The medallion architecture

A standardized approach to ELT?

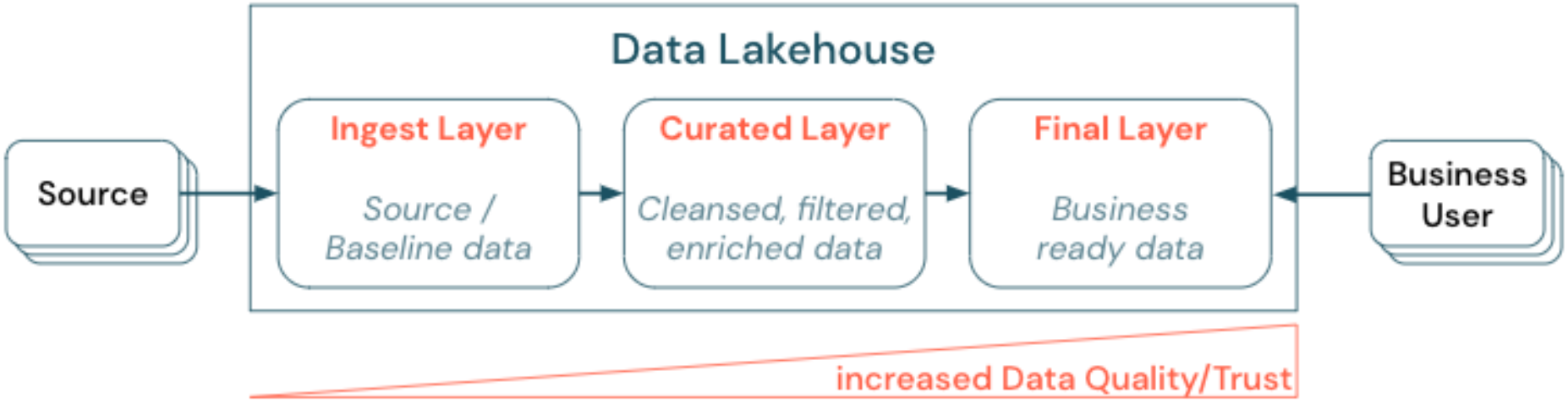
Abdel Dadouche
DJZ Consulting

adadouche@hotmail.com
@adadouche

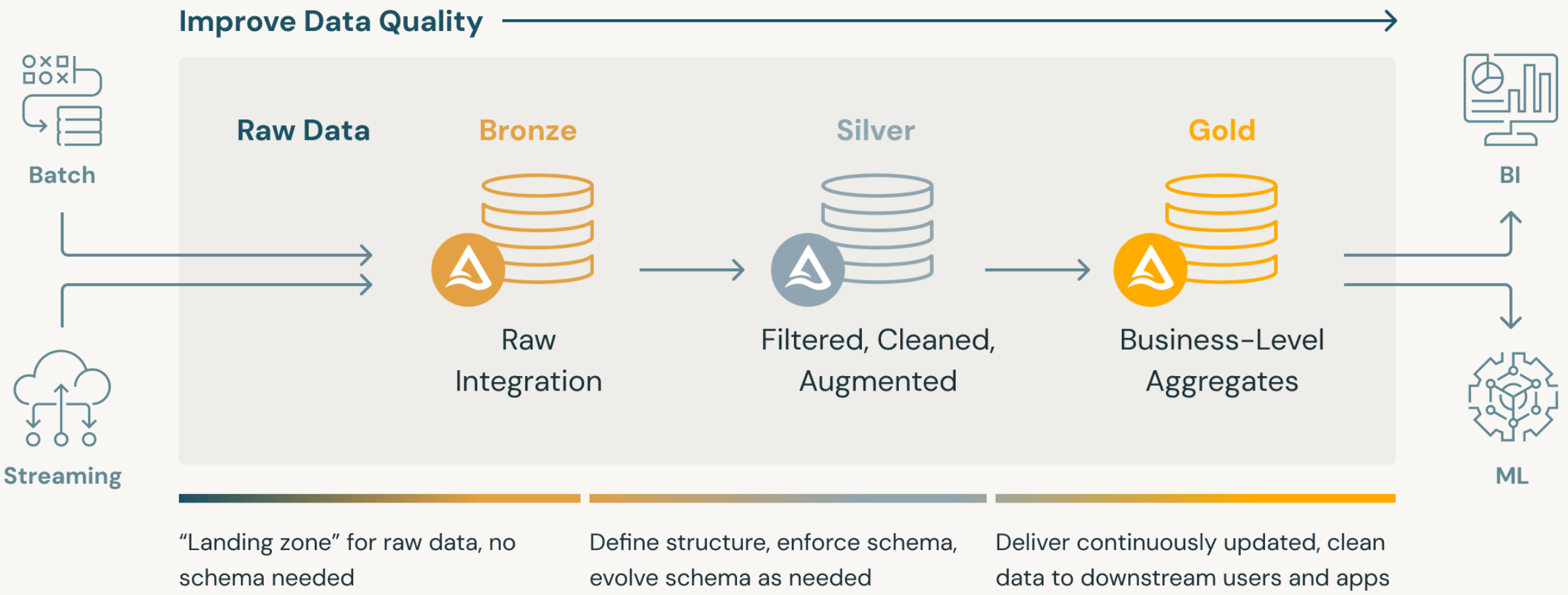
What is a medallion architecture?

What is a medallion architecture?

- Curating data is essential to creating a high-value data lake for BI and ML/AI.
- Treat data like a product with a clear definition, schema, and lifecycle
- Ensure semantic consistency and that the data quality improves from layer to layer so that business users can fully trust the data.



What is a medallion architecture?



Source: https://www.databricks.com/glossary/medallion-architecture

Medallion architecture – Bronze Layer

- The **Bronze layer** is where lands all the data from external source systems
- The table structures in this layer correspond to the source system table structures "as-is," along with any additional metadata columns that capture the load date/time, process ID, etc.
- The focus in this layer is quick “Change Data Capture” (CDC) and the ability to provide an historical archive of source (cold storage), data lineage, auditability, reprocessing if needed without rereading the data from the source system.

Medallion architecture – Silver Layer

- In the **Silver layer** of the lakehouse, the data from the Bronze layer is matched, merged, conformed and cleansed ("just-enough")
- Provide an "Enterprise view" of all its key business entities, concepts and transactions. (e.g. master customers, stores, non-duplicated transactions and cross-reference tables).
- Enables self-service analytics for ad-hoc reporting, advanced analytics and ML.
- Only minimal or "just-enough" transformations and data cleansing rules are applied while loading the Silver layer.
- From a data modeling perspective, the Silver Layer has more “3rd-Normal Form” like data models.

Medallion architecture – Gold Layer

- Typically organized in consumption-ready "project-specific" databases.
- Used for reporting and uses more de-normalized and read-optimized data models with fewer joins.
- The final layer of data transformations and data quality rules are applied here.
- Often use Kimball style star schema-based data models or Inmon style Data marts fit in the Gold Layer of the lakehouse.
- Final presentation layer of projects such as Customer Analytics, Product Quality Analytics, Inventory Analytics, Customer Segmentation, Product Recommendations, Marking/Sales Analytics etc. fit in this layer.

Benefits of a data lakehouse medallion architecture

Improved Data Quality & Consistency

- **Bronze Layer:** Raw, ingested data (often messy or semi-structured).
- **Silver Layer:** Cleansed, de-duplicated, and conformed data.
- **Gold Layer:** Curated, business-ready datasets.
- Ensures **clean, reliable data** for downstream analytics and ML.

Benefits of a data lakehouse medallion architecture

Supports Multiple Workloads

- Bronze: For data scientists needing raw, detailed data.
- Silver: For analysts needing cleansed, normalized datasets.
- Gold: For BI dashboards, executive reporting, ML features.
- Enables **self-service analytics** without compromising raw data access.

Benefits of a data lakehouse medallion architecture

Simplified ETL/ELT Processes

- Data transformations are **incremental** across layers.
- Developers can **apply "just-enough" transformations** in the Silver layer, and complex business rules in the Gold layer.
- Encourages **modular pipelines** and easier debugging.

Benefits of a data lakehouse medallion architecture

Scalability & Flexibility

- Each layer can scale independently.
- Raw data can be retained indefinitely in Bronze, enabling **historical replay or reprocessing**.
- Supports **structured, semi-structured, and unstructured data** seamlessly.

Benefits of a data lakehouse medallion architecture

Auditability & Governance

- Layered approach provides a **clear lineage** from raw to curated data.
- Easier to implement **access control**, data quality monitoring, and compliance tracking.

Medallion architecture example