimal latency. Optimal query processing is essential for real-time analytics and dynamic business intelligence frameworks. To improve QPS, ETL infrastructures should include advanced indexing methods, strategic partitioning, and caching optimizations, ensuring that users access insights with speed, accuracy, and scalability.

Cost efficiency ($ per TB):

Cost efficiency in data processing assesses the financial cost associated with handling each terabyte of information, with the lowest cost being the optimal scenario. A well-designed ETL framework strategically allocates resources to minimize unnecessary costs while maintaining operational efficiency. Judicious use of cloud services, compute power, and storage solutions plays a key role in reducing costs. Organizations seek a balance between economic viability and performance excellence to ensure that large-scale data operations are financially and computationally sustainable.

Execution Time (hours):

Execution time refers to the overall amount of time required to complete an ETL process, with a lower value indicating higher performance. A streamlined execution time improves system responsiveness, facilitating faster insights into business operations. Conversely, extended execution times can introduce analysis delays, which can hinder timely decision-making. Improving execution efficiency requires mechanisms such as parallel computing, optimal scheduling strategies, and balanced workload distribution. Reducing ETL execution time is critical for deploying high-performance data pipelines, especially in real-time analytics ecosystems.

Data Latency (seconds):

Data latency measures the time it takes for data to be generated and then made available for analysis, with lower latency reflecting improved performance. Excessive latency disrupts real-time decision-making processes and impacts overall operational agility. Mitigating latency requires improvements in data ingest algorithms, reduction of network congestion, and refinement of processing logic. Within financial and transactional infrastructures, maintaining minimal latency is essential to guarantee instant insights and immediate system responsiveness. Modern ETL frameworks incorporate stream processing methods to reduce latency and increase overall system performance.

Resource utilization (%):

Resource utilization measures the proportion of computing assets engaged during data processing, with lower percentages indicating improved performance. High resource consumption can indicate system stress, accelerating performance degradation and driving up operational costs. Optimally designed ETL frameworks streamline resource allocation through strategic workload distribution, scalable infrastructure adaptation, and advanced caching methods. Effective resource management ensures that processing power, memory, and storage are used judiciously, mitigating unnecessary stress, while strengthening system resilience and economic viability in data-intensive environments.

MOORA Method:

The MOORA (Multi-Objective Optimization by Ratio Analysis) method serves as a key decision-making framework, widely used in various domains to evaluate and rank multiple alternatives in the presence of conflicting criteria. Its importance is particularly pronounced in complex decision-making situations, where it is necessary to strike a balance between competing objectives. MOORA is distinguished by its straightforward and powerful methodology, MOORA facilitates a systematic approach to multi-criteria decision-making, which ensures both clarity and efficiency in the evaluation process [13]. Essentially, the MOORA method is expressed through a series of structured phases that organize the decision-making journey. The process begins with the selection of relevant evaluation criteria, either quantitative or qualitative, depending on the specific decision-making context, and the identification of alternatives that require evaluation. Following this, a decision matrix is precisely designed, integrating the performance of each alternative against the predefined criteria. This matrix forms the analytical backbone of the method, laying the foundation for subsequent computational steps that drive the optimization process [14-15]. The MOORA method adopts a unique strategy for normalizing decision matrix data, which is a critical step in ensuring comparability across criteria that may be on different measurement scales. This normalization process reconstructs the values onto a unified scale, facilitating the equal

evaluation of competing alternatives. Once normalization is complete, the method compares the ratio of each alternative's performance to the criteria, effectively illustrating the relative strengths and weaknesses inherent in each option [16]. A defining characteristic of the MOORA method is its ability to accommodate benefit and cost criteria within its analytical framework. While benefit criteria favor higher values, cost criteria necessitate lower values for optimal outcomes. By using systematic adjustments, the method efficiently reconciles these opposing views while preserving the integrity of the decision-making process. This two-pronged approach ensures that the overall performance of each alternative is assessed in a complete and balanced manner [17]. Once the ratios have been determined, the next step involves generating a composite score for each alternative. This score is typically calculated by summing the ratios across all established criteria. The alternatives are then ranked according to their composite scores, helping decision makers identify the most favorable choice. This ranking mechanism is intuitive and useful, providing explicit insights into the superiority of alternatives based on predefined parameters [18-19]. The MOORA method excels in situations where transparency and accuracy are required in decision-making. Its structured algorithm reduces ambiguity, thereby strengthening the legitimacy of the selection process. Furthermore, its adaptability allows it to be used in a wide variety of applications, from environmental assessments to supplier evaluations in supply chain management [20-21]. Serving as a robust framework for multi-objective decision-making, the MOORA method reconciles conflicting criteria. Its formal structure, with the ability to accommodate both cost and benefit factors, provides a preferred strategy for navigating complex decision-making scenarios. By establishing a well-defined mechanism for evaluating alternatives, the MOORA method empowers decision makers to make well-informed choices, ultimately resulting in better outcomes across multiple domains [22].

Analysis and Discussion

Table 1: A comparative analysis of the six ETL frameworks, illustrated							
Alternative	Processing Speed (min)	Data Consistency (%)	Query Performance (QPS)	Cost Efficiency ($ per TB)	Execution Time (hrs)	Data Latency (sec)	Resource Utilization (%)
PySpark-Optimized ETL Framework	35.00	98.00	2500.00	1.21	4.24	15.00	70.00
AWS Glue-Based Data Pipeline	25.00	99.00	2700.00	1.10	3.90	12.00	65.00
Hybrid ETL with Spark & Redshift	35.00	96.00	2300.00	1.30	5.20	19.00	80.00
Athena-Driven Serverless Analytics	22.00	99.50	2800.00	1.00	3.80	10.00	68.00
Data Lakehouse with Delta Lake	40.00	95.00	2100.00	1.40	5.50	20.00	85.00
Hadoop-Based Batch Processing	28.00	97.00	2400.00	1.20	4.70	14.00	73.00

A comparative analysis of the six ETL frameworks, illustrated in Table 1, reveals significant differences in processing efficiency, data reliability, and resource management. Notably, the Athena-Driven Serverless Analytics solution outperforms its peers, achieving a peak query rate of 2800 QPS, high data consistency of 99.5%, and a minimum latency of 10 seconds. Furthermore, with a cost efficiency of $1.00 per terabyte, it emerges as the most resource-optimized framework for real-time analytics workloads and business intelligence solutions. Similarly, the AWS Glue-based data pipeline demonstrates strong performance metrics, maintaining high data integrity at 99% while executing queries at 2700 QPS. With a significantly lower operational duration of 3.90 hours and a resource utilization rate of 65%, it efficiently manages workload distribution. In parallel, the PySpark-Optimized ETL framework exhibits commendable processing efficiency, completing jobs in 35 minutes while ensuring a 98% data consistency rate. However, it lags slightly in both query performance and cost-effectiveness. In contrast, frameworks such as Data Lakehouse with Delta Lake and Hybrid ETL combining Spark & Redshift deliver much lower data consistency at 95% and 96%, respectively. These frameworks exhibit long processing times and high data latency, suggesting limitations in their ability for real-time data processing. Meanwhile, the Hadoop-based batch processing model delivers balanced to moderate performance, maintaining consistent query processing, but struggles to compete in terms of processing speed and latency optimization.

Table 2: MOORA method facilitates comparative analysis by evaluating the performance metrics of the six ETL frameworks						
0.4540	0.4106	0.4119	0.4087	0.3762	0.3972	0.3871
0.3243	0.4148	0.4448	0.3716	0.3461	0.3178	0.3594
0.4540	0.4023	0.3789	0.4391	0.4614	0.5031	0.4424
0.2854	0.4169	0.4613	0.3378	0.3372	0.2648	0.3760
0.5189	0.3981	0.3460	0.4729	0.4880	0.5296	0.4700
0.3632	0.4064	0.3954	0.4053	0.4170	0.3707	0.4037

In Table 2, the normalized matrix obtained through the MOORA method facilitates comparative analysis by evaluating the performance metrics of the six ETL frameworks. The Athena-Driven Serverless Analytics framework boasts the highest normalized values for query performance (0.4613) and data consistency (0.4169), indicating its strong data processing capabilities and excellent integrity. In addition, it records the lowest figures for execution time (0.3372) and data latency (0.2648), underscoring its effectiveness in real-time analytics. Similarly, the AWS Glue-based data pipeline exhibits commendable performance, achieving significant query efficiency (0.4448) and data consistency

(0.4148), while also achieving relatively low processing time (0.3461) and data latency (data latency of 0.3178). Exhibiting well-balanced performance, the PySpark-Optimized ETL framework excels in particular in processing speed (0.4540) and data consistency (0.4106), ensuring reliable operational efficiency. However, it falls short in cost efficiency (0.4087) and query efficiency (0.4119), reducing its overall cost-effectiveness. In contrast, Hybrid ETL with Spark & Redshift and Data Lakehouse with Delta Lake outperformed Big Data workloads with sustained execution times (0.4614 and 0.4618 in Opal performance) in terms of data consistency (0.4023 and 0.3981, respectively) and query performance (0.3789 and 0.3460). On the other hand, the Hadoop-based batch processing architecture delivers moderate results across all evaluated parameters, maintaining stable values in processing speed (0.3632), data consistency (0.4064), and query performance (0.3954). However, its relatively elevated processing time (0.4170) and data latency (0.3707) indicate potential delays, making it a less viable option for real-time applications.
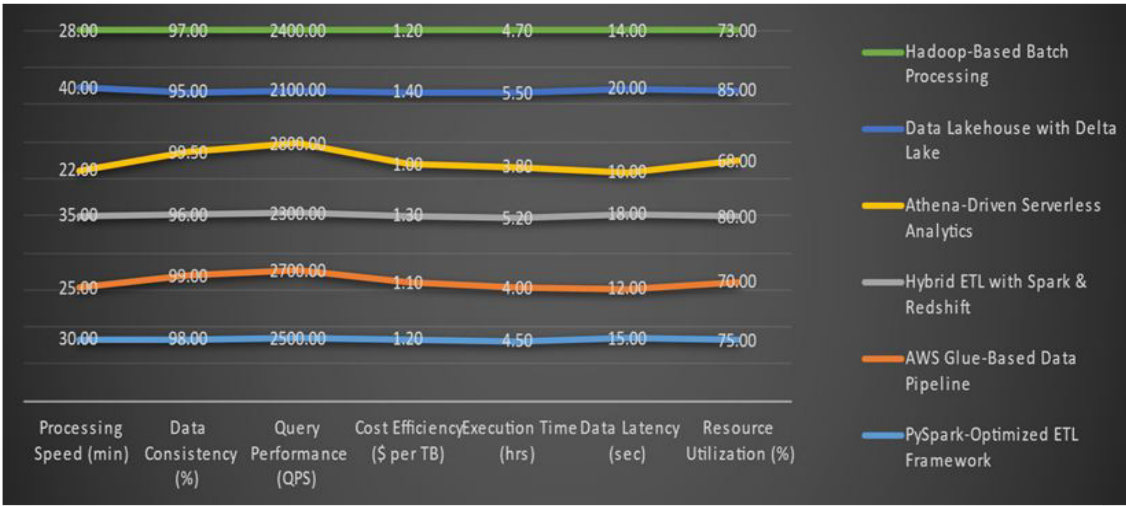

Figure 1: presents a comparative assessment of six ETL frameworks, analyzing their performance on key metrics such as processing capacity, data reliability, and resource allocation

Figure 1 presents a comparative assessment of six ETL frameworks, analyzing their performance on key metrics such as processing capacity, data reliability, and resource allocation. Notably, significant discrepancies emerge among the evaluated systems. Athena-Driven Serverless Analytics demonstrates exceptional query performance, peaking at 2800 QPS, while ensuring remarkable data consistency at 99.5% and achieving a minimum latency of just 10 seconds. Furthermore, with a cost efficiency of $1.00 per TB, it positions itself as an optimal solution for real-time analytics and business intelligence applications. Similarly, AWS Glue-based Data Pipeline boasts an impressive query processing capacity of 2700 QPS with a data consistency rate of 99%, delivering commendable performance. With a resource consumption rate of just 65% and a resource consumption rate of just 3.90 hours, it efficiently manages workload distribution. The PySpark-Optimized ETL Framework, on the other hand, maintains fast execution (35 minutes) and strong data integrity (98%), although it lags slightly in terms of query efficiency and economic viability. On the other hand, Data Lakehouse with Delta Lake and Hybrid ETL with Spark & Redshift record the lowest data consistency, marked at 95% and 96%, respectively. These frameworks struggle with prolonged execution times and high data latency, signaling inefficiencies in real-time data processing. Meanwhile, the Hadoop-based batch processing paradigm achieves a consistent query execution rate, but lags in terms of processing speed and latency reduction, with very modest performance.

Table 3: follows an equal allocation method, in which each evaluation criterion receives the same weight of 0.1429 across the six ETL frameworks

0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429
0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429
0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429
0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429
0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429
0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429

The parameter weight distribution described in Table 3 follows an equal allocation method, in which each evaluation criterion receives the same weight of 0.1429 across the six ETL frameworks. This strategy ensures that no single parameter disproportionately influences the decision-making process, promoting an unbiased evaluation of each framework. Because all parameters are considered equally important, the evaluation relies on overall performance rather than emphasizing individual aspects such as query efficiency, processing speed, or execution time. This uniform weighting approach is particularly advantageous in contexts where each evaluation criterion is considered equally important in measuring the performance and reliability of an ETL framework. It mitigates bias toward specific performance dimensions, ensuring that improvements in one area do not unnecessarily distort the overall ranking. However, this approach does not always align with practical priorities, as some factors—such as execution time and data latency—may carry more weight in real-time analytics scenarios than in cost-effectiveness. By maintaining equal weights, the evaluation remains neutral and objective, facilitating a thorough comparison of the relative advantages and limitations of each architecture. For example, if one architecture exhibits

superior query processing but performs poorly in execution speed, the influence is distributed equally, preserving a balanced evaluation. However, in application decision-making, reordering the weights based on specific business needs can provide a more accurate and context-sensitive ranking, aligning the evaluation process with operational objectives and performance criteria.

Table 4: refines the comparative evaluation of the six ETL frameworks using the MOORA method						
0.0649	0.0587	0.0588	0.0584	0.0537	0.0567	0.0553
0.0463	0.0593	0.0635	0.0531	0.0494	0.0454	0.0513
0.0649	0.0575	0.0541	0.0627	0.0659	0.0719	0.0632
0.0408	0.0596	0.0659	0.0483	0.0482	0.0378	0.0537
0.0741	0.0569	0.0494	0.0676	0.0697	0.0757	0.0671
0.0519	0.0581	0.0565	0.0579	0.0596	0.0530	0.0577

The weighted normalized matrix depicted in Table 4 refines the comparative evaluation of the six ETL frameworks using the MOORA method. By integrating the weight assignments from Table 3, this matrix reweights the normalized values, ensuring a more equitable assessment of each framework's capabilities across multiple performance parameters. The findings highlight the differences in performance, stability, and resource consumption, facilitating a nuanced understanding of the most optimal solution for data processing requirements. Athena-Driven Serverless Analytics exhibits significant strengths in data consistency (0.0596) and query performance (0.0659), confirming its effectiveness in high-speed analytics workloads. Furthermore, it achieves very short processing times (0.0482) and very low data latency (0.0378), making it a favorable choice for real-time processing scenarios. Similarly, AWS Glue-based Data Pipeline achieves commendable results in query performance (0.0635) and data consistency (0.0593), maintaining well-rounded performance across various evaluation metrics. Conversely, Data Lakehouse with Delta Lake records higher processing speed (0.0741) but is hampered by prolonged processing time (0.0697) and elevated data latency (0.0757), indicating inefficiency in real-time operations. Hybrid ETL with Spark & Redshift architecture performs strongly in specific dimensions but exhibits increased execution time (0.0659) and resource consumption (0.0632), which can impact cost-effectiveness.

Table 5: The evaluation metrics and rankings obtained through the MOORA method		
Alternative	Assessment value	Rank
PySpark-Optimized ETL Framework	0.07496	2
AWS Glue-Based Data Pipeline	0.07604	1
Hybrid ETL with Spark & Redshift	0.03820	5
Athena-Driven Serverless Analytics	0.07477	3
Data Lakehouse with Delta Lake	0.03545	6
Hadoop-Based Batch Processing	0.05414	4

The evaluation metrics and rankings obtained through the MOORA method, outlined in Table 5, provide a solid assessment of the six ETL frameworks, measuring their overall performance across multiple performance criteria. Dominating the rankings, AWS Clue-based Data Pipeline achieves the highest evaluation score

(0.07604), which is characterized by its excellent data consistency, high query performance, and optimized processing efficiency, positioning it as the top choice for large-scale data workflows. Close behind, PySpark-Optimized ETL Framework takes second place with a rating of 0.07496. While it demonstrates strong processing speed and strong data integrity, its cost-effectiveness and resource optimization are slightly lower when combined with AWS Clue. Meanwhile, Athena-Driven Serverless Analytics takes third place (0.07477), excelling in query performance and reducing data latency, making it a strong option for real-time analytics. However, its somewhat reduced processing speed can be a hindrance in high-performance environments. Occupying the fourth tier, the Hadoop-based batch processing framework registers a rating of 0.05414, which strikes a moderate balance between operational efficiency and performance. Following in fifth place is Hybrid ETL with Spark & Redshift (0.03820), which, despite providing satisfactory query performance, has extended processing times and sub-resource management. At the lowest level, Data Lakehouse with Delta Lake (0.03545) is hampered by data consistency, long execution times, and high latency, making it the least viable contender for streamlined ETL operations. These rankings establish a distinct performance hierarchy, serving as a decisive reference for selecting the most appropriate ETL framework in alignment with specific processing demands.
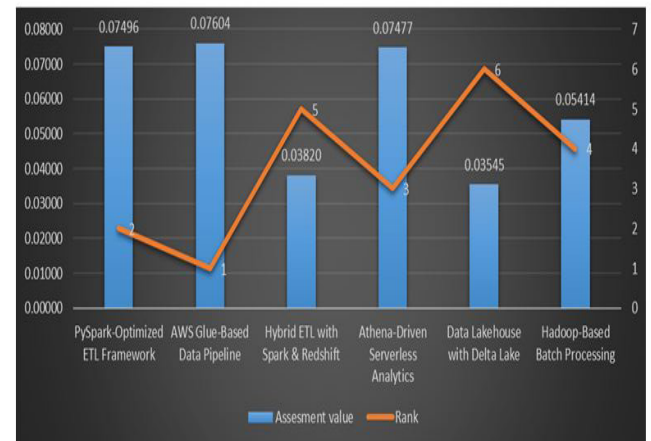

Figure 2: The evaluation results obtained through the MOORA method provide a rigorous comparative analysis of six distinct ETL frameworks

As illustrated in Figure 2, the evaluation results obtained through the MOORA method provide a rigorous comparative analysis of six distinct ETL frameworks, each evaluated on a multi-faceted set of performance criteria. With a score of 0.07604, AWS Clue-based Data Pipeline emerges as the most efficient alternative, achieving the highest ranking. Its exceptional data consistency, fast query execution, and optimized processing power position it as a highly viable option for large-scale data operations. Closely following in second place is PySpark-Optimized ETL Framework, which receives a score of 0.07496. While excelling in processing speed and data integrity, it lags slightly behind AWS Clue in cost-effectiveness and resource allocation. Meanwhile, Athena-Driven Serverless Analytics ranks third (0.07477), offering significant query performance and minimal latency, making it a compelling solution for real-time analytics workloads. However,

its slightly reduced processing speed may pose limitations in high-demand scenarios. Hadoop-based batch processing ranks fourth (0.05414), reflecting a well-balanced and moderate level of performance across various performance dimensions. Hybrid ETL combining Spark & Redshift follows in fifth place (0.03820), which is constrained by extended processing times and significant resource consumption, although it maintains appreciable query performance. Finally, Data Lakehouse with Delta Lake has the lowest ranking (0.03545), primarily due to suboptimal data consistency, prolonged execution times, and increased latency, which provide a less-than-performing ETL approach. This ranking structure defines a clear performance hierarchy, helping to select the optimal ETL architecture to suit specific data processing demands. The MOORA-driven assessment places AWS's Clue-based Data Pipeline at the top, due to its unmatched execution speed, data consistency, and query performance. PySpark-Optimized ETL and Athena-Driven Serverless Analytics continue to demonstrate competitive capabilities. In contrast, Data Lakehouse with Delta Lake occupies the lowest ranking.

Conclusion

In a rapidly evolving digital landscape, financial institutions face formidable challenges in handling vast amounts of data. The refinement of big data workflows and the optimization of ETL (Extract, Transform, Load) processes have emerged as essential pillars for financial institutions striving to maintain market dominance, comply with stringent regulatory mandates, and extract strategic insights from their data warehouses. An organization's ability to rapidly process and interpret financial datasets directly impacts its ability to adapt to changing market dynamics. A detailed assessment of ETL frameworks designed for financial data warehousing underscores the profound changes in the fintech industry. Leveraging the MOORA multi-objective optimization framework, this study identifies AWS Glue-based Data Pipeline as the most efficient ETL mechanism, with PySpark-Optimized ETL Framework and Athena-Driven Serverless Analytics lagging far behind. These frameworks demonstrate exceptional performance in key performance indicators such as data integrity, query execution speed, and operational efficiency. The results underscore the critical need to balance various performance factors when evaluating ETL solutions for financial organizations. While fast processing and efficient query processing are essential for real-time analytics, unwavering data consistency is fundamental to accurate financial reporting and compliance with regulatory mandates. High-performance architectures effectively strike this balance, ensuring seamless data integration without compromising operational agility. The infusion of artificial intelligence and machine learning into ETL workflows represents a paradigm shift in financial data management. AI-driven automation alleviates manual intervention, refines data accuracy through anomaly detection, and unlocks predictive insights that facilitate proactive risk mitigation.

These findings have particular significance in the finance domain, where data integrity underpins strategic decision-making and comprehensive risk assessment. Cloud-driven ETL frameworks have gained prominence due to their adaptability, cost-effectiveness, and ability to manage the vast and diverse range of financial data. However, organizations must carefully evaluate architectural frameworks, security measures, and governance structures when integrating cloud-based solutions. This study highlights the critical role of rigorous data governance in mitigating

issues such as redundancy, inconsistencies, and quality deficiencies that often challenge financial institutions. This research describes a systematic framework for evaluating ETL solutions according to unique business needs for financial institutions aiming to improve their data warehousing systems. The MOORA technique facilitates an objective and structured decision-making paradigm, balancing both cost and benefit parameters, thereby empowering them to adopt solutions that are compatible with the organizations' operational objectives and technology ecosystems. As the fintech ecosystem undergoes relentless change, the efficiency of ETL processes is a key aspect of maintaining a competitive edge. Organizations that prioritize streamlined data integration frameworks are preparing themselves to leverage sophisticated analytics, refine operational workflows, and sharpen decision-making accuracy. The path of ETL improvements is poised to emphasize real-time data manipulation, enhance data integrity via AI-powered validation algorithms, and facilitate frictionless interoperability with new technologies such as blockchain and quantum computing. Navigating the landscape of financial data warehousing requires constant reevaluation and reengineering of ETL methodologies to align with changing business needs and technological frontiers.

References

1. Santosh Kumar, Singu. "Maximizing Financial Intelligence-The Role Of Optimized Etl In Fintech Data Warehousing." INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING AND TECHNOLOGY (IJCET) 15, No. 4 (2024): 464-471.

2. Raghavendra Sunku. (2024). AI-Powered Forecasting and Insights in Big Data Environments. Journal of Business Intelligence and Data Analytics, 1(2), 254. https://doi.org/10.55124/jbid.v1i2.254

3. Gurubasannavar, S. D. (2024) Resource-Aware End-to-End Test Optimization: A Comprehensive Framework for Automated Test Generation and Performance Prediction in Modern Software Development J Comp Sci Appl Inform Technol. 9(2): 1-8.

4. Badgujar, Pooja. "Optimizing ETL Processes For Large-Scale Data Warehouses." Journal Of Technological Innovations 2, No. 4 (2021).

5. Divya Soundarapandian. (2024) Algorithmic Framework for Retail Media Optimization and Consumer Engagement Enhancement. Journal of Business Intelligence and Data Analytics, 1(3), 1-7. https://doi.org/10.55124/jbid.v1i3.259

6. VERMA, PRACHI, And RKGIIT GHAZIABAD. "Optimizing ETL Processes For Financial Data Warehousing." (2023).

7. Blake, Harrison. "Enhancing Decision-Making In Fintech Through Advanced ETL Processes For Data Warehousing." (2024).

8. Tran, Trung. "In-Depth Analysis And Evaluation Of ETL Solutions For Big Data Processing." (2024).

9. Zdravevski, Eftim, Petre Lameski, Ace Dimitrievski, Marek Grzegorowski, And Cas Apanowicz. "Cluster-Size Optimization Within A Cloud-Based ETL Framework For Big Data." In 2019 IEEE International Conference On Big Data (Big Data), Pp. 3754-3763. IEEE, 2019.

10. Divya Soundarapandian (2024). Machine Learning-Driven Reliability Engineering for E-Commerce Sites: A Study on Browsing, Cart, and Checkout Phases. Journal of Business Management and Entrepreneurship, 3(1), 1-9. doi: https://doi.org/10.55124/jbme.v3i1.239

11. Gadde, Hemanth. "AI-Enhanced Data Warehousing: Optimizing ETL

Processes For Real-Time Analytics." Revista De Inteligencia Artificial En Medicina 11, No. 1 (2020): 300-327.

12. Katari, Abhilash, And Anjali Rodwal. "NEXT-GENERATION ETL IN FINTECH: LEVERAGING AI AND ML FOR INTELLIGENT DATA TRANSFORMATION."

13. Kiran Kumar Mandula Samuel, Venkata Pavan Kumar Aka. (2023). AI for Special Education Closing Gaps in Inclusive Learning Using MOORA Method. International Journal of Computer Engineering and Technology (IJCET), 14(1), 249-267.

14. Ma, S., Wang, H., Xu, B., Xiao, H., Xie, F., Dai, H. N., ... & Wang, T. (2018, October). Banking Comprehensive Risk Management System Based On Big Data Architecture Of Hybrid Processing Engines And Databases. In 2018 IEEE Smartworld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet Of People And Smart City Innovation (Smartworld/SCALCOM/UIC/ATC/Cbdcom/IOP/SCI) (Pp. 1844-1851). IEEE.

15. Fikri, Noussair, Mohamed Rida, Noureddine Abghour, Khalid Moussaid, And Amina El Omri. "An Adaptive And Real-Time Based Architecture For Financial Data Integration." Journal Of Big Data 6 (2019): 1-25.

16. Sienkiewicz, Mariusz, And Robert Wrembel. "Managing Data In A Big Financial Institution: Conclusions From A R&D Project." In EDBT/ICDT Workshops. 2021.

17. Varun Venkatesh Dandasi, Suresh Deepak Gurubasannavar, Raghavendra Sunku. (2024). Data Mart Design and Predictive Modeling at Palandir Foundry: from Ontology to Operational Intelligence Using Machine Learning. International Journal of Computer Engineering and Technology (IJCET), 15(2), 274-294.

18. Ali, Syed Muhammad Fawad, And Robert Wrembel. "Towards A Cost Model To Optimize User-Defined Functions In An ETL Workflow Based On User-Defined Performance Metrics." In Advances In Databases And Information Systems: 23rd European Conference, ADBIS 2019, Bled, Slovenia, September 8–11, 2019, Proceedings 23, Pp. 441-456. Springer International Publishing, 2019.

19. Raja, Chandrasekar, M. Ramachandran, Sathiyaraj Chinnasamy, And Vimala Saravanan. "Competitiveness And Sustainable Development Analysis Of Alternative Energy Exploitation Using MOORA Method." Journal On Materials And Its Characterization 3 (2024): 2.

20. Chakraborty, Santonab, Himalaya Nirjhar Datta, Kanak Kalita, And Shankar Chakraborty. "A Narrative Review Of Multi-Objective Optimization On The Basis Of Ratio Analysis (MOORA) Method In Decision Making." Opsearch 60, No. 4 (2023): 1844-1887.

21. Sinaga, Novendraadisaputra, Heru Sugara, Ewin Johan Sembiring, Melva Epymardianamanurung, Harsudiantosilaen, Pipin Sumantrie, And Victor Marudut Mulia Siregar. "Decision Support System With MOORA Method In Selection Of The Best Teachers." In AIP Conference Proceedings, Vol. 2453, No. 1. AIP Publishing, 2022.

22. Siregar, Victor Marudut Mulia, Mega Romauly Tampubolon, Eka Pratiwi Septaniaparapat, Eve Ida Malau, And Debora Silvia Hutagalung. "Decision Support System For Selection Technique Using MOORA Method." In IOP Conference Series: Materials Science And Engineering, Vol. 1088, No. 1, P. 012022. IOP Publishing, 2021.

23. Gundala, Tirumala Rao. (2025). Predictive Analytics for SSO Performance: Improving Authentication Response Times in Oracle Enterprise Environments Using Linear Regression, Random Forest Regression. International Journal of Computer Science and Data Engineering, 2(4), 1–6. doi: https://dx.doi.org/10.55124/csdb.v2i4.265

24. Emovon, Ikuobase, Oghenenyerovwho Stephen Okpako, And Edith Edjokpa. "Application Of Fuzzy MOORA Method In The Design And Fabrication Of An Automated Hammering Machine." World Journal Of Engineering 18, No. 1 (2021): 37-49.

25. Nagababu. K. (2025). Enterprise-Level Emphasis Operational Analytics: Predicting Healthcare Encounter Resolution Success Using Machine Learning.International Journal of Robotics and Machine Learning Technologies, 1(2), 1-6, doi: https://doi.org/10.55124/ijrml.v1i2.242

26. Singh, Ramanpreet, Vimal Kumar Pathak, Rakesh Kumar, Mithilesh Dikshit, Amit Aherwar, Vedant Singh, And Tej Singh. "A Historical Review And Analysis On MOORA And Its Fuzzy Extensions For Different Applications." Heliyon 10, No. 3 (2024).

27. Anggrawan, Anthony, Christofer Satria, And Lalu Ganda Rady Putra. "Scholarship Recipients Recommendation System Using AHP And Moora Methods." International Journal Of Intelligent Engineering & Systems 15, No. 2 (2022).

28. Sinaga, Dedi Candroparulian, Preddy Marpaung, And Baringinsianipar. "The Application Of The MOORA Method In The Decision Making System For The Selection Of The Best Employees At CV. Lautan Mas." IJISTECH (International Journal Of Information System And Technology) 5, No. 2 (2021): 233-239

29. Gundala, Tirumala Rao. (2024). Oracle OIPA Cloud Migration Analysis: Machine Learning Models for Predicting Resource Utilization and Success Outcomes. International Journal of Artificial intelligence and Machine Learning, 2(3), 1-8. https://doi.org/10.55124/jaim.v2i3.284

30. Venkata Pavan Kumar Aka and Kiran Kumar Mandula Samuel. (2024). Adoption of SAP FSCM – Enhancing Collections and Dispute Processes in Spain, Portugal, and UK Operations. International Journal of Information Technology and Management Information Systems (IJITMIS), 15(2), 148–161. DOI: https://doi.org/10.34218/IJITMIS_15_02_012

31. Cakranegara, Pandu Adi, Desty Endrawati Subroto, Yunita Dwi Wikandari, And Ahmad Jurnaidi Wahidin. "Selection Outstanding Student Using Moora Method." INFOKUM 10, No. 4 (2022): 33-40.

32. Nagababu Kandula. (2025). Evaluating Research Methods for DMR Reporting Using the Balanced Scorecard Approach. International Journal of Computer Engineering and Technology (IJCET), 16(3), 330–352.

33. Rajamanickam, Jaganathan, M. Ramachandran, Kurinjimalar Ramu, And Chandraseker Raja. "Morphological Characterization And Assessment Of Genetic Variability, Character Association Using MOORA Method."

34. Suresh Deepak Gurubasannavar, Varun Venkatesh Dandasi, Raghavendra Sunku. (2025). Enhancing Smart Home Automation with AI And Topsis-Based Decision Making. International Journal of Information Technology and Management Information Systems (IJITMIS), 16(6), 1-22.