Data Warehousing

Exploring Levels of Modeling Abstraction

Karolyn Duncan
KRDuncan@att.net
Our Purpose Today

- Understand levels of modeling abstraction
- Review types of data models
- Learn the various warehousing models at all levels of abstraction
# Levels of Modeling Abstraction

## The Zachman Framework

<table>
<thead>
<tr>
<th>Scope</th>
<th>Data (what?)</th>
<th>Function (how?)</th>
<th>Network (where?)</th>
<th>People (who?)</th>
<th>Time (when?)</th>
<th>Motivation (why?)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enterprise Model</td>
<td>list of things</td>
<td>list of processes</td>
<td>list of locations</td>
<td>list of organizations</td>
<td>list of events</td>
<td>list of business goals</td>
</tr>
<tr>
<td>System Model</td>
<td>semantic model</td>
<td>Business process model</td>
<td>Logistics network</td>
<td>Workflow model</td>
<td>Master schedule</td>
<td>Business plan</td>
</tr>
<tr>
<td>Technology Model</td>
<td>logical data model</td>
<td>Application architecture</td>
<td>Distribution architecture</td>
<td>Human interface architecture</td>
<td>Processing schedule</td>
<td>business rule model</td>
</tr>
<tr>
<td>Detailed Representations</td>
<td>data design</td>
<td>System design</td>
<td>System architecture</td>
<td>Presentation architecture</td>
<td>Control structure</td>
<td>Business rule design</td>
</tr>
<tr>
<td>System</td>
<td>data definition</td>
<td>data definition</td>
<td>Network architecture</td>
<td>Security &amp; access architecture</td>
<td>Timing definition</td>
<td>Rule specification</td>
</tr>
</tbody>
</table>

The Zachman Framework provides a framework for understanding and organizing knowledge about a system or enterprise. It includes six perspectives: **Scope**, **Enterprise Model**, **System Model**, **Technology Model**, **Detailed Representations**, and **System**. Each perspective is further divided into **Data**, **Function**, **Network**, **People**, **Time**, and **Motivation** dimensions. This framework helps in systematically addressing various aspects of a system or enterprise, from a high-level abstraction to detailed representations.
Levels of Modeling Abstraction

- Source Data Files
- Source Data File Descriptions
- Source Data Files
- Source Data Structure Model
- Integrated Source Data ERM
- Source Composition
- Source Subjects
- Business Drivers, Goals, & Information Needs
- Facts & Qualifiers
- Target Configuration
- Business Questions
- Warehouse Subjects
- Staging & Warehouse ERMs
- Entity Lifecycle
- Warehouse DDMs
- Peacock Charts
- Staging Structure
- Warehouse Relational & Dimensional Structures
- Staging Phys. DB Design
- Warehouse Phys. DB Design
- Implemented Warehousing Databases
- Structural Models
- Logical Models
- Conceptual Models
- Contextual Models

Adapted from © The Data Warehousing Institute
## Types of Data Models

<table>
<thead>
<tr>
<th>Level</th>
<th>Source</th>
<th>Target</th>
<th>Metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contextual Model</td>
<td>Drivers, Goals, Information needs</td>
<td>Drivers, Goals, Information needs</td>
<td>Metadata Goals</td>
</tr>
<tr>
<td>Conceptual Model</td>
<td>Subject area models</td>
<td>Business questions, fact/qualifier matrix, subject model, target configuration</td>
<td>Requirements matrix</td>
</tr>
<tr>
<td>Logical Model</td>
<td>3NF E-R Model of business elements</td>
<td>E-R and/or dimensional Model, Entity Lifecycle Model, Peacock Chart</td>
<td>Requirements matrix (extended)</td>
</tr>
<tr>
<td>Structural Model</td>
<td>Subject/Venn or Matrix model of data stores</td>
<td>E-R and/or dimensional models extended for warehousing</td>
<td>Determined by tool set or application</td>
</tr>
<tr>
<td>Physical Model</td>
<td>Exists in many forms, but must be mapped to structural model</td>
<td>Star, snowflake or relational schema</td>
<td>Determined by tool set or application</td>
</tr>
</tbody>
</table>
Types of Contextual Data Models

<table>
<thead>
<tr>
<th>Level</th>
<th>Source</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contextual Model</td>
<td>Drivers, Goals, Information needs</td>
<td>Drivers, Goals, Information needs</td>
</tr>
<tr>
<td>Conceptual Model</td>
<td>Subject area models</td>
<td>Business questions, fact/qualifier matrix, subject model, target configuration</td>
</tr>
<tr>
<td>Logical Model</td>
<td>3NF E-R Model of business elements</td>
<td>E-R and/or dimensional Model, Entity Lifecycle Model, Peacock Chart</td>
</tr>
<tr>
<td>Structural Model</td>
<td>Subject/Venn or Matrix model of data stores</td>
<td>E-R and/or dimensional models extended for warehousing</td>
</tr>
<tr>
<td>Physical Model</td>
<td>Exists in many forms, but must be mapped to structural model</td>
<td>Star, snowflake or relational schema</td>
</tr>
</tbody>
</table>

“Contextual modeling provides a view of the scope of the planned data warehousing program. These models communicate understanding of the business environment, establishing a context for analysis.”
Company Context

Business Context

Karolyne’s Kritters™

- Retail sales of dog and cat training, grooming, and dietary products.
- Products include training videos, dog and cat foods, dog and cat dietary supplements.
- Services include dog and cat grooming, boarding, and training.
- Owners are required to become members (membership is free). Membership benefits include a monthly newsletter, coupons, and tips for proper animal care.
- All sales data is associated with membership data.
"Business Driver: An external event to which a corporation chooses to respond"

**Business Drivers**

- Stricter legislative rules governing boarding conditions.
- Different states have different regulations. It's difficult to keep track of these.
- Stricter legislative rules governing dietary supplements.
- Supplements have been a large seller. New regulations are increasing the cost of our dietary supplements.
- Increased competition from discount pet retail stores.
- Discount stores are adding more products and services, like self grooming. It's difficult to stay competitive.
- Increased competition from internet based companies.
Business Goals

“Business goals describe the desired outcomes of taking action.”

Corporate Goals

Karolyn’s Kritters™

- To offer quality boarding for dogs and cats at an affordable price.
- To enhance the lives of dogs and cats with quality foods and dietary supplements.
- To provide quality grooming services.
- To be the premier supplier for dog training.
- To assist owners in solving dog and cat behavioral problems through training and diet.

Functional Goals

Karolyn’s Kritters™ ~ Owner Service

- Provide high quality service to owners.
- Decrease the number of owner complaints by 5% every year.
- Strive for a 100% fill rate on 95% of orders.
- Apply accurate and timely credits to owners.
Information Needs/Biz Questions

“An Information Need uses business terms to describe necessary characteristics for a set of related business questions.”

Information Needs Glossary

Owner Satisfaction

This need is designed to help us better understand owner satisfaction through our knowledge of the sales order lifecycle including orders filled, return of products, and owner complaints. The overall value of the owner (how much money they've spent with us over the years) is not a factor in understanding owner satisfaction at large. Further, an understanding of the customer and their characteristics is not needed to gauge owner satisfaction. Basic franchise information (such as name) is helpful but franchise details (such as owner) aren't needed. Product type is important, however, product characteristics aren't required.
## Types of Conceptual Data Models

<table>
<thead>
<tr>
<th>Level</th>
<th>Source</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contextual Model</td>
<td>Drivers, Goals, Information needs</td>
<td>Drivers, Goals, Information needs</td>
</tr>
<tr>
<td>Conceptual Model</td>
<td>Subject area models</td>
<td>Business questions, fact/qualifier matrix, subject model, target configuration</td>
</tr>
<tr>
<td>Logical Model</td>
<td>3NF E-R Model of business elements</td>
<td>E-R and/or dimensional Model, Entity Lifecycle Model, Peacock Chart</td>
</tr>
<tr>
<td>Structural Model</td>
<td>Subject/Venn or Matrix model of data stores</td>
<td>E-R and/or dimensional models extended for warehousing</td>
</tr>
<tr>
<td>Physical Model</td>
<td>Exists in many forms, but must be mapped to structural model</td>
<td>Star, snowflake or relational schema</td>
</tr>
</tbody>
</table>

“Conceptual models describe data requirements from a business point of view without the burden of technical details.”

© Information Strategies, Inc. 2003 Adapted from TDWI
Information Needs Glossary

Owner Satisfaction
This need is designed to help us better understand owner satisfaction through our knowledge of the sales order lifecycle including orders filled, return of products, and owner complaints. The overall value of the owner (how much money they've spent with us over the years) is not a factor in understanding owner satisfaction at large. Further, an understanding of the customer and their characteristics is not needed to gauge owner satisfaction. Basic franchise information (such as name) is helpful but franchise details (such as owner) aren't needed. Product type is important, however, product characteristics aren't required.

Business Questions:
1. What is the number of owner complaints due to late or inaccurate orders?
2. Show the percent of returns by owner, by product, by reason code.
3. What is the number of calls, per month, received by our 800 service lodging a complaint? Show by complaint code.
4. What percentage of orders are completely fillable at the time they are placed? Show this by month, and product.
Functional Subject Area Diagram Example

S.A. Diagram:
- Represents high level groupings of data
- Is related to dimensions in the dimensional model
- Is associated with fact tables in the dimensional model
The Fact/Qualifier Matrix

- A data model that decomposes business questions into discrete data items.
- Prevents hasty selection of a specific design (relational or dimensional)
- Includes:
  - **Facts**: Those data items the business needs or wants to know.
  - **Qualifiers**: Those data items by which facts are organized and analyzed.
**Conceptual Modeling**

**The Fact/Qualifier Matrix**

1. What is the number of owner returns due to late or inaccurate orders? Include owner. Show this by month.

2. Show percent of returns by owner, by month.

3. Show all owners and their location who have never returned items.

4. Show the percentage of returns due to late orders by owner. Show the list of owners. What is the return dollar value?

5. Show the number of returns by month. Drill through to show number by owner.

<table>
<thead>
<tr>
<th></th>
<th>Late Orders</th>
<th>Inaccurate Orders</th>
<th>No Returns</th>
<th>Month</th>
<th>Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Returns</td>
<td>1</td>
<td>1</td>
<td>1, 5</td>
<td>1, 5</td>
<td></td>
</tr>
<tr>
<td>Owners</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locations</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent of Returns</td>
<td>4</td>
<td></td>
<td>2</td>
<td>2, 4</td>
<td></td>
</tr>
<tr>
<td>Return Dollar Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>
Conceptual Modeling

Targets Configuration – Roles

Objective: Support future, unknown needs
Role: Intake

Objective: Support data movement
Role: Distribution

Objective: Support user needs
Role: Information Delivery

Source Data

Base Tables

Standard Collections

Data Marts

ETL
Hub & Spoke Targets Configuration

- Owners
- Sales
- Products
- Complaints
- Owners
- Data Acquisition ETL Processing
- Base Tables
- Data Mart ETL Processing
- Grooming Analysis Mart
- Data Mart ETL Processing
- Owner Satisfaction
- Owner Satisfaction
- ETL Processing
- Analysis Mart
- Intake
- Access
- Distribution
Data Warehouse Bus Configuration

Requires the use of Conforming Dimensions

Data Sources

Access

Data Staging Tables

Intake & Distribution

© The Data Warehousing Institute
## Logical Modeling

### Types of Logical Data Models

<table>
<thead>
<tr>
<th>Level</th>
<th>Source</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contextual Model</td>
<td>Drivers, Goals, Information needs</td>
<td>Drivers, Goals, Information needs</td>
</tr>
<tr>
<td>Conceptual Model</td>
<td>Subject area models</td>
<td>Business questions, fact/qualifier matrix, subject model, target configuration</td>
</tr>
<tr>
<td>Logical Model</td>
<td>3NF E-R Model of business elements</td>
<td>E-R and/or dimensional Model, Entity Lifecycle Model, Peacock Chart</td>
</tr>
<tr>
<td>Structural Model</td>
<td>Subject/Venn or Matrix model of data stores</td>
<td>E-R and/or dimensional models extended for warehousing</td>
</tr>
<tr>
<td>Physical Model</td>
<td>Exists in many forms, but must be mapped to structural model</td>
<td>Star, snowflake or relational schema</td>
</tr>
</tbody>
</table>

“Logical models are technology oriented designs yet still platform independent.”
Business Questions

Business Question for Owner Satisfaction Information Need

Business Questions:

1. What is the number of owner complaints due to late or inaccurate orders?
2. Show the percent of returns by owner, by product, by reason code.
3. What is the number of calls, per month, received by our 800 service lodging a complaint? Show by complaint code.
4. What percentage of orders are completely fillable at the time they are placed? Show this by month, and product.

Expansion of Questions During the Release:

5. What is the number of owner complaints due to inaccurate orders? Show this by owner.
6. What percentage of our total owner base have ever lodged a complaint. Show this by complaint code.
7. Show all owners and their location who have never complained.
8. What products are bought by the owners who complain the most? Show this by geographic region.
9. Show a side by side comparison, by month, for the last two years, of the number of owners, the number of orders filled, the number of returns, and the number of complaints.
10. Which franchise has the highest number of internal complaints (franchise originated)? Show all franchise data, by franchise, ordered from most complaints to least.
11. Show the top fifty owners complaining due to inaccurate orders. How does this list compare to the owners who complain because of late orders?
12. What is the complete list of complaint codes? What is the most common complaint? Show the most common complaint by franchise.
13. Which franchise has the highest number of external complaints (800 number originated)? Show all franchise data, by franchise, ordered from most complaints to least.
### Decomposed List of Biz Questions

The deliverable should look something like this:

<table>
<thead>
<tr>
<th>Priority</th>
<th>Corporate Goal</th>
<th>Functional Goal</th>
<th>Constituent Group</th>
<th>Biz Question</th>
<th>Dimensions</th>
<th>States</th>
<th>Measures</th>
<th>Time Summary</th>
<th>Source Notes</th>
<th>Additional Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium</td>
<td>To enhance the lives of dogs and cats with quality foods and dietary supplements.</td>
<td>Manage vendor purchasing to ensure quality materials go into our products.</td>
<td>Purchasing Pattern Analysis by Vendor Management and Purchasing Analysts</td>
<td>How many active vendors do we have right now? What has been the trend over the past three years?</td>
<td>Vendor</td>
<td>Active Vendor</td>
<td>Count of active vendors</td>
<td>Calendar Year</td>
<td>Vendor Master Access DB</td>
<td>Multiple sheets; one per franchise; need to look across all franchise data.</td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td></td>
<td></td>
<td>How many vendors do we have by line of business? How has this changed over the past three years?</td>
<td>Vendor Line of Business</td>
<td>All Vendors</td>
<td>Count of all vendors</td>
<td>Calendar Year</td>
<td>Purchasing Excel Spreadsheets</td>
<td>Multiple sheets; one per franchise; need to look across all franchise data.</td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
<td>Which vendors have placed us on credit hold more than three times in the past two years?</td>
<td>Vendor</td>
<td>Purchase Order Credit Hold</td>
<td>Calendar Year / Calendar Month</td>
<td>Purchasing Excel Spreadsheets</td>
<td>Multiple sheets; one per franchise; need to look across all franchise data.</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td>Which vendors have assessed us purchasing penalties?</td>
<td>Vendor</td>
<td>All Vendors</td>
<td>Calendar Month</td>
<td>Purchasing Excel Spreadsheets</td>
<td>Multiple sheets; one per franchise; need to look across all franchise data.</td>
<td></td>
</tr>
</tbody>
</table>
Modeling Example

Owner Subject Area
- Owner
  - Owner Last Name
  - Owner First Name
  - Gender
  - Date of Birth
  - Social Security Number
- Owner Address
  - Street One
  - Street Two
  - City
  - State
  - Zip Code
- Owner Contact
  - Email Type
  - Email Address
  - URL
- Owner Phone
  - Phone Type
  - Area Code
  - Prefix
  - Suffix
  - Extension

Order Subject Area
- Order Header
  - Order Number
  - Order Placed Date
- Order Line
  - Order Line Number
  - Product ID
  - Quantity
  - Price

Product Subject Area
- Product
  - Product ID
  - Product Description
  - Color
  - Height
  - Length
  - Width

Subject Areas

Attributes
State Transition Modeling

Owner
  Owner Last Name
  Owner First Name
  Gender
  Date of Birth
  Social Security Number

Owner Phone
  Phone Type
  Area Code
  Prefix
  Suffix
  Extension

Owner Address
  Street One
  Street Two
  City
  State
  Zip Code

Owner Contact
  Email Type
  Email Address
  URL

Order Header
  Order Number
  Order Placed Date

Order Line
  Order Line Number
  Product ID
  Quantity
  Order Line Status Code
  Shipped Date
  Return Reason Code
  Price

Product
  Product ID
  Product Description
  Color
  Height
  Length
  Width

Product Cost
  Cost Type
  Cost

Process Based Entity
Logical Modeling

State Transition Diagram

States

Order Placed \rightarrow Order Accepted \rightarrow Order on Hold \rightarrow Order Accepted \rightarrow Order Available

Implied Actions

Order Received \rightarrow Order Passed Inspection \rightarrow Order in Inventory

Order Not Delivered \rightarrow Order Canceled

Order Received \rightarrow Order Failed Inspection

Order Returned \rightarrow Order Credited

Order Discarded \rightarrow Order Credited

Order Kept \rightarrow Order in Inventory

Alternate Vendor Found \rightarrow Order Accepted

Order Available \rightarrow Order Canceled

Order Canceled

END
Logical Modeling

Dimension Level Example

Karolyn’s Kritters
Organization Structure

Level

Corporate

Region

East

South

Northeast

Southeast

Central

District

North Central

Mid Central

South Central

Franchise

Wheaton

Naperville

Atlanta

Southwest

Southeast

Tulsa

Members
The Peacock Dimension Chart

All Dimensions Represented

Material
- All
- Class
- Family
- Variety

Vendor Market
- All
- Segment
- Class
- Variety

Organization
- Corporate
- Region
- District

Time
- All
- Year
- Qtr
- Month
- Day

Product
- All
- Class
- Family
- Variety
- Product/SKU

Owner
- All
- Segment
- Class
- Variety
- Owner

Material/SKU
- Vendor
- Franchise
- Event

Logical Modeling
Logical Modeling

Relational or Dimensional Design?

Characteristics of the Data

Metric items of interest

OR

Textual items of interest

User Interaction with the Data

Hunters

Data Miners

Slicing

Dicing

OR

Listing

Reporting

© Information Strategies, Inc. 2003 Adapted from TDWI
“The key to understanding the relationship between DM and ER is that a single ER breaks down into multiple DM diagrams. Think of a large ER diagram as representing every possible business process in the enterprise. … Thus the first step in converting an ER diagram to a set of DM diagrams is to separate the ER diagram into its discrete business processes and to model each one separately.”

* Ralph Kimball, “A Dimensional Modeling Manifesto”; August 1997
Dimensional Modeling at the Logical Level

Logical Modeling

Complaint Dimension
- Complaint-Set-Code
- Complaint-Set-Desc
- Complaint-Code
- Complaint-Reason

Complaint-Count
Order-Count
Return-Order-Count
Return-$-Value

Return Analysis

Time Dimension
- Year
- Quarter
- Month
- Day
- Week

Owner Dimension
- Owner-ID
- Owner-First-Name
- Region-ID
- Region-Name
- Area-ID
- Area-Name

Dimension Levels

Dimension Hierarchy

Developed by TDWI
Identifying Conformed Dimensions

Organization Dimension

Food Sales Paid Fact Table

Time Dimension

Supplement Sales Paid Fact Table

Which Owner Dimension?

Which Owner Dimension?

Training Sales Paid Fact Table

Training Owners

Food Owners

Supplement Owners

Karolyn’s Kritters Owner Master

© Information Strategies, Inc. 2003
### Types of Structural Data Models

<table>
<thead>
<tr>
<th>Level</th>
<th>Source</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Contextual Model</strong></td>
<td>Drivers, Goals, Information needs</td>
<td>Drivers, Goals, Information needs</td>
</tr>
<tr>
<td><strong>Conceptual Model</strong></td>
<td>Subject area models</td>
<td>Business questions, fact/qualifier matrix, subject model, target configuration</td>
</tr>
<tr>
<td><strong>Logical Model</strong></td>
<td>3NF E-R Model of business elements</td>
<td>E-R and/or dimensional Model, Entity Lifecycle Model, Peacock Chart</td>
</tr>
<tr>
<td><strong>Structural Model</strong></td>
<td>Subject/Venn or Matrix model of data stores</td>
<td>E-R and/or dimensional models extended for warehousing</td>
</tr>
<tr>
<td><strong>Physical Model</strong></td>
<td>Exists in many forms, but must be mapped to structural model</td>
<td>Star, snowflake or relational schema</td>
</tr>
</tbody>
</table>

“Structural models specify the design necessary for the environment to maintain history, distribute data, and provide for ease of use.”

© Information Strategies, Inc. 2003 Adapted from TDWI
Data Design Inputs

- Structural Modeling
- Business Questions
- Enterprise Model
- Subject Areas
- Report Rollup Structures
- State Transition Analysis
- Entity Lifecycle Events
- Dimensional Analysis
- Base
- SC
- DMs
- Elements Available
- Elements to be Acquired
- Access
- Consumer
- History
- Source Systems
- Subjects Being Addressed

© Information Strategies, Inc. 2003
Maintaining History in the Intake Layer

<table>
<thead>
<tr>
<th>Last Name</th>
<th>First Name</th>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taylor</td>
<td>Elizabeth</td>
<td>27-Feb-32</td>
<td>Born in Hampstead London UK</td>
</tr>
<tr>
<td>Hilton</td>
<td>Liz</td>
<td>06-May-50</td>
<td>Married Nicky Hilton (Nicholas Conrad Hilton Jr. Hotelier)</td>
</tr>
<tr>
<td>Taylor</td>
<td>Liz</td>
<td>29-Jan-51</td>
<td>Divorced from Nicky Hilton</td>
</tr>
<tr>
<td>Wilding</td>
<td>Liz</td>
<td>21-Feb-52</td>
<td>Married Michael Wilding (actor)</td>
</tr>
<tr>
<td>Taylor</td>
<td>Liz</td>
<td>01-Jan-57</td>
<td>Divorced from Michael Wilding (actual date &amp; event unknown)</td>
</tr>
<tr>
<td>Taylor</td>
<td>Liz</td>
<td>01-Jan-57</td>
<td>Abducted by aliens (actual date &amp; event unknown)</td>
</tr>
<tr>
<td>Todd</td>
<td>Liz</td>
<td>02-Feb-57</td>
<td>Married Mike Todd (Producer)</td>
</tr>
<tr>
<td>Taylor</td>
<td>Liz</td>
<td>23-Mar-58</td>
<td>Widowed from Mike Todd</td>
</tr>
<tr>
<td>Fisher</td>
<td>Liz</td>
<td>01-Jan-59</td>
<td>Married Eddie Fisher (Singer) (not sure of exact date)</td>
</tr>
<tr>
<td>Taylor</td>
<td>Liz</td>
<td>06-Mar-64</td>
<td>Divorced from Eddie Fisher</td>
</tr>
<tr>
<td>Burton</td>
<td>Liz</td>
<td>15-Mar-64</td>
<td>Married Richard Burton</td>
</tr>
<tr>
<td>Taylor</td>
<td>Liz</td>
<td>01-Jan-74</td>
<td>Divorced from Richard Burton (actual date unknown)</td>
</tr>
<tr>
<td>Burton</td>
<td>Liz</td>
<td>01-Jan-75</td>
<td>Remarried Richard Burton (actual date unknown)</td>
</tr>
<tr>
<td>Taylor</td>
<td>Liz</td>
<td>01-Aug-76</td>
<td>Divorced from Richard Burton (summer 1976)</td>
</tr>
<tr>
<td>Warner</td>
<td>Liz</td>
<td>04-Dec-76</td>
<td>Married John Warner (Republican Senator)</td>
</tr>
<tr>
<td>Taylor</td>
<td>Liz</td>
<td>07-Nov-82</td>
<td>Divorced from John Warner</td>
</tr>
<tr>
<td>Fortensky</td>
<td>Liz</td>
<td>06-Oct-91</td>
<td>Married Larry Fortensky (Truck Driver)</td>
</tr>
<tr>
<td>Taylor</td>
<td>Liz</td>
<td>01-Nov-96</td>
<td>Divorced from Larry Fortensky</td>
</tr>
</tbody>
</table>
Maintaining History in the Intake Layer

In the Base, because we keep history, all changes are recorded as new rows in the Base Owner table.

<table>
<thead>
<tr>
<th>Owner ID</th>
<th>Start DT/TM</th>
<th>Last Name</th>
<th>First Name</th>
<th>Action Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT0227</td>
<td>1-Jun-51</td>
<td>Taylor</td>
<td>Liz</td>
<td>Add</td>
</tr>
<tr>
<td>LT0227</td>
<td>21-Feb-52</td>
<td>Wilding</td>
<td>Liz</td>
<td>Change</td>
</tr>
<tr>
<td>LT0227</td>
<td>1-Jan-57</td>
<td>Taylor</td>
<td>Liz</td>
<td>Change</td>
</tr>
<tr>
<td>LT0227</td>
<td>15-Jan-57</td>
<td>Taylor</td>
<td>Liz</td>
<td>Delete</td>
</tr>
<tr>
<td>LT0227</td>
<td>1-Feb-57</td>
<td>Taylor</td>
<td>Liz</td>
<td>Add</td>
</tr>
<tr>
<td>LT0227</td>
<td>2-Feb-57</td>
<td>Todd</td>
<td>Liz</td>
<td>Change</td>
</tr>
</tbody>
</table>
Distributing Data

2nd Normal Form

Base Design: 3rd Normal Form

Owner
- Owner Last Name
- Owner First Name
- Gender
- Date of Birth
- Social Security Number

Owner Address
- Street One
- Street Two
- City
- State
- Zip Code

Owner Contact
- Email Type
- Email Address
- URL

Owner Phone
- Phone Type
- Area Code
- Prefix
- Suffix
- Extension

Owner All
- Owner ID
- Age
- Gender
- Home Street One
- Home Street Two
- Home City
- Home State
- Home Zip Code
- Work City
- Work State
- Work Zip Code
- Home Email Address
- Work Email Address
- Home Phone Complete
- Work Phone Complete
- Load Date Time Stamp
Ease of Use Design Techniques

Derived Data
Aggregation (pre-joining)
Summarization
Vertical Summarization
Data Arrays
Creative “Index”
Star Schema
Types of Physical Data Models

<table>
<thead>
<tr>
<th>Level</th>
<th>Source</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contextual Model</td>
<td>Drivers, Goals, Information needs</td>
<td>Drivers, Goals, Information needs</td>
</tr>
<tr>
<td>Conceptual Model</td>
<td>Subject area models</td>
<td>Business questions, fact/qualifier matrix, subject model, target configuration</td>
</tr>
<tr>
<td>Logical Model</td>
<td>3NF E-R Model of business elements</td>
<td>E-R and/or dimensional Model, Entity Lifecycle Model, Peacock Chart</td>
</tr>
<tr>
<td>Structural Model</td>
<td>Subject/Venn or Matrix model of data stores</td>
<td>E-R and/or dimensional models extended for warehousing</td>
</tr>
<tr>
<td>Physical Model</td>
<td>Exists in many forms, but must be mapped to structural model</td>
<td>Star, snowflake or relational schema</td>
</tr>
</tbody>
</table>

“Physical models represent the detailed specification of what is physically implemented using specific technology.”
### Logical/Structural to Physical Conversion

#### Owner All
- Owner ID
- Age
- Gender
- Home Street One
- Home Street Two
- Home City
- Home State
- Home Zip Code
- Work City
- Work State
- Work Zip Code
- Home Email Address
- Work Email Address
- Home Phone Complete
- Work Phone Complete
- Load Date Time Stamp

#### Age/Gender Sales Trend
- Date Key
- Age Range Key
- Gender ID
- Paid Sales Dollars
- Paid Sales Quantity
- Age Range Description
- Gender Description
- Load Date Time Stamp

#### Vendor All
- Vendor ID
- Street One
- Street Two
- City
- State
- Zip Code
- Vendor Contact Name
- Email Address
- Contact Phone Complete
- Cell Phone Complete
- Vendor Confidence Scale
- Load Date Time Stamp
Logical/Structural to Physical Conversion

Fact Table

Sales Trend Fact Table
Collection Date
KK Franchise ID
Owner ID
SKU ID
Order Quantity
Order Extended Price

Product Dimension Table
SKU Product ID
Material Variety ID
Material Family ID
Material Class ID
SKU Description
Product Group Desc

Owner Dimension Table
Owner ID
Age
Gender
Home Street One
Home Street Two
Home City

Org Dimension Table
KK Franchise ID
KK Region ID
KK Country ID
Franchise Description
Region Description

Time Dimension Table
Date
Month
Quarter
Year

Dimensions

Facts
Physical Modeling

Relational Design

What is our sales trend in quantity and dollar amounts by owner age and gender, for the past 2 years?

**Age/Gender Sales Trend Entity**
- Date Key
- Age Range Key
- Gender ID
- Paid Sales Dollars
- Paid Sales Quantity
- Age Range Description
- Gender Description
- Date Description

Relational Design: All Needed Data in a Single Table

Employs a Concatenated Key

What if no relationship is discovered and the user wished to explore the sales data by income range or hobby?
Physical Modeling

Dimensional Design

What is our sales trend in quantity and dollar amounts by owner age and gender, for the past 2 years?

Owner Dimension
- Owner ID
- Name
- Age
- Date of Birth
- Gender
- Income Range
- Primary Hobby
...

Time Dimension
- Date ID
...

Sales Trend Fact Table
- Date ID
- Owner ID
- Organization ID
- Sales Dollars
- Sales Quantity
...

A dimensional design, such as this Star Schema, provides greater analytical flexibility.
1. What % of orders are discarded, by vendor, by month?  

State = Placed & Discarded

2. Which vendors place us on credit hold? Show by vendor and month.  

State = Credit Hold

3. What % of purchase orders don’t pass inspection, by vendor?  

State = Placed & Passed Inspection

4. How many vendors do we have available to us?  

State = None; Dimension Count
A Four Step Process with Inputs

1. **Model A Business Process**
   - Sales Orders, Invoices, Shipments, Inventory, Purchase Orders
   - Expand with State Transition Diagram

2. **Determine Business Process Grain**
   - Individual Transactions, Individual Daily Snapshots, Individual Monthly Snapshots
   - Use Business Questions

3. **Assign Dimensions**
   - Time, Product, Customer, Promotion, Organization
   - Use Source Systems to determine Dimension Elements

4. **Define Measured Facts**
   - Order Count, Order Quantity, Order Dollar Amount
   - Use Business Questions

## Checklist

### Review of Deliverables

<table>
<thead>
<tr>
<th>Deliverable</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>√ F/Q Matrix</td>
<td>Decomposition of business questions. Helps identify descriptive or metric facts.</td>
</tr>
<tr>
<td>√ Subject Area Model</td>
<td>Identification of business data groups. Helps determine areas for data integration and can drive data mart design.</td>
</tr>
<tr>
<td>√ State Transition Diagram</td>
<td>Expansion of entities found in the ERM. Necessary for understanding business questions.</td>
</tr>
<tr>
<td>√ Peacock Chart</td>
<td>Documents dimension hierarchies. Necessary for understanding dimensional data.</td>
</tr>
<tr>
<td>√ Entity Relationship Model</td>
<td>Represents the necessary business attributes. Used to understand data independent of physical sources.</td>
</tr>
<tr>
<td>√ Star Schema</td>
<td>Implementation of a physical design. Best suited for business questions with metric facts.</td>
</tr>
</tbody>
</table>