CHAPTER 4

Learning from the Data Model

In the previous chapter, we attempted to extract the essential tasks involved in a real-world problem and express them with use cases. We also made a first attempt at determining the data that are necessary to support those tasks and formed an initial data model, which we depicted with a class diagram. In this chapter, we look more closely at the data model to see how it can further our understanding of a database system.

A data model is a precise description of the data stored for a real-world problem, in much the same way that a mathematical equation describes a real-world physical event, or an architectural drawing describes the plan of a building. However, like a mathematical equation or an architectural plan, the data model is neither a complete nor an exact description of a real situation. It will always be based on definitions and assumptions, and it has a finite scope. For example, a high school student’s simple mathematical equation to describe the path of a ball tossed into the air will probably make assumptions about the constancy of the gravitational force and the absence of air resistance, and will likely assume low speeds where relativistic effects can be ignored. The equation is precise and correct for the assumptions that have been made, but it does not reflect the real problem exactly. It is, however, a good, pragmatic, and extremely useful description that captures the essentials of the real physical event.

A data model has similar benefits and limitations to a mathematical equation. It is a model of the relationships among the data items that are being stored about a problem, but it is not a complete model of the real problem itself. Constraints on money, time, and expertise will always mean that problems will need to be scoped and assumptions made in order to extract the essential elements. It is crucial that the definitions and assumptions are clearly expressed so that the client and the analyst are not talking at cross-purposes.

In the early stages of the analysis, as client and developer are trying to understand the problem (and each other), the details will necessarily be vague. In this chapter, we look at how the initial data model can be used to discover where definitions and scope may need to be more rigorously expressed.

Review of Data Models

The essential aspects of a data model were defined in Chapter 2. We will revisit these by way of an example that will highlight some additional features. Think about a small hostel that provides a number of single rooms for school groups visiting a national park. The hostel has a small database to keep track of its rooms and the people currently in residence. Because the hostel primarily deals with groups of students with a single point of contact, the idea of a group is central to their business model. It is still important to know which rooms particular students or teachers have been allotted. An initial data model to capture this information is shown in Figure 4-1.

You can see that there is a 1–Many relationship between the Group and Guest classes. Reading from left to right in Figure 4-1, we have that a particular group is related to one or more guests, and from right to left that a particular guest is associated with exactly one group. Figure 4-1 also depicts a 1–1 relationship between Guest and Room. Reading left to right, we have that each guest must be associated with one room and in the other
direction that a room can be associated with at most one guest but maybe none. In normal speak, we have that groups consist of a number of guests, and each guest has a room. Rooms are for one guest only, and they may not all be full. Some possible instances of these objects and relationships are shown in Figure 4-2. We have

**Figure 4-1.** Initial data model for the current occupancy of a small hostel

**Figure 4-2.** Objects and relationship instances consistent with Figure 4-1