Chapter 1

Introduction

The topic of databases, and especially relational databases, may sound pretty far removed from what you need or can even comprehend. You might be surprised to know how heavily our society depends on databases. Businesses, governments, and organizations of all types would quickly grind to a halt if the world's databases were to suddenly disappear.

Databases store critical business information for small, medium-sized, and large organizations. They store scientific information for just about every field you can imagine, from the Human Genome Project, to sub-atomic physics experiments, to spacecraft exploring the Solar System. Sports teams store information about their own and opposing teams in databases. Even people with hobbies, such as those tracing their ancestry through the study of genealogy, keep the data they collect about their ancestors in databases. Wherever large quantities of facts are accumulated, those facts are stored in databases.

Knowing how to design and build a useful and reliable database system is an important skill for anyone who deals with quantities of data. Database development is an interesting and rewarding career, but you don't have to be a professional to benefit from knowledge of the process. If you are a user of data or a manager of such users, you need to know what constitutes a good design, what the major development pitfalls are, and what must be done to keep an existing database healthy. Good database design is not easy. However, Beckwith Consulting, inc. has a solid grounding in how to build robust and reliable database systems.

The recent explosive growth of the Internet and of the World Wide Web has made databases even more important. Unimaginably huge quantities of data are transmitted across the Web every day. Practically all of that data is stored in a database somewhere, often in multiple databases. Understanding databases is the key to understanding one of the major components of E-business, E-commerce, and E-everything.

Chapter 2

Dealing with the Data Deluge

Current technology has made it possible to collect mind-boggling amounts of data about every topic imaginable. Even the lowest cost personal computers can store billions of bytes, or gigabytes, of data. Business and scientific computers can store much more.

With current communications technology you can transmit data at near the speed of light to people anywhere in the world. As a result, organizations as well as individuals have more data at their disposal than they could possibly analyze. Buried in that mountain of data is a relatively small fraction that would actually answer a question you might have. The problem of how to extract only the information you want from the mountain of data that you don't want is a bigger problem today than it has ever been. With each passing day the problem gets worse.

Database Structure

The only hope for making large quantities of data manageable is to impose a structure on it that separates it into meaningful pieces. With structure, you don't have to wade through all the data to find what you want. You only have to look at the specific pieces that are directly related to your question.

Giving structure to data means organizing it according to its meaning. For example, if you had a business and wanted to keep track of your customers, you would want to keep their records separate from employee records, product records, etc. Structuring the data so that all the customer records are always in one place, all the employee records in another, and so on, means that you don't have to sift through all the data to answer a question. You only need to examine the data that is directly relevant.

Database Systems

NASA spearheaded the move from flat file systems to database systems in the early 1960s, to meet the needs of the Apollo Moon landing program. The Apollo spacecraft, including the Saturn V launch vehicle, the Command and Service Module, and the Lunar Excursion Module, was the most complex machine ever built. It was composed of hundreds of thousands of components, every one of which had to be tested to a higher level of reliability than any machine component had ever been tested before.

Apollo flight director Gene Kranz is famous for saying, 'Failure is not an option.' after the Apollo 13 spacecraft was crippled by an explosion on the way to the Moon. Running low on oxygen, fuel, and everything else, that philosophy brought the three Apollo 13 astronauts home alive and well. During that era, the 'Failure is not an option.' ethos permeated thinking at NASA.

NASA contracted with IBM to build and supply the Apollo project with a database system that offered the required high capacity and reliability. IBM delivered the system that made it possible to keep track of and verify the reliability of all the components of the Apollo spacecraft, which in turn made the Moon landings possible. It even made possible the rescue of the crew of the ill-fated Apollo 13 mission. By quickly retrieving an inventory of the exact items available on the damaged ship, engineers on the ground were able to design a jury-rigged carbon dioxide scrubbing device that saved the astronauts from suffocation. IBM later sold a commercial version of the Apollo database, called MS. MS was an important factor in IBM's dominance of the computer market throughout the 1960s and 1970s.

Relational Database Systems

Since the early 1980s, practically all new database systems that don't have to be compatible with older legacy systems, adhere to the relational database model. A legacy system is any system that was designed and built before relational databases came into common use. As a result, just about any database you are likely to encounter will be a relational database.

Relational databases are fast, reliable, and flexible, and allow for minimum duplication of data, which is usually referred to as data redundancy. The flexibility comes from the fact that data is stored in multiple tables, which are largely independent of each other. You can add new tables, delete obsolete ones, or change the structure of a table while having minimal impact on the others. This flexible architecture makes it easy to change the database to reflect the changes that inevitably occur in any growing business.

Databases can be big. They can contain millions, billions, or even trillions of bytes of data. Those capacities require massive storage space, typically on multi-gigabyte hard disk

drives. The database management system, or DBMS, is that part of a total system that interfaces directly with the database.

A database application is a program you write to make it easy for computer users to communicate with their database. The DBMS obeys instructions it receives from the application you develop, which in turn interacts with the user.

The application makes it easy for computer users to add new data to the database, change existing data, and delete obsolete data, as well as retrieve the information they want. In a sense, the database application is a bridge between the human user and the database. It translates human actions into commands that the database can understand, and translates database responses into a form that humans can understand.

Chapter 3

Databases, Database Applications and SQL

As we have seen, databases are repositories of potentially huge quantities of data. Retrieving the small quantity of data that is the answer to whatever question you currently have, is arguably the most important database operation. After all, if you can't get the information you want when you want it, why bother to store it?

There are two ways to get information out of a database:

First, the user can type a question, or query, from his or her computer keyboard and wait for the answer to appear on the screen. This is convenient if you have a onetime question that you are not likely to want to ask again.

Second, a database application developer can write a program that gives the user a set of predetermined options. When the user selects an option, the program sends a query to the database and receives whatever the database returns, possibly printing out a report, or displaying a form on the screen. This method is best for questions that are likely to be asked repeatedly.

The first option above, the user typing a question, rarely happens in practice because most computer users do not know SQL, the language that database queries must be written in. The second option requires Beckwith Consulting, Inc., as the database application developer, to know SQL, but the user need not. Since building database applications is our job, you are not going to need to know SQL anyway.

The purpose of a database application is to provide you the user with an easy-to-understand interface to the database. Ideally, a database application is not only easy to understand, but also foolproof. We want to design and build your applications in such a way that you and your users can easily tell what to do in order to accomplish their tasks. We also want to make it impossible for them to add invalid data to the database, or corrupt it in some other way.

The DBMS understands instructions it receives from the application, which are in the form of SQL statements. SQL is an internationally recognized standard language, specifically designed to communicate with database management systems. Virtually all relational database management systems work with some form of SQL

SQL has three primary functions:

- Creating and maintaining a database structure
- Manipulating data within the structure

- Protecting the data and the structure from harm

SQL Main Components

There are three parts of SQL, each corresponding to one of the primary functions. The Data Definition Language, or DDL, takes care of creating and maintaining database structures. The Data Manipulation Language, or DML, deals with the data that goes into those structures. The Data Control Language, or DCL, is responsible for protecting both the structure and the data from either inadvertent or intentional harm.

SQL is all you need to create, maintain, manipulate, and protect a database, but is rarely used to perform all those functions. Commercial grade database management systems provide tools that make it easy to create and maintain database structures. They provide other tools to help you add, change, or delete the data that you keep in your tables. In general, database developers will use these tools to create databases rather than using SQL. For small systems, you may even teach data entry operators how to use the tools for adding, changing, and deleting table data. However, for more complex systems, the possibility of data entry errors is too great. It is better to create and use database applications.

Chapter 4

Where and How Databases are Used

Today there are three primary classes of databases: enterprise databases, personal databases, and workgroup databases. The classes differ in the size of their user bases as well as the size of the databases they support.

Enterprise Databases

The enterprise databases were the first to be developed. When databases first appeared in the early 1960s computers were big and expensive. Only large organizations such as NASA, the Department of Defense, and American Airlines could justify the cost of having them. These computers hosted database applications that were critical to the success of the organizations that had them. Today, enterprise databases are even more vital to large organizations around the world. They typically support hundreds or thousands of users who access them from remote client computers.

Personal Databases

The growing power of the personal computer in the early I 980s gave rise to the personal database. The personal database serves a single user, sitting at a single computer, and resides on the computer’s hard disk drive. This type of database serves the needs of individuals and small organizations that do not have to share information among multiple people. This is the environment in which products such as Microsoft Access are designed to run. Personal databases are generally the least complex and least powerful, as well as the easiest to learn and use. They are designed for people with a minimum of

training in database use. They are not appropriate for major database development projects.

Workgroup Databases

Workgroup databases rose to prominence in the 1990s as more and more personal computers in organizations were linked together in local area networks. Linking computers made it much easier for coworkers to share information, greatly enhancing productivity. Database management systems designed to operate in a workgroup environment are significantly more complex than personal database management systems. They must protect the database from being corrupted when two people try to access the same data at the same time. They also must protect it from malicious damage. It is much harder to control physical access to all the computers connected to a network than it is to control such access to a single computer. Microsoft SQL Server, IBM DB2, and Oracle are examples of DBMS products in the workgroup class. These products also scale up to support enterprise database systems.

Tracking Business Information

Probably the most widespread use of databases today is in maintaining business information. Any business that is more complex than the corner produce market wants to keep track of its customers. Once you have a system for maintaining customer information, you might as well keep track of your products and your employees too. It's only a short step from there to monitoring sales and purchase transactions, maintaining accounts payable, accounts receivable, profit and loss, and inventory records. Businesses that are performing any of these functions electronically, whether they know it or not, are using databases.

Larger businesses have even more use for databases. They use them for market analysis and as an aid to planning what new products to introduce. They maintain information on their competitors, and whether or not they are losing market share to them. They keep reams of data on product tests. The recall of a poorly designed product can cost a company millions or even billions of dollars. It's better to test products exhaustively before they are sold, recording all results in a database, and then fix any defects before they turn into lawsuits.

Databases in E-commerce

E-commerce represents a relatively new business use of databases. Businesses make transactions over the Internet. Amazon.com is an example of a business whose primary assets are its databases. Amazon has advertised itself as Earth's largest bookstore, based on the millions of books they list in their database. They stock only a fraction of the books they list, which dramatically lowers their operating costs. Amazon also sells a variety of other products from music CDs to lawnmowers. The database keeps track of source inventory levels and processes sales transactions.

Another Amazon distinctive is the amount of information it keeps on its customers. In addition to the usual name and address information, Amazon keeps track of what products customers have bought in the past, and even of what they have shown an interest in but not bought. With this information, they automatically construct a profile of each customer.

Whenever a repeat customer visits the Amazon site, a personalized screen appears that advertises products they are most likely to buy, based on the preferences indicated by their profile. This profile is stored in a database, which is updated every time a customer returns to the site. Every time I return to the Amazon Web site, it has new recommendations for me. Often I am surprised by how close those recommendations are to what interests me. The system works. I have bought recommended items that I otherwise would not have known existed.

Databases in Scientific Research

Scientific enterprises store huge volumes of data in databases. In the biological sciences, the genomes of humans, mice, yeast, and fruit flies have been recorded in databases. Sequencing the genomes of many more plants and animals is currently underway. Spacecraft exploring the solar system are returning huge quantities of data about the sun, Mars, Jupiter, and other planets, asteroids, comets, and moons. To be most useful, this data must be stored in a database so that the parts of interest can be quickly accessed and examined.

Databases in Government

Government entities, such as the Social Security system, the Internal Revenue Service, and the Selective Service, or draft board, also maintain massive databases. Many smaller databases are used at the federal, state, county, and city levels. All of these databases and their associated applications, whether used by businesses, scientific enterprises, governments, non-profit organizations, or individuals are conceived, designed, and built by database developers.

Chapter 5

Conclusion

Even small organizations have so much computer data today that the only hope for finding the small pieces you want is to store it in a structure that makes locating, retrieving, or synthesizing those small pieces easy. All kinds of people and organizations use databases to store data that is of value to them. The databases come in all sizes, and run on a wide variety of computers, some standing alone, some connected to local area networks, and some connected to the Internet.

Over the last 40 years databases have evolved into the powerful and flexible relational databases that dominate the market today. Relational databases store data records in tables. You can create and manipulate those databases using the SQL data sublanguage or with tools provided by database management system vendors