Data Warehouse & Business Intelligence Fundamentals

Todor Kichukov

todor.kichukov@bipartner.biz

https://www.facebook.com/groups/SUDWBI2022/

Faculty of Mathematics and Informatics

Sofia University

2022

Data Warehouse & Business Intelligence Fundamentals

Course Scope

- DW Concept
- DW Architecture
- DW Data Modeling
- Data Integration
- Gathering and Analyzing Requirements
- Business Intelligence
- Deployment, Support and Maintenance

Data Warehouse Data Modeling Part II

- Dimensional Modeling Techniques
- Slowly Changing Dimensions (SCD)
- Common Dimensional Modeling Mistakes
- A Dimensional Modeling Manifesto
- Terminology
- Project Work

Dimensional Modeling Techniques

- Date Dimension – Date as PK, level-based hierarchy
- Conformed Dimensions – single dimension implementation, multiple usage
- Dimension Roles
- Conformed Facts – a discipline in facts naming, calculation and consistent usage in the enterprise
- Degenerate Dimension – for Trx Number, Order Number, etc.
- Junk Dimension – grouping of low-cardinality flags
- Surrogate keys – single integer, based on sequence (or hash key)

Slowly Changing Dimensions (SCD)

Defined by Ralph Kimball as Types 0 to 7

- Type 0 – no change at all (useful for non-changing dimensions)
- Type 1 – overwrite values
- Type 2 – keep history of changes, adding new dimension row

Useful links:

- [Kimball - SCD Type 1](#)
- [Kimball - SCD Types 2, 3](#)
- [Kimball - SCD Types 0, 4, 5, 6, 7](#)

Common Dimensional Modeling Mistakes

1. Fail to conform facts and dimensions across the data warehouse
2. Expect users to query the lowest-level data in a normalized model
3. Design the dimensional model based on a specific report
4. Neglect to declare and then comply with the fact table's grain
5. Use operational or smart keys to join dimensions to facts
6. Solve all query performance problems by adding more hardware
7. Ignore the need to track dimension attribute changes
8. Split hierarchies and their levels into multiple dimensions
9. Limit verbose descriptive attributes in dimensions to save space
10. Place text attributes used for constraining and grouping in fact tables

A Dimensional Modeling Manifesto

- DBMS Journal, Volume 10 Issue 9, Aug 1997
- Ralph Kimball - A Dimensional Modeling Manifesto - Drawing the Line Between Dimensional Modeling and ER Modeling Techniques

Terminology

- Conformed Dimensions, Conformed Facts
- Dimension Role
- Degenerate Dimension, Junk Dimension
- Slowly Changing Dimension (SCD)

Project Work

- Today
  - CBS (source) logical model
  - Staging Area logical model (draft)
  - Clarify all open questions
- Next steps
  - Justification of selected DWH building approach (draft)