Labs 1–3 Summary: Docker Stack, Delta Lake, and Apache Iceberg

Z5008 Big Data Lab · Even Semester 2026 · Dr. Innocent Nyalala

1. Medallion Lakehouse Architecture

Data progresses from raw immutable **Bronze** landing files through validated, deduplicated **Silver** tables to business-ready **Gold** aggregates. Partitioning Parquet tables on low-cardinality filters (such as `region` or `event_date`) enables partition pruning during scans.

2. Delta Lake and Apache Iceberg ACID Guarantees

Both table formats layer optimistic concurrency control and immutable metadata logs (`_delta_log` in Delta; manifest lists and snapshots in Iceberg) atop object storage (MinIO/S3), enabling `MERGE INTO` upserts, schema evolution, hidden partitioning, and reproducible time-travel queries.