4. Conceptual Overview

I don't pretend to understand the Universe—it's a great deal bigger than I am.
—THOMAS CARLYLE

TOPICS COVERED IN THIS CHAPTER

Understanding how to design a relational database isn't quite as hard as understanding the universe; in fact, it's much easier. It is important, however, for you to have an overall idea of the way the database design process works and a general idea of the steps involved within the process. The purpose of this chapter is to provide an overview of the database design process.

For the purpose of this overview, I've consolidated all of the techniques in the design process into seven phases and I discuss each phase in general terms. This discussion provides a good overall picture of thedatabase design process and I hope it will give you a much clearer understanding of each design technique covered in this part of the book.

You can use the design methodology in this book to design a new database completely from scratch, refine an existing database, or help you analyze an existing database so that you can design a new database based on the results of your analysis.

Note

A database can be designed by a single individual or a design team composed of two or more individuals. Throughout the remainder of the book, I use the phrase *database developer* and the word *developer* to refer to the person or group designing the database.

THE IMPORTANCE OF COMPLETING THE DESIGN PROCESS

One thing I want to make perfectly clear from the very beginning is the importance of completing the design process. I'm often asked if it's truly necessary to go through the entire design process. My answer is always a resounding "Yes!" I'm then asked whether it's still necessary if someone is only going to create a "simple" database. (*Simple* is one of the most dangerous words known to database developers. *Nothing* is ever "simple.") Again, my answer is yes, it's *still* necessary. The type, size, or purpose of the database is

totally irrelevant to the value of undertaking a fully developed design. You should implement and follow the database design process from beginning to end.

It is a well-known and proven fact that it is a bad idea to attempt to design a database without employing a thorough database design process. Many database problems are caused by poor database design, and *partially* following the design process is just about as bad as not using it at all. An incomplete design is a poor design. Only if you follow through with a whole, unabbreviated design process are you assured a sound structure and data integrity.

An important point to keep in mind is that the level of structural integrity and data integrity in your database is directly proportional to how thoroughly you follow the design process. The less time you spend on the design process, the greater the risk you run of encountering problems with the database. Thoroughly following the database design process may not eliminate all of the problems you might encounter when designing a database, but it will greatly help to minimize them. As you work with your RDBMS software, you'll find that a well-designed database is easier to implement than a poorly designed one.

Databases are not hard to design; it just takes a little time to design them properly. Don't allow yourself to take shortcuts when it seems as if the design process is taking too long—just be patient and remember what a wise old sage once said:

There's *never* time to do it right, but there's *always* time to do it over!

DEFINING A MISSION STATEMENT AND MISSION OBJECTIVES

The first phase in the database design process involves defining a *mission statement* and *mission objectives* for the database. The mission statement establishes the purpose of the database and provides you with a distinct focus for your design work.

Every database is created for a specific purpose, whether it's to solve a particular business problem, to manage the daily transactions of a business or organization, or to be used as part of an information system. You identify the purpose of your database and define it within a *mission statement*. This will help ensure that you develop an appropriate database structure and that you collect the data necessary to support the intended purpose of the database.

You'll also define *mission objectives* in this phase. These are statements that represent the general tasks your users can perform against the data in the database. You use these objectives to support your mission statement and to help you determine various aspects of the database structure.

There are two separate groups of people who will be involved in defining the mission statement and the mission objectives. The first group includes the database developer (you), the owner or head of the organization, and management personnel, and it is responsible for defining the mission statement. The second group includes the database developer (you again), management personnel, and end users, and it will be responsible for defining the mission objectives.

ANALYZING THE CURRENT DATABASE

The second phase in the database design process involves analyzing the current database, if one exists. Depending on your organization, the database will typically be a *legacy database* or a *paper-based database*. A legacy database (also known as an *inherited* database) is one that has been in existence and in use for several years. A paper-based database, as you may already know, is a loose collection of forms, index cards, manila folders, and the like. Whatever the database type or condition, analyzing it will yield valuable information about the way your organization is currently using and managing its data. The analysis also involves reviewing the way your organization is currently collecting and presenting the data. You look at how your organization uses paper to collect data (via forms) and present data (via reports). If your organization uses some software application program to manage and manipulate the data in the database, you study the way it collects and presents the data on-screen. Finally, you take into account how (if at all) your organization is using its data on the Web, and you review any browser-based applications that work with the database.

Another part of the analysis involves conducting interviews with users and management to identify how they interact with the database on a daily basis. As the database developer, you ask users how they work with the database and what their information requirements are at the current time. You then interview management personnel and ask them about the information they currently receive and their perception of the overall information requirements for the organization. These interviews are an important component of your analysis because the questions you ask (or don't ask) will have a great impact on your final database structure. You must conduct full and complete interviews if you are to design a database that truly meets your organization's information needs.

Next, you use the information you've gathered from the analysis and the interviews to compile an initial list of fields. You then refine this list by removing all calculated fields and placing them on their own list—you'll use these calculated fields later in the design process. The refined list constitutes your organization's fundamental data requirements and provides a starting point for the design of a new database. (As you know, nothing is ever truly final. Rest assured that you'll extend and refine this field list further as you develop your design.)

Once your initial field list is complete, you send it to your users and management for a brief review and possible refinement. You encourage feedback and take their suggestions for modifications into consideration. If you think the suggestions are reasonable and well supported, you make the appropriate modifications, record the list in its current state, and move on to the next phase.

CREATING THE DATA STRUCTURES

Creating the data structures for the database is the third phase in the database design process. You define tables and fields, establish keys, and define field specifications for every field.

Tables are the first structures you define in the database. You determine the various subjects that the tables will represent from the mission objectives you wrote during the first phase of the design process and the data requirements you gathered during the second phase. Then you establish these subjects as tables and associate them with fields from the field list you compiled during the second phase of the design process. After you've completed this task, you review each table to ensure that it represents only one subject and that it does not contain duplicate fields.

Now you go on to review the fields within each table. You refine all multipart or multivalued fields in the table so that they each store only a single value, and you move or delete fields that do not represent distinct characteristics of the subject the table represents. When you complete this review, you then review and refine the table structures. This involves checking the work you performed on the fields to ensure that you didn't accidentally miss anything, and ensuring that each table structure is properly defined. Next, you establish the appropriate keys for each table. Your main task is to ensure that each table has a properly defined primary key; this particular key uniquely identifies each record within a table.

The final step in this phase is to establish field specifications for each field in the database. Here you conduct interviews with users and management to help you identify the specific field characteristics that are important to them and review and discuss any characteristics with which they may be unfamiliar. After you've completed these interviews, you define and document field specifications for each field. You then review the table structures and field specifications with users and management once more for possible refinements. The table structures are ready for the next phase once you complete the refinements (if any) that you identified during the review.

DETERMINING AND ESTABLISHING TABLE RELATIONSHIPS

The fourth phase of the database design process involves establishing table relationships. You conduct interviews with users and management once again, identify relationships, identify relationship characteristics, and establish relationship-level integrity.

Working with users and management is a prudent exercise because they can assist you in identifying relationships among the data. You cannot possibly be familiar with every aspect of the data your organization uses, so leveraging whatever knowledge they have about the data they use will be very beneficial to you.

After you've identified the relationships, you establish a logical connection between the tables in each relationship with a primary key or with a linking table. What you actually use depends upon the type of relationship you're establishing between the tables. Next, you determine the type of participation and degree of participation for the tables in each relationship. In some cases, these participation characteristics will be obvious to you due to the nature of the data stored in the tables. In other cases, you'll base the participation characteristics on specific business rules.

DETERMINING AND DEFINING BUSINESS RULES

Determining and defining business rules is the fifth phase of the database design process. During this phase, you'll hold interviews, identify limitations on various aspects of the database, establish business rules, and define and implement validation tables.

The manner in which your organization views and uses its data will determine a set of limitations and requirements that you must build into the database. Your interviews with users and management will help you identify the specific constraints you will impose on the data, data structures, or relationships. You then establish and document these specifications as business rules.

The interviews you conduct with users will reveal *specific* limitations on various aspects of the database. For example, a user working with an order processing database is very aware of specific details, such as the fact that a ship date must occur later than an order date; that there must always be a daytime phone number; and that a shipping method should always be indicated. Your interviews with management, on the other hand, reveal *general* limitations on various aspects of the database. For example, the office manager for an entertainment agency is familiar with general issues such as the fact that an agent can represent no more than 20 entertainers and that promotional information for each entertainer must be updated every year.

Next, you define and implement validation tables as necessary to support certain business rules. Suppose you find that certain fields have a finite range of values because of the manner in which your organization uses them. You can use validation tables to ensure the consistency and validity of the values stored in those fields.

The level of integrity that business rules establish at this point is significant because it relates directly to the way your organization views and uses its data. The organization's perspective on the data will change as the organization grows, which means that the business rules must change as well. Determining and establishing business rules is an ongoing, iterative process, and you must be constantly diligent if you are going to maintain this level of integrity properly.

DETERMINING AND DEFINING VIEWS

The sixth phase of the design process involves determining and defining views. Here you'll conduct interviews (once again), identify various ways of working with the data, and establish the views.

You identify the types of views you need to build in the database by interviewing users and management and determining how they work with their respective data. You may find, for example, that many users require detailed information to perform their work, while others need only summary information to help them make strategic decisions for the organization. Each group of users must access information in very specific ways, and you can use views to accommodate these situations.

Next, you define the views you've identified during the interview process using the appropriate tables and fields, and establish criteria for those views that are required to retrieve specific information. For instance, you would establish criteria for a view that

must list all customers located in Texas or a view that must display the total number of authorized vendors (by city) in Washington.

REVIEWING DATA INTEGRITY

The seventh and final phase in the database design process involves reviewing the final database structure for data integrity.

First, you review each table to ensure that it meets the criteria of a properly designed table and you check the fields within each table for proper structure. You then resolve any inconsistencies or problems you encounter and review the structures once more. After you've made the appropriate refinements, you check table-level integrity.

Second, you review and check the field specifications for each field. You make necessary refinements to the fields and then check field-level integrity. This review reaffirms the field-level integrity you identified and established earlier in the database design process.

Third, you review the validity of each relationship, confirm the relationship type, and confirm the participation characteristics for each table within the relationship. You then review relationship integrity to ensure that there are matching values between shared fields and that there are no problems inserting, updating, or deleting data in either of the tables within the relationship.

Finally, you review the business rules that you identified earlier in the database design process and confirm the constraints you've placed on various aspects of the database. If there are any other limitations that have come to your attention since the last set of personnel interviews, you establish them as new business rules and add them to the existing set of business rules.

You're ready to implement your logical database structure in an RDBMS program once you've completed the entire database design process. However, the process is never *really* complete because the database structure will always need refinement as your organization evolves.

SUMMARY

We began this chapter with a discussion of the importance of completing the design process, and you learned that designing a database without the benefit of a good design method leads to poor and improper design. We also discussed the fact that the level of structural and data integrity is in direct proportion to how thoroughly you follow the design process. You then learned that inconsistent data and inaccurate information are two problems typically associated with poorly designed databases.

Next we looked at an overview of the entire database design process. The process was consolidated into the following phases in order to provide you with a clear picture of the general steps involved in designing a database.

1. *Define a mission statement and mission objectives for the database.* The mission statement defines the purpose of the database, and the mission objectives define the tasks that are to be performed by users against the data in the database.

2. *Analyze the current database.* You identify your organization's data requirements by reviewing the way your organization currently collects and presents its data and by conducting interviews with users and management to determine how they use the database on a daily basis.

3. *Create the data structures.* You establish tables by identifying the subjects that the database will track. Next, you associate each table with fields that represent distinct characteristics of the table's subject, and you designate a particular field (or group of fields) as the primary key. You then establish field specifications for every field in the table.

4. *Determine and establish table relationships.* You identify relationships that exist between the tables in the database and establish the logical connection for each relationship using primary keys and foreign keys or by using linking tables. Then you set the appropriate characteristics for each relationship.

5. *Determine and define business rules.* You conduct interviews with users and management to identify constraints that must be imposed upon the data in the database. The manner in which your organization views and uses its data typically determines the types of constraints you must impose on the database. You then declare these constraints as business rules, and they will serve to establish various levels of data integrity.

6. *Determine and establish views.* You interview users and management to identify the various ways they work with the data in the database. When your interviews are complete, you establish views as appropriate. You define each view using the appropriate tables and fields, and you establish criteria for those views that must display a limited or finite set of records.

7. *Review data integrity.* This phase involves four steps. First, you review each table to ensure that it meets proper design criteria. Second, you review and check all field specifications. Third, you test the validity of each relationship. Fourth, you review and confirm the business rules.

REVIEW QUESTIONS

1. Why is it important to complete the design process thoroughly?

2. True or False: The level of structural integrity is in direct proportion to how thoroughly you follow the design process.

3. What is the purpose of a *mission statement?*

4. What are *mission objectives?*

5. What constitutes your organization's fundamental data requirements?

6. How do you determine the various subjects that the tables will represent?

7. True or False: You establish field specifications for each field in the database during the second phase of the database design process.

8. How do you establish a logical connection between the tables in a relationship?

9. What determines a set of limitations and requirements that you must build into the database?

10. What can you design and implement to support certain business rules?

11. How do you determine the types of views you need to build in the database?

12. When can you implement your logical structure in an RDBMS program?