Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The Medallion Architecture in Data Engineering: A Layered Approach to Data Quality and Governance

Sreedhar Pasupuleti
Independent Researcher, USA

ARTICLE INFO

Received: 30 Sept 2025

Revised: 05 Nov 2025

Accepted: 13 Nov 2025

ABSTRACT

The medallion architecture is a new way to handle data engineering. This method simplifies how companies deal with different types of data by refining that data step by step. It's built with three levels: bronze, silver, and gold. Each level has a specific job in the data transformation process. The bronze level takes in raw data and keeps it safe, making sure to record where the data came from and other important details. Cleaning, checking, and standardization of the data is done at the silver level. This fixes any quality issues by removing duplicates and ensuring it meets business rules. The gold level then offers data that's ready for assessment and business intelligence tools to use. The way this architecture is set up allows for data to be processed quickly and in stages. It also provides governance to help with company-wide rollouts. Compared to older data warehouse systems, this architecture is more adaptable. It also provides better organization and quality control than simple data lake methods. It's a mix of both, giving flexibility in storage and structured quality management for various company needs.

Keywords: Data Architecture, Progressive Refinement, Multi-Tier Processing, Data Quality Management, Enterprise Scalability

1. Introduction

Today's data engineering world is full of hard problems. Organizations deal with data systems that are growing fast. These systems need well-designed solutions. Organizations face lots of difficulties when handling different sources of data. The amount of data is growing, so companies need solid systems to manage it [1]. Moving from old ways of handling data to better designs shows that there is a need for ways to organize data. These ways should handle different needs while keeping things running well.

Data designs have become important for facing the many problems in today's data handling situations. Research indicates that groups using organized designs see progress in how well they process data and how reliably things work, compared to just handling data without a plan [1]. The medallion design is a change to using layers for data handling. It has ways to change data, make it better, and get value from it through steps of making it better. This design tackles key worries in data engineering by using steps for organizing and processing data. Computers at the edge, layers in the fog, and cloud platforms all act differently when working with big data. Studies show that using layers to process data in different places has clear pluses [2]. The medallion design makes use of these processing skills through its bronze, silver, and gold layers. This allows teams to efficiently use resources in different steps, making sure data is secure and can be accessed when needed.

Studies on making data processing better explain that layered designs do better than single-step ways, mainly when handling different data and processing needs [2]. The medallion design's way of making data better fits with ideas of how data should grow. It adds current ways such as schema-on-read, step-by-step data handling, and using storage in different places. These design ideas allow groups to stay open to getting data while making sure quality improves along the way.

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The basics of this design come from data handling ideas, but it grows those ideas with current engineering methods. Real-world uses show that layered data processing gives better scaling, data control, and quality control than old, single-step ways [1]. The design's focus on improving data step-by-step lets groups keep raw data while making datasets that are ready for use. This meets the need for keeping all data and making it easy to study.

Today's data engineering problems need smart fixes that can work with different data, needs, and organizational limits. The medallion design gives a full way to face these problems. It has a way to organize data, handle quality, and create business value [2]. As organizations depend more on making choices based on data, designs that give organized, reliable, and well-controlled data pipelines become key. They help keep an advantage and run well in changing markets.

2. Conceptual Framework and Theoretical Foundations

The medallion architecture establishes its conceptual foundation through progressive data refinement principles that transform raw information assets into strategic business resources through systematic processing methodologies. Modern data architecture frameworks emphasize the critical importance of mining raw data, refining information through structured processes, and ultimately converting data assets into valuable digital resources that drive organizational success [3]. This transformation paradigm aligns with established data management theories while incorporating contemporary processing techniques that address the complexities of heterogeneous data environments.

Medallion architecture's theory is based on data maturity modeling's key ideas and how to improve data quality in a systematic way. Progressive refinement approaches demonstrate superior effectiveness in handling diverse data types and sources compared to traditional monolithic processing methods [3]. The architectural framework extends conventional Extract, Transform, Load methodologies by incorporating modern engineering practices such as schema-on-read capabilities, incremental processing logic, and distributed storage optimization techniques that enable organizations to maintain processing flexibility while ensuring data integrity throughout transformation pipelines.

Separation of concerns represents a cornerstone principle within medallion architecture implementations, ensuring that each processing tier maintains distinct operational characteristics and responsibility boundaries. Deep neural network models integrated within big data processing platforms demonstrate enhanced performance when processing layers operate independently while maintaining coordinated data flow mechanisms [4]. This architectural separation enables development teams to optimize individual layer performance characteristics without compromising overall system coherence, resulting in more maintainable and scalable data processing implementations.

Data lineage maintenance throughout the transformation process provides comprehensive traceability mechanisms that enable effective auditing and debugging capabilities across complex data processing workflows. Big data platforms incorporating advanced neural network architectures demonstrate superior lineage tracking performance when processing large-scale datasets through multiple transformation stages [4]. The medallion framework leverages these tracking capabilities to ensure complete visibility into data transformation processes, enabling organizations to maintain regulatory compliance and operational transparency requirements.

Incremental processing principles within medallion architectures enable efficient handling of large-scale datasets by focusing computational resources on modified or newly arrived data elements rather than complete dataset reprocessing. Modern data architecture implementations demonstrate significant efficiency gains when processing only delta changes, particularly in environments handling

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

substantial data volumes with frequent update cycles [3]. This approach reduces computational overhead while maintaining data freshness and accuracy across all architectural layers.

Domain-driven design concepts integrated within medallion architectures treat each processing layer as a distinct domain with specific business rules and transformation logic. Neural network-based processing platforms demonstrate enhanced performance when domain boundaries are clearly defined and processing logic is optimized for specific layer requirements [4]. This domain separation facilitates parallel development processes while ensuring that business rules and data quality standards are consistently applied throughout the transformation pipeline.

The architectural framework embraces data lake principles by enabling native format storage for structured, semi-structured, and unstructured data within bronze layer implementations. Advanced processing platforms demonstrate superior storage efficiency and query performance when data is preserved in original formats during initial ingestion phases [3]. This method keeps data connected and intact. It also allows the needed ability to handle different analyses and operations across various parts of an organization.

Component	Characteristic
Progressive Refinement	Superior effectiveness over monolithic methods
Schema-on-Read	Modern engineering practice integration
Incremental Processing	Efficient large-scale dataset handling
Domain Separation	Parallel development facilitation
Data Lake Principles	Native format storage capability
Neural Network Integration	Enhanced lineage tracking performance

Table 1: Medallion Architecture Theoretical Foundations [3,4]

3. Layer-by-Layer Analysis: Bronze, Silver, and Gold Tiers

3.1 Bronze Layer: Raw Data Ingestion and Preservation

Data quality problems, such as correctness, completeness, consistency, and validity, may arise in Big data. The bronze layer sets up the basic system for taking in and keeping raw data. It uses multi-tier architectural rules for processing different data sources in real-time and batch setups. Multi-tier designs handle many data sources at once well, mostly in systems needing both soft and hard real-time processing. The medallion architecture uses data maturity modeling and quality improvement. The bronze layer tries to keep all data as original, so it can be changed later.

The bronze layer takes in organized data from databases, semi-organized data like JSON and XML files, and unorganized data from different places. The design allows for flexible schemas and fast data intake to deal with different data types without limiting the structure at the start. This makes sure important info in the original data is still available for analysis and keeps processing fast for all data types.

Data governance in bronze layers focuses on capturing metadata, tracking data origins, and checking data quality, like completeness and format consistency. Keeping data in its original format lets groups keep complete audit trails and supports future analysis that might need the original data structures. Partitioning strategies based on time, source system, and data type improve storage use and retrieval speed for later processing.

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3.2 Silver Layer: Data Cleaning and Standardization

The silver layer employs a detailed process to refine data. It addresses the complexities of examining and raising data quality within big data environments. In large datasets, data quality can be impacted in terms of correctness, completeness, consistency, and validity. These issues call for assessment methods that can deal with the huge amounts, speeds, and types of today's data flows. The silver layer is a key link where raw data from the bronze layer is carefully cleaned, validated, and standardized, so it can be reliably processed later on.

Data quality assessment frameworks within silver layers address fundamental challenges, including duplicate record identification, format inconsistency resolution, missing value imputation, and business rule validation across heterogeneous data sources. The complexity of data quality management increases significantly when processing large-scale datasets with diverse structural characteristics and varying quality standards [6]. Silver layer implementations incorporate advanced deduplication algorithms, standardization routines for data format harmonization, and comprehensive validation mechanisms that ensure data integrity before progression to gold layer processing.

To merge info effectively, silver layers make data uniform. This includes standardizing data formats, names, and how data is coded from different sources. Silver layers overcome the concerns that may otherwise arise while checking data quality using sources that vary in quality and completeness by means of data profiling, measuring quality, and automatic fixing of errors [6]. This keeps track of where data comes from and boosts how reliable the overall data is.

3.3 Gold Layer: Business-Ready Data Assets

The gold layer represents the culmination of the medallion architecture's data refinement process, delivering curated datasets optimized for analytical consumption and business intelligence applications. Multi-tier processing architectures demonstrate enhanced performance characteristics when final processing layers are specifically optimized for query performance and user accessibility requirements [5]. The gold layer implements sophisticated data modeling techniques, including dimensional modeling approaches, aggregation strategies, and performance optimization mechanisms that support diverse analytical workloads while maintaining data governance standards.

Business-ready datasets within gold layers incorporate advanced modeling techniques that organize information according to logical business entities and operational relationships reflecting organizational analytical requirements. The architectural design emphasizes query performance optimization through pre-computed aggregations, indexing strategies, and data mart implementations that serve specific business domains [5]. Access control mechanisms, data privacy implementations, and comprehensive audit logging ensure that sensitive information receives appropriate protection while maintaining complete traceability of data usage patterns across different user communities and analytical applications.

4. Implementation Strategies and Best Practices

A good medallion architecture needs careful planning. This includes picking the right tech, getting everyone on board, and setting up clear steps for how things will work. This makes sure it can grow and run well [7]. When designing a data system for a big business, organizations need to think about how powerful the computers are, how much data the storage can hold, and how well everything is managed. This will help handle different jobs from different parts of the company.

When choosing tech, focus on computers that can grow to handle lots of company data. They should also be able to work with different types of data and tasks. Company data systems should be able to spread the work across many computers while keeping the data consistent and reliable [7]. Modern

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

systems use tools such as Apache Spark for changing data in large amounts, Delta Lake for making sure data changes are correct, and cloud storage that can grow cheaply as needed.

To manage large business data volumes using medallion architectures, incremental data processing is necessary. Data pipelines should monitor changes, process only new information, and have checkpoints to ensure accuracy [8]. This approach processes only changed data each time, saving resources and time, and maintains current data across all architecture layers.

Dividing data according to business needs and usage patterns is critical for fast queries and low storage costs. Research shows that good data partitioning can improve query speeds and cut storage use [8]. Effective data partitioning should consider creation date, business unit, and access frequency to place data correctly for rapid queries.

For a medallion architecture to work well, organizations need good data quality standards, ways to watch what's happening, and clear instructions for how data is processed in each layer. Managing data pipelines means checking quality, watching performance, and writing down how things work so the system can be maintained and relied on [8]. Companies should have rules for who owns the data, how to check quality, and what to do if there are data problems.

When using medallion architectures, especially when data structures change, it's important to have ways to manage changes and keep track of versions. Company data systems should be able to handle data structure changes without messing up analysis or business processes [7]. This can involve making sure old systems still work, coordinating changes across layers, and testing to make sure data stays correct. Version control systems should track changes to data structures, processing steps, and settings. This helps ensure changes can be undone if something goes wrong.

Strategy Component	Implementation Focus
Technology Selection	Scalable processing capabilities
Incremental Processing	Modified data element handling
Data Partitioning	Query performance optimization
Quality Standards	Comprehensive monitoring systems
Version Control	Schema evolution management
Change Management	Structural modification coordination

Table 3: Medallion Architecture Implementation Strategies [7,8]

5. Benefits, Challenges, and Comparative Analysis

The medallion architecture offers real advantages for organizations. It uses a systematic way to handle data, which tackles current problems that big companies face and improves how they operate. Big data architectures designed for large organizations demonstrate significant advantages in terms of maintainability, scalability, and operational efficiency when implementing layered processing approaches compared to monolithic data management systems [9]. The clear separation of concerns inherent in medallion architecture enables development teams to work independently on different processing layers while maintaining overall system coherence, resulting in improved development velocity and reduced coordination overhead across distributed development environments.

Scalability represents a fundamental advantage of medallion architecture implementations, with each processing tier capable of independent scaling based on specific workload characteristics and resource requirements. Large organizational big data architectures benefit from the ability to allocate computational resources dynamically across different processing layers, enabling optimized resource utilization and cost management [9]. The bronze layer can prioritize high-throughput data ingestion capabilities, the silver layer can focus on complex transformation and quality assurance processes, and

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

the gold layer can optimize query performance for analytical workloads, creating a balanced approach to resource allocation across diverse processing requirements.

However, medallion architecture implementations present notable challenges, particularly regarding processing latency and data consistency management across multiple architectural layers. Stream processing frameworks deployed as microservices in cloud environments demonstrate varying performance characteristics when handling real-time data processing requirements, with latency considerations becoming critical for time-sensitive applications [10]. With the multi-layered approach, additional processing stages are inherently introduced, which impact end-to-end data availability times. This potentially affects use cases requiring immediate data access or real-time analytical capabilities.

Data duplication across architectural layers represents another significant challenge, potentially increasing storage costs and the complexity of data synchronization processes. Big data architectures for large organizations must carefully balance data redundancy requirements with storage optimization strategies to maintain cost effectiveness while ensuring data availability and processing performance [9]. Managing consistency across multiple data copies requires sophisticated synchronization mechanisms and careful coordination of schema changes, data corrections, and structural modifications that affect multiple processing layers simultaneously.

Comparative analysis against traditional data warehouse architectures reveals that medallion implementations provide superior flexibility and schema evolution capabilities while potentially sacrificing some query performance optimization. Stream processing benchmarks demonstrate that microservices-based architectures can achieve comparable performance to traditional systems while providing enhanced scalability and deployment flexibility [10]. Traditional data warehouses offer optimized indexing and partitioning strategies specifically designed for analytical workloads, potentially delivering superior query performance for well-defined business intelligence use cases compared to medallion gold layer implementations.

When evaluated against data lake architectures, Medallion frameworks provide enhanced data organization and quality assurance mechanisms while maintaining reasonable storage flexibility. Large organizational data architectures benefit from structured approaches to data quality management and governance that medallion architectures provide, addressing common challenges associated with data lake implementations, including data discovery, quality assurance, and governance enforcement [9]. The medallion approach effectively combines the flexibility advantages of data lakes with the structured quality assurance mechanisms traditionally associated with data warehouse implementations, creating a hybrid solution that addresses diverse organizational requirements while maintaining operational efficiency and data governance standards across different processing layers and analytical use cases.

Architecture Type	Key Characteristics
Medallion Architecture	Enhanced flexibility and schema evolution
Traditional Data Warehouse	Optimized indexing and partitioning
Data Lake Architecture	Maximum storage flexibility
Microservices Framework	Comparable performance with enhanced scalability
Multi-layered Approach	Balanced resource allocation capability
Hybrid Solution	Combined flexibility and quality assurance

Table 4: Data Architecture Performance Comparison [9,10]

Conclusion

The medallion architecture is a good way to deal with current data engineering problems. It gives companies a step-by-step way to change basic data into useful business tools. Because of its layered

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

way of processing, this architecture handles key issues like making sure data is good, being able to grow as needed, and following rules while still being able to change within different company settings. The way this three-level system improves data bit by bit allows companies to keep all their data safe while also making data sets ready for business use. This meets the need to keep raw data and make it easy to obtain for processing and evaluation. How it's set up stresses picking the right tech, adding processing power slowly, and having strong rules to help big companies use it while keeping data safe as it changes. When compared to others, the medallion architecture is better at changing schemas than old-style data warehouses. It also has better ways to arrange and check data quality than normal data lakes. The mix-and-match nature of this setup combines storage that can be changed with organized quality control. This creates a balanced fix that handles different company needs across different processing levels and uses. As businesses keep seeing data as a key tool, the medallion architecture will be more and more important in helping them make good data-based decisions and stay ahead in fast-changing markets.

References

[1] Sanath Chilakala, "Enterprise Data Architectures: A Comprehensive Analysis of Modern Solutions, Market Trends, and Implementation Frameworks", ResearchGate, Feb. 2025. Available: https://www.researchgate.net/publication/389633154_Enterprise_Data_Architectures_A_Comprehensive_Analysis_of_Modern_Solutions_Market_Trends_and_Implementation_Frameworks

[2] Thanda Shwe and Masayoshi Aritsugi, "Optimizing Data Processing: A Comparative Study of Big Data Platforms in Edge, Fog, and Cloud Layers", MDPI, 2024. Available: https://www.mdpi.com/2076-3417/14/1/452

[3] Forvis Mazars, "Modern Data Architecture: Mine, Refine, & Turn Data Into Digital Gold", 1st Jul. 2025. Available: https://www.forvismazars.us/forsights/2025/7/modern-data-architecture-mine-refine-turn-data-into-digital-gold

[4] Sheng Huang, "Big data processing and analysis platform based on deep neural network model", ScienceDirect, 2024. Available: https://www.sciencedirect.com/science/article/pii/S277294192400036X

[5] Suman De and Vinod Vijayakumaran, "A Multi-tier Architecture for Soft and Hard Real-Time Systems Involving Multiple Data Sources for Efficient Data Processing", Springer Nature, 2020. Available: https://link.springer.com/chapter/10.1007/978-981-15-7961-5_1

[6] Mounika Kothapalli, "The Challenges of Data Quality and Data Quality Assessment in the Big Data", ResearchGate, 2023. Available: https://www.researchgate.net/publication/370129565_The_Challenges_of_Data_Quality_and_Data_Quality_Assessment_in_the_Big_Data

[7] Syed Ziaurrahman Ashraf, "Designing Scalable Data Architectures for Enterprise Data Platforms", IJFMR, 2023. Available: https://www.ijfmr.com/papers/2023/1/23892.pdf

[8] Elon Oliveira Albuquerque et al., "Data Pipelines Implementation and Management for Data Engineering: A Case Study Applied to the Public Sector", Springer Nature, 27th Jul. 2025. Available: https://link.springer.com/chapter/10.1007/978-3-031-93106-2_18

[9] Fathima Nuzla Ismail et al., "Big Data Architecture for Large Organizations", arXiv, May 2025. Available: https://arxiv.org/html/2505.04717v1

[10] Sören Henning and Wilhelm Hasselbring, "Benchmarking scalability of stream processing frameworks deployed as microservices in the cloud", ScienceDirect, 2024. Available: https://www.sciencedirect.com/science/article/pii/S0164121223002741

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.