Database Concepts

Substantially adapted from Capron, Computers, 6th edition, 2000.

Motivation

A database is an organized collection of related data. In a loose sense you are using a database when you use a phone book or when you take papers from a file cabinet. Unfortunately, as the amount of data increases, creating, storing, changing, sorting, and retrieving data become overwhelming tasks.

Suppose you had a collection of names and addresses, each on a separate index card, stored in an index-card file. If you had only 25 cards, sorting the cards into alphabetical order or even finding all the people who have the same zip code would be fairly easy. But what if you had 100, or 1000, or 10,000 cards? What if you had several different boxes, one organized by names, one by cities, and one by zip codes? What if different file clerks added more cards each day, not knowing whether they were duplicating cards already in the file? And what if another set of clerks was trying to update the data on the same cards? As you can see, things might get out of hand. Enter computers and database management software.

A database management system (DBMS) is software that helps you organize data in a way that allows fast and easy access to the data. In essence, the program acts as an efficient and elaborate file system. With a database program you can enter, modify, store, and retrieve data in a variety of ways.

Databases are different

It should be noted that database systems are different from word processing or spreadsheet software. Generally, most users have a good understanding of both word processing and spreadsheets. They enter and use the data in the same form as that in which it resides on disk. Data in databases, however, can reside on the disk in ways unknown to a user. In particular, sophisticated database systems, particularly those designed for a mainframe computer environment, are complex and must be planned and managed by computer professionals.

Several advantages are generally associated with databases

Reduced redundancy. Data stored in separate files, as opposed to in a database, tends to repeat some of the same data over and over. A college, for example, needs to have various kinds of information for a student--perhaps financial, academic, and career data. If each of these sets of data is in a separate file, some repetition is inevitable--such as each student's name, address, and Social Security number or other identification. In a database most of this information would appear just once.
Integrated data. Rather than being in separate and independent files, data in a database is considered integrated because any item of data can be used to satisfy an inquiry or a report. This advantage is related to the reduced redundancy advantage: Since data can be retrieved from any place in the database, many specific data items need not be repeated.

Integrity. People who maintain any kind of file hope that it has integrity, that is, that the file is accurate and up-to-date. Integrity concerns increase as the

sophistication of the data increases. Reduced redundancy increases the likelihood of data integrity.

General Database Concepts

Database Models

The way a database organizes data depends on the type, or model, of the database. There are three main database models: hierarchical, network, and relational. Each type structures, organizes, and uses data differently. Hierarchical and network databases are usually used with large computers. However, relational databases are used with personal computers as well as mainframes.

A relational database organizes data in a table format consisting of related rows and columns.

Each box in a table contains a single piece of data, known as a data item, such as the name of one city. Each column of a table represents a field, which is a type of data item. The specific data items in a field may vary, but each field contains the same type of data item--for example, first names or zip codes. The full set of data in any given row is called a record. Each record has a fixed number of fields. The fields in a particular record contain related data--for example, the name and address of a person. A collection of related records makes up a file. In a relational database a file is also called a relation. There can be a variable number of records in a given relation. There can also be more than one file in a database.

Database PowerNow that you know what *field, record, file,* and *relation* mean, you are ready to glimpse the real power of databases. The power is in the connection: A relational system can relate data in one file to data in another file, allowing a user to tie together data from several files. A database can store data relationships so that files can be integrated.

CREATING A DATABASE

After you have considered your needs, such as what reports and inquiries you will want to make, there are two steps to creating a database file: (1) designing the structure of the file and (2) entering the data into the file.

File Structure

Sketch on paper the **file structure**--what kind of data you want in each column. To create the file structure, you must choose meaningful fields. The fields you choose should be based on the data you will want to retrieve from the database. Let us take a look at each type.

Field Name: Names of the types of data you want to use are called field names. Each field must have a unique name.

Field Type Although different software packages offer different field types, there are four commonly used types of fields: character fields, numeric fields, date fields, and logical fields. **Character fields** contain descriptive data, such as names, addresses, and telephone numbers. **Numeric fields** contain numbers used for calculations, such as rate of pay. **Date fields** are usually limited to eight characters, including the slashes used to separate the month, day, and year**. Logical fields** are used to keep track of true/false, yes/no conditions.

Key Fields One or more key fields can be designated as a field on which an inquiry to the database can be based. A key field is sometimes called an *index field.*

Access presents the *design view* to accept the file structure.

Entering the data

When it is time to key in the records in the file, Access presents a table called *datasheet view*.

After you have filled in all the data for the first record, the database program automatically displays another blank input line so that you can enter the data items for the fields in the second record.

The following are descriptions of operations available with any database software package.

List the records. You could ask for a list of all existing records, either displayed on the screen or printed out on paper. If you are displaying the records on-screen, the software displays only as many records as will fit on the screen. Scrolling up or down displays additional records. If there are a large number of fields in a record

List specific fields. In addition to printing all records, you have the option of printing just certain fields of each record. Perhaps, to satisfy a customer request, you could print only the Description and Cost fields for each record. The software also offers the option of printing the fields in any order requested, not just the order in which they appear in the record. For example, you could request a list of these fields in this order: Description, Tour-ID, Walk, and Hours.

Query. You can make a query--ask a question--about the records in the file. You will need to use a relational operator when entering instructions that involve making comparisons. These operators are particularly useful when you want to locate specific data items.

Add new records. You can add records for new Tours at any time

Modify existing records. You may need to change an existing record.

Delete records. Sometimes a record must be removed, or deleted, from a database file. Database software provides this option.

Review

A database is an organized collection of related data. A database management system (DBMS) is software that creates, manages, protects, and provides access to a database.

Advantages of databases are reduced redundancy, integrated data, and integrity

A database can store data relationships so that files can be integrated. The way the database organizes data depends on the type, or model, of database. There are three main database models--hierarchical, network, and relational.

A relational database organizes data in a table format consisting of related rows and columns. Each location in the table contains a single piece of data, known as a data item. Each column of the table represents a field, which consists of data items.

The full set of data in any given row is called a record. Related records make up a file. In a relational database, a file is also called a relation.

Computer activities are considered transparent if a user is unaware of them as they are taking place.

The power of databases is in the connection: A relational system can relate data in one file to data in another file, allowing a user to tie together data from several files.

There are two steps to creating a database file: (1) designing the file structure and (2) entering the data.

When a file structure is defined, many database programs require the user to identify the field types, field names, and field widths. Field names are used to describe the data you want to use. There are four commonly used types of fields: character fields, numeric fields, date fields, and logical fields.

Once a file structure is defined, it is presented to the user as an input form so that data for each record may be entered.

A relational operator is needed when making comparisons or when entering instructions to query the database.

EXAMPLES

Consider these workers as possible users of a database: a crime lab technician handling evidence, a tulip-bulb grower who produces 47 varieties for more than 200 customers, a runners club that tracks meets and member data. What data might such users want to look up? What data would they need to store? What fields might be key fields?

An environmental organization concerned about preserving undeveloped land keeps a database of its donors with these fields: last name, first name, street address, city, state, zip code, phone number, amount of last donation, date of last donation, amount of highest donation, date of highest donation, amount of average donation, code for special interests (M for mountains, R for rivers, and so forth). The organization regularly sends out form letters soliciting donations. The form letters are keyed to a particular donor population, for example, those who have not sent a donation in six months or those who might want to give to a special shorebirds preserve. Among the fields listed, which might be key fields?