

Data Modeling Master Class

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

Description

The Master Class is a complete course on requirements gathering and data modeling, containing four days of practical techniques for producing solid relational and dimensional data models. After learning the styles and steps in gathering 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**. Challenging exercises and workshops will 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® as registered under Steve Hoberman & Associates. Earn 24 Continuing Development Units (CDU) through the IIBA, and 24 Professional Development Units (PDUs) through the Project Management Institute!

IIBA and BABOK are trademarks owned by the International Institute of Business Analysis

Top 10 Objectives

- You will know which requirements gathering techniques are most effective for each situation
- You will be able to clearly explain core data modeling terminology
- You will be able to read a data model of any size and complexity with the same confidence as reading a book
- You will be able to validate any data model with the Data Model Scorecard®
- You will be able to build a fully normalized relational data model as well as an easily navigatable dimensional model
- You will be able to apply techniques to turn a logical data model into an efficient physical design
- You will know when to use abstraction, and when it should never be used
- You will be able to leverage a series of templates for capturing and validating requirements
- You will be able to write clear, complete, and correct definitions
- You will be able to explain the critical factors that must be in place for a successful enterprise data model

Prerequisite(s)

This course assumes no prior data modeling knowledge and, therefore, there are no prerequisites. Analysts, architects, business users, developers, managers and modelers have all been successful in 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. We will answer the following questions:

- 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?
- How does a data model improve communication during the analysis process and after the model is complete?
- What two situations can degrade a data model's precision?
- 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 four subtype types?
- 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 is recursion and why is it such an emotional topic?
- Why is the line between data and metadata starting to blur?

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 answer the following questions:

- 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 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 really is a Slowly Changing Dimension?

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. We will answer the following questions:

- What is the Requirements Lifecycle?
- What are the most useful ways of gathering requirements?
- What are the proper ways to phrase an interview question?
- When is brainstorming an effective way to gather 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 Family Tree, Grain Matrix, and Interview templates?

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. We will answer the following questions:

- What techniques can you use to avoid scope creep?
- What is metadata and how do you play “Metadata Bingo”?
- What type of metadata is most abused?
- Are all amounts really facts in a dimensional model?
- How can you estimate how long it will take to complete the data models?

Part 6: Following acceptable modeling principles

We will focus on techniques for building sound designs. We will answer the following questions:

- 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. We will answer the following questions:

- 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?
- What are some modeling components that can be reused across models?

Part 8: Applying consistent naming standards

We will focus on techniques for applying correct and consistent naming standards. We will discuss the following:

- 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. We will answer the following questions:

- 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. We will answer the following questions:

- 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 model complements the “big picture”. We will answer the following questions:

- 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?
- What are the three approaches to building an enterprise data model?

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? We will answer the following questions:

- How can the Data Quality Validation Template help us with catching data surprises early?
- What are the some of the challenges in early detection?
- What is unstructured data and how will it impact our world?

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 Bio

Steve Hoberman is a trainer, consultant, and writer in the field of data modeling. He balances the formality and precision of data modeling with the realities of building software systems with severe time, budget, and people constraints. Steve focuses on templates, tools, and guidelines to reap the benefits of data modeling with minimal investment.

After graduating Phi Beta Kappa from Queens College and completing a Master of Science in Information Networking at Carnegie Mellon University, Steve joined Bell Communications Research (Bellcore) in 1990 and started his data modeling career on an information engineering project. He was an analyst and data modeler on a team building an enterprise data model for the entire telephone industry. This project offered broad exposure to many cutting-edge disciplines of the 90's – object-oriented modeling techniques, Computer Aided Software Engineering (CASE) tools, and first-generation metadata repositories.

Building on the Bellcore foundation, Steve developed an interest in the human side of data modeling and the “next generation” techniques. In 1994 he went to work on Wall Street, performing data modeling work for many financial applications. Many of his “do it fast but right” modeling techniques came from the high pressure projects of Wall Street. In 1997 Steve joined Mars, Inc. as their Data Warehouse Architect. During his nine year tenure at Mars, Steve filled a variety of roles including Lead Data Modeler, Developer Team Lead, and SAP Functional Analyst. Each of these roles provided opportunities to grow and evolve his unique, experience-based approach to data modeling.

Steve taught his first data modeling class in 1992 and has taught over 10,000 people data modeling and business intelligence techniques since then. He has presented at over 50 international conferences in every format from short presentations to full-day classes, and has been selected to deliver keynote addresses at major industry conferences in North America and in Europe.

Steve is a columnist and frequent contributor to industry publications. He is the author of several data modeling books including *Data Modeling Made Simple*, *Data Modeler's Workbench*, and *Data Modeling for the Business*. With interest in building a data modeler's community, he founded the Design Challenges group, which today boasts more than 3,000 data management practitioners who tackle monthly data modeling puzzles. (Add your email address to join this group at www.stevhoberman.com.) Steve is an innovator in data modeling and the inventor of the Data Model Scorecard®, which has quickly become the standard for data model quality.