1. Azure Data Architecture with Databricks
   1. Raw Zone/Streaming Platform
   2. Delta Lake (Medallion Architecture)
      1. Gradually smaller inputs via filtering
      2. Reasons for a globally non-redundant Bronze Zone
         1. Always easier to read it than the Raw Zone
         2. Faster to read it than the Raw Zone
         3. Doesn’t need to be redeployed because of changing needs
         4. Unifies Streaming Platforms and Data Lake Raw Zones
      3. Reasons for a (Data Mart) locally non-redundant Silver Zone
         1. Deduplication performance problem -> should be non-redundant
         2. Deduplication subjectivity (pre-filtering, choosing fields) -> should be redundant
      4. Reasons for a Gold Zone
         1. To minimize Silver Zone redundancy and frequency of redeployment.
   3. Databricks SQL, Photon and Hive UDFs (Self-Service & Late Arriving Data)
      1. Terrible join performance otherwise
      2. Somewhat terrible aggregation performance otherwise
      3. Either data loss, unscalable batch pipelines or inefficient client-side ETL
2. Data Mesh
   1. The Mesh for Decentralization
      1. Domain-oriented data (Source-aligned)
      2. Cross-domain consumption (Consumer-aligned)
   2. Microservice-like ETL for Collaboration
      1. Multiple Projects
      2. Polyrepo
   3. Data Governance:
      1. Extensions & Tools
      2. Standardization via composition
      3. Best practices, sample project & code review
   4. Why the Bronze Zone is the best Data Product:
      1. High stability & low frequency of change compared to Data Marts
      2. Minimal information loss/ETL subjectivity compared to Data Marts
      3. Always easier to read it than the Raw Zone
      4. Faster to read it than the Raw Zone