Data Modeling 101
Donna Burbank and Steve Hoberman
Session Code ED03
Session Abstract

— Learn the basics of data modeling. This session will cover a practical working knowledge of data modeling concepts and best practices, and how to apply these principles with CA ERwin Data Modeler r8. Conceptual, Logical, and Physical data models will be discussed, as well as the proper use-case for each. Basic principles of relational data modeling will be covered such as entities, relationships, keys, and more.

— This session is based on Donna and Steve's recent book, Data Modeling Made Simple with CA ERwin Data Modeler r8. The first two people to register and attend will receive a FREE copy of the book.
Speaker Bios

**Donna Burbank** is a recognized industry expert and author, with more than 15 years of experience in data management, metadata management, and enterprise architecture. Donna currently is the senior director of product marketing for CA’s data modeling solutions. She has worked with dozens of Fortune 500 companies worldwide in the U.S., Europe, Asia, and Africa and speaks regularly at industry conferences.

**Steve Hoberman** is the most requested data modeling instructor in the world. Steve taught his first data modeling class in 1992 and has educated more than 10,000 people about data modeling and business intelligence techniques since then, spanning every continent except Africa and Antarctica. Steve’s **Data Modeling Master Class** is recognized as the most comprehensive data modeling course in the industry. More at [www.stevehoberman.com](http://www.stevehoberman.com)
Donna and Steve have co-authored two books together:

- **Data Modeling for the Business**
- **Data Modeling Made Simple with CA ERwin Data Modeler r8**, on which this presentation is based
Donna and Steve have co-authored two books together:

- **Data Modeling for the Business**
- **Data Modeling Made Simple with CA ERwin Data Modeler r8**, on which this presentation is based

20% Off for ERworld participants using promo code P20ERWIN
www.technicspub.com
Agenda

— What is a Data Model?
— Basic Logical Data Modeling Components
— Logical Data Modeling with CA ERwin Data Modeler
— Demo
What is a Data Model?
Models are everywhere

A set of symbols and text used to make a complex concept easier to grasp

Jack
Enterprise Architect

Mary
Enterprise Modeler

Bob
Enterprise Analyst
Models are everywhere

- A set of symbols and text used to make a complex concept easier to grasp

Jack
Enterprise Architect

Mary
Enterprise Modeler

Bob
Enterprise Analyst
Models are everywhere

— A set of symbols and text used to make a complex concept easier to grasp
Models are everywhere

— A set of symbols and text used to make a complex concept easier to grasp
Data Model Definition

— A set of symbols and text used to make the actual data easier to grasp
— Includes both data elements and business rules

Customer

Own

Account

Savings Account

Brokerage Account

Checking Account

Each Customer can own one or many Accounts.
Each Account must be owned by one and only one Customer.

Each Account can be a Savings, Brokerage, or Checking Account.
## Data Model Settings

### - Format -

<table>
<thead>
<tr>
<th>Conceptual (Proof sheet)</th>
<th>Logical (Negative)</th>
<th>Physical (Instantiation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High level business solution</td>
<td>Detailed business solution</td>
<td>Detailed technical solution</td>
</tr>
<tr>
<td>Scoping tool</td>
<td>Normalization</td>
<td>Denormalization, indexing, views, partitioning</td>
</tr>
<tr>
<td>Only basic and critical concepts</td>
<td>Dimensionality</td>
<td>Star schema and snowflake</td>
</tr>
<tr>
<td></td>
<td>“Essence”</td>
<td>“Incarnation”</td>
</tr>
<tr>
<td>Data Model Settings - Format -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conceptual (Proof sheet)</th>
<th>Logical (Negative)</th>
<th>Physical (Instantiation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High level business solution</td>
<td>Detailed business solution</td>
<td>Detailed technical solution</td>
</tr>
<tr>
<td>Scoping tool</td>
<td>Normalization</td>
<td>Denormalization, indexing, views, partitioning</td>
</tr>
<tr>
<td>Only basic and critical concepts</td>
<td>Dimensionality</td>
<td>Star schema and snowflake</td>
</tr>
<tr>
<td></td>
<td>“Essence”</td>
<td>“Incarnation”</td>
</tr>
</tbody>
</table>
Basic Logical Data Modeling Components
An entity is a collection of information about something that the business deems important and worthy of capture.
Data Element

A data element is a property of importance to the business whose values contribute to identifying, describing, or measuring instances of an entity.

Employee

- Employee Identifier
- Employee Last Name
- Employee First Name
- Employee Hire Date
- Employee Signed Employment Contract
- Employee Drivers License Photo
A key helps you find entity instances

Student

- Student_Identifier
- Student_Last_Name (IE1.1)
- Student_First_Name (IE1.2)
- Student_Social_Security_Number (AK1.1)

Class

- Class_Identifier

Class_Identifier (FK)
- Student_Identifier (FK)
- Semester_Identifier (FK)

Semester

- Semester_Identifier

Student_Grade

- Final_Grade
A key helps you find entity instances

Candidate key
Primary key
Surrogate key

Candidate key
Alternate key
Natural key

Student

Student_Identifier

Student_Last_Name (IE1.1)
Student_First_Name (IE1.2)
Student_Social_Security_Number (AK1.1)

Student_Grade

Class

Class_Identifier

Class_Identifier (FK)
Student_Identifier (FK)
Semester_Identifier (FK)

Semester

Semester_Identifier

Semester_Identifier (FK)

Candidate key
Primary key
Surrogate key

Candidate key
Primary key
Surrogate key
A key helps you find entity instances

Candidate key
Primary key
Surrogate key

Non-unique index
Composite key

Candidate key
Alternate key
Natural key

Class
Class_Identifier

Student
Student_Identifier
Student_Last_Name (IE1.1)
Student_First_Name (IE1.2)
Student_Social_Security_Number (AK1.1)

Student_Grade

Class
Class_Identifier

Student
Student_Identifier (FK)
Semester_Identifier (FK)

Semester
Semester_Identifier

Final_Grade
A key helps you find entity instances

Candidate key
Primary key
Surrogate key

Non-unique index
Composite key

Candidate key
Alternate key
Natural key

Class
Class_Identifier

Candidate key
Primary key
Surrogate key

Student
Student_Identifier
Student_Last_Name (IE1.1)
Student_First_Name (IE1.2)
Student_Social_Security_Number (AK1.1)

Student_Grade

Class_Identifier (FK)
Student_Identifier (FK)
Semester_Identifier (FK)

Final_Grade

Candidate key
Primary key
Surrogate key

Semester
Semester_Identifier

Candidate key
Primary key
Surrogate key

Class
Class_Identifier (FK)
Student_Identifier (FK)
Semester_Identifier (FK)

Final_Grade
A rule is an instruction about how to behave in a specific situation. A static rule is represented on a model via a relationship.

### Static Rules

#### Structure
- Each product can appear on one or many order lines.
- Each order line must contain one and only one product.

#### Referential Integrity (RI)
- An order line cannot exist without a valid product.
- A student cannot exist without a valid student number.

### Action Rules
- Freshman students can register for at most 18 credits a semester.
- Take 10% off an order if the order contains more than five products.
Identifying vs. Non-Identifying Relationships

Non-Identifying

Customer
- Customer Id
- Customer Name

Account
- Account Code
- Customer Id (FK)
- Account Name

Identifying

Customer
- Customer Id
- Customer Name

Account
- Account Code
- Customer Id (FK)
- Account Name
Supertypes/Subtypes

— What is subtyping?
  – Subtyping is grouping together the common data elements and relationships of entities, while keeping what’s unique within each entity.

— Other names for subtyping:
  – Supertyping
  – Generalization
  – Inheritance
Logical Modeling in CA ERwin Data Modeler
Baker Cakes Example
Baker Cakes, Inc.

— Baker Cakes is a family-run business whose main stakeholder is Bob Baker, the owner/operator of Baker Cakes. Bob is in charge of making most decisions, from database design to icing color selection.

— In our example, we’re building a data model for Baker Cakes, who is looking to build a new application to manage their data.
Step 1: Understanding Our Customers

— The first business concepts (entities) we need to describe are the customers for Baker Cakes.

— Mr. Baker sells to both retail and wholesale customers

— How do we represent this in a data model?
In this case, we can use a supertype/subtype relationship to show the two types of customers.
Understanding Our Business

— We ask Mr. Baker what the differences are between a Retail Customer and a Wholesale Customer.

— One key difference is that Wholesale Customers are managed by a Sales Rep.

— We’ll need to show this on our data model.
  – Creating an entity for **Sales Rep**
  – Creating a relationship between **Sales Rep** and **Wholesale Customer**
Business Rules

— Our business rules for the relationship between Sales Rep and Wholesale Customer are as follows:
  – Each Sales Rep must call upon one or more Wholesale Customers
  – Each Wholesale Customer must be called upon by one Sales Rep.
Business Rules

— Our business rules for the relationship between Sales Rep and Wholesale Customer are as follows:
  – Each Sales Rep must call upon one or more Wholesale Customers
  – Each Wholesale Customer must be called upon by one Sales Rep.
Our business rules for the relationship between Sales Rep and Wholesale Customer are as follows:

- Each Sales Rep must call upon one or more Wholesale Customers.
- Each Wholesale Customer must be called upon by one Sales Rep.
Our business rules for the relationship between Sales Rep and Wholesale Customer are as follows:

- Each Sales Rep must call upon one or more Wholesale Customers.
- Each Wholesale Customer must be called upon by one Sales Rep.
Identifying or Non-Identifying?

— Should we make this relationship an identifying or non-identifying relationship?
Identifying or Non-Identifying?

— Should we make this relationship an identifying or non-identifying relationship?

— Key question: Do we need wholesale customer information to retrieve a given sales rep?
Identifying or Non-Identifying?

— Should we make this relationship an identifying or non-identifying relationship?

— Key question: Do we need wholesale customer information to retrieve a given sales rep?

— No. Mr. Baker tells us that Sales Reps are each identified by an Employee ID.
Identifying or Non-Identifying?

— Should we make this relationship an identifying or non-identifying relationship?

— Key question: Do we need wholesale customer information to retrieve a given sales rep?

— No. Mr. Baker tells us that Sales Reps are each identified by an Employee ID.

— Therefore, this is a non-identifying relationship (dashed line).
The relationship between **Sales Rep** and **Wholesale Customer** will appear as the following on our data model.

- Each **Sales Rep** must call upon one or more **Wholesale Customers**
- Each **Wholesale Customer** must be called upon by one **Sales Rep**.
The relationship between **Sales Rep** and **Wholesale Customer** will appear as the following on our data model.

- Each **Sales Rep** must call upon one or more **Wholesale Customers**
- Each **Wholesale Customer** must be called upon by one **Sales Rep**.

Verb Phrases generally read "clockwise"
Defining Keys and Attributes

— Now that we have our high-level entities and relationships defined, we need to add more detail about our Customers and Sales Reps.

— First, let’s identify what information uniquely retrieves an instance of each (primary key).

— In this case, it’s easy. Baker Cakes uses:
  
  – A Customer ID to uniquely identify Customers (both retail and wholesale)
  
  – An Employee ID to uniquely identify Sales Reps.
Primary and Foreign Keys

- Our Model now looks like this.
- Primary Keys are listed “above the line”.
- Note that Foreign Keys (FK) were automatically created for us. (Remember Referential Integrity/RI!)

![Diagram of Primary and Foreign Keys]
Attributes

— In addition to the key/identifying attributes, there are other important attributes to define for Customers and Sales Reps.

— For instance, let’s define an Employee First Name and Employee Last Name for Sales Reps.

— And a Company Name for Wholesale Customers, as shown below.
Demo using CA ERwin Data Modeler

- Creating Entities
- Creating Relationships
- Creating Primary Keys
- Viewing how Foreign Keys are Inferred
- Defining Attributes
Try it yourself – Free Software

CA ERwin Data Modeler Community Edition

www.erwin.com

The Market Leader in Data Modeling for Over 20 Years
CA ERwin Modeling - Why trust your data to anyone else?

CA ERwin Data Modeler Community Edition

The CA ERwin Data Modeler Community Edition is a FREE, entry-level data modeling tool that is a subset of CA's flagship CA ERwin Data Modeler Standard Edition product. It is a great way for students and those new to modeling to get started with an industry-leading data modeling tool.

Visualize the Power of Your Data

CA ERwin Data Modeler r6 Overview

Listen to Product Manager Neil Buchwalter describe how CA ERwin Data Modeler r6 provides state-of-the-art visualization capabilities, coupled with a dynamic new user interface and enhanced set of workflow capabilities as well as comprehensive database support.

Data Modeler Editions »
Additional Products »
Product Bundles »

Download Now!
Live Chat Now!
Have Us Call You

Product Information

• Product Family Brief (PDF)
• Feature Comparison Matrix (PDF)
• Top 10 Reasons to Choose CA ERwin (PDF)
• Comparison Guide of Top

Have Us Call You
— A data model is a “blueprint” for your data assets
— There are three levels of data models: conceptual, logical, and physical
— A logical data model (LDM) represents a detailed business solution. Logical model objects include:
  – Entities define the “who”, “what”, “where”, “when” and “why”
  – Relationships define business rules around data
  – Keys help identify instances of data
  – Attributes provide detailed information about entities.
— CA ERwin Data Modeler helps automate the logical model design process
Thank You

Donna Burbank
donna.burbank@ca.com
Twitter: @donnaburbank

Steve Hoberman
me@stevehoberman.com
www.stevehoberman.com

© Copyright CA 2011. All rights reserved. All trademarks, trade names, service marks and logos referenced herein belong to their respective companies. No unauthorized use, copying or distribution permitted.
And the Book Winners are….

• To view the book winners:
  • Visit the Prize Center
  • Follow @ERwinModeling on Twitter

• If you didn’t win, you can receive 20% off by ordering via www.technicspub.com using promotion code P20ERWIN.
Legal notice

© Copyright CA 2011. All rights reserved. All trademarks, trade names, service marks and logos referenced herein belong to their respective companies. No unauthorized use, copying or distribution permitted.

THIS PRESENTATION IS FOR YOUR INFORMATIONAL PURPOSES ONLY. CA assumes no responsibility for the accuracy or completeness of the information. TO THE EXTENT PERMITTED BY APPLICABLE LAW, CA PROVIDES THIS DOCUMENT “AS IS” WITHOUT WARRANTY OF ANY KIND, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NONINFRINGEMENT. In no event will CA be liable for any loss or damage, direct or indirect, in connection with this presentation, including, without limitation, lost profits, lost investment, business interruption, goodwill, or lost data, even if CA is expressly advised of the possibility of such damages.

Certain information in this presentation may outline CA’s general product direction. This presentation shall not serve to (i) affect the rights and/or obligations of CA or its licensees under any existing or future written license agreement or services agreement relating to any CA software product; or (ii) amend any product documentation or specifications for any CA software product. The development, release and timing of any features or functionality described in this presentation remain at CA’s sole discretion.

Notwithstanding anything in this presentation to the contrary, upon the general availability of any future CA product release referenced in this presentation, CA may make such release available (i) for sale to new licensees of such product; and (ii) in the form of a regularly scheduled major product release. Such releases may be made available to current licensees of such product who are current subscribers to CA maintenance and support on a when and if-available basis.