

# Data Modeling Master Class

## Steve Hoberman's Best Practices Approach to Understanding and Applying Fundamentals Through Advanced Modeling Techniques

The Master Class is a complete course on requirements elicitation and data modeling, containing three days of practical techniques for producing solid relational and dimensional data models. After learning the styles and steps in capturing and modeling requirements, you will apply a **best practices approach** to building and validating data models through the Data Model Scorecard®. You will know not just how to build a data model, but also how to build a data model **well**. Three case studies and many exercises reinforce the material and enable you to apply these techniques in your current projects.

This course has recently received world recognition by the International Institute of Business Analysis (IIBA): The Data Modeling Master Class is an endorsed course by the IIBA V2.0 of the BABOK®. Earn 24 Continuing Development Units (CDU) through the IIBA, and 24 Professional Development Units (PDUs) through the Project Management Institute!

### Top 10 Objectives

1. Apply requirements elicitation techniques including interviewing and prototyping
2. Explain data modeling constructs and employ the “6 Questions” approach to ensure model precision
3. Demonstrate reading a data model of any size and complexity with the same confidence as reading a book
4. Validate any data model with key “settings” (scope, abstraction, timeframe, function, and format) as well as through the Data Model Scorecard®
5. Practice finding structural soundness issues and standards violations
6. Build relational and dimensional subject area, logical, and physical data models
7. Recognize situations where abstraction would be most valuable and situations where abstraction would be most dangerous
8. Use a series of templates for scoping and validating requirements, and for data profiling
9. Express how to write clear, complete, and correct definitions
10. Describe the two reasons an enterprise data modeling project can fail, and the factors that must be in place for the enterprise data model to succeed

### Prerequisite(s)

This course assumes no prior data modeling knowledge and, therefore, there are no prerequisites. Business analysts, data modelers, data architects, business users, database administrators, developers, and project managers are the frequent roles that take this class.

## Topics

### Part 1: Modeling Basics

Assuming no prior knowledge of data modeling, we will begin this section with an entertaining exercise that will illustrate an important gap filled by data models. Next, we will explain data modeling concepts and terminology. We will also explore each component on a data model and practice reading business rules. You will be able to answer the following questions by the end of this section:

- What is a data model and how can a piece of paper with boxes and lines be such a valuable wayfinding tool to our organizations?
- What six questions must be asked to translate ambiguity into precision?
- What two situations can ruin a data model's credibility?
- What are five key skills every data modeler should possess?
- What do a data model and a camera have in common?
- What are entities, data elements, domains, and relationships?
- Why subtype and what are the differences between exclusive and non-exclusive subtypes?
- What are the different types of keys on a model?
- What are the perceived and actual benefits of surrogate keys?
- What is cardinality and how are the relationships on a data model read?
- What are the different ways to model hierarchies and networks?
- What is recursion and why is it such an emotional topic?

### Part 2: Overview to the Data Model Scorecard®

The Scorecard is a set of ten categories for validating a data model. We will explore best practices from the perspectives of both the modeler and reviewer, and you will be provided with a template to use on your current projects. Each of the following categories heavily impacts the usefulness and longevity of the model. Our discussion of them will be accompanied by many examples.

- Understanding subject area, logical, and physical data models
- Ensuring the model captures the requirements
- Validating model scope
- Following acceptable modeling principles
- Determining the optimal use of generic concepts
- Applying consistent naming standards
- Arranging the model for maximum understanding
- Writing clear, correct and consistent definitions
- Matching the model with the enterprise
- Comparing the metadata with the data

### Part 3: Understanding subject area, logical, and physical data models

The subject area model captures a business need within a well-defined scope; the logical data model captures an application-independent business solution; and the physical data model captures the technical solution by focusing on factors such as performance and security. Each of these models will be explained in detail in this section. We will also

practice building several data models and you will be able to answer the following questions by the end of this section:

- How do relational and dimensional models differ?
- What are the three types of subject area models and how are they built?
- What is normalization and how do you apply it?
- What are the differences between transaction, snapshot and accumulating measures?
- What are conformed dimensions, junk dimensions, degenerate dimensions, and behavioral dimensions?
- What are the different navigation paths needed to drill down, up, and across?
- What are some dimensional modeling do's and don'ts?
- What is the difference between a star schema and a snowflake?
- Where should denormalization be performed on your models?
- What are the five ways of denormalizing?
- What is the difference between aggregation and summarization?
- What are views, indexing, and partitioning and how can they be leveraged to improve performance?
- What are the three different types of Slowly Changing Dimensions?

#### **Part 4: Ensuring the model captures the requirements**

We will focus on techniques such as the use of spreadsheets and business assertions to ensure the data model meets the business requirements. You will be able to answer the following questions by the end of this section:

- What is the Requirements Lifecycle?
- What are the most useful ways of eliciting requirements?
- What are the proper ways to phrase an interview question?
- When is brainstorming an effective way to capture requirements?
- What are three creative prototyping techniques for the non-techie?
- What does optionality reveal on a data model?
- How can you validate that a data model captures the requirements without showing the data model?
- How can you leverage the Interview Template and Family Tree to validate requirements?

#### **Part 5: Validating model scope**

We will focus on techniques for validating that the scope of the requirements matches the scope of the model. If the scope of the model is greater than the requirements, we have a situation known as “scope creep.” If the model scope is less than the requirements, we will be leaving information out of the resulting application. You will be able to answer the following questions by the end of this section:

- Why is the line between data and metadata starting to blur?
- What techniques can you use to avoid scope creep?
- How do you play “Metadata Bingo”?
- What type of metadata is most abused?
- How can you use the Grain Matrix to scope requirements?

## **Part 6: Following acceptable modeling principles**

We will focus on techniques for building sound designs. You will be able to answer the following questions by the end of this section:

- What tools exist to automate checking model structure?
- What are circular relationships and why are they evil?
- What are the most common structural violations on a data model?
- Can an alternate key ever be empty?

## **Part 7: Determining the optimal use of generic concepts**

We will focus on techniques for capturing the ideal use of generic concepts such as Party and Event. You will be able to answer the following questions by the end of this section:

- Why are “what if” scenarios so important to document?
- What three questions must be asked prior to abstracting?
- Why are Roles so important to Business Intelligence projects?
- What are metadata entities?
- How do different modeling notations handle subtyping?
- What are some common modeling patterns?

## **Part 8: Applying consistent naming standards**

We will focus on techniques for applying correct and consistent naming standards. You will be able to answer the following questions by the end of this section:

- Explain name structure and give examples
- Explain term and give examples
- Explain syntax and give examples
- Learn why class words are so important

## **Part 9: Arranging the model for maximum understanding**

We will focus on techniques for arranging the entities, data elements, and relationships to maximize readability. You will be able to answer the following questions by the end of this section:

- How do you improve model readability at a model level?
- How do you improve model readability at an entity level?
- How do you improve model readability at a data element level?
- How do you improve model readability at a relationship level?

## **Part 10: Writing clear, correct, and consistent definitions**

We will focus on techniques for writing useable definitions. You will be able to answer the following questions by the end of this section:

- How do you play Definition Bingo?
- Why are definitions so much more important now than they were in the past?
- What are some techniques for writing a good definition?
- How do you validate a definition?
- Which types of data elements require sample values in their definitions?

## **Part 11: Matching the model with the enterprise**

We will focus on techniques for ensuring the data model complements the “big picture”. You will be able to answer the following questions by the end of this section:

- What is an enterprise data model and why have one?
- What are the secrets to achieving a successful enterprise data model?
- What are industry data models and how can they be leveraged?

## **Part 12: Comparing the metadata with the data**

We will focus on techniques for confirming the data elements and their rules match reality. Does the data element Customer Last Name really contain the customer’s last name, for example? You will be able to answer the following questions by the end of this section:

- How can the Data Quality Validation Template help us with catching data surprises early?
- What are the some of the challenges in conducting an early data quality assessment?
- How can I quickly identify potential data quality issues using the data model?

## **Quotes on the Data Modeling Master Class**

*In my long professional career, I have participated in many training seminars but I have never encountered a class in which the subject had been so thoroughly considered and presented in such a clear and engaging manor. As a fairly new, but full time data modeler, I expect to use the things that I learned in this class every day. My only regret is that the class had to end.*

- G. Schmid, Travelers Insurance

*A must class if you want to become a serious and good modeler and a fun class for not so much fun topic. You definitely get your money’s worth and exceedingly above and beyond – awesome, two thumbs up!*

- R. Lee, UBS

*An excellent course with focus on the basics. I can only imagine the effort that went into developing this course. Lucky that I happened to find it.*

- R. Sampath, Deloitte

*Having never done data modeling before, I can now say, I am excited about implementing the skills I’ve learned.*

- L. Felder, Johns Hopkins HealthCare

*This was the most comprehensive, informative, energetic, interesting and just plain FUN class I have ever taken on the subject of Data Modeling.*

- G. Werner, Long Island Railroad

*Truly enjoyed this class even though I have been modeling databases for 24 years. Thought the baseball and concentration metaphors were interesting.*

- M. Austin, Wells Fargo

## Speaker

Steve Hoberman is the most requested data modeling instructor in the world. Introduced at over 50 international conferences as everything from a “data modeling guru” to “data modeling rock star”, Steve balances the formality and precision of data modeling with the realities of building software systems with severe time, budget, and people constraints. In his consulting and teaching, he focuses on templates, tools, and guidelines to reap the benefits of data modeling with minimal investment. He 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 is known for his entertaining, interactive teaching and lecture style (watch out for flying candy!), and organizations around the globe have brought Steve in to teach his **Data Modeling Master Class**, which is recognized as the most comprehensive data modeling course in the industry. Steve is the author of five books on data modeling, including the bestseller *Data Modeling Made Simple*. He is the founder of the Design Challenges group, inventor of the Data Model Scorecard®, and CEO of Technics Publications. You can learn more about his books at Technics Publications' website, [www.technicspub.com](http://www.technicspub.com). To receive Steve's monthly Design Challenges, visit his website at [www.stevhoberman.com](http://www.stevhoberman.com). Steve can be reached at [me@stevhoberman.com](mailto:me@stevhoberman.com).