



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 I

- Data Modeling Levels
- Data Modeling Techniques
- Data Types / Data Categorization
- Dimensional Modeling Fundamentals
 - Star schema
 - Fact table
 - Dimension table
 - Hierarchies
 - From 3NF to Dimensional
 - Four-Step Dimensional Design Process
- Terminology

Data Modeling Levels 1/2

- Conceptual Data Model
- Logical Data Model
- Physical Data Model

| Feature | Conceptual | Logical | Physical |
|----------------------|------------|---------|----------|
| Entities | ✓ | ✓ | |
| Entity Relationships | ✓ | ✓ | |
| Attributes | | ✓ | |
| Primary Keys | | ✓ | ✓ |
| Foreign Keys | | ✓ | ✓ |
| Tables | | | ✓ |
| Columns | | | ✓ |
| Column Data Types | | | ✓ |

Data Modeling Levels 2/2

- Conceptual Data Model
- Logical Data Model
- Relational Data Model
- Physical Data Model

| Feature | Conceptual | Logical | Relational | Physical |
|-----------------------|------------|---------|------------|----------|
| Concepts | ✓ | | | |
| Concept Relationships | ✓ | | | |
| Entities | | ✓ | | |
| Entity Relationships | | ✓ | | |
| Attributes | | ✓ | | |
| Unique Identifiers | | ✓ | | |
| Views | | ✓ | | |
| Tables | | | ✓ | |
| Columns | | | ✓ | |
| Indexes | | | ✓ | |
| Table Relationships | | | ✓ | |
| Relational Views | | | ✓ | |
| Clusters | | | | ✓ |
| Contexts | | | | ✓ |
| Dimensions | | | | ✓ |
| Directories | | | | ✓ |
| Disk Groups | | | | ✓ |
| External Tables | | | | ✓ |
| Indexes | | | | ✓ |
| Roles | | | | ✓ |
| Rollback Segments | | | | ✓ |
| Segments | | | | ✓ |
| Sequences | | | | ✓ |
| Stored Procedures | | | | ✓ |
| Synonyms | | | | ✓ |
| Tables | | | | ✓ |
| Tablespaces | | | | ✓ |
| Users | | | | ✓ |
| Views | | | | ✓ |

Data Modeling Techniques

- Entity-Relationship Diagram?
- 1-2-[3]-4-5-6 NF (Normal Form) Normalized Data Model
- Dimensional (Denormalized, Star-Schema) Data Model
- Multidimensional (Cube) Data Model
- Data Vault Model

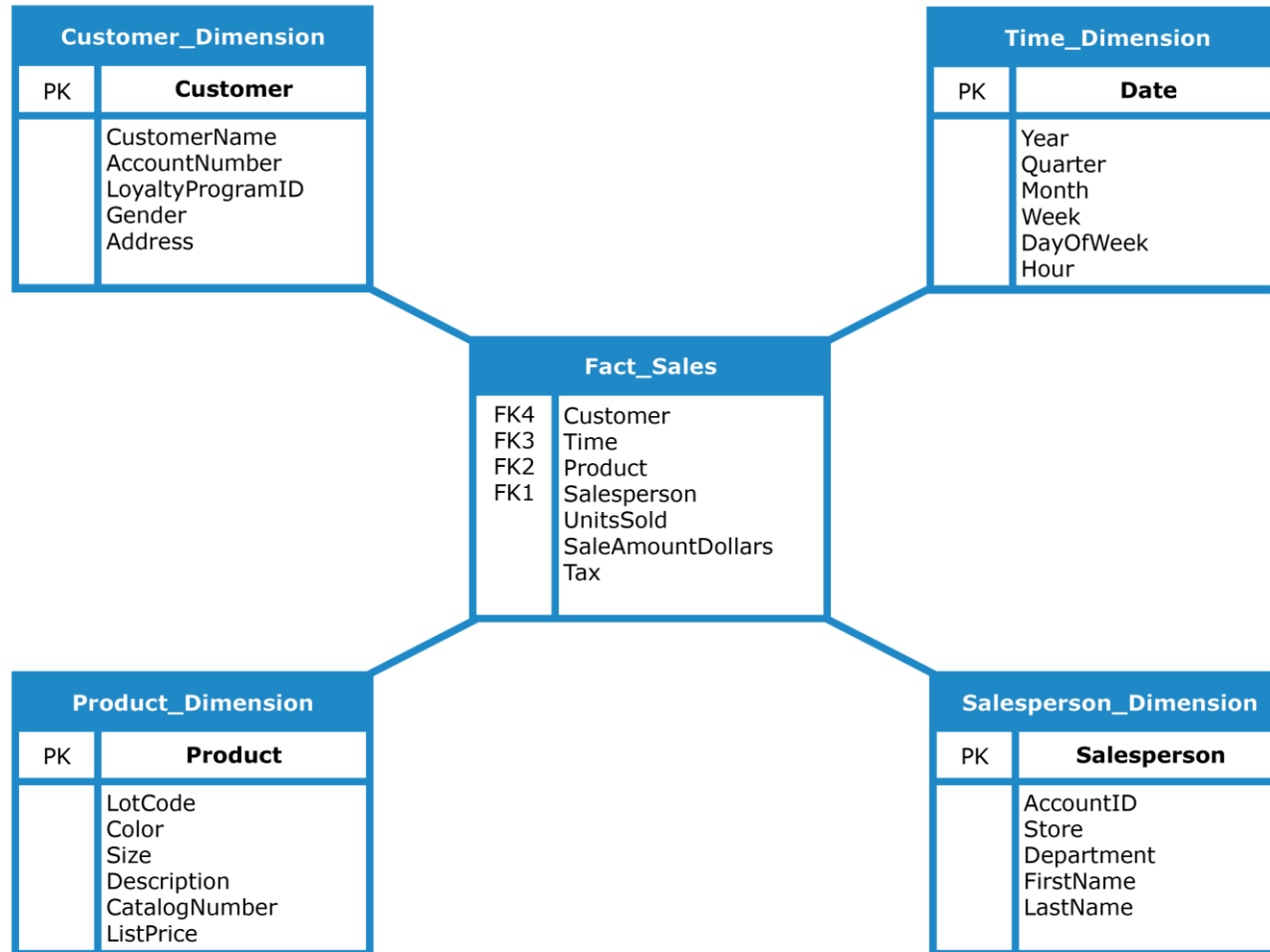
Data Types / Data Categorization

- Number - Text - Date
- Quantitative - Qualitative
- Base - Derived
- Detailed - Aggregated
- Data Categories
 - **Reference Data** – Nomenclatures, List of Values, Hierarchies, Flags, ...
 - **Master Data** – People (Customers, Prospects, Employees, Vendors, Suppliers, ...), Places (Branches, Sales Territories, ...), Things (Products, Packages, Accounts, ...).
 - **Transactional Data** – Events, Logs, Snapshots (Periodic, Accumulating) like Sales Orders, Invoices, Purchase Orders, Shipping Documents, Insurance Claims, ...

Dimensional Modeling Fundamentals

Star Schema - Facts & Dimensions Together

- Simplicity, Symmetry
- Performance benefits



Legend: PK = primary key, FK = foreign key

Dimensional Modeling Fundamentals

Fact Table

- Stores numerical measurements
- Row in the fact table = one measurement
- Grain = level of detail
- Fact = business measure (indicator); sometimes the term is used to represent the whole measurement (which usually is called event or snapshot)
- Additive – Semi-Additive – Non-Additive Facts
- Foreign Keys, Primary Key
- Fact Table Types
 - Transactions (Events)
 - Periodic Snapshots
 - Accumulating Snapshots

Dimensional Modeling Fundamentals

Dimension Table

- Contains textual descriptors of the business; entry point to the fact tables
- Serves for grouping, filtering and report labeling
- Dimensions usually have less rows than fact tables, but more columns
- Single PK (artificially created, no business meaning, few exceptions)
- Usually 0 / 1 / many hierarchies built on every dimension

Dimensional Modeling

Fundamentals

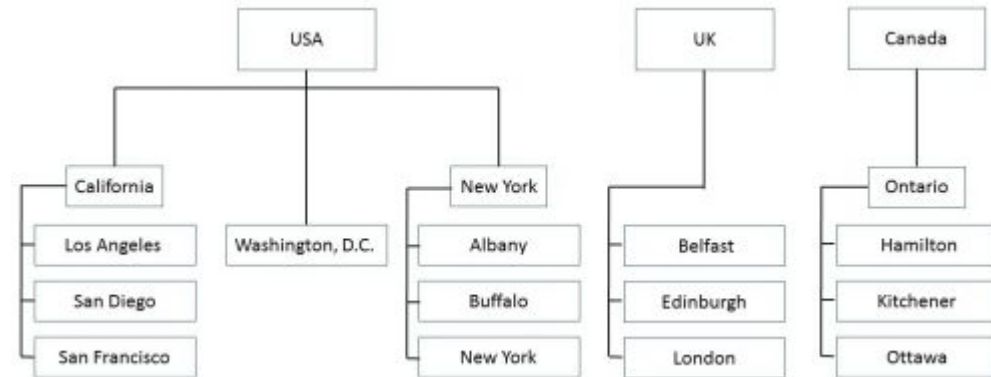
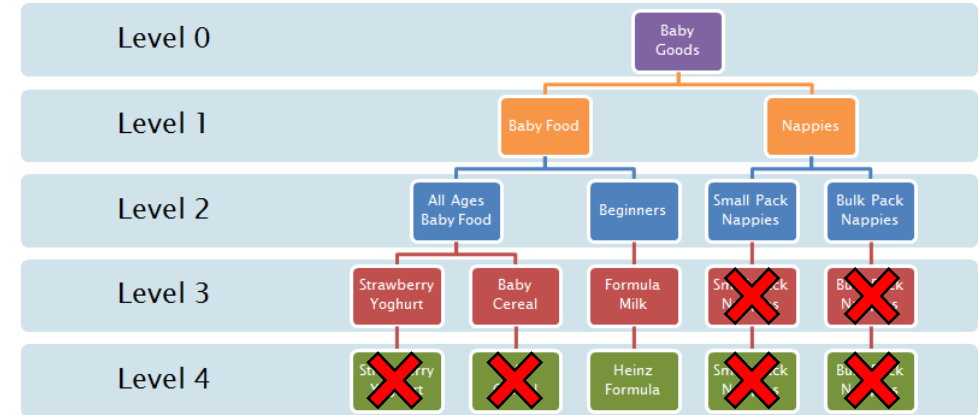
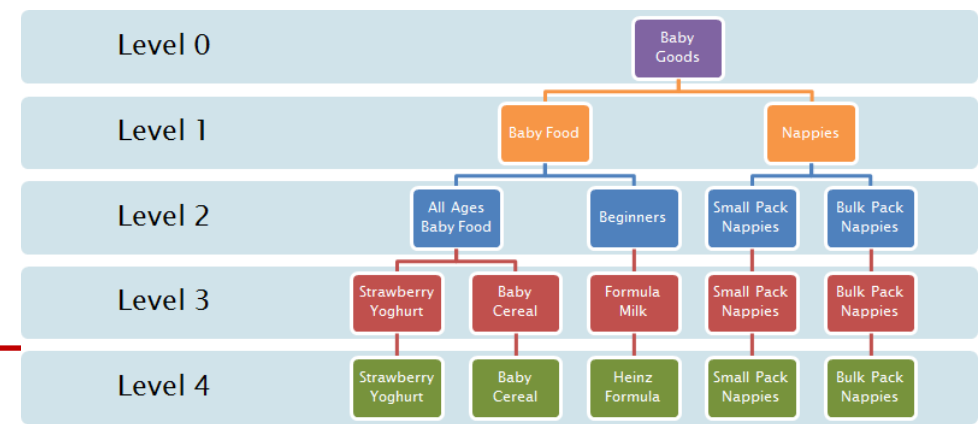
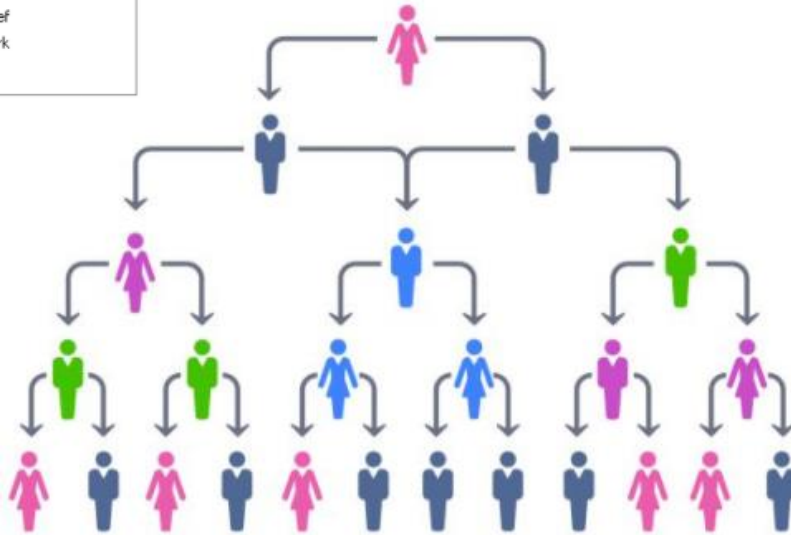
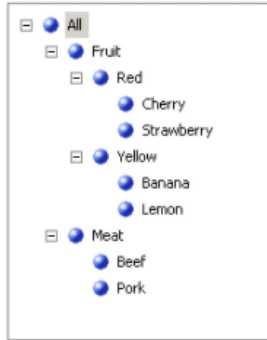
Hierarchies

- Value-based, Level-based
- Balanced, Unbalanced, Ragged

`select * from Products`

Results Messages

| | ProductID | ParentID | Name |
|----|-----------|----------|------------|
| 1 | 1 | NULL | All |
| 2 | 2 | 1 | Fruit |
| 3 | 3 | 2 | Red |
| 4 | 4 | 3 | Cherry |
| 5 | 5 | 3 | Strawberry |
| 6 | 6 | 2 | Yellow |
| 7 | 7 | 6 | Banana |
| 8 | 8 | 6 | Lemon |
| 9 | 9 | 1 | Meat |
| 10 | 10 | 9 | Beef |
| 11 | 11 | 9 | Pork |



Dimensional Modeling Fundamentals

From 3NF to Dimensional

1. Separate its discrete business processes and model them separately
2. Select the N:M relationships that contain numeric facts and design them as fact tables
3. Denormalize the remaining tables into flat tables that join directly to the fact tables – the dimensions

Dimensional Modeling Fundamentals

Four-Step Dimensional Design Process (Kimball)

1. Select the business process (it usually leads to one fact table)
2. Declare the grain – what exactly a fact row represents
3. Choose the dimensions
4. Identify the facts (measures, indicators)

Dimensional Modeling Fundamentals

First Dive

- Factless Fact Table
- Date Dimension
 - Year
 - Half
 - Quarter
 - Month
 - (Week)
 - Day

Terminology

- Reference / Master / Transactional Data
- Fact table, fact, measure, indicator, measurement, grain
- Dimension table, hierarchy types

Project Work

- Today
 - Did you analyze the source model and the data?
 - Did you draft the CBS (source) logical model (on paper or electronically)?
 - Clarify all open questions
- Next steps
 - CBS (source) logical model
 - Staging Area logical model (draft)