

# Medallion Architecture

The **Medallion Architecture** is a structured **data lakehouse design** that organizes data into multiple layers to improve **quality, performance, and governance**.

## 1. Bronze Layer (Raw Data)

- **Purpose:** Acts as a landing zone for raw, unprocessed data.
  - **Characteristics:**
    - Contains **raw ingested data** from various sources (e.g., databases, APIs, IoT devices, logs).
    - No transformations or validations applied.
    - Retains historical data for auditing and reprocessing if needed.
  - **Storage:** Typically stored in **cheap, scalable storage** (e.g., cloud storage like S3, ADLS, or HDFS).
  - **Example:** A table storing raw JSON logs from an application.
- 

## 2. Silver Layer (Cleansed & Transformed Data)

- **Purpose:** Cleans, standardizes, and enriches raw data.
  - **Characteristics:**
    - **Deduplicated** and **filtered** to remove bad records.
    - Data types are corrected (e.g., timestamps formatted properly).
    - **Joins & enrichments** may be applied (e.g., adding metadata).
  - **Storage:** Still in a **data lake**, but optimized for querying (e.g., using Delta Lake or Parquet).
  - **Example:** A table with application logs where errors have been categorized, missing values imputed, and redundant data removed.
- 

## 3. Gold Layer (Business-Ready Data)

- **Purpose:** Provides high-quality, aggregated data for analytics and reporting.
  - **Characteristics:**
    - Data is **modeled** for business use cases (e.g., star schema, OLAP).
    - Used for **BI dashboards, machine learning models, or executive reports**.
    - May contain **aggregations** (e.g., daily revenue per product).
  - **Storage:** Optimized for performance, often in a **data warehouse** (e.g., Snowflake, BigQuery, Redshift).
  - **Example:** A table showing monthly revenue trends per customer segment.
- 

## Summary of Differences

### *1. Bronze Layer*

- **Data Type:** Raw, unprocessed
- **Processing:** Minimal
- **Use Case:** Storage & auditing

### *2. Silver Layer*

- **Data Type:** Cleaned, structured
- **Processing:** Standardized
- **Use Case:** Analytics & ML models

### *3. Gold Layer*

- **Data Type:** Aggregated, optimized
- **Processing:** Business logic applied
- **Use Case:** BI, dashboards, reports