Data Warehouse Design

Knowledge Objectives

1. Distinguish demand and data driven approaches
2. Enumerate project phases
3. Enumerate steps in dimensional modeling
4. Enumerate factual requirements
5. Explain compatibility of facts

Design method

Approaches

- Data-driven/Supply-driven
- Requirement-driven/Demand-driven

Project phases

I. Technology
II. Data
    1. Dimensional modeling
        - Logical schema
        - Relationships with sources
    2. Physical design
        - Deployment
        - Optimization
            - Indexing
            - Partitioning
            - Materializing
    3. Data staging design and implementation
        - Extraction
        - Transformation
            - Quality improvement
            - Data preparation
        - Load
III. Applications

Dimensional modeling

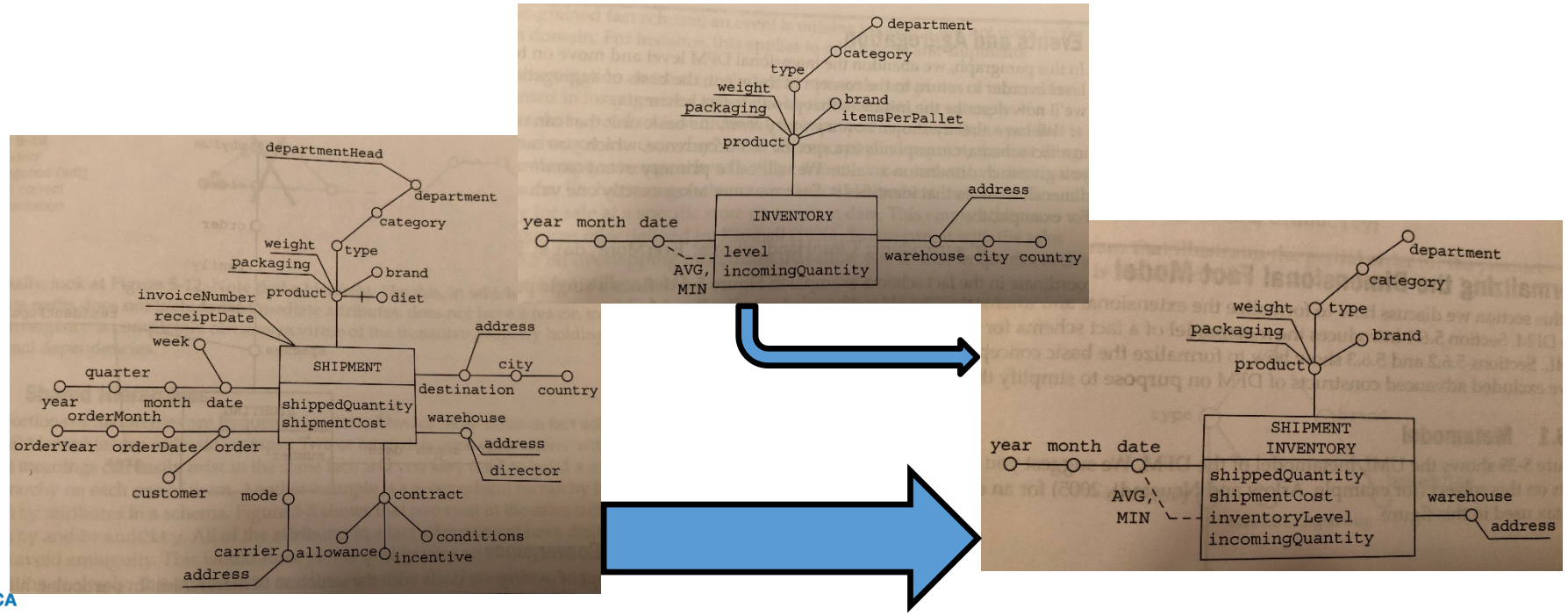
1. Analyse sources
   a. Understand available source schema
   b. Rengineer them to discover unexpressed relationships
   c. Identify data useful for decision making
   d. Assess data quality
   e. Align different sources
2. Requirement analysis
   a. Identify facts
      - Determine granularities
   b. Determine volumes and workloads
3. Create a star schema for each fact
   - Includes measures, dimensions and hierarchies
4. Translate the schema into relational tables
5. Validate the queries

Factual requirements

- Measures
  - Additivity
- Dimensions
  - Hierarchies
- Queries
  - Granularities
- History length

Overlapping Compatible Fact Schemata

- Conditions:
  - a) Comparable fact schemata
    - Share a piece of multidimensional space
  - b) Shared dimensions
    - Conformed hierarchies
- Result:
  - a) Unite measures
  - b) Merge hierarchies
  - c) Merge dimensional instances



Closing

Summary

- Demand-driven vs Data-driven
- Project phases
- Dimensional modeling

Bibliography

- M. Golfarelli and S. Rizzi. *Data Warehouse Design*. McGrau-Hill, 2009