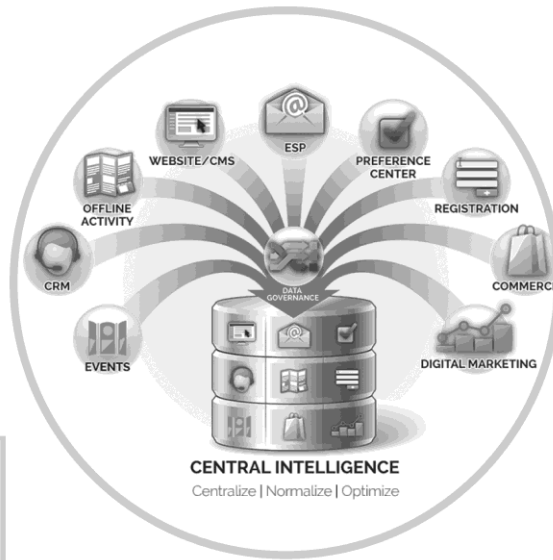CENTRAL INTELLIGENCE
Centralize | Normalize | Optimize

1
UNIT

DATABASE MANAGEMENT SYSTEM

Learning Objectives

After studying Unit 01, the students will be able to:

- Definition of data and information.
- The various database architecture models.
- The role of database administrator.
- The normalization.
- The issues in database security.
- The various database design methodology.

1.1 INTRODUCTION

DATA:

Data is raw fact or figure or entity. When activities in the organization take place, the effect of these activities need to be recorded which is known as data. Thus, the quantities, symbols, or characters on which operations are performed by a computer, which may be stored and transmitted in the form of electromagnetic or electrical signals and recorded on storage devices such as magnetic, optical, or mechanical recording media.

Eg. Chanakya, 32, 601 → Raw facts

Data collection is usually done with software, and there are many different data collection procedures, strategies, and techniques. Most data collection is centered on electronic data, and since this type of data collection encompasses so much information, it usually crosses into the realm of big data.

So why is data collection important? It is all the way through data collection that a business or management has the quality information they need to make informed decisions from further analysis, study, and research.

INFORMATION:

Processed data is called information. The purpose of data processing is to generate the information required for carrying out the business activities. In other words, information is knowledge that derived from data, which have been transformed, interpreted, stored and used for some purposes. With that information we can find out what we don't know before and it will affect what we already know.

Besides, we can make a right decision. Decisions are impossible without information and users are constantly seeking more and better information to support decision making. It also can reduce the sense of doubt and a sense of uncertainty about the event. For example, accounting information is very important for a company to be able to determine the profit or loss of the business.

Eg

Name	Roll	Id Number
Chanakya	32	601

= Information

For example:

The temperature of last months is

13, 12, 14, 12, 17, 10, 11, ...

Above numbers have no meaning. Hence it is data. But using those data following information can be obtained.

Highest temperature is 17

Lowest temperature is 10

DATABASE:

A database is a repository for collection of data which is organized in such a way that it is can be accessed by wide variety of different application programs. It is actually a single organized collection of structured data. The data in the database are always common to all the users of the system. A very simple example of database is telephone diary. Thus database is a collection of related data.

FUNCTION OF DATABASE :

There are following functions of database

i.   Storing, manipulating and managing information.

ii.   Reducing repeated storage of data.

iii. Keeping latest and correct information.

iv.   Providing data security.

v.   Improving disk optimization.

vi.   Fast accessing of data.

DATABASE MANAGEMENT SYSTEM (DBMS) :

DBMS is a collection of program that enables user to create and maintain a database. DBMS is a general purpose software system that facilities the process of defining, constructing and manipulating database for various application. A DBMS handles user request for database actions and allows for control of security and data integrity requirement. DBASE, MS Access, SQL, Oracle etc are some software package to work with DBMS.

OBJECTIVES OF DBMS

Following are the objectives of DBMS

1.   Providing very easy access to data for user.

2.   Providing very fast response to user request for any data.

3.   Removing any duplicate data.

4.   To provide huge space for storage of relevant data.

5.   To facilitate more than 1 user to use same data at a same time.

6.   To give protection to the data against unwanted access, unauthorized access and   any

kind of physical harm.

7. To preserve the data content i.e. to maintain data integrity.

8. To freely allow the growth of the database system.

9. To give the user the latest modification of the database.

1.2 BASIC COMPONENTS OF DBMS

Object

A database object is any defined object in a database that is used to store or reference data. Database objects can be made using the create command. These database objects are used for holding and manipulating the data in the database. The examples of database objects includes view, table, sequence, synonym, index etc.

Table

Tables are the main units of data storage in a database. A table is a collection of data about a specific topic; it is made up of one or more fields.

Field

A field is a column in a table and defines a data type for a set of values in a table. For example, a mailing list table might include fields for first name, last name, address, city, state, zip code, and telephone number.

Record

A record is a row in a table and is a set of values defined by fields. In a mailing list table, each record would contain the data for one person as specified by the intersecting fields.

Ms. Access database consists of several different components. Each component listed is called an *object*.

Queries

Queries are basically questions about the data in a database. A query consists of specifications indicating which fields, records, and summaries you want to see from a database. Queries allow you to extract data based on the criteria you define.

Forms

Forms are designed to ease the data entry process. For example, you can create a data entry form that looks exactly like a paper form. People generally prefer to enter data into a well-designed form, rather than a table.

Reports

When you want to print records from your database, design a report. Ms. Access even has a wizard to help produce mailing labels.

Pages

A data access page is a special type of Web page designed for viewing and working with data from the Internet or an intranet. This data is stored in a Microsoft Access database or a Microsoft SQL Server database.

Macros

A macro is a set of one or more actions that each performs a particular operation, such as opening a form or printing a report. Macros can help you automate common tasks. For example, you can run a macro that prints a report when a user clicks a command button.

Modules

A module is a collection of Visual Basic for Applications declarations and procedures that are stored together as a unit.

Data type

Data types are the properties of each field. A field only has one data type, such as Character, Number or Date.

KEYS IN DBMS:

Generally keys are used for proper identification of an entity. Key is an attribute or set of attribute which uniquely identifies each rows of data in a table or establishes relationship with another table. Following are the types of keys:

a.   Candidate key:

An attribute or set of attribute is called candidate key if it uniquely identifies records in a table.

Table : Student

Serial No.	Regd.	Name	Address
1	1073	Rohan	Pokhara
2	1072	Sohan	Pokhara
3	1071	Mohan	Kathmandu

In the table student the value of serial no., regd. and name are all unique in all rows. Hence all these 3 (serial no. regd. And name) are candidate key.

b.   Alternate key:

Alternate Key or Secondary Key is the key that has not been selected to be the primary key, but are candidate keys.

In above table, Serial No., Regd. and Name are the candidate keys. They are considered candidate keys since they can uniquely identify the student record. Select any one of the

candidate key as the primary key. Then rest of the two keys would be Alternate or Secondary Key.

c.   Primary key:

A primary key is a value that can be used to identify a unique record in a table. It is the candidate key which is chosen of identifying each row in a table. In the above table student; serial nos. can be chosen as primary key. Primary key doesn't allows duplicate values and Null values.

d.   Foreign key:

A Foreign Key is a database key that is used to link two tables together. When a primary key of one table appears as an attribute in another table, then that is foreign key in the second table. It is used to relate 2 tables.

Table Faculty member		
Dep_code	Dep_name	Location
PHY	Physics	P Block
CHE	Chemistry	C Block
MAT	Math	M Block

Table Department		
Facult_mem_code	fac. Name	Dep_code
001	Rohan	PHY
002	Sohan	CHE
003	Kabita	MAT

Here, if Dep_code was primary key in department table. Then as it appears as an attribute in faculty member table, so Dep_code is foreign key in the faculty member table.

1.3 ADVANTAGES OF DBMS

There are following advantages of DBMS

    i.   Data redundancy and data inconsistency can be reduced.

    ii.   Data can be shared among all authorized user.

    iii.   Data integrity is improved.

    iv.   User can get better service.

    v.   Cost of developing and maintaining system is cheaper.

    vi.   Great deal of programming time can be saved.

    vii.   Report can be generated for a particular requirement easily.

DISADVANTAGES OF DBMS:

There are following disadvantages of DBMS

i. There is always a possibility of loosing confidentiality and privacy.

ii. In countries like Nepal, using computerized database system is too expensive.

iii. DBMS can be some time too complex to understand and implement due too many rules.

iv. The necessary trained manpower is unavailable.

1.4 COMPONET OF DBMS:

Structured Query Language(SQL) is the database language by the use of which we can perform certain operations on the existing database and also we can use this language to create a database. SQL uses certain commands like Create, Drop, Insert, etc. to carry out the required tasks. These SQL commands are mainly categorized into four categories as: Data Definition Language (DDL), Data Query Language (DQL), Data Manipulation Language (DML) and Data Control Language (DCL).

Two components of DBMS are there

A. **Database Management Language**

B. **Data Dictionary**

Each of above are discussed in topics below:

A. **Database Management Language:**

We have following type of database management language.

i. Data Definition Language (DDL):

It is the language which defines all attributes and properties of a database, specially record layouts, field definitions, key fields, file locations and storage capacity. Using this language the logical structure and the file within the database may be defined. After compiling DDL statements we get a set of table which is stored in data dictionary. Data Definition Language serves like our interface for application program that uses the data. Let's look an example, suppose a payroll program needs employment number of an employee, then the DDL is used to define the logical relationship between employment number and take other data in the database and it also acts as an interface between payroll program and the file that contain employment number. Examples of DDL command are:

CREATE - is used to create the database or its objects (like table, index, function, views, store procedure and triggers).

ALTER - is used to alter the structure of the database.

DROP - is used to delete objects from the database.

TRUNCATE – is used to remove all records from a table, including all spaces allocated for the records are removed.

COMMENT – is used to add comments to the data dictionary.

RENAME – is used to rename an object existing in the database.

ii. Data Manipulation Language (DML):

DML is the database language which mainly does processing and manipulating the data in the database. Actually the DML provides a comprehensive set of commands to allow modification of the data within a data base. We can use DML to insert data in, update and query a database. DML are often capable of performing mathematical and statistical calculation that facilitates generating reports. Nowadays, in 4GLS DML is seen to be integrated with other programming language to implement sophisticated data base function. Examples of DML commands are:

INSERT – is used to insert data into a table.

UPDATE – is used to update existing data within a table.

DELETE – is used to delete records from a database table.

iii. Data Inquiry Language:

These language are used for querying the data base and we receive the filtered data which can be used for various different application.

B. DATA DICTIONARY:

Data dictionary is a master table which contains data about all the databases in a data base system. They are also called data dictionaries. Data dictionary contains information such as:

i.   What tables and column are included in the present structure.

ii. Name of current table and column.

iii. The characteristics of each item of data, such as its length and data type.

iv. Any restriction on the value of certain columns.

v.   The meaning of any data field that are not self evident for example a field such as "course type".

vi. The relationship between items of data.

vii. Which program access which item of data and whether they merely read the data or change it.

Data dictionary allows DBMS to keep track of the data and helps the developers and users in finding the necessary data. A good data dictionary has to ensure that there is consistent definition of data across different data bases. An example is MS Access holds data dictionary as a set of system table called MS SYS object.

1.5 DATABASE MODEL:

Database model depicts the way in which data in the database are structured(represented). Furthermore it also shows relationship among the data elements. Following two type of database model are there.

1.   Conceptual model   2.   Implementation model

Conceptual model focuses on how the data element in the database is to be grouped. Implementation model focuses on how the data are represented in the database. Both of these model has types which are discussed as follows:

1.   Conceptual data model:

Entity Relationship (ER) model is a conceptual data model.

ER model represents real world situation using concepts commonly used by people. ER model shows graphical representation of entities and their relationship in a database structure. ER model is the basis for the design of a database by the database administrator. The elements of ER model are as follows:

A.   Entity :

An entity is a thing of interest to an organization about which data is to be held. Example is customer, employee supplier. It is denoted by rectangle.
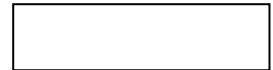
B.   Attribute :

Attribute is a property or characteristic of an entity. Example of attribute associated with a customer include customer id, customer name, title, address etc. It is denoted by an oval symbol.
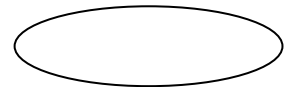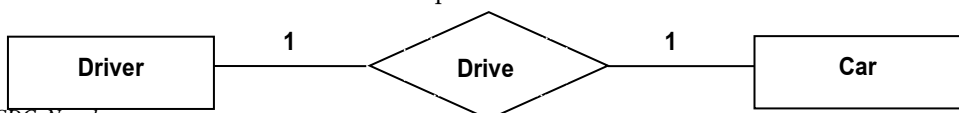
C.   Relationship :

A relationship is a link of association between entities. An example is the link between dentist and patients. It is denoted by diamond or rhombus symbol.
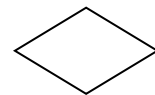
There are following type of relationship :

a.   One to one relationship:
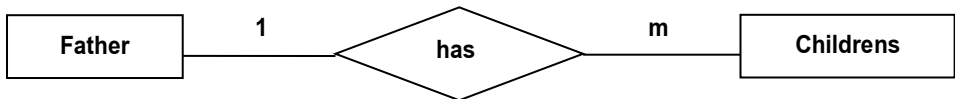
If one instance of an entity is associated with one instance of another entity we say that there exists one to one relationship between the entities

| Driver | —1— | Drive | —1— | Car |

1 driver can drive only one car at a time and a car can only be driven by a driver so here one to one relationship exists.
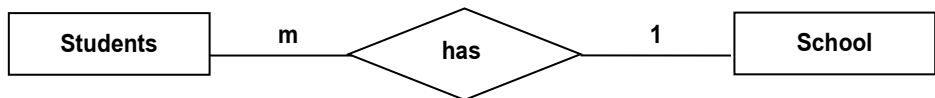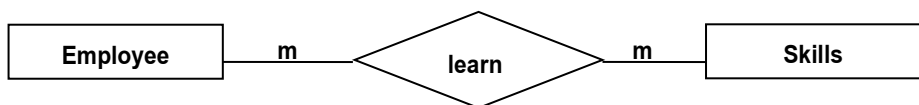
b.   One to many relationship:

If one instance of an entity is associated with many instance of another entity we say that one to many relationship exists between the entities.

Father	1 --- has --- m	Childrens

Here 1 father can have many children but a child can be of only 1 father so one to many relationship exists.

c.   Many to one relationship :

If many instance of an entity is associated with one instance of another entity then we say that many to one relationship exists between the entities.

Students	m --- has --- 1	School

Here, a school has many students but a student can be of only one school. Hence many to one relationship exists in this case.

d.   Many to many relationship :

If many instance of an entity is associated with many instance of another entity then we say that many to many relationship exists between the entities.

Employee	m --- learn --- m	Skills

Many instance of employee can learn a skill and many instance of skill scan be learned by an employee. Hence many to many relationship exists in this case.

Advantages of ER model :

i.   ER model is very well integrated with relational database model.

ii.   Database's main entity and their relationship are easily viewed and understood through ER model.

Disadvantage of ER model :

i.   No data manipulation language or commands are available in ER model.

ii.   This model becomes crowded due to huge presence of entity.

2. Implementation model :

It has 4 types hierarchical database model, network database model, relational database model and object oriented database model. Each of these database model are described as follows:

A.   Hierarchical database model :

It is a database model in which records are grouped in such a way that their relationships form a branching, tree like structure. In very simple language this type of database model organizes data in top down structures that resembles a tree.
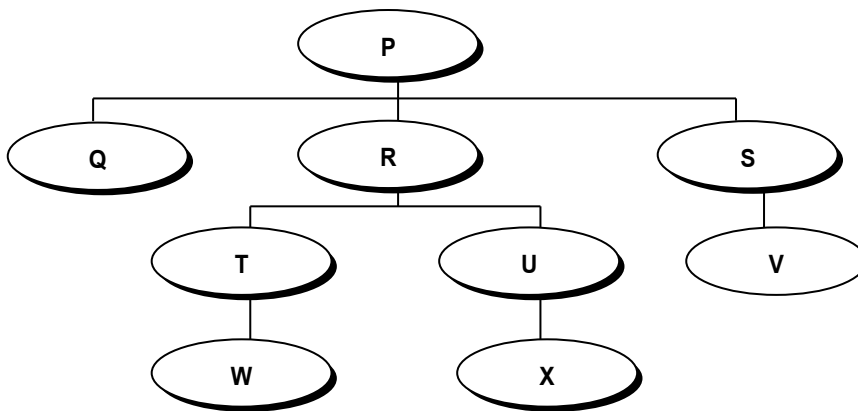
Fig: Hierarchical database model

Top file is called root and bottom files are called leaves. Any file other than root has to be connected to superior file (called as parent). In above diagram P is root, W and X are leaves. R is parent of T and U.

Advantages of hierarchical database model:

i.   It is most simplest and easiest database model.

ii.   If parent is known then searching is easy and fast.

iii.   Database security is said to be good because we can't modify, delete a child without consulting its parent.

iv.   It is efficient to handle 1 to many relationship.

Disadvantages of hierarchical database model

i.   This model nowadays has become an outdated model.

ii.   This model cannot handle many to many relationship.

iii.   When parent mode is deleted, child are automatically deleted.

B.   Network database model :

It is a DBMS model where a single file can be associated to one to more than one number of files. When this association among the files are made, then a network is said to form.
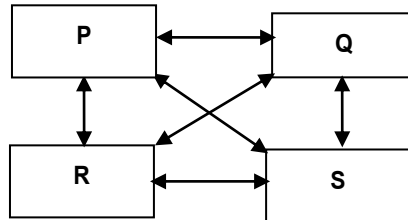
Fig: Network database model

In the above figure P, Q, R, S files are associated with each other forming a network, hence network database model is formed.

Advantages of network model:

i.   It is able to handle many to many relationship.

ii.   Searching is generally fast due to the presence of multidirectional pointer.

iii.   Data redundancy is reduced because same data is not needed again and again.

Disadvantages of network model

i.   It is too complex, because database administrator, programmer must be familiar with internal structure in order to access the database.

ii.   There is need of long programs to handle relationship.

C.   Relational Database model :

It is the type of database model where data re-represented using two dimensional table which are made up of rows and columns. Each row represents a record and each column represents a field. Record is also referred as tuple and field is also referred as attribute.

Employee	Account
ename	Eid
eid	balance
eaddress	
ephone	

Table Account
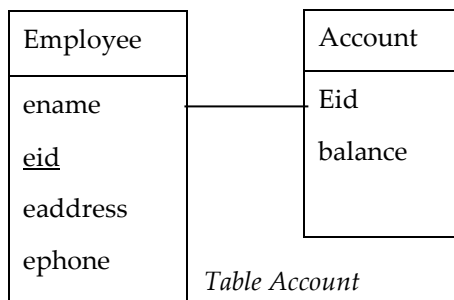
Table Employee

ename	eid	Eaddress	ephone
Rohan	E101	Ktm	4427494
Sohan	E102	Brt	0152854
Mohan	E103	Pkr	0222345

eid	Balance
E101	4091
E102	3994
E103	2973

The above figure shows relational database comprising of 2 tables viz. employee and account. One table is related to other by means of a common field (eid).

Some terms associated with relational database model are as follows:

   i. Field:

   It is a piece of information about an element. It is also called attribute.

   ii. Record:

   One full set of all field is record.

   iii. Table:

   Complete collection of records is table. It consists of rows and columns.

   iv. Domain:

   Area of database record and field.

Advantages of relational database model:

   i.   There are few redundancy in this database model.

   ii.   Normalization of database is possible.

   iii.   The database processing is also very fast.

Disadvantages of relational database system

   i.   It is the most complex database model.

   ii.   Here are too many rules which makes this database model non-user friendly.

D.   Object oriented database model :

It is a flexible database that supports the use of abstract data types, objects and classes and that can store a wide range of data, often including sound, video and graphics, in addition to text and numbers. Some object oriented databases allow data retrieval procedures and rules for processing data to be stored along with the data or in place of the data. This allows the data to be stored in areas other than in the physical database, which is often desirable when the data files are large, such as for video files.

In short object oriented databases stores data and function (methods) as objects which can

then be automatically retrieved and shared.

Advantages of object oriented database model:

    i.   Redundant code can be eliminated by making use of concept called inheritance.

    ii.   Easy upgrading of a data system from small to large system.

    iii.   They can use more and more variety of data than other type of data model and also can access data much faster.

Disadvantage of object oriented database model

    i.   Poor performance in comparison to relational database model.

    ii.   Benefit is only on long run.

    iii.   This database model is very complex and it needs trained manpower to be operated.

1.6 NORMALIZATION:

Normalization is the process of organizing the data in the data base in such a way that:

    i.   no data is unnecessary duplicated.

    ii.   data is consistent throughout the data base

    iii.   structure of each table is flexible enough to allow us to enter as many or as few items as required.

    iv.   the structure should enable a user to make all kinds of complex queries relating data from different tables.

    v.   ambiguity are eliminated.

There are three types of anomalies that occur when the database is not normalized. These are – Insertion, update and deletion anomaly. Normalization helps to bring out good data base system.

We will look at three stages of normalization known as first normal form, second normal form and third normal form.

1.   First Normal Form (1NF):

A table is in first normal form if it contains no repeating attributes or group of attributes.

Let's take an un-normalized table as follows.

Course nos.	Course Name	Lecturer Name	Student Id	Student Name	Student Address	Student Gender
coo1	English	Rohan	13	Rita	Brt	F
			24	Gita	Pkr	F
			17	Kabita	Ktm	F
coo2	Nepali	Sohan	22	Ram	Brt	M
coo3	Computer	Mohan	32	Shyam	Brt	M
			25	Prabha	Ktm	F

This table is **not in 1NF** as the rule says "each attribute of a table must have single values.

To make the table complies with 1NF we should have the data like this:

Course Nos.	Course Name	Lecturer Name
coo1	English	Rohan
coo2	Nepali	Sohan
coo3	Computer	Mohan

Student Id	Student Name	Student Address	Student Gender
13	Rita	Brt	F
24	Gita	Pkr	F
17	Kabita	Ktm	F
22	Ram	Brt	M
32	Shyam	Brt	M
25	Prabha	Ktm	F

Second normal form (2NF)

A table is said to be in 2NF if both the following conditions hold:

- Table is in 1NF (First normal form)
- No non-prime attribute is dependent on the proper subset of any candidate key of table.

An attribute that is not part of any candidate key is known as non-prime attribute.

Following tables are in 2 NF of above 1 NF:

Student Id	Student Name	Student Address	Student Gender
13	Rita	Brt	F
24	Gita	Pkr	F
17	Kabita	Ktm	F
22	Ram	Brt	M
32	Syham	Brt	M
25	Prabha	Ktm	F

Course Nos.	Course Name	Lecturer Name
coo1	English	Rohan
coo2	Nepali	Sohan
coo3	Computer	Mohan

Third Normal form (3NF)

A table design is said to be in 3NF if both the following conditions hold:

- Table must be in 2NF
- Transitive functional dependency of non-prime attribute on any super key should be removed.

An attribute that is not part of any candidate key is known as non-prime attribute.

In other words 3NF can be explained like this: A table is in 3NF if it is in 2NF and for each functional dependency X—>Y at least one of the following conditions hold:

- X is a super key of table.
- Y is a prime attribute of table.

An attribute that is a part of one of the candidate keys is known as prime attribute.

Following tables are in 3 NF of above 2 NF:

Student name	Student Address	Student Gender
Rita	Brt	F
Gita	Pkr	F
Kabita	Ktm	F
Ram	Brt	M
Syham	Brt	M
Pabha	Ktm	F

Student Id	Student Name
13	Rita
24	Gita
17	Kabita
22	Ram
32	Syham
25	Prabha

1.7 DATABASE SYSTEM :

Database system generally refers to the way the components are organized which defines and regulate the collection, storage, management and use of data with in a data base. The type of computer system that data base can run on are of 4 categories viz. centralized, personal, client/server and distributed. Each of these is discussed as follows:

i. Centralized data base system:

In this type of database system all program run on a main host computer. Also all DBMS including the application that access the data base and the communications facilities that send and receive data from user terminal are also in that main host computer. Main host computer can be either minicomputer or mainframe computer. User can access database through terminal.

Terminals are of two type as follows:

    a.   **Local terminal:** They are simply work stations.

    b.   **Remote terminal:** Here user accessing data base are not locally connected.
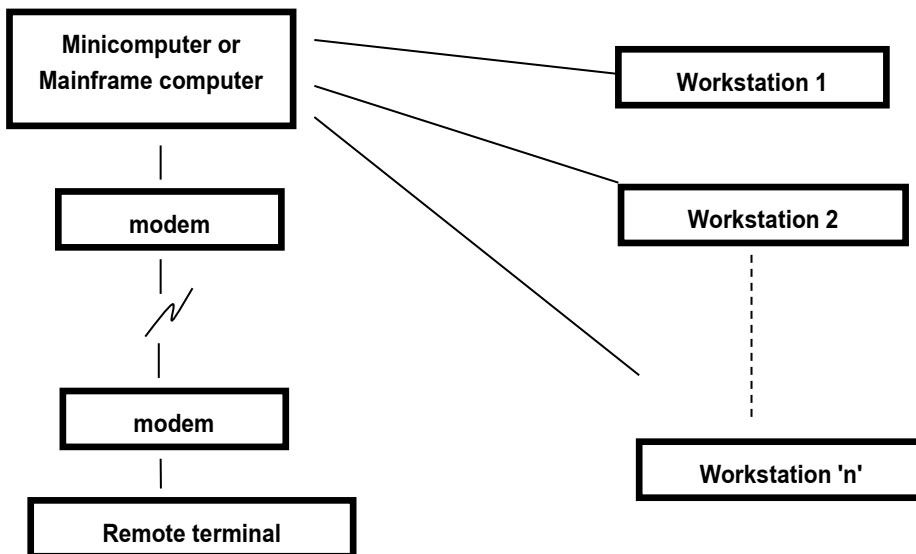
Figure. Block diagram of centralized data base system

The terminals have little or no processing power of their own and they consists of only screen, keyboard and necessary hardware to communicate with host.

Advantages of centralized database system

- Since all data is stored at a single location only thus it is easier to access and co-ordinate data.
- The centralized database has very minimal data redundancy since all data is stored at a single place.
- It is cheaper in comparison to all other databases available.

Disadvantages centralized database system

- The data traffic in case of centralized database is more.
- If any kind of system failure occurs at centralized system then entire data will be destroyed.

ii.   Personal Computer Data base System:

When the DBMS runs on a PC then the PC acts as both host computer and terminal computer. This condition is personal computer data base system.

iii.   Client Server Database System:

In case of client server DBMS, DBMS server and DBMS client software runs on individual clients. The server software processes the request for data searches, sorts and reports query that originate from individual client running DBMS client software.
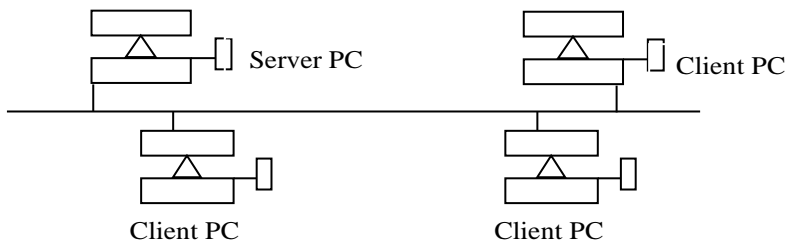
Figure. Client Server Database System

Advantages of Client-Server model

- Centralized system with all data in a single place.
- Cost efficient requires less maintenance cost and Data recovery is possible.
- The capacity of the Client and Servers can be changed separatel

Disadvantages of Client-Server model
- Clients are prone to viruses, Trojans and worms if present in the Server or uploaded into the Server.

- Server are prone to Denial of Service (DOS) attacks.

- Data packets may be spoofed or modified during transmission.

- Phishing or capturing login credentials or other useful information of the user are common and MITM (Man in the Middle) attacks are common.

iv. Distributed data base system:

In case of distributed data base system several numbers of computers connected to a network store a part of data which can be accessed by any computer on the network.
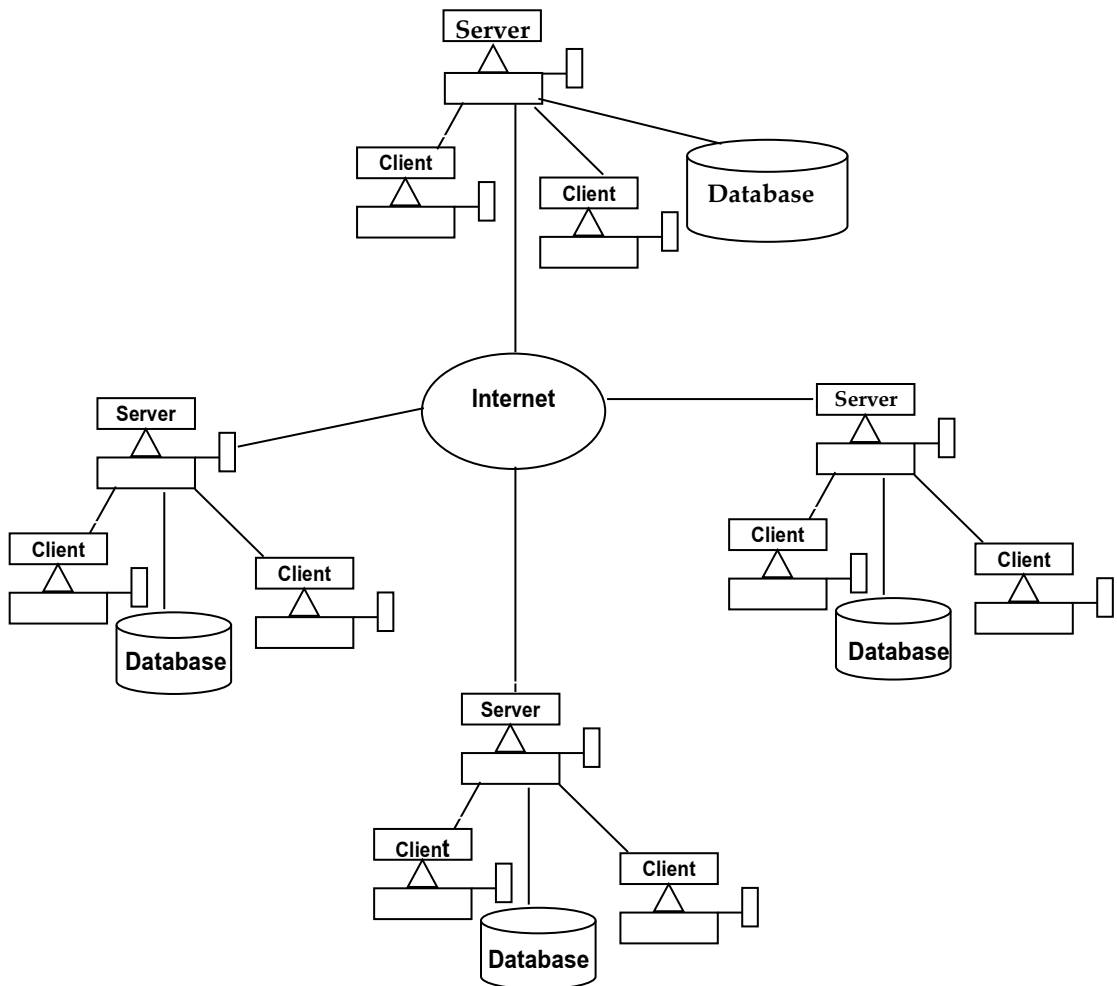
Location independence is striking feature of this kind of database system. If we make a collection of number of different centralized and client server data base system of different location we say a distributed data base system is formed.

Advantages of distributed database system

- Management of data with different level of transparency – Network transparency, replication transparencies and fragmentation transparency database should be distribution transparent in the sense of hiding the details of where each file is physically stored within the system.

- Increased Reliability and availability.

- Easier Expansion.

- Improved Performance.

- The distributed database can easily be accessed from different networks.

Disadvantages of distributed database system

- This database is very costly and it is difficult to maintain because of its complexity.

- In this database, it is difficult to provide a uniform view to user since it is spread across different physical locations.

1.8 Prerequisite of good database system:

Every database system is not necessarily a good database system. For a database system to be a good database system following are the prerequisites

i.   Data integrity:

It is the accuracy of data and its conformity to its expected value. In simple word data integrity means data is accurate, consistent and updated. Data integrity becomes important in cases when data is kept unused for a long period of time. Following are its type.

a.   Entity integrity:

Entity integrity guarantees that each row can be uniquely identified by primary key. Entity integrity prevents primary key from taking (accepting) NULL values. This will help us to distinguish one record from another.

b.   Referential Integrity:

Referential integrity ensures that for every value of foreign key there is a matching value of a primary key. This type of data integrity keeps us away from making data entry mistake.

    c.   Domain integrity:

        Domain integrity ensures that there are only valid entries in the domain.

ii.   Data independence:

This means that data in a data base are separated from the program that manipulates it. The term data independence also refer to that storage and retrieval of data is not affected by other application in the computer.

iii.   Avoid data redundancy:

Data redundancy has to be avoided as far as possible.

iv.   Database security:

This term relates to protection of database. Maintaining data security means keeping database safe from various hazards to which it may be subjected.

These hazards include following:

    a.   Natural hazards such as fire, floods, hurricane or earthquakes.
    B.   Deliberate corruption or destruction of data by malicious or terrorist acts.
    C.   Illegal access to data by " hackers".
    D.   Accidental destruction of data by hardware failure or program or operator error.

Following measures can be taken to ensure the database security.

    i.   Dates can be protected physically. There has to be a proper mechanism to restrict the access of unauthorized person to database .

    ii.   There can be use of password in data to prevent unauthorized use of computer terminals or unauthorized access to online file.

    iii.   Using backup copies of risks or files.

    iv.   Constant checks of security against computer virus since through internet the computer virus may easily attack the database.

    v.   Fully reliable OS has to be used since weakness in OS may cause the accidental loss in data.

v. Data maintenance:

Good database management involves procedures for updating records, adding records and deleting the records.

1.9 Concept of RDBMS (Relational Database Management System)

There are several types of DBMS available. The most common and popular type of DBMS is RDBMS. It is widely used in all system from microcomputer to mainframe computer. In RDBMS, data are held in table and the tables are linked by means of some common field. Conceptually, one row of a table holds one record. Each column in the table holds one field or attribute. The common field which relates table is called foreign key.
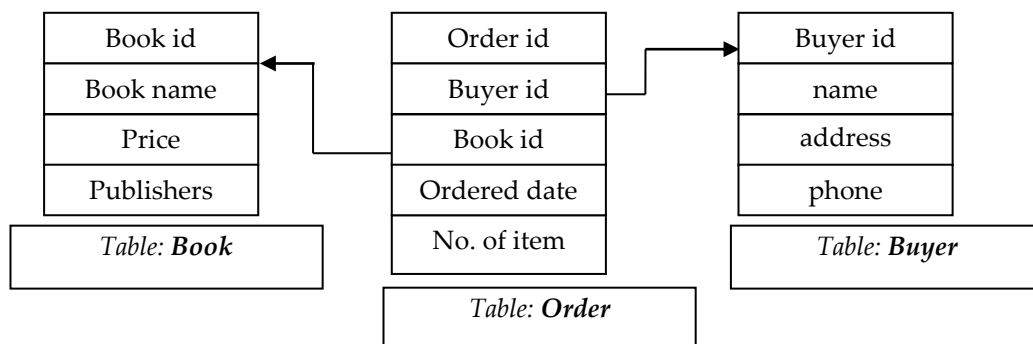
Book id		Order id		Buyer id
Book name		Buyer id		name
Price		Book id		address
Publishers		Ordered date		phone
*Table: **Book***		No. of item		*Table: **Buyer***
		*Table: **Order***		

Figure. Block diagram of distributed data base system.

As seen in above figure there are three tables viz. book, order and buyer. Order is the main table. When a buyer orders for some book; it is entered in order table as the order id, buyer id, book id, and so on. In the order table, there is just buyer id or book id. To get detail information of book or buyer we have to go to the book table or buyer table with particular id. This scheme thus reduces the size of database.

1.10 SQL (Structured Query Language)

SQL is a popular query language. It is a database sublanguage which is used in querying, updating and managing relational databases the de fecto standard for data base products. It is a non procedural language. It was developed by IBM in 1970. It is declarative in nature. It was first standardized in 1986.

Although it is possible to extract a great deal of information from a data base such as Access using Query by Example. There are occasions when complex queries cannot be formulated using this technique. This is when we need SQL. SQL is not a programming language but a data access language.

SQL commands can be categorized as follows:

    i.   DDL such as create, alter, drop etc.

    ii.   DML such as insert, update, delete etc.

There are about 60 commands in SQL.

1.11 DBA (Data Base Administrator)

Data base administrator is the person who manages the data base. The DBA determine the content, internal structure and access strategy for a data base, defines security and integrity and monitors performance. A DBA has to possess the following

    i.   Knowledge of query language(SQL).

    ii.   Knowledge of various operating system on which data base server can run and is running.

iii. Knowledge in designing the data base.

iv. Knowledge about network architecture.

1.12 Responsibility of DBA (Role)

There are following responsibility of a database administrator

i. DBA should give idea to an organization on deciding which department will be looking on maintenance and update of data in the database.

ii. DBA has to assure 24 hour access to each department in the organization that needs the data.

iii. DBA has to install and timely upgrade the data base server.

iv. DBA has to use storage space available for data in an effective manner.

v. DBA has to make proper backup and also to develop recovery procedure in case the DBMS crashes/malfunctions.

vi. DBA also has to make sure that database server is giving optimum performance.

vii. DBA has to work with the developers and need to assist in designing the overall data base.

viii. DBA has to manage all the user who use data base and determine proper security level for each user.

1.13 Types of database software/language:

First successful data base product for micro computer was DBASE for MS DOS/PCDOS. The first commercial data base software is Oracle, DBZ. In early days DBASEIII, DBASEIII+, FOXPRO etc. were used. Now a days MSACCSS, SQL, Oracle8i, Oracle9i etc. are the data base software which are run.

1.14 Data base Design:

Data base design is just a structural design will be used to store and manage data. If a data base is well designed; it will facilitate data management and will become a valuable information generator. Following four guiding principles are used as general data base design methodology.

i. Representing the user view as collection of tables:

After some sort of requirement is viewed by user; then there is a need of developing collection of table that will support it. We perform following steps in making collection of tables.

a. The entities involved are determined and a separate table is created for each

type of entity.

b.   Determine primary key for each of the table created.

c.   Determine properties for each of the entities. These properties along with primary key become column in appropriate table.

ii.   Normalize each table:

Normalization avoids repeating groups of data and data redundancy. It also helps to make sure that the storage area is used efficiently and the data are retrieved in an easy way. Generally we normalize the table up-to third normal form. First normal form removes the repeating group of attributes or the repeating attributes itself. Second normal form removes the partial dependency that may be in the data in the database. Third normal form removes any transitive dependency that may exist in the data in the database.

iii. Identify the necessary keys:

All the necessary keys needs to be properly identified. Generally we need to identify the primary key, foreign key, alternate key and secondary key. Primary key is the attribute which can uniquely identify each single instance of the entity. Alternate key is the primary key which could have been chosen as primary key but was not chosen. Secondary key is the attribute that are of interest strictly for the purpose of retrieval of a group of data. Foreign key is an attribute for a table which appears as the primary key in some other table.

iv.   Merge the results of previous steps in to the design:

The last step is merging the above user views into overall design. Top down approach and bottom up approach of data base design are also very popular. These two approaches are dealt as follow:

1.   Top down data base design:

It is the one which starts with general data base design that models the overall enterprise and repeatedly refines the model until a design is achieved that supports all necessary applications. This data base design approach gives a more global feel to the project; at least we have some idea about where we are heading during the design.

2.   Bottom up data base design:

Here we begin with set of attributes. These attributes are grouped into entities and relationships among entities are formed. We go on to attempt to find

higher level entities by grouping them and locating relationship at this higher level. This process goes on till a required design is obtained.

1.15 Introduction to Microsoft Access

Microsoft Access is a powerful database application software, yet easy to learn, relational database application for Microsoft Windows.

Objectives

By using this, you will learn to perform the following operations in Access.

> ❖ Launch Access and identify the parts of the screen.
>
> ❖ Define fields and field properties constructing table structures.
>
> ❖ Enter and edit records in a table.
>
> ❖ Find, sort data.
>
> ❖ Design custom queries to display data.
>
> ❖ Import and export data between Excel and Access.

1.16 Access vs. Excel

Access databases and Excel spreadsheets although originally very different, have evolved to have similar functions. Although either type of program can do many, but not all, of the things that the other one does, there are some things which are easier in one program than the other.

Excel is not database management systems. Excel spreadsheets store data in rows and columns called worksheets. The areas within the worksheet where the rows and columns intersect are called cells. The most common use of spreadsheets is to manage basic information such as telephone numbers, employee names, financial data, calculations, etc.

Access stores data in tables that look similar to worksheets but function quite differently. Tables, the foundation of all relational databases function both independently and interdependently with other tables to allow the database user to combine data from multiple sources and analyze and/or report information in almost limitless ways.

1.17 The Access Interface.

This section introduces the Access interface and discusses how to use menus and tools bars, create databases and switch between database components.

1.18 Starting Access and Opening a Database.

1. Click on the START button on your computer and position your cursor over the PROGRAMS menu to view a list of installed programs. Once you see the list, position the cursor over Microsoft Office and then click on Microsoft Access.

2.   After opening Access, you will be presented with the window shown in Figure 1.
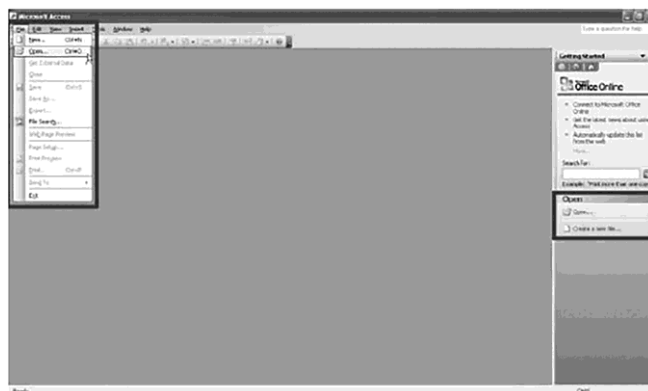
Figure - 1

3.   If you want to edit an existing database, you can either go to **File—Open**, or move your cursor to the lower right-hand side to click on Open.

4. If you are creating a new database, you can also either go to File—New, or move your cursor to the lower right-hand side to click on Create a new file. You will be presented with five options on the right side menu as shown in Figure 2.
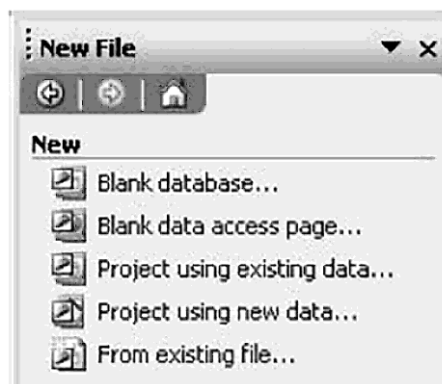
5. Unlike Word documents, Excel spreadsheets, and Power Point presentations, you must save an Access database before you start working on it. After selecting "Blank database", you will first be prompted to specify a location and name for the database. In the New Database dialog box, type in a name for the database and a location and click on **Create** as shown in Figure 3.

Figure - 2

Figure - 3

6. Now a new database is created as shown in Figure 4.
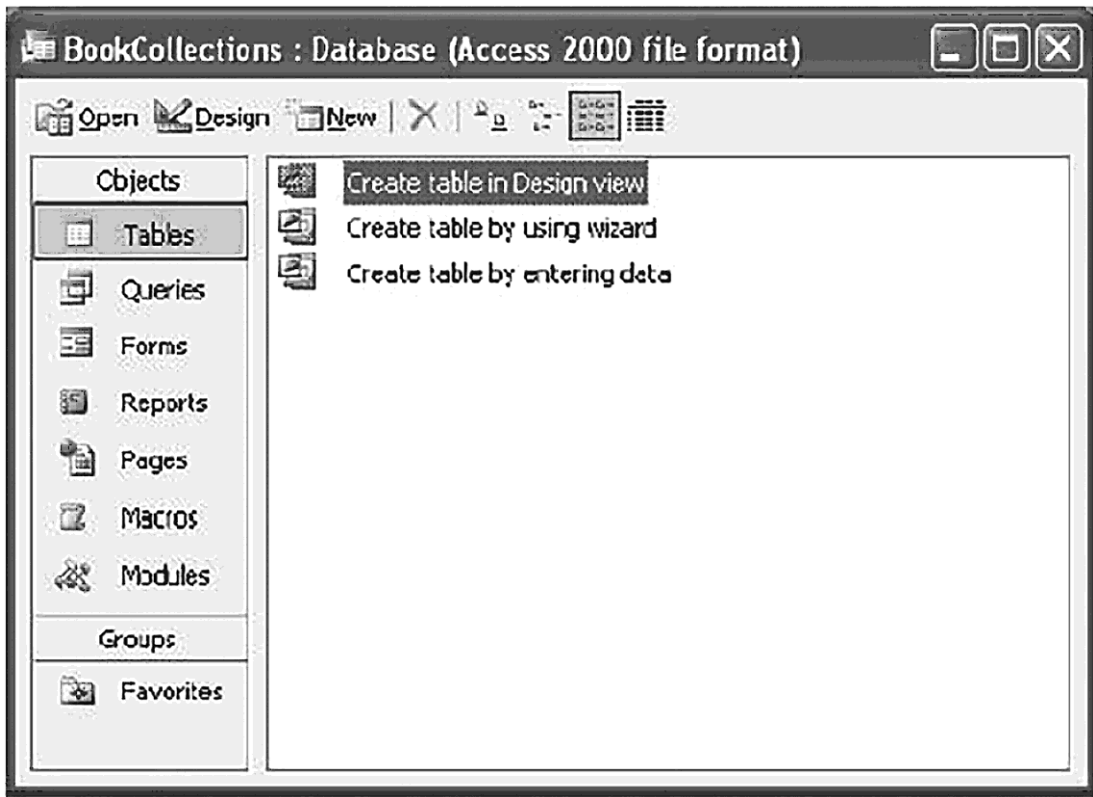
Figure - 4

1.19 Building Database Tables.

This section describes how to construct data tables, enter and edit data, sort records, and find records that meet certain conditions.

1.20 Creating Tables in Design View

1. By clicking the Table tab on the left hand side, you will find Access provides three ways to create a table for which there are icons in the Database window.

- Create Table in Design view will allow you to create the fields of the table. It allows you to define the fields in the table before adding any data to the datasheet. This is the most common way of creating a table and is explained in detail below.

- Create table by using wizard will step you through the creation of a table.

- Create table by entering data will give you a blank datasheet with unlabelled columns that looks much like an Excel worksheet.

2. Double clicking on Create table in Design view displays the Table Design screen where you define fields for your table. The screen is divided into two parts: a top pane for entering the field name, data type, and an option description of the field and a bottom pane for specifying field properties.
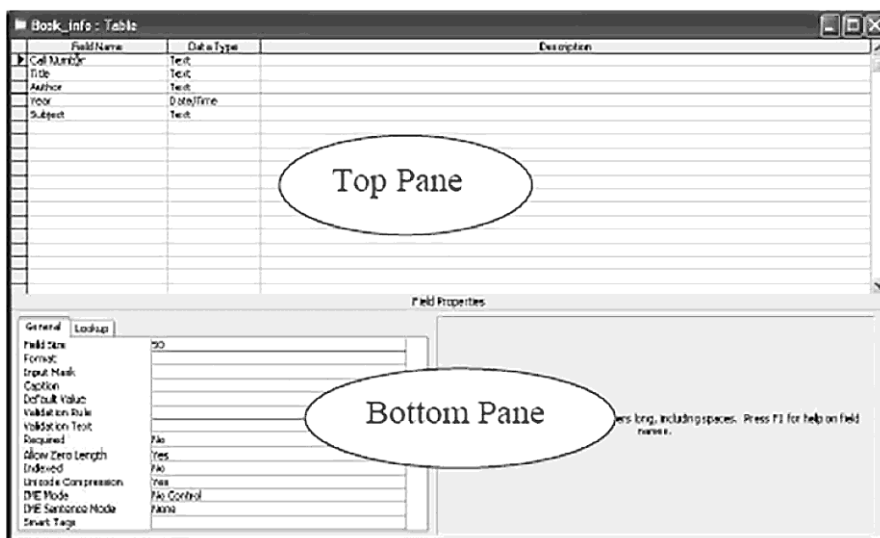
Figure - 5

3. Every table consists of fields. For each field, specify the name of the field, the type of data, and any description needed to determine what data the field contains. Pressing the **Tab** key moves the cursor from one column to the next in the Table Design screen. You can select the data type from the drop-down list in the Data Type column as shown in Figure 6.
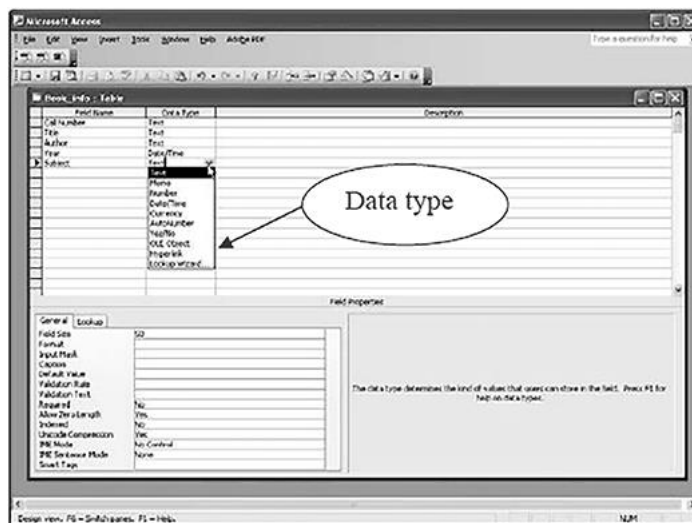
Figure - 6

4.   Each table in your database should have a "primary key." A primary key is a field that uniquely identifies each record in the database. In a database of book collection, there might be two books with the same author so author is not a good primary key. However, every book has a unique "call number", so "call number" would be a good choice for a primary key. To set the primary key for your table, highlight the key field and choose **Primary Key** from the Edit menu.

Figure - 7

5.   When the primary key is set, you should find a little key icon next to the field name on the left side. Note: To remove a primary key, simply repeat this procedure to toggle the primary key off.

Figure- 8

6.   As a final step, the table must be saved. Pull down the File Menu to choose **Save**. A dialogue box will pop up where the name of the new table should be specified. Access gives a default name such as "Table1" or "Table2." Simply type over this default name with the name of the table. For this example, name the table "Book_Info", then click **OK**.
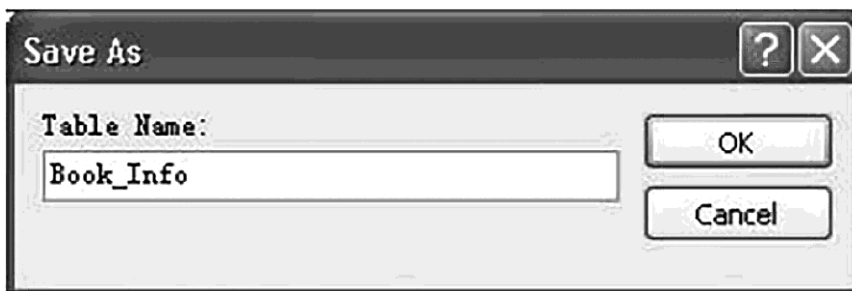
Figure -9

7.   At this point, the new table has been created and saved. Switch back to the Access main screen by pulling down the File menu and choosing **Close**. This will *close* the Design View for the table and display the Access main screen. Notice that the new Book_Info table appears below the Table tab.
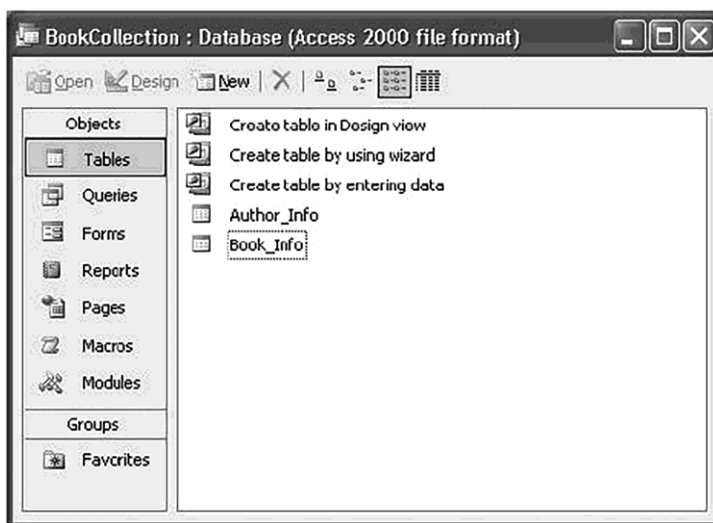
Figure - 10

1.21 Adding and Deleting Fields to Tables

Even the best planned database will sometimes require you to add or delete a field in the Design View. To add a new field, select the row you want to insert above and choose "Row" from the Insert menu. If you need to delete a field, select the row containing the field and choose "Delete Row" from the Edit menu.

You can also move fields in the table design. Select the field you want to move by clicking on its row. A black triangle will appear to indicate the field is selected. Then, drag the field to the new location.

1.22 Cautions to Modifying Table Design Structure

- Backup your table before you make any modifications to it.
- Deleting a field will delete the records in the field.

- If you delete a field and that field is used in any of your forms, queries or reports, you must also delete that field from these other documents.

- Renaming a field will have the same effect on a table as deleting the field.

- Resizing a field (making it smaller) may truncate the data in your forms or reports.

1.23 Entering and Editing Data

1. After you have defined fields in the table, you can enter data. Highlight the table, choose Open from the database window menu bar. This displays a grid with field names across the top and a row for each record of the database as shown in Figure 11.
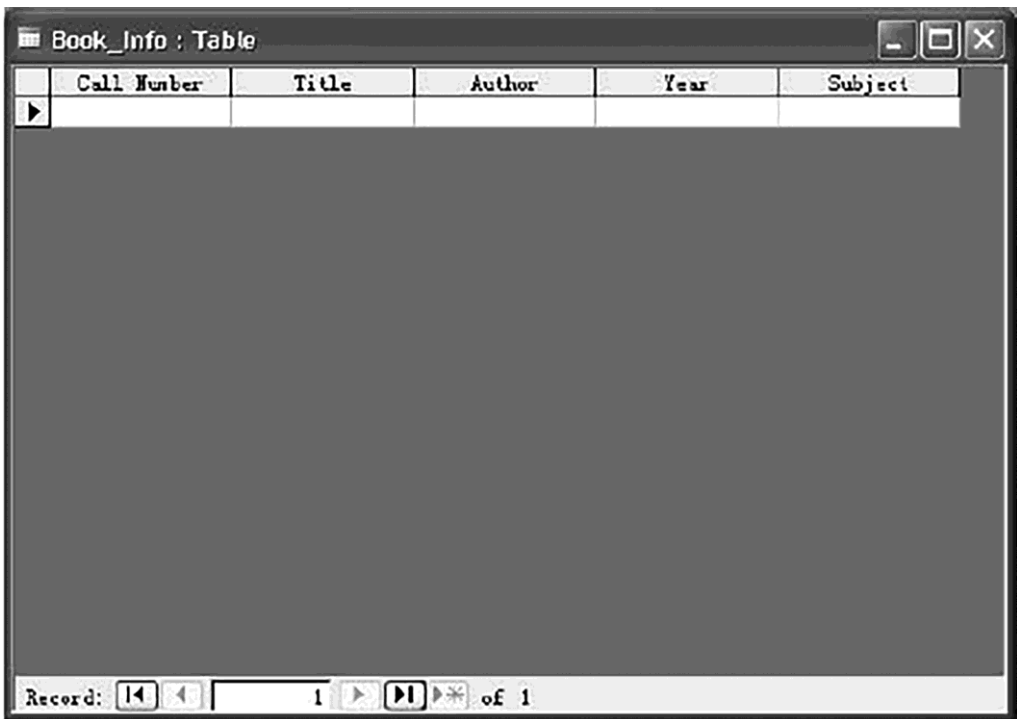
Figure 11

2. To enter data, simply click the mouse in the field you want to enter and type a value. Press the **Tab** key to move from field to field. When you are in the last field of a record, pressing the Tab key automatically creates a new record for you.

3. The table in Figure 12 is shown in Datasheet View. Use the **Datasheet View** to add, delete, and move records.

Figure 12

4. To save the new data, pull down the **File** menu and choose **Save**.

5. To navigate to other records in the table, use the navigation bar at the bottom of the Screen:

Sorting and Finding Records: Two of the most common database tasks are sorting records and searching for records that meet a particular condition. Fortunately, both of these operations are easy to perform while viewing the table datasheet.
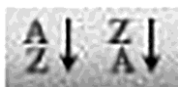
Sorting Records

To sort records by a particular field, select the field you want to sort. From the **Records** menu select **Quick Sort** then choose either **Ascending** or **Descending** order. Access quickly sorts the records by the selected field.

Finding Records

If you want to locate a record that contains a certain name or date use the **Find** command on the **Edit** menu. Type the search string in the **Find What** box and click **Find First**. Access will highlight the first record that contains the search string.

Table Relationships

To prevent the duplication of information in a database by repeating fields in more than one table, table relationships can be established to link fields of tables together. Follow the steps below to set up a relational database:

1. Click the **Relationships** button on the toolbar:

2. From the **Show Table** window (click the **Show Table** button on the toolbar to make it appear), double click on the names of the tables you would like to include in the relationships. When you have finished adding tables, click **Close**.
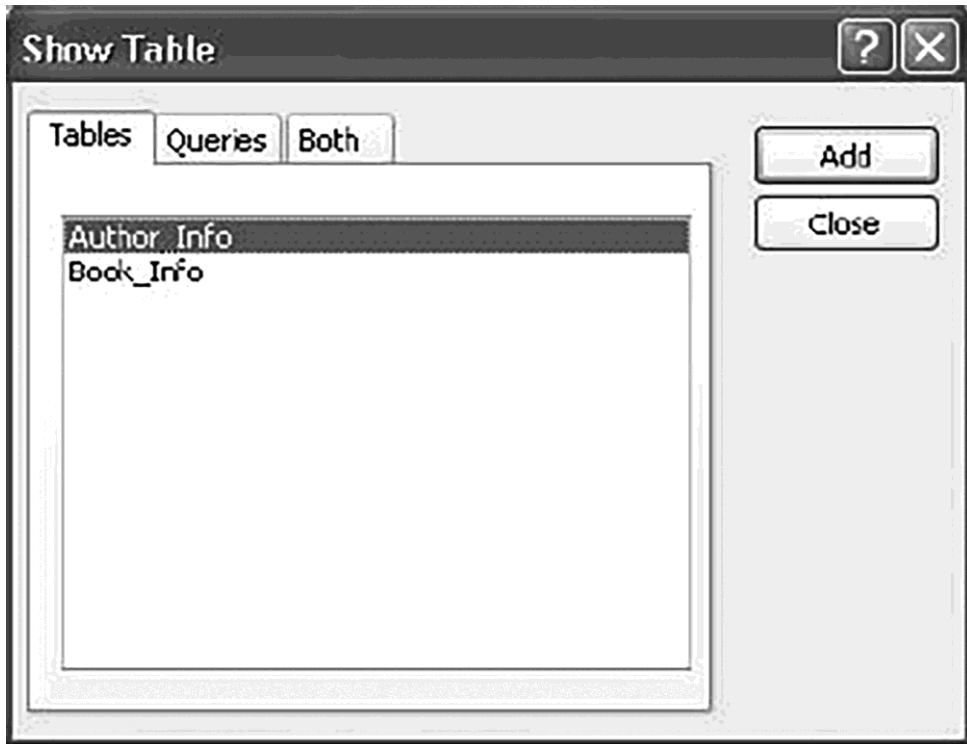
Figure 13

3. To link fields in two different tables, click and drag a field from one table to the corresponding field on the other table and release the mouse button. The **Edit Relationships** window will appear as shown in Figure 14. From this window, select different fields if necessary and select an option from **Enforce Referential Integrity** if necessary. These options give Access permission to automatically make changes to referential tables if key records in one of the tables
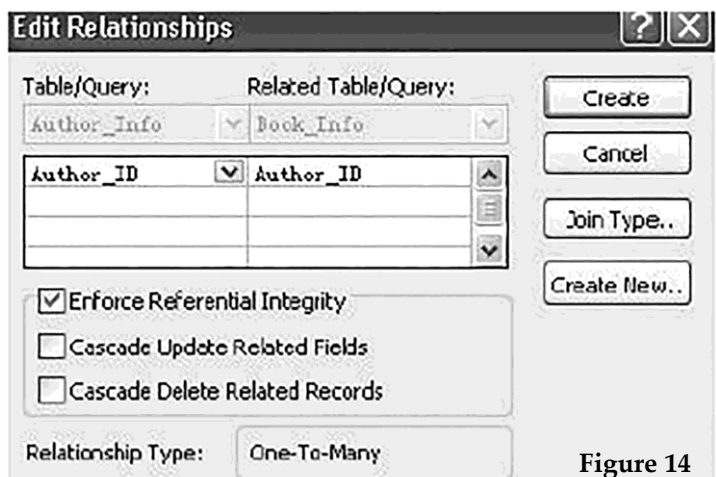
Figure 14

are deleted. Check the **Enforce Referential Integrity** box to ensure that the relationships are valid and that the data is not accidentally deleted when data is added, edited, or deleted. Click **Create** to create the link.

4. A line now connects the two fields in the Relationships window as shown in Figure 15. Note the symbols "1" (indicating the "One" side) and the infinity symbol (indicating the "Many" side) on the relationship. Close the relationships screen and select **Yes** to save the changes to the Relationships layout.
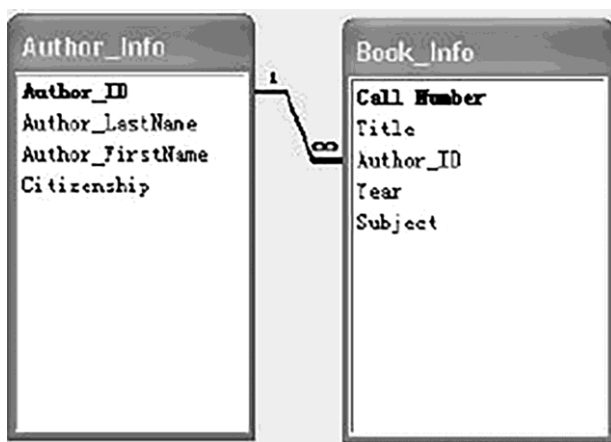
Figure 15

1.24 Creating Queries.

A query is simply a question you ask a database. How many books are written by a certain author? Who wrote a book in a certain year? Queries select records from one or more tables in a database that match the criteria you set, they can be viewed, analyzed, and sorted on a common datasheet. The resulting collection of records, called a **dynaset** (short for dynamic subset), is saved as a database object and can therefore be easily used in the future. The query will be updated whenever the original tables are updated.

1. To construct a query, Click on the **New** button in the database window as shown in Figure 16. Choose **Design View**, click **OK**.

2. In the **Show Table** dialog box as shown in Figure 17, you will be asked to choose a table/tables for the query. Select the tables you want to query and click **Add**.
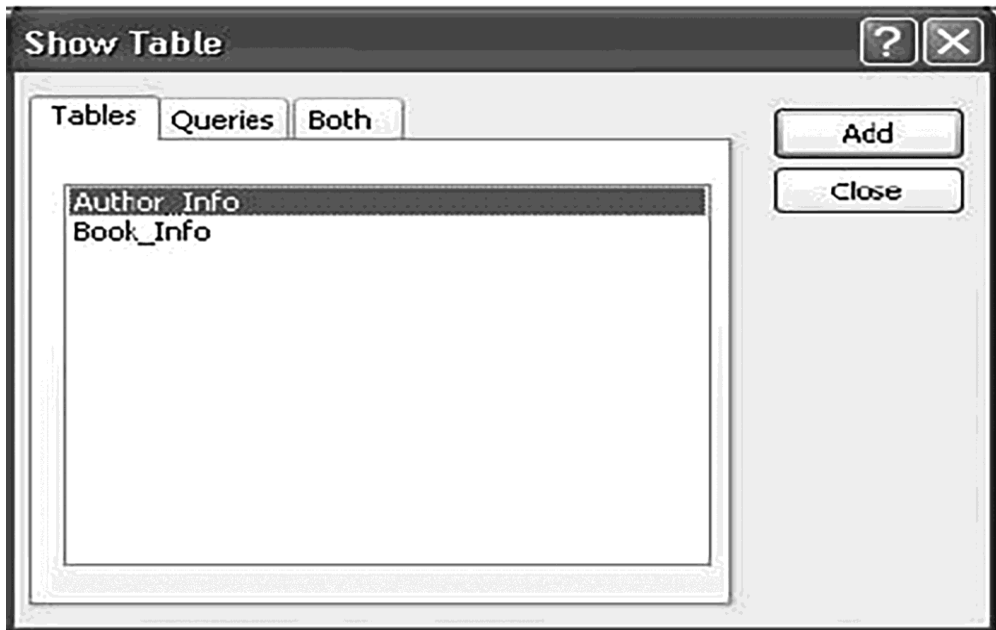
Figure 16
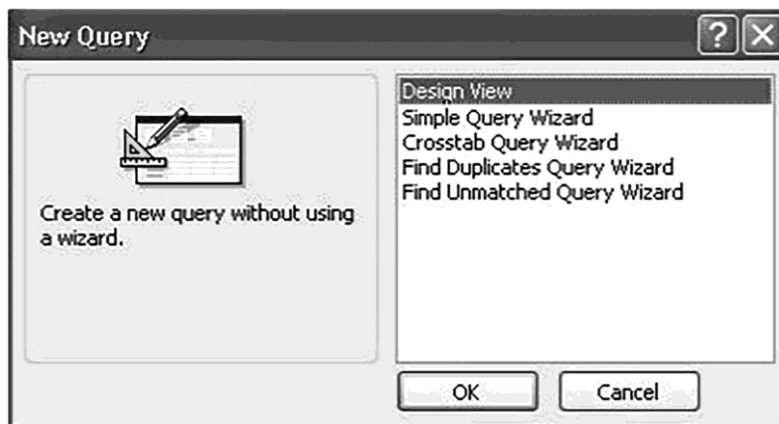
Figure 17

3. Add fields from the tables to the new query by double-clicking the field name in the table boxes or selecting the field from the **Field** and **Table** drop-down menus on the query form as shown in Figure 18. Specify sort orders if necessary.
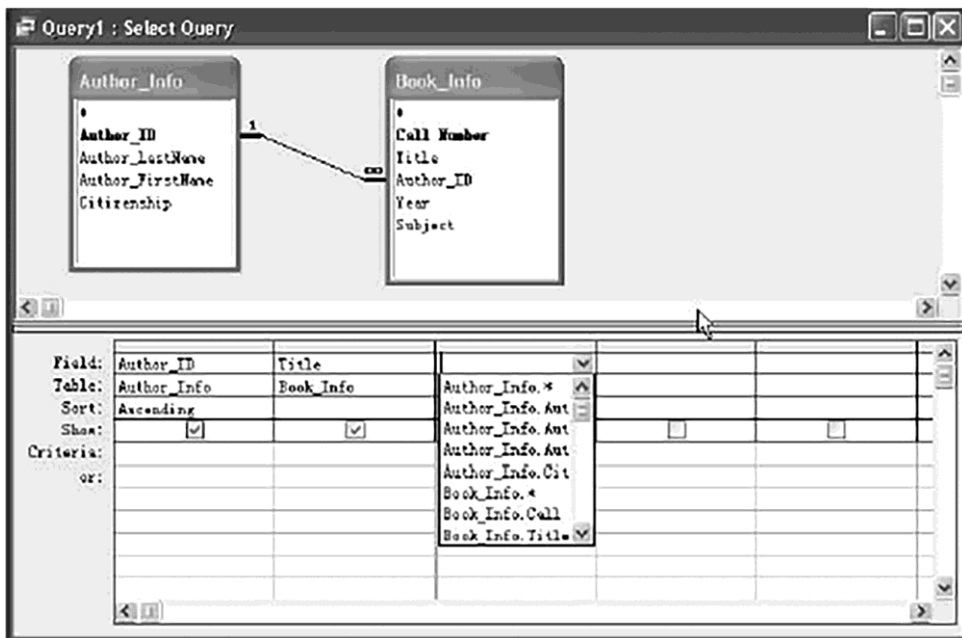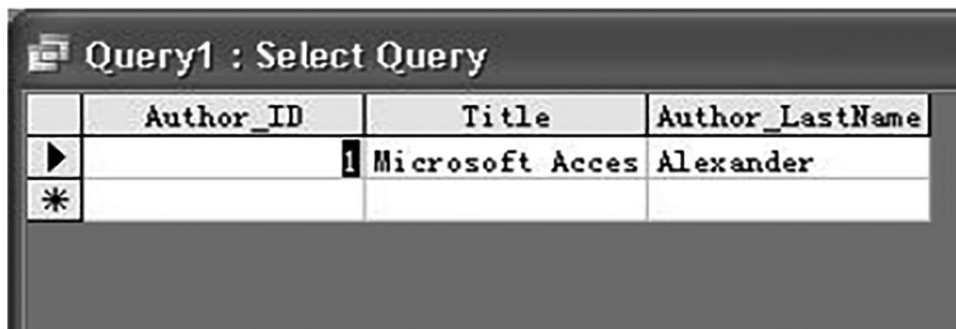

Figure 18

4. Enter the criteria for the query in the **Criteria** field. The **Expression Builder** can also be used to assist in writing the expressions in the **Criteria** field.

5. After you have selected all of the fields and tables, click the **Run** button on the toolbar to execute the query. Figure 19 shows the query result, records that match the criteria you set.

Figure 19

6. Choose **Save** from the **File** menu to save a query for later execution.

1.25 Importing Excel data to Access.

1. To format data in an Excel spreadsheet, you need to make sure the data is list format. This means that each column has a label in the first row and contains similar facts, and there are no blank rows or columns within the list. Then, close the Excel workbook that contains the data you want to use in Access.

2. In Access, open the database (if you already have one) or create a new database where you want to copy the Excel data.

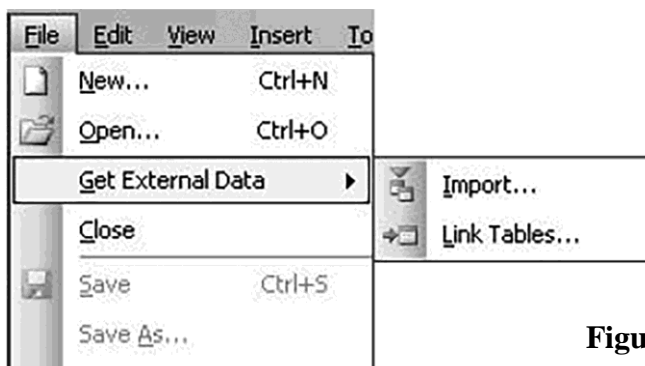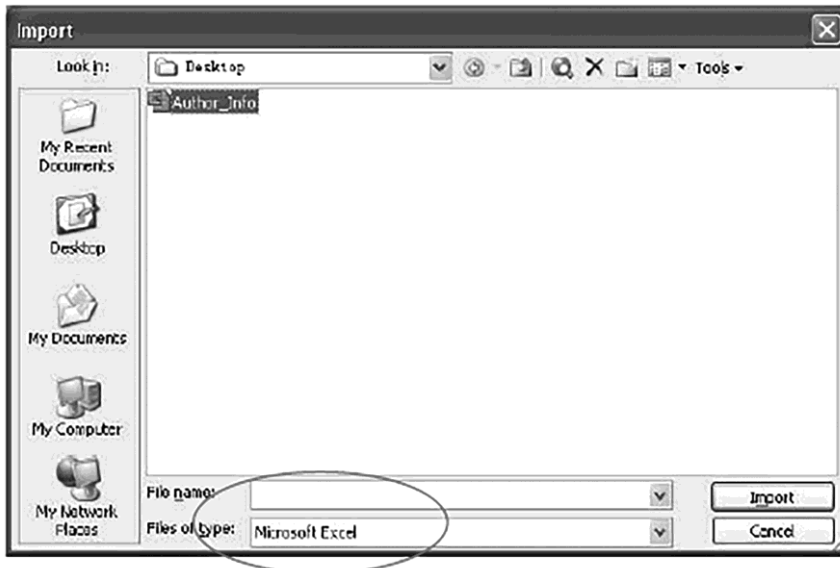3. On the Access **File** menu, point to **Get External Data**, and then click **Import**.


Figure 20

4.   In the **Import** dialog box, click **Microsoft Excel** in the **Files of type** box. Then locate the file you want to import, and then double-click the file.

Figure 21

5.   After you select the spreadsheet to import, you'll be presented with a new screen asking you to select the Worksheet or Named Range that you wish to import. In the example here, there is only one worksheet. Select the worksheet and click **Next**.

6.   Specify whether column headings are present**.** Often, Excel users utilize the first row of their spread sheet to provide column names for their data. In our example file, we did this to identify Author_ID, etc. In the window shown below, ensure that the **First Row Contains Column Headings** box is checked. This will tell Access to treat the first row as names, rather than actual data to be stored in the list. Click the **Next** button to continue.

Figure 22
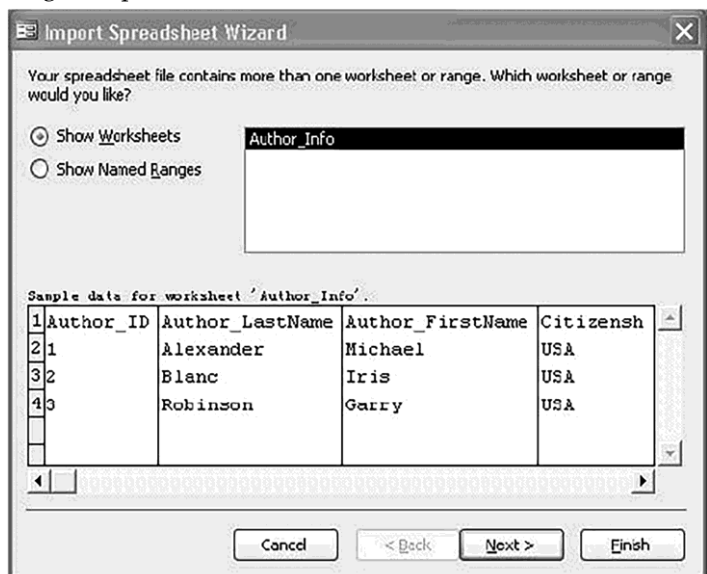
7. Choose a destination type. If you're working with a brand-new database, you'll want to check the **In a New Table** radio box. If you are importing data into an existing database, you could specify the destination table here. Click the **Next** button to continue.
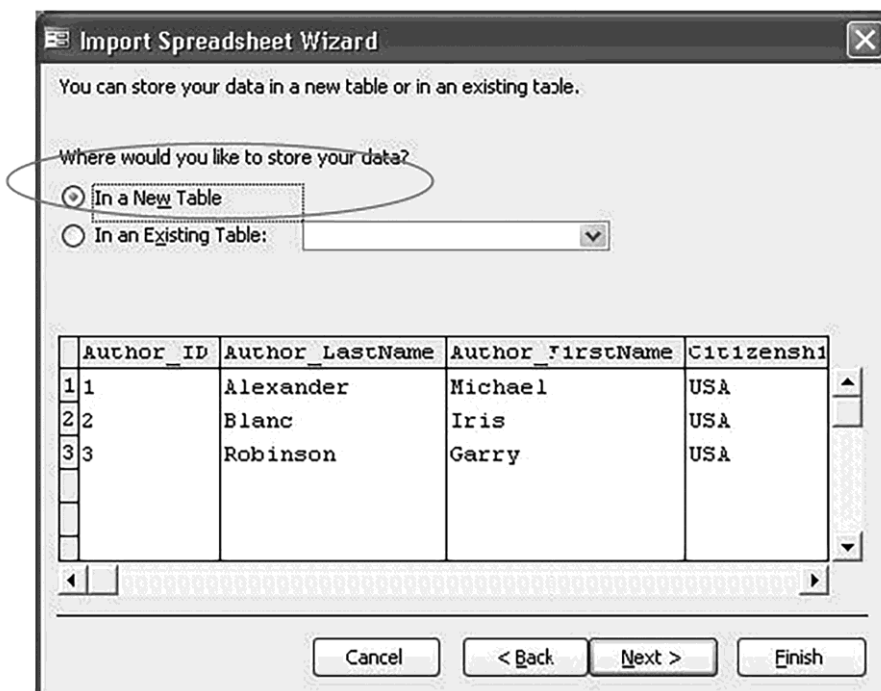

Figure 23

8. Create any desired indices. Database indices are an internal mechanism that can be used to increase the speed at which Access can find information in your database. You can apply an index to one or more of your database columns at this step. Simply click the **Indexed** pull-down menu and select the appropriate option.

Figure 24

9. Select a primary key. You can either let Access generate a primary key for you, or you can define one of the fields in your worksheet as primary key.
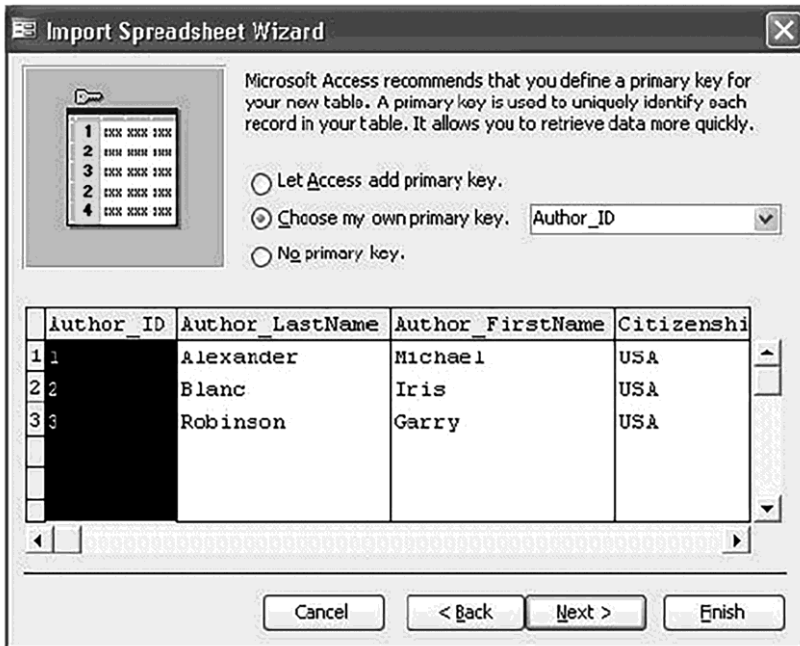

Figure 25

10. The last step is to name your table and click **Finish**. A dialogue box will appear to inform you that your spreadsheet has been imported successfully.
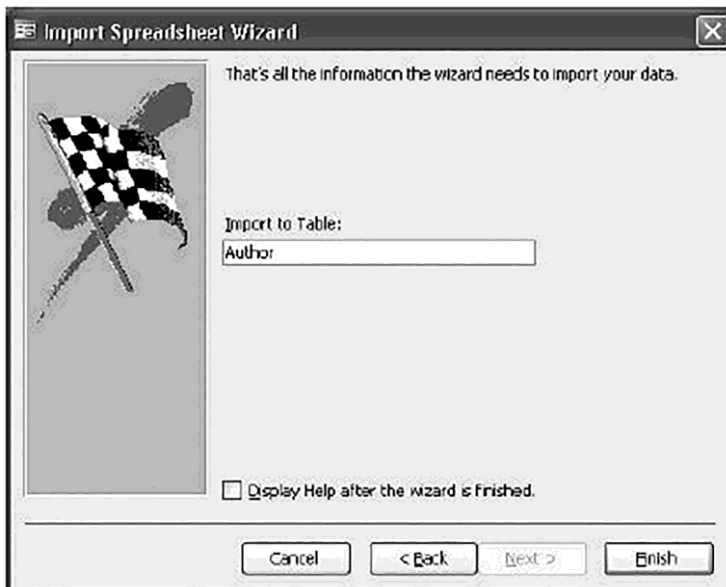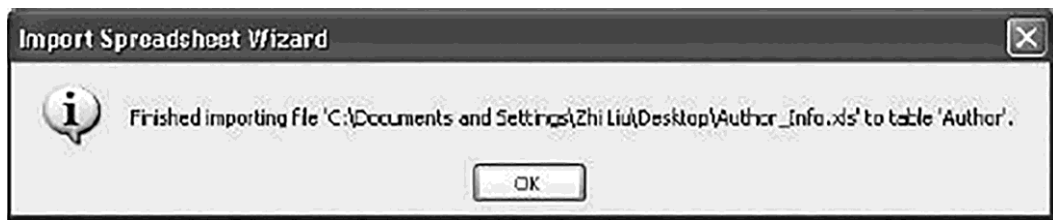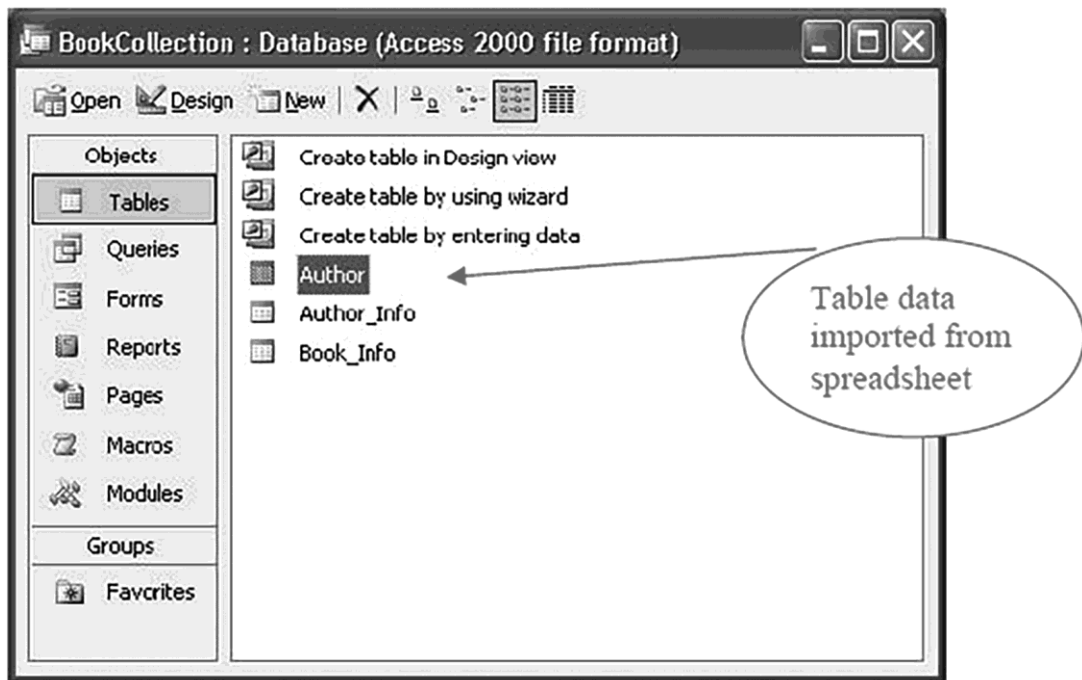

Figure 26

Figure 27

11. Return to the database window, you will find the new table a part of your database now.

Figure 28

Summary

- Data is raw facts or figure or entity.
- Information is processed data.
- Database is collection of related data.
- DBMS is a collection of program which enables the user to create and maintain a database.
- Database model depicts the way in which data in the database are structured (represented).
- Conceptual data model focuses on how the data element in the database is to be grouped.
- Implementation data model focuses on how the data are represented in the database.
- Entity relationship model shows graphical representation of entities and their relationship in a database structure.
- Entity, attribute and relationship are the elements of ER model
- Entity is a thing of interest to an organization about which data is to be held.
- Attribute is a property or characteristic of an entity.
- Relationship is a link or association between entities.
- Hierarchical database model groups the record in such a way that their relationship form a branching.
- Network database model is a DBMS model where a single file can be associated to one or more than one number of files forming a network.
- In relational database model the data are separated using two dimensional table which are made up of rows and columns.
- Flat file database is the database which takes the form of the table, where only one table can be used for each database. A flat file database can work only with one file at a time.
- An attribute or set of attribute is called candidate key; if it uniquely identifies records in a table.
- Primary key is the candidate key which is chosen by database designers as a principle means of identifying each row in a table.
- When a primary key of one table appears as an attribute in another table then that is foreign key in the second table.
- In the era of traditional DBMS, applications were developed independently for different departments.
- Data redundancy means same data was repeated in different files.
- Data integrity is the accuracy of data and its conformity to its expected value.
- Data independence means that data in a database are separated from the program that manipulates it.
- Normalization organizes the data in a database in a proper manner.

- In first normal form we make the table have no repeating attributes or group of attributes.
- In second normal form we remove all partial dependencies.
- In third normal form we remove all transitive dependencies.
- Data dictionary is a master table which contains data about all the databases in a database system.
- Database administration is the person who manages the database.
- Microsoft Access is a powerful database application software, yet easy to learn, relational database application for Microsoft Windows.

Exercises

1. What do you mean by Data, database and DBMS?
2. Explain various type of DBMS model.
3. Write short notes on:
   a. Database security
   b. Primary key
   c. Foreign key
4. Explain about DDL, DML.
5. Discuss various advantages and disadvantages of DBMS.
6. Write a short note on data base administrator.
7. Explain normalization.
8. What is entity relationship model?
9. What are the prerequisite of good data base system?
10. Starting from any unnormalized table make 1NF, 2NF and 3NF tables
11. Who is database administrator?
12. List out the responsibilities of database administrator.
13. Explain the four principle guidelines used during the database design.
14. Explain the bottom up and top down approach to database design.
15. Explain the various types of database system.
16. Explain Relational database management system.
17. What is ms-access? What are the basic components of Microsoft access? Define.
18. Explain the major features of Microsoft Access.
19. What are the uses of ms-access? List out.

LAB Assignment:

I. Prepare a database of your colleagues in your class with field having name, address, roll, sex, age and contact number. Normalize the table in 1 NF, 2 NF and 3 NF.

II. Create a query of ten students of your class of personal information by asking them. Insert primary key, foreign key and candidate key to the field and record. Also secure your database by applying database security features. It is highly recommend to use SQL for your database.

►►►