SIGMOID

Best practices for **data management** using medallion architecture

The medallion architecture is a design pattern for data lakehouses that helps organizations effectively manage, and analyze data at scale. A structured yet flexible arrangement of the data layers evolves to meet changing requirements and optimizes performance. However, some best practices must be followed at each step of this multi-layered approach:

1. Data ingestion and extraction

- Allow seamless push-based data upload from source systems to internal storage with whitelisted IP addresses
- Enable business users to upload data directly to a single internal storage location
- Use storage-agnostic connectors for pulling data from various databases

2. Data storage

- Store raw data in blob storage with pre-determined retention periods based on business needs
- Apply consistent naming conventions and coding norms for database objects
- Add audit parameters to Bronze layer tables and ensure schema adherence to the source
- Transform data in the Bronze layer to the Silver layer by standardizing formats, removing duplicates, and documenting transformation logic with important assumptions
- Choose appropriate star or snowflake schemas for Gold layer tables based on business or project needs.

3. Data management and security

- Enforce schema evolution, track data lineage, and maintain clear documentation for data processing, transformation, and governance processes across layers
- Avoid storing PII in any layer, encrypt sensitive data, and implement granular access controls to ensure end-to-end security

4. Data consumption and monitoring

- Leverage dashboards with off-the-shelf tools for quick data consumption and analysis. Refresh data occasionally to optimize cost and computation time
- Optimize query performance with techniques like partitioning, indexing, and caching. with auditing tools

The Medallion Architecture offers a structured, streamlined, and efficient way to manage data. With the right best practices in place, this framework can be a game-changer for any organization looking to harness the power of their data.

Contact our experts to implement best practices for Data Engineering, DataOps, and Data Management across all your initiatives.

Contact us