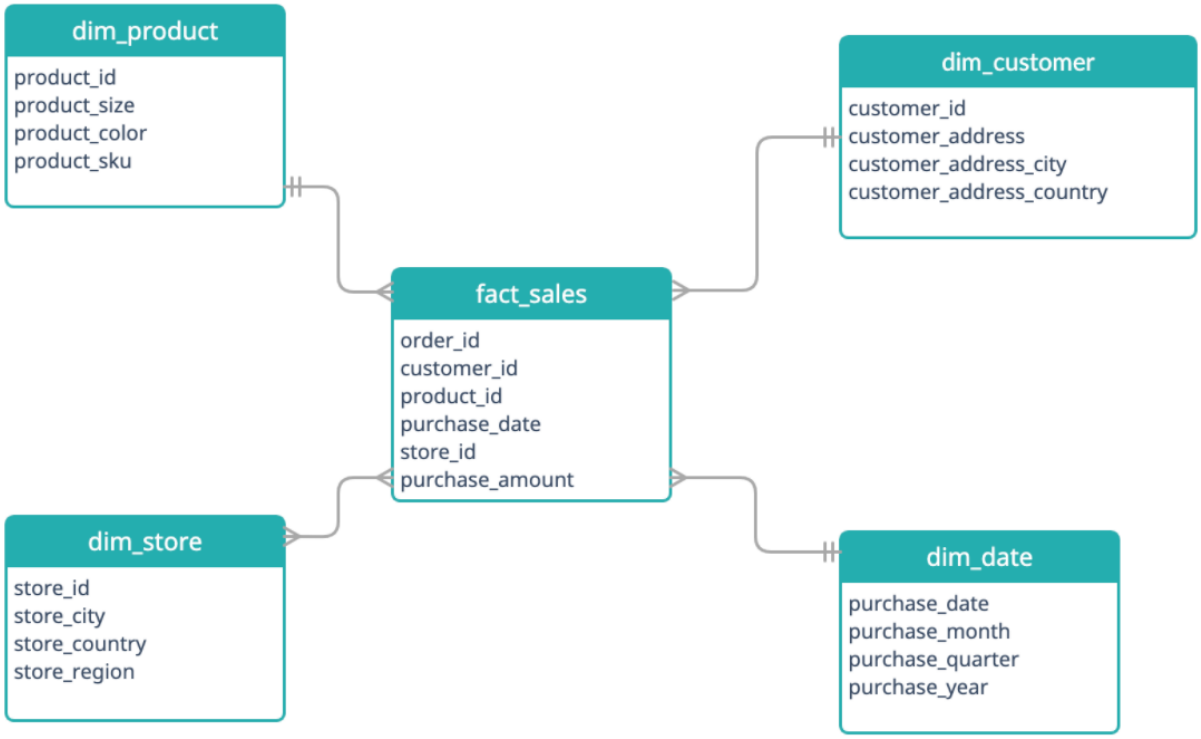
Chapter 5 - DATA WAREHOUSE DESIGN METHODOLOGIES

Kimball Methodology

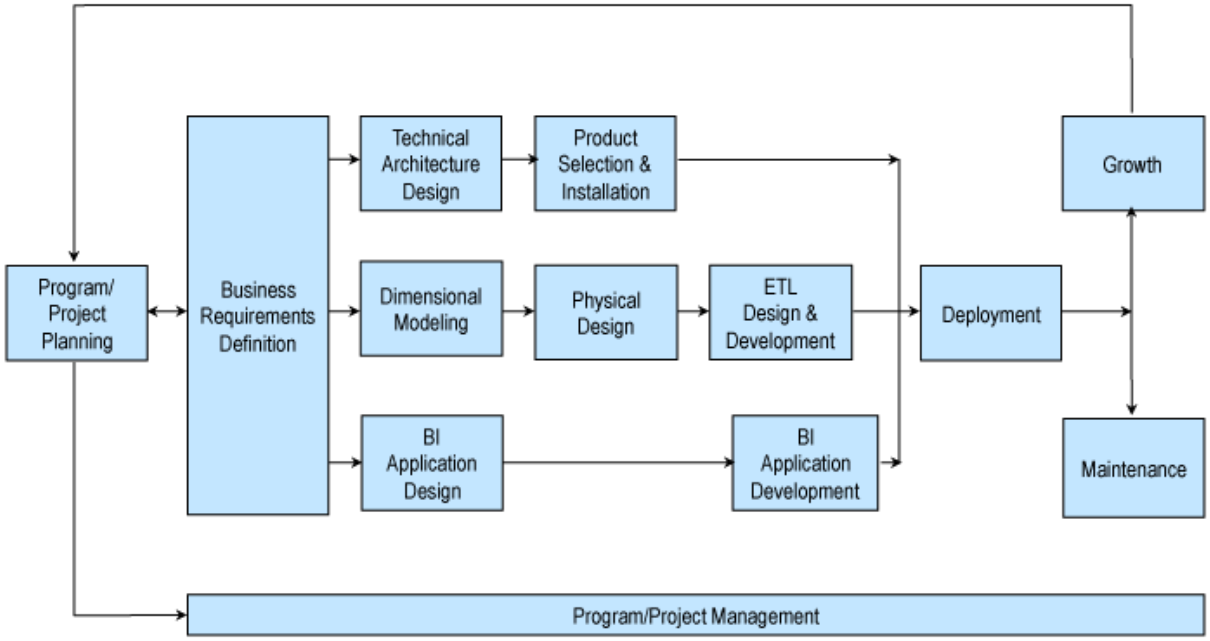
I. Overview

- Defines data warehouse as “a copy of transaction data specifically structured for query and analysis”
- Starts with identifying key business process and requirements (Bottom-Up)
- Focus of this approach is to enable Business Intelligence fast
- Data Marts are created first instead of enterprise data warehouse
- Dimensional Model – STAR SCHEMA (denormalized)
- The model design is built on fact and dimension tables

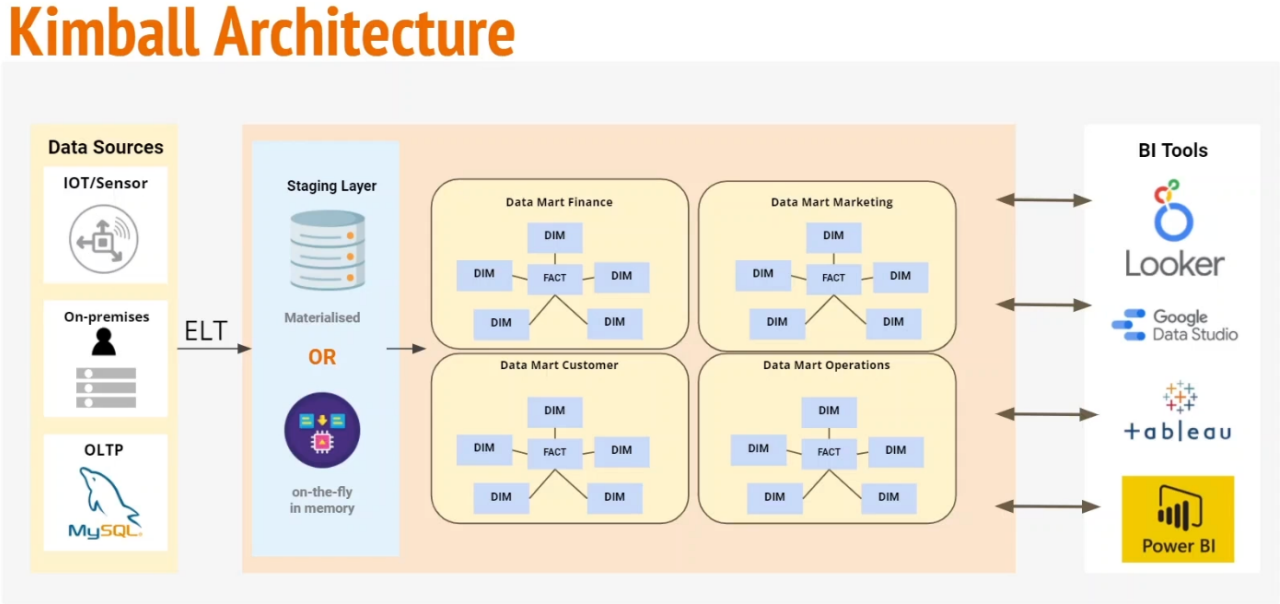
II. Processes of Kimball Methodology



- Follows dimensional modelling technique
- Conformed data dimensions (shared attributes) are used to create dimensional data warehouse
- Multiple Star Schema can exist in same model



III. Kimball Architecture



Data Vault Methodology

I. Overview

- Hybrid Approach (3NF + Dimensional Model/Star Schema)
- Store all your RAW data and then use it for the specific business need
- Every entity table (hubs) are connected through a link key table
- Insert Only Architecture
- Hash keys are used instead of composite or surrogate keys
- Data Marts and built on top of data vault as views to enable reporting
- Complex data model but easier built incrementally
- Handles streaming data, structured, and unstructured data at scale

II. Data Vault Components

- Hub
  - Represents business entity (customer, products, employees)
  - Contain unique business keys (Hash)
  - Lost business keys result in losing all information associated
- Link
  - Establish relationship between different business keys
  - Many-to-many relationships (3NF)
  - Link makes easier to add new course
- Satellite
  - Contains attributes of original tables (descriptive data)
  - Similar to Dimensions
  - Subject to change over time
  - Stores change at granular level (Similar to SCD Type 2)