Mastering Data Modeling

A User-Driven Approach

John Carlis
Joseph Maguire

Praise for *Mastering Data Modeling: A User-Driven Approach*

This is without doubt the most complete and worthy "common sense" approach to data modeling that I have read. It spans the 'stupidity wars' of differences of notation and nomenclature with a clear and easy elegance that espouses common sense over dogma. . . . An excellent work which deserves to be a foundation of every programmer and software designer's bookshelf. Without understanding of data, there can be no intelligent software design.

Paul Irvine
Senior Manager
Emerald Solutions

I will also be able to use this book to make myself not only a better data modeler and database designer, but a better overall systems analyst, project planner, and business communicator. A second (and perhaps third) reading is definitely in order!

Matthew Keranen
Consulting Database Administrator

An invaluable resource for anyone who wants to learn about data modeling, this book is exceptionally powerful because its developed from real world experiences—its a pragmatic guide for data modeling practitioners.

Peter O'Kelley
Senior Consultant/Analyst
The Patricia Seybolt Group

This book presents a detailed, structured approach to the process of data modeling with emphasis on interacting with users, a unique feature which impressed me.

David McGoveran
Alternative Technologies, Inc.

I hope the book is well received. It deserves to be, as it provides extremely clear, practical thinking on an important but often misunderstood subject.

Mitch Kapor
Founder of the Lotus Development Corporation

Shape-recognition skills

- Finding and focusing on shapes within a large LDS
- Recognizing the differences between shapes that are similar but not identical
- Recognizing the similarity between seemingly dissimilar shapes
- Distinguishing between legitimate shapes and syntactically invalid LDS fragments

Shape-evolution skills

- Knowing how shapes are likely to evolve
- Asking questions that help users choose between two similar shapes
- Knowing when to ask questions of users
- Knowing when and how to modify the LDS to make a shape evolve
- Understanding the relative frequency of the various shapes

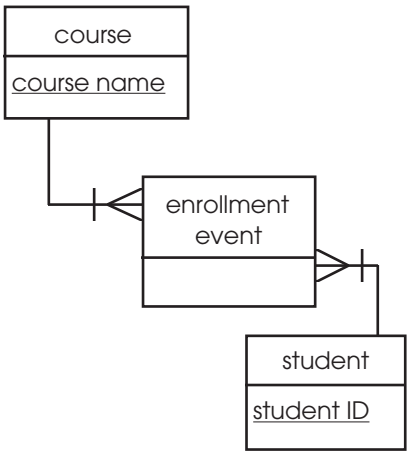
Shape-studying skills

- Referring to each fundamental shape by its name

The remainder of this chapter elaborates on these aspects of shape mastery.

Reading a Shape Aloud in Several Ways

There will be times when you or the users become confused by an in-progress data model. You should not be discouraged when confusion arises. In fact, you should expect it, so you can prepare for it and conquer it. One excellent way to combat confusion is to be able to read an LDS fragment in several ways.



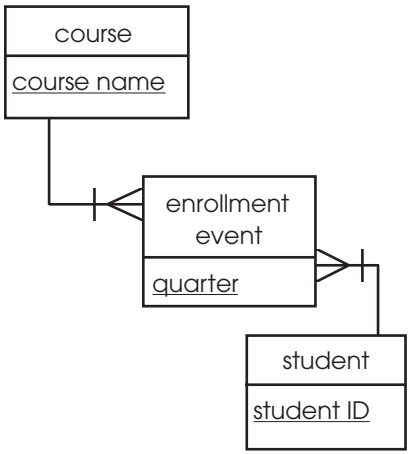
Fragment 7-6

Facing confusion, you'll naturally want to help the user by paraphrasing the LDS in more conversational diction. If you're not careful, however, you'll resort to vague statements and imprecise hand waving. Remember, there is no room for imprecision. For each shape, you need to be able to say in several ways exactly what the shape means.

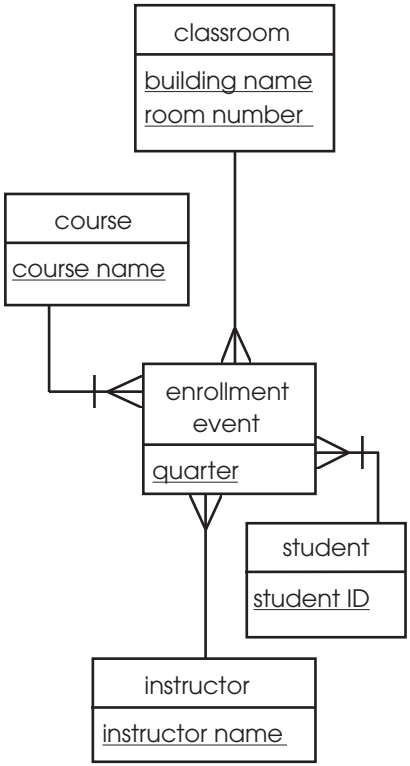
For example, consider LDS Fragment 7-6. You can read this shape according to the guidelines for reading LDSs presented in Chapter 3.

- *About each course, we can remember its name and its enrollment events. Each course is identified by its name.*
- *About each enrollment event, we can remember its course and its student. Each enrollment event is identified by its course and its student.*
- *About each student, we can remember its student ID. Each student is identified by its student ID.*

Because these sentences conform to the guidelines for reading LDSs, we know they are accurate. However, they effectively read the shape piecemeal—in installments.



Fragment 7-7



Fragment 7-8

They fail to give an overall picture of the shape. You can say other accurate sentences about this shape, such as the following:

- *Each course can have many enrolled students* and *Each student can have many enrolled-in courses.* These sentences together stress that the enrollment event entity effectively constitutes a many-many relationship between student and course.
- *Each enrollment event indicates that a particular student enrolls in a particular course.* This sentence stresses the meaning of the word enrollment. It also stresses that each enrollment event is about a (*course, student*) pair.

Note that these conversational sentences supplement the original sentences but do not substitute for them. For example, these conversational sentences say nothing about how to distinguish instances from one another.

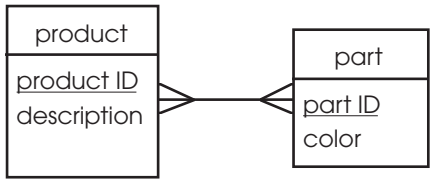
Looking back at the conversation about creatures and skills in the previous chapter, we can see several times when this conversational diction might come in handy. For example, if you read Fragment 6-5 with this diction, you can help users understand the similarity between Fragments 6-5 and 6-3.

What's more, as shapes grow more complex, these conversational sentences grow more enlightening to confused persons. For example, consider the two successively more complex shapes in Fragments 7-7 and 7-8 and the corresponding sentences:

- *Each enrollment event indicates that a particular student enrolls in a particular course during a particular quarter.*
- *Each enrollment event indicates that during a particular quarter, a particular student enrolls in a particular course from a particular instructor in a particular classroom.* When you say this sentence for users, you can further help them understand the diagram by pointing to portions of the diagram as you say the corresponding portions of the sentence.

Visualizing Sample Data in Several Formats

Just as some users find some diction more helpful than other diction, some users prefer to view sample data in particular formats, that is, with different HOWs. Thus you need to be able to visualize data in several different formats. Throughout this book, you'll notice that we use different formats depending on what we're trying to illustrate. And sometimes



Fragment 7-9

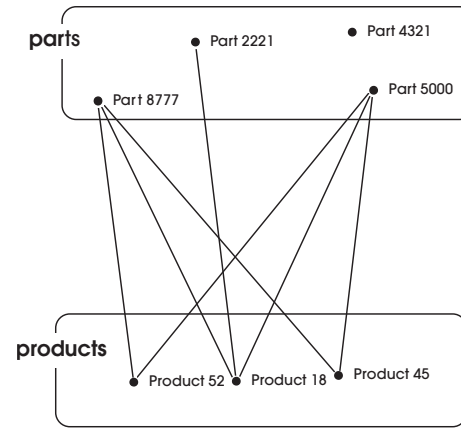
we use a less typical format to remind you that such a format exists or to remind you that you have choices about how you visualize the data.

Although there are many formats for illustrating any ramification of a shape, each format offers particular shortcuts for illustrating particular ramifications. For example, suppose you want to illustrate the left chicken foot from the shape in Fragment 7-9. That is, you want to illustrate that each *part* can have more than one *product.*

If you use the format in the accompanying table, you illustrate the effect of the chicken foot by saying, *Notice that within any particular column, there can be several* X*'s* (you can point to the column for Part 8777 when you say this). *That's because a part can have many products* (point to the left chicken foot).

In a diagram format, you illustrate the same ramification by saying, *Notice that from a particular part, there can be several emerging lines* (point to part 8777). *That's because a part can have many products* (point to the left chicken foot).

	Part 2221	**Part 8777**	**Part 4321**	**Part 5000**
Product 45		X		X
Product 52	X	X		X
Product 18		X		X



In the conversation about creatures and skills, notice that the data modeler used one of these visualization techniques to discuss Fragment 6-4.

Discussing and Illustrating Noteworthy Disallowed Data

Remember that as the LDS gradually evolves, the database boundary gradually moves. Each incremental change to the LDS connotes a correspondingly small expansion or contraction of the boundary. Until the LDS matures, most of these modifications are expansions of the boundary, as the users describe more and more of their data to you.

You can cue the users by anticipating how the boundary will expand and explicitly asking them about what you anticipate. That is, you can talk about interesting data that is just outside the boundary of an in-progress LDS.

For example, consider Fragment 6-14. You anticipated that the LDS could evolve into Fragment 6-15 because you realized that Fragment 6-14 could not accommodate some potentially noteworthy data. Specifically, you realized that a role could be worth remembering even if no creature has ever fulfilled that role in any achievement.

In that conversation, you merely talked about the disallowed data. But sometimes it is helpful to illustrate such data. This can get tricky because you need to make it clear to the users that you are illustrating the impossible. One good way to do this is to choose a format that illustrates the allowed data, jot down some sample data, then try to add the disallowed data. If you choose your format carefully, it will be clear that there is "no room" for the data.

Finding and Focusing on Shapes Within a Large LDS

After finding a shape, you can focus on it, temporarily ignoring the rest of the LDS. For example, in Fragments 6-18 through 6-23, you and the users focused on a small portion of the LDS, ignoring the entities named *contribution, role,* and *skill.* Such focus is possible because in Fragment 6-18, you recognized a shape, anticipated how it would evolve, and began to lead the users through a discussion that revealed the evolution.

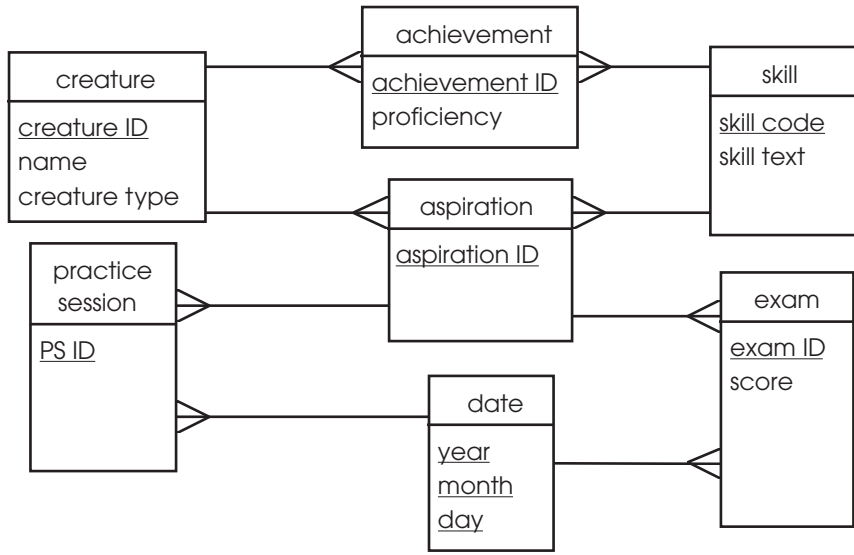
Shapes overlap. That is, when you focus your attention on a small LDS fragment, it can participate in several shapes. That's okay. In fact, you should expect that several shapes will apply to any particular LDS fragment. Because a fragment participates in several shapes, you can apply your knowledge about all the applicable shapes to it. You do not have to choose which shape is "most important" or which one is the "best match" for the fragment.

Look at Fragment 7-10 and consider two shapes:

- **Chicken feet in.** Shape in which an entity has several one-many relationships with the chicken feet near the entity. *Achievement* participates in the chicken-feet-in shape.
- **Chicken feet out.** Shape in which an entity has several one-many relationships with the chicken feet away from the entity. *Date* participates in the chicken-feet-out shape.

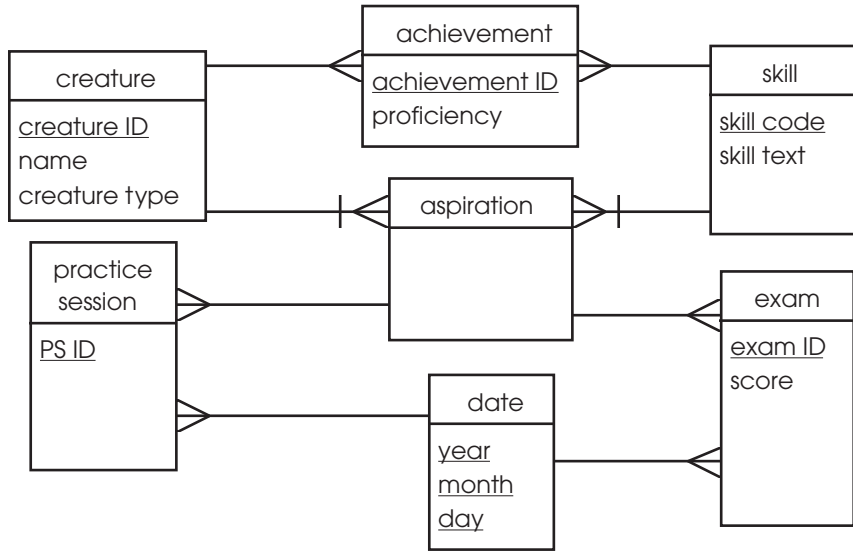
Notice that both shapes apply to the *aspiration* entity. You can see this easily from the diagram: *creature-aspiration-skill* looks like *creature-achievement-skill* (feet in), and *practice session-aspiration-exam* looks like *practice session-date-exam* (feet out).

Fragment 7-10 makes it easy to see that *aspiration* is like both *achievement* and *date,* even though *achievement* and *date* are not alike at all. Speaking realistically, things won't always be this easy. Keep the following in mind.



Fragment 7-10

- **Shapes are independent of layout.** You detect shapes by detecting patterns of entities, descriptors, and identifiers. Variations in layout might obscure these patterns. For example, *creature-aspiration-skill* is parallel to *date-exam-aspiration;* both exhibit the chicken-feet-in shape.
- **Fragments need not share any entities to share a shape.** For example, *practice session-date-exam* and *aspiration-skill-achievement* both exhibit the chicken-feet-out shape.
- **Fragments can be compared to canonical shapes.** Comparing fragments to one another is useful, and you will do this a lot while you are a novice modeler. You will make analogies between a new fragment and another fragment you have already scrutinized. But as you become a master, you will develop an abstract, content-neutral appreciation of the shapes. You won't always compare a new fragment to an old one; you will perceive a new fragment according to your masterful expectation of how fragments and shapes behave. This is what happened during the conversation about creatures and skills in Fragments 6-3 through 6-5. There was no other many-many relationship available for comparison, but the modeler knew in general how to work with many-many relationships.
- **Recognizing shapes is not a raw talent or gift; it requires knowledge.** You need to learn what the shapes are and how to recognize them. In particular, this means learning what to observe and what to ignore (temporarily) on an in-progress LDS. In Fragment 7-10, for example, you can perceive that *aspiration* participates in the chicken-feet-in shape by temporarily ignoring the lower relationships to *practice session* and *exam.* Likewise, in Fragment 7-11, you can appreciate that *aspiration* participates in the chicken-feet-in shape despite the



Fragment 7-11

change in the identifier. Part of the knowledge you need about the chicken-feet-in shape is that identifiers have no bearing on the shape. This is exactly the kind of information in Chapters 8 through 12.

You will find that every entity participates in many shapes. Consider the *contribution* entity of Fragment 6-22. It conforms to at least two shapes, *intersection entity* (see Chapter 8) and chicken feet in. Thus, when you discuss this fragment with users, you can leverage your skill with both of these shapes. Any data-visualization technique that applies to chicken feet in can apply to this fragment. Any conversational diction that can apply to intersection entities can apply to this fragment.

Recognizing the Differences Between Shapes That Are Similar but Not Identical

For any pair of similar but not identical shapes, you need to be able to articulate immediately the differences between them. One good way to articulate these differences is with sample data. That is, you design some sample data that is accommodated by one of the shapes but not by the other.

For example, consider the two similar shapes of Fragments 7-12 and 7-13. The top shape shows a two-part identifier for the *enrollment event* entity, whereas the bottom shape shows a three-part identifier. To appreciate the differences between these two shapes, you can design some sample data that highlights them.

For example, the data in the accompanying table shows that a student (ID number 123456) can take the same course (history of Peru) in several different quarters.