B.Com (Computers) II Year
RELATIONAL DATABASE MANAGEMENT SYSTEM
Unit- I

1. What is Data?
A. Data is a collection of raw information.

2. What is Information?
A. Information is a collection of processed data

3. What is Database?
A. Database is a collection of inter-related data items that can be processed by one or more application systems.

4. What is DBMS?
A. Database Management System is a collection of interrelated data and set of programs to access those data. The DBMS is a general purpose software system that facilitates the process of defining constructing and manipulating databases for various applications.

5. What are the disadvantages of File Oriented System?
A. The typical file-oriented system is supported by a conventional operating system. Permanent records are stored in various files and a number of different application programs are written to extract records from and add records to the appropriate files.

The following are the disadvantages of File-Oriented System:

(i). Data redundancy and Inconsistency: Since files and application programs are created by different programmers over a long period of time, the files are likely to be have different formats and the programs may be written in several programming languages. Moreover, the same piece of information may be duplicated in several places. This redundancy leads to higher storage and access cost. In addition, it may lead to data inconsistency, i.e. the various copies of same data may no longer agree.

(ii). Difficulty in accessing data: the conventional file processing environments do not allow needed data to be retrieved in a convenient and efficient manner. Better data retrieval system must be developed for general use.

(iii). Data isolation: Since data is scattered in various files, and files may be in different formats, it is difficult to write new application programs to retrieve the appropriate data.

(iv). Concurrent access anomalies: In order to improve the overall performance of the system and obtain a faster response time, many systems allow multiple users to update the data simultaneously. In such an environment, interaction of concurrent updates may result in inconsistent data.

(v). security problems: Not every user of the database system should be able to access all the data. For example, in banking system, payroll personnel need only that part of the database that has information about various bank employees. They do not need access to information about customer accounts. It is difficult to enforce such security constraints.

(vi). Integrity problems: The data values stored in the database must satisfy certain types of consistency constraints. For example, the balance of a bank account may never fall below a prescribed amount. These constraints are enforced in the system by adding appropriate code in the various application programs. When new constraints are added, it is difficult to change the programs to enforce them. The problem is compounded when constraints involve several data items for different files.

(vii). Atomicity problem: A computer system like any other mechanical or electrical device is subject to failure. In many applications, it is crucial to ensure that once a failure has occurred and has been detected, the data are restored to the consistent state existed prior to the failure.

6. What are the advantages of DBMS over File Oriented System?
A. The following are the advantages of DBMS:

I. Data Redundancy: A major difficulty was that many applications used their own special files of data. Thus, some data items were common to several applications. In a bank, for example, the same customer name might appear in a checking account file, a savings account file and an installment loan file. Moreover, even though it was always the customer name, the related field often had a different name in the various account files. Thus, CNAME in the checking account file became SNAME in the savings account file and INAME in the installment loan file. The same field also has a different length in the various files. For example, CNAME could be up to 20 characters, but SNAME and INAME might be limited to 15 characters. This redundancy increased the overhead costs of maintenance and storage. Data redundancy also increased the risk of inconsistency among the various versions of common data.

Suppose a customer's name was changed. The name field might be immediately updated in the checking account file, updated next week in the savings account file and updated incorrectly in the installment loan file. Over time, such discrepancies can cause serious degradation in the quality of information contained in the data files.

Database systems can eliminate data redundancy, since all applications share a common pool of data. Essential information such as customer name will appear just once in the database.

Thus, we can enter a name or change once and know that applications will be accessing consistent data.

II. Poor Data Control: In the file system, there was no centralized control at the data element level. It was very common for the same data element to have multiple names, depending on the file it was in.

At a more fundamental level, there is always the chance that the various departments of a company will be inconsistent in their terminology.

III. Inadequate Data Manipulation Capabilities: Indexed sequential files allow the applications to access a particular record by a key such as ProductID. For example, if we knew the ProductID for the table, it is easy to access a record in the table. Suppose we want a set of records. It is not possible to obtain a set of records using file system because they are unable to provide strong connections between data in different files. Database systems were specifically developed to make the interrelating of data in different files.

IV. Excessive Programming Effort: A new application program often required an entirely new set of file definitions. Even though an existing file may contain some of the data needed, the application often required a number of other data items. As a result, the programmer had to recode the definitions of needed data items from the existing file as well as definitions of all new data items. Thus, in file-oriented systems, there was a heavy interdependence between programs and data.

Database provides a separation between programs and data, so that programs can be somewhat independent of the details of data definition. By providing access to a pool of shared data and by supporting powerful data manipulating languages, database systems eliminate a large amount initial and maintenance programming.

7. What is Instance and Schema?

A. Instance: The collection of information stored in the database at a particular moment is called an instance of the database.

Schema: The overall design of the database is called the database schema.

8. What is Data Independence?

A. Data Independence: The ability to modify a schema definition in one level without effecting a schema definition in the next level is called Data Independence.

There are two levels of data independence:

(i). Physical Data Independence: The ability to modify the physical schema without causing application programs to be rewritten.

(ii). Logical Data Independence: The ability to modify the conceptual schema without causing application programs to be rewritten.

Logical Data Independence is more difficult to achieve than physical data independence since application programs are heavily dependent on the logical structure of the data they access.

9. What is a data model?

Data Model: A conceptual method of structuring data is called Data Model.

The development of systems based on three principal data models. These three models are the Hierarchical, the Network and the Relational.

10. Explain the components of Database System.

A. A complete database system in an organization consists of four components.

(i). Hardware: The hardware is the set of physical devices on which a database resides. It consists of one or more computers, disk drives, CRT terminals, printers, tape drivers, connecting cables, etc.

The computers used for processing the data in the database may be mainframe, mini computers or personal computers. Mainframe and mini computers have traditionally been used on a stand-alone basis to support multiple users accessing a common database. Personal computers are often used with stand-alone databases controlled and accessed by a single user.

Disk drivers are the main storage mechanism for databases. Desktop computers, CRT terminals and printers are used for entering and retrieving information from the database.

The success of the database system has been heavily dependent on advances in hardware technology. A very large amount of main memory and disk storage is required to maintain and control the huge quantity of data stored in a database.

(ii). Software: A database system includes two types of software:

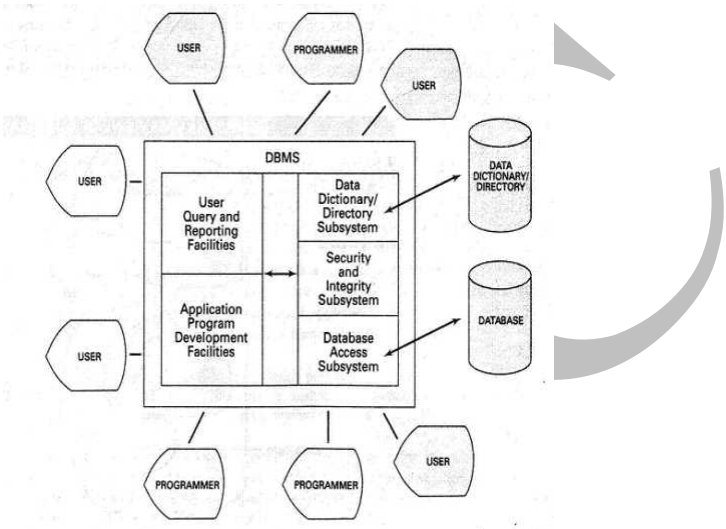
a. General Purpose database management software usually called the database management system (DBMS).
b. Application software that uses DBMS facilities to manipulate the database to achieve a specific business functions.

Application software is generally written by programmers to solve a specific company problem. It may be written in languages like COBOL or C or it may be written in a language supplied by DBMS like SQL. Application software uses the facilities of the DBMS to access and manipulate data in the database providing reports or documents needed for the information and processing needs of the company.

The DBMS is system software similar to an operating system. It provides a number of services to end users and programmers.

DBMS typically provides most of the following services.

1. A central data definition and data control facility known as a data dictionary/directory or catalog.
2. Data security and integrity mechanisms.
3. Concurrent data access for multiple users.
4. User-oriented data query, manipulation and reporting capabilities.
5. Programmer-oriented application system development capabilities.

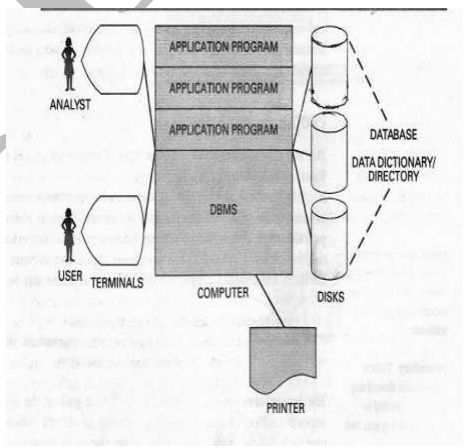


(iii). Data: No database system can exist without data. Data can be collected and entered into the database according to the defined structure.

(iv). People: Two different types of people concerned with the database.

They are:

1. Users: Executives, Managers, Staff, Clerical personnel.
2. Practitioners: Database Administrators, Programmers.



11. What is Data Dictionary?

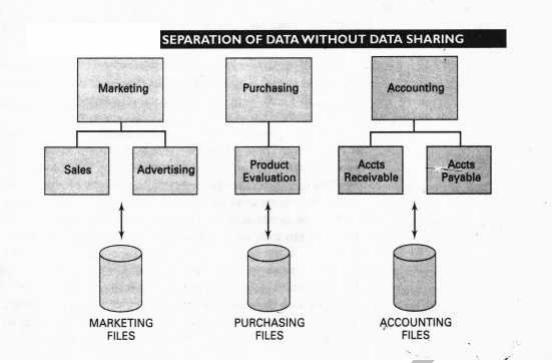
A. A data dictionary / directory subsystem keeps track of the definitions of all the data items in the database. This includes elementary-level data items (fields), group and record-level data structures and relational tables. It keeps track of relationships that exist between various data structures. It

maintains the indexes that are used to access data quickly. It also keeps track of screen and report format definitions that may be used by various application programs.

12. Explain Data Sharing

A.

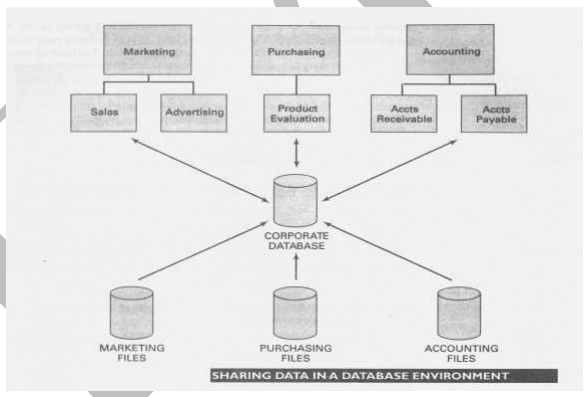
Data without sharing:



The most significant difference between a file-based system and a database system is that data are shared.

There are three types of data sharing:

(i). Sharing between Functional Units: The data sharing suggests that people in different functional areas use common pool of data, each of their own applications. Without data sharing, the marketing group may have their data files, the purchasing group theirs, the accounting group theirs and so on. Each group benefits only from its own data. The combined data are more valuable than the sum of the data in separate files. Not only does each group continue to have access to its own data but, within reasonable limits of control, they have access to other data as well. The concept of combining data for common use is called data integration.



(ii). Sharing data between Different Levels of Users: Different levels of users need to share data. Three different levels of users are normally distinguished: operations, middle management and executive. These levels correspond to the three different types of automated business systems that have evolved during the past three decades:

a. Electronic Data Processing (EDP): EDP was first applied to the lower operational levels of the organization to automate the paperwork. Its basic characteristics include:

- A focus on data, storage, processing and flows at the operational level.
- Efficient transaction processing.
- Summary reports for management.

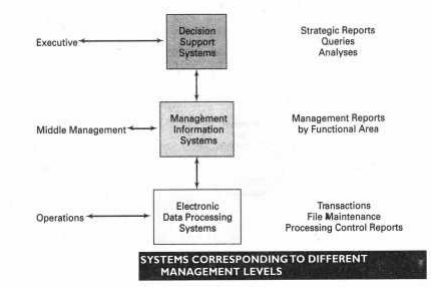
b. Management Information System (MIS): The MIS approach elevated the focus on information systems activities with additional emphasis on integration and planning of the information systems function. This includes:

- An information focus aimed at the middle mangers.
- An integration of EDP jobs by business function such as production MIS, marketing MIS, personnel MIS, etc.
- Inquiry and report generation usually with a database.

c. Decision Support System: DSS is focused still higher in the organization with an emphasis on the following characteristics:

- Decision focused, aimed at top managers and executive decision makers.

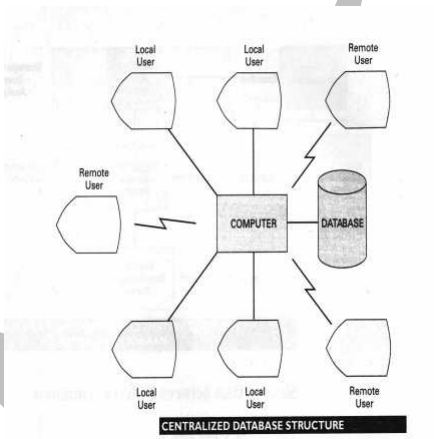
- Emphasis on flexibility, adaptability and quick response.
- Support for the personnel decision-making styles of individual mangers.



SYSTEMS CORRESPONDING TO DIFFERENT MANAGEMENT LEVELS

(iii). Sharing data between Different Locations: A company with several locations has important data distributed over a wide geographical area. Sharing these data is a significant problem.

A centralized database is physically confined to a single location, controlled by a single computer. Most functions for the databases are created are accomplished more easily if the database is centralized. That is, it is easier to update, back up, query and control access to a database if we know exactly where it is and what software controls it.



CENTRALIZED DATABASE STRUCTURE

A distributed database system is made up of several database systems running at local sites connected by communication lines. A query or update is then no longer a single process controlled by one software module, but a set of cooperating processes running at several sites controlled by independent software modules. For a distributed database system to function efficiently, adequate communication technology must be available and the DBMS in the system must be able to communicate while interacting with the communications facilities.

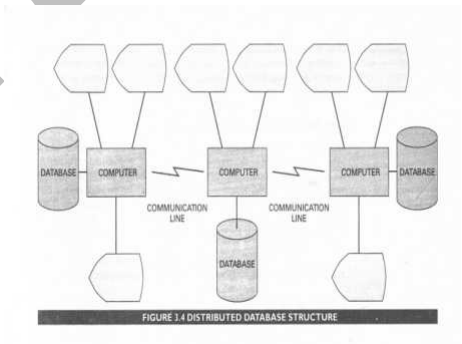


FIGURE 1.4 DISTRIBUTED DATABASE STRUCTURE

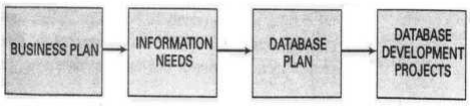
13. Explain Strategic Database Planning.

A. Database Planning is a strategic corporate effort to determine information needs for an extended period. A successful database planning project will precede operational projects to design and implement new databases to satisfy the organization's information needs.

The Need for Database Planning: Database planning has significant advantages:

- It expresses management's current understanding of the information resource.
- It identifies and justifies resource requirements.

- It identifies opportunities for effective resource management including collaboration among departments or divisions within the organizations.
- It specifies action pans for achieving objectives.
- It can provide a powerful stimulus and sense of direction to employees at all levels, focusing their efforts, increasing their productivity and making them feel that they are a genuine part of the enterprise.



The Database Planning Project: Strategic Database Planning is initiated by senior management. They allocate resources and identify personnel to participate in the project. With their commission from management, team members have resources needed to carry out a successful project.

The project team should be extensive experience in information systems and other functional areas of the company. A group of four full-time members, two from information systems and two acquainted with most other areas of the company. All team members should be skilled and respected employees, since their work will have a major impact on the organization for many years. If they are not skilled in a methodology for carrying out the study, an outside consultant should be employed as an advisor to train the team in a suitable methodology. The project team leader should be a consultant but a permanent employee and possibly the head of the database administration.

During the project, the team interacts with senior managers from all the primary user areas. The senior end users identify the principal processes, activities, and entities used in manual or automated information processing. The project team synthesizes these data into a corporate information model included as part of the comprehensive database plan.

A report covering at least the next five should be delivered to senior management. This report will include analyses of the following:

- Information needs of the functional areas.
- Information needs of different management levels.
- Information needs of the geographical locations.
- A model of this information needs.
- Anticipated data volumes by geographical location projects for the period under study.
- A preliminary estimate of costs associated with system upgrades.
- Recommendations for detailed development of new or enhanced databases with schedules.

14. Explain the functions of DBA.

A. Database Administrator is a person with the responsibility of controlling and protecting the data. The DBA should coordinate the design of the database, guide the development and implementation of data security procedures, protect the integrity of data values and make sure system performance is satisfactory.

In a small organization, one person carries out all these responsibilities. Often, these functions are assigned to a group of people. This is most likely in a large organization where DBA responsibilities are divided among several people managed by a chief administrator.

The functions of DBA include:

(i). Database Design: Conceptual Database Design consists primarily of defining the data elements to be included in the database, the relationship that exists between them and the value constraints apply. A value constraint is a rule defining the permissible values for a specific data items. Physical Database Design determines the physical structure of the database and includes such decisions as what access methods will be used to retrieve data and what indexes will be built to improve the performance of the system.

(ii). User Training: The DBA is responsible for educating users in the structure of the database and in its access through the DBMS. This can be done in formal training sessions by interacting with users to create database views, through user's manuals and periodic memos and through company information centers. An information center is an area where users are provided with facilities to do their own computing.

(iii). Database Security and Integrity: The concept of combining an organization's data into one common pool accessible to all has both advantages and disadvantages. The obvious advantage of data sharing is offset by the disadvantage that data can be misused or damaged by the users who do not have original responsibility and authority over the data. The DBA provides procedures and controls to prevent the abuse of data.

Access of database is ultimately controlled by a password mechanism, whereby a user attempting access gives a system-validates password. The system allows the validated user only those access rights recorded in the data dictionary. The DBA is responsible for assigning passwords and controlling privileges. Data integrity refers to the problems of maintaining the accuracy and consistency of data values. Security mechanisms such as passwords and data views protect data integrity.

(iv). Database System Performance: A database system being simultaneously accessed by many users may respond very slowly at times because the physical problems associated with users competing for the same resources are not trivial. Thus, the DBA staff should include technically skilled

personnel who can diagnose and solve system response-time problems. Problem solution may hardware acquisition, physical rearrangement of data on disk, construction of indexes for rapid access to high-volume data or the writing of special software to improve access time. The DBA may decide to maintain redundant copies of data to improve system performance. Such redundancy must be controlled; however problems of data inconsistency will be avoided.

15. What are the Risks and Costs of databases?

A. Database systems have drawbacks.

The following are the Risks & Costs of a database:

(i). Organizational Conflicts: Pooling data in a common database may not be politically feasible in some organizations. Certain user groups may not be willing to relinquish control over their data to the extent needed to integrate data. Moreover, the risk involved in data sharing – for example, that one group may damage another group's data – and the potential system problems that may limit a group's access to its own data may be viewed as more troublesome than beneficial. Such people problems could prevent the effectual implementation of a database system.

(ii). Development Project Failure: For a variety of reasons, the project to develop a database system may fail. Sometimes management was not fully convinced of the value of the database system in the first place. A database project that seems to be taking too long may be terminated.

A project too large in scope may be almost impossible to complete in a reasonable time. Again, management and users become disenchanted and the project fails.

During the course of a project, key personnel may unexpectedly leave the company. If replacement personnel cannot be found, then the project might not be successfully completed.

(iii). System Failure: When the system goes down, all users directly involved in accessing must wait until the system is functional again. This may require a long wait. Moreover, if the system or application software fails, there may be permanent damage to the database. It is very important, therefore to carefully evaluate all software that will have a direct effect on the database to be certain that it is as free as errors as possible. If the organization does not use a database, it is not exposed to this risk, since the data and its software are distributed.

(iv). Overhead Costs: The database approach may require an investment in both hardware and software. The hardware to run large DBMS must be efficient and will generally require more main memory and disk storage than simpler file-based system. Tape drivers for rapidly backing up the database are also required. In addition, the DBMS itself may be quite expensive.

The DBMS may also need increase operating costs, since it requires more execution time. For example, an application system using a DBMS will usually execute more slowly than a system not using a DBMS.

(v). Need for Sophisticated Personnel: The database administration function requires skilled personnel who are capable of coordinating the needs of different user groups, designing views, integrating those views into a single schema, establishing data recovery procedures and fine tuning the physical structure of the database to meet acceptable performance criteria. There is a risk involved in identification of personnel for the DBA, since if no person having the requisite skills can be found, the DBA function may not be properly performed. This could result in significant problems and may even result in the failure of a database implementation.

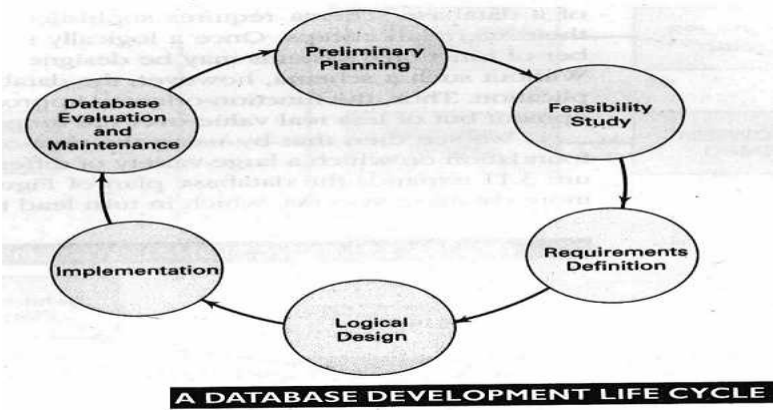
16. List and explain the different stages of DDLC.

A. DDLC (Database Development Life Cycle):

It is a process for designing, implementing and maintaining a database system.

It consists of six stages:

1. Preliminary design
2. Feasibility design
3. Requirements definition
4. Conceptual design
5. Implementation
6. Database evaluation and maintenance.



A DATABASE DEVELOPMENT LIFE CYCLE

Preliminary Planning: It is a specific database system takes place during the strategic database planning project. After the database implementation project begins, the general information model produced during database planning is reviewed and enhanced if needed. During this process, the firm collects information to answer the following questions:

1. How many application programs are in use, and what functions do they perform?
2. What files are associated with each of these applications?
3. What new applications and files are under development?

This information can be used to establish relationships between current applications and to identify uses of application information. It also helps to identify future system requirements and to assess the economic benefits of a database system.

Feasibility Study: A feasibility study involves preparing report on the following issues:

1. Technological feasibility: Is suitable technology available to support database development?
2. Operational feasibility: Does the company have personnel, budget and internal expertise to make a database system successful?
3. Economic feasibility: Can benefits be identified? Will the desired system be cost-beneficial? Can costs and benefits be measured?

Requirements Definition: It involves defining the scope of the database identifying management and functional area information requirements and establishing hardware/software requirements. Information requirements are determined from questionnaire responses, interviews with managers and clerical users and reports and forms currently being used.

Conceptual Design: The conceptual design stage creates the conceptual schema for the database. Specifications are developed to the point where implementation can begin. During this stage, detailed models of user view are created and integrated into a conceptual data model recording all corporate data elements to be maintained in the database.

Implementation: During database implementation, a DBMS is selected and acquired. Then the detailed conceptual model is converted to the implementation model of the DBMS, the data dictionary built, the database populate, application programs developed and users trained.

Database Evaluation & Maintenance: Evaluation involves interviewing users to determine if any data needs are unmet. Changes are made as needed. Over time the system is maintained via the introduction of enhancements and addition of new programs and data elements as business needs change and expand.

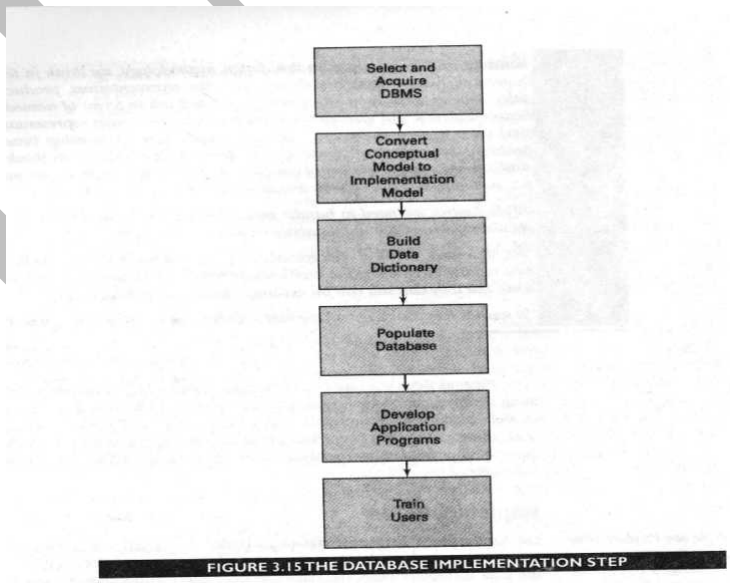


FIGURE 3.15 THE DATABASE IMPLEMENTATION STEP