

Understanding Data Layers in the Medallion Architecture

Optimize your data pipelines by leveraging Source, Bronze, Silver, and Gold layers.

The Medallion Architecture organizes data into layers for better management and processing. Each layer has a specific purpose in the data pipeline.

Building reliable, performant data pipelines with  **DELTA LAKE**



Source Layer

Purpose: Ingest raw data from various systems.

- **Data Type:** Unstructured or semi-structured (CSV, JSON, XML, etc.).
- **Key Features:**
 - Direct input from source systems (files, APIs, databases).
 - No transformations applied, only ingestion.

Bronze Layer

Purpose: Store raw ingested data with minimal processing.

- **Key Features:**
 - Captures full history of data, including metadata (e.g., timestamps).
 - Append-only for auditability and data lineage.
 - Stored in Delta Lake format for reliability and performance.

Use Case: Serve as the base for further processing in downstream layers.



Silver Layer

Purpose: Ingest, validate and enrich data for analytics and reporting.

- **Key Features:**

- Cleansed and deduplicated data.
- Schema enforcement for data quality.
- Contains enriched, transaction-level data.

Use Case: Provides a single source of truth for analytics.

Gold Layer

Purpose: Deliver refined data for business intelligence (BI) and machine learning (ML).

- **Key Features:**

- CapAggregated and pre-computed data for faster queries.
- Optimized for dashboarding, reporting, and ML models.
- Fine-grained access control for sensitive data.

Use Case: Powers decision-making through real-time analytics and dashboards.



Data Flow Between Layers

- **Source to Bronze:** Raw data ingestion.
- **Bronze to Silver:** Data cleansing, deduplication, and schema validation.
- **Silver to Gold:** Data aggregation and preparation for analytics.

Advantages of Medallion Architecture

- **Scalability:** Handle increasing volumes of data with ease.
- **Data Lineage:** Track changes across layers.
- **Performance Optimization:** Query-ready data at every stage.
- **Flexibility:** Supports batch and streaming data seamlessly.

Use Case Example

Retail Data Pipeline:

- **Source:** Ingest data from POS systems, customer feedback.
- **Bronze:** Store raw transactional data.
- **Silver:** Enrich with product details and customer profiles.
- **Gold:** Generate sales performance dashboards and ML-driven insights.



Conclusion

Organizing your data pipelines with the Medallion Architecture ensures clean, reliable, and high-performance data for your analytics and machine learning needs.



**Follow for more
content like this**



Abhishek Agrawal

Azure Data Engineer

