
UNIT 5: DATABASE MANAGEMENT SYSTEM

Contents

- 5.0 Aims and Objectives
- 5.1 Introduction
- 5.2 File Organization Terms and Concepts
- 5.3 Accessing Records from Computer Files
- 5.4 Problems with The Traditional File Environment
- 5.5 Database Management System (DBMS)
 - 5.5.1 Components of DBMS
 - 5.5.2 Designing Database
 - 5.5.3 Requirements of Database System
- 5.6 Summary
- 5.7 Answer to Check Your Progress Exercise

5.0 AIMS AND OBJECTIVES

After studying this unit, you will be able to explain

- the meaning of database
- what database management system is
- the components of database system
- the merits and demerits of DBMS
- the requirements of DBMS

5.1 INTRODUCTION

An effective information system provides users with timely, accurate, and relevant information. This information is stored in computer files. When the files are properly arranged and maintained, users can easily access and retrieve the information they need.

Well-managed, carefully arranged files make it easy to obtain data for business decisions whereas poorly managed files lead to chaos in information processing, high costs, poor

performance, and little, if any flexibility. Despite the use of excellent hardware and software, many organizations have inefficient information systems because of poor file management.

Database technology can cut through many of the problems created by traditional file organization. A more rigorous definition of a database is a collection of data organized to serve many applications efficiently by centralizing the data and minimizing redundant data. Rather than storing data in separate files for each application, data are stored physically to appear to users as being stored in only one location. A single database services multiple applications.

5.2 FILE ORGANIZATION TERMS AND CONCEPT

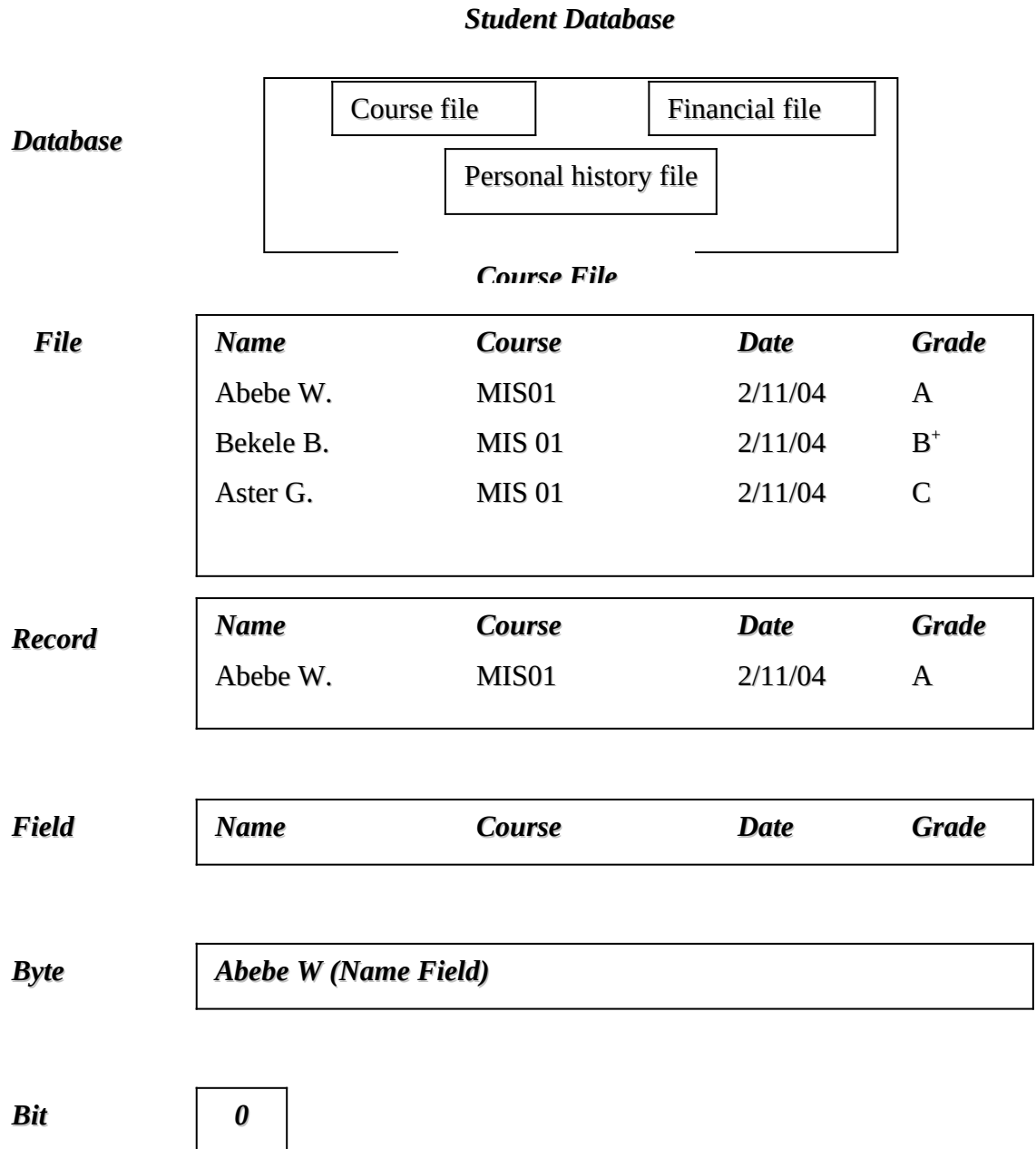
A computer system organizes data in a hierarchy that starts with bits and bytes and progresses to fields, records, files, and database. A bit represents the smaller unit of data a computer can handle.

A group of bits, called a byte, represent a single character, which can be a letter, a number or another symbol. A grouping of characters into a word, a group of words, or a complete number (such as a person's name or age), is called a field. A group of related fields, such as the student's name, the course taken, the date and the grade make up a record. A group of records of the same time is called a file. A group of related files make up a database.

A record describes an entity. An entity is a person, place, thing, or event on which we maintain information. An order is a typical entity in a sales order file, which maintains information on a firm's sales orders. Each characteristic or quality describing a particular entity is called an attribute. For example, order number, order date, order amount, item number, and item quantity would each be an attribute of the entity order.

Every record in a file should contain at least one field that uniquely identifies that record so that the record can be retrieved, updated or sorted. This identifier field is called a key field.

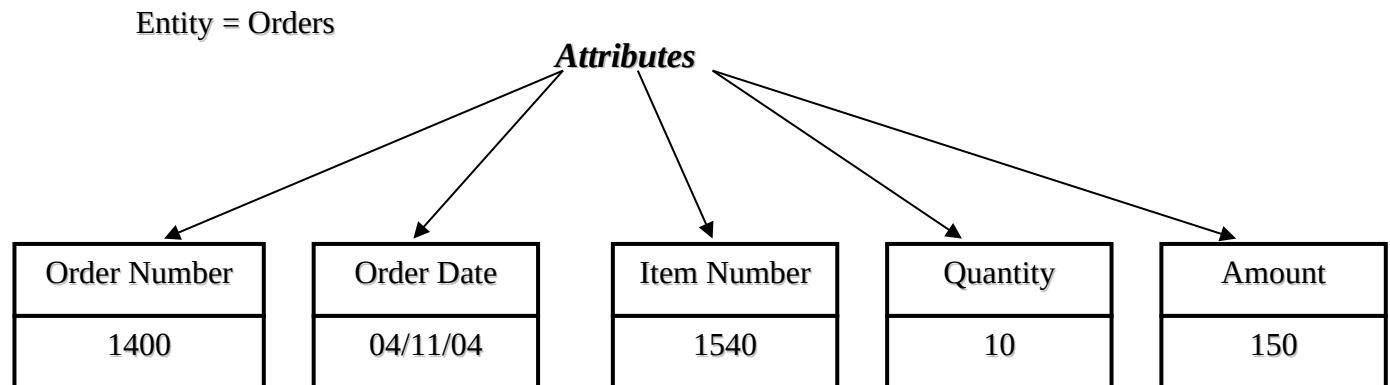
Figure5.1 : Hierarchy of Data



A computer system organizes data in a hierarchy that starts with the bit, which represents either a 0 or a 1. Bits can be grouped to form a byte to represent one character, number or symbol. Bytes can be grouped to form a field and related fields can be grouped to form a record.

Related records can be collected to form a file and related files can be organized into a database.

Figure 5.2 Entities and Attributes



The above record describes the entity called order and its attributes. The specific values for order and its attributes. The specific vales for order number, order date, item number, quantity, and amount for this particular order are the fields for this record. Order number is the key field because each order is assigned a unique identification number.

5.3 ACCESSING RECORDS FROM COMPUTE FILES

Computer system store files on a secondary storage devices. Records can be arranged in several ways on storage media, and the arrangement determines the manner in which individual records can be accessed or retrieved.

One way to organize records is sequentially. In sequential file organization, data records must be retrieved in the same physical sequence in which they are stored. In contrast, direct or random file organization allows users to access records in any sequence they desire without regard to actual physical order on the storage media.

Sequential file organization is the only file organization method that can be used on magnetic tape. This file organization method is no longer popular, but some organization still use it for batch processing applications in which they access and process each record sequentially. A typical application using sequential files is payroll, where all employees in a firm must be paid one by one and issued a cheque. Direct or random file organization is utilized with magnetic disk technology (although records can be stored sequentially on disk if desired). Most computer applications today utilize some method of direct file organization.

Although records may be stored sequentially on direct access storage devices, individual records can be accessed directly using the Indexed Sequential Access Method (ISAM). This access method relies on an index of key fields to locate individual records. An index to a file is similar to the index of a book, as it lists the key field of each record and where that record is physically located in storage to expedite location of that record. Records are stored on disk in their key sequence. ISAM is employed in applications that require sequential processing of large numbers of records but that occasionally require direct access of individual records.

The direct file access method is used with direct file organization. This method employs a key field to locate the physical address of a record. However, the process is accomplishing using a mathematical formula called a transform algorithm to translate the key field directly into the record's physical storage location on disk. The algorithm performs some mathematical computation on the record key, and the result of that calculation is the records physical address.

This access method is most appropriate for applications where individuals records must be located directly and rapidly for immediate processing only. A few records in the file need to be retrieved at one time, and the required records are found in no particular sequence, i.e., on line hotel reservation system.

5.4 PROBLEMS WITH THE TRADITIONAL FILE ENVIRONMENT

Most organizations began information processing on a small scale automating one application at a time. Systems tended to grow independently, and not according to some grand plan. Typically, each division of a multi-division company developed its own applications. Within each division, each functional area tended to develop systems in isolation from other functional areas. Accounting, finance, manufacturing, and marketing all developed their own systems and data files.

In the company as a whole, this process led to multiple master files created, maintained and operated by separate divisions or departments. The traditional file environment is a way of collecting and maintaining data in an organization that leads to each functional area or division creating and maintaining its own data files and programs.

Under this file environment, there is no central listing of data files, data elements or definition of data. The organization is collecting the same information on far too many documents. The resulting problems are data redundancy, program data dependence, inflexibility, poor data security, and inability to share data among applications.

i) Data redundancy

It is the presence of duplicate data in multiple data files. Data redundancy occurs when different divisions, functional areas, and groups in an organization independently collect the same piece of information.

ii) Program data dependence

It is the tight relationship between data stored in files and the specific programs required to update and maintain those files. Every computer program has to describe the location and nature of the data with which it works. These data declarations can be longer than the substantive part of the program. In a traditional file environment, any change in data requires a change in all of the programs that access the data.

iii) Lack of flexibility

A traditional file system can deliver routine scheduled reports after extensive programming efforts, but it cannot deliver ad hoc reports or respond to unanticipated information requirements in a timely fashion. The information required by ad hoc requests is "somewhat in the system" but is too expensive to retrieve. Several programmers would have to work for weeks to put together the required data items in a new file.

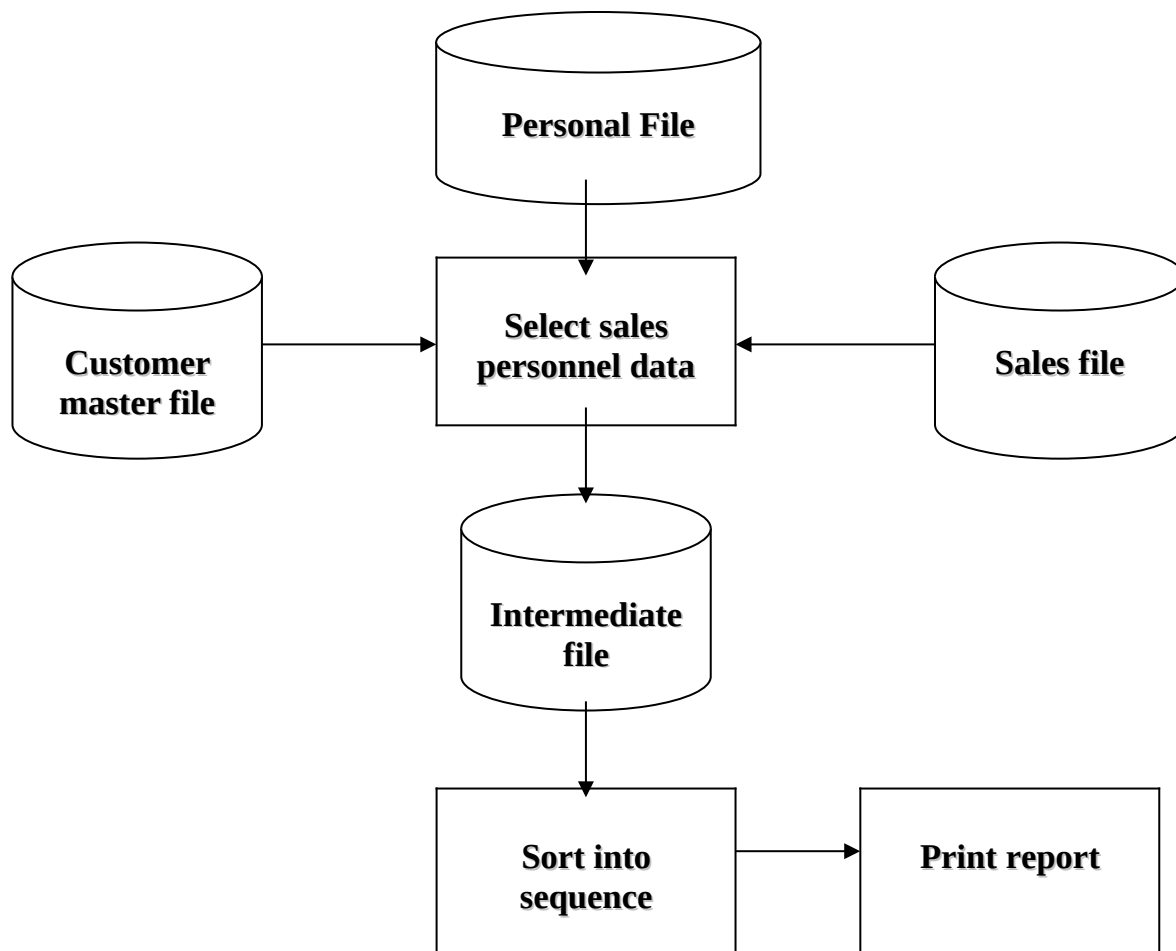
iv) Poor security

Because there is little control or management of data, access to and dissemination of information are virtually out of control. What limits on access exist tend to be the result of habit and tradition, as well as of the sheer difficulty of finding information.

v) Lack of data sharing and availability

The lack of control over access to data in this confused environment does not make it easy for people to obtain information. Because pieces of information in different files and different parts of the organization cannot be related to one another, it is virtually impossible for information to be shared or accessed in a timely manner. The following figure illustrates the traditional file processing.

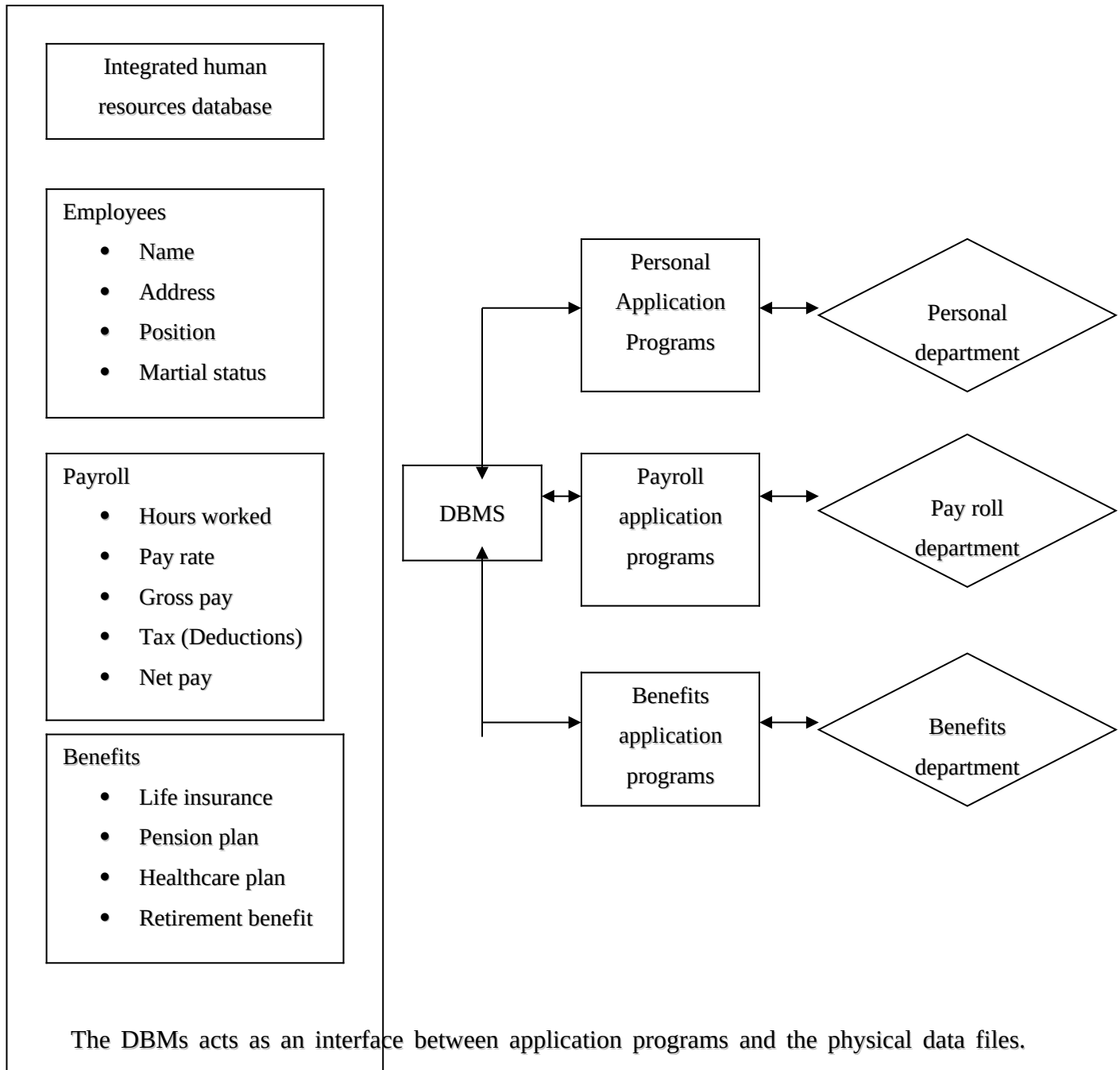
Figure 5.3 Traditional File Processing



5.5 DATABASE MANAGEMENT SYSTEM (DBMS)

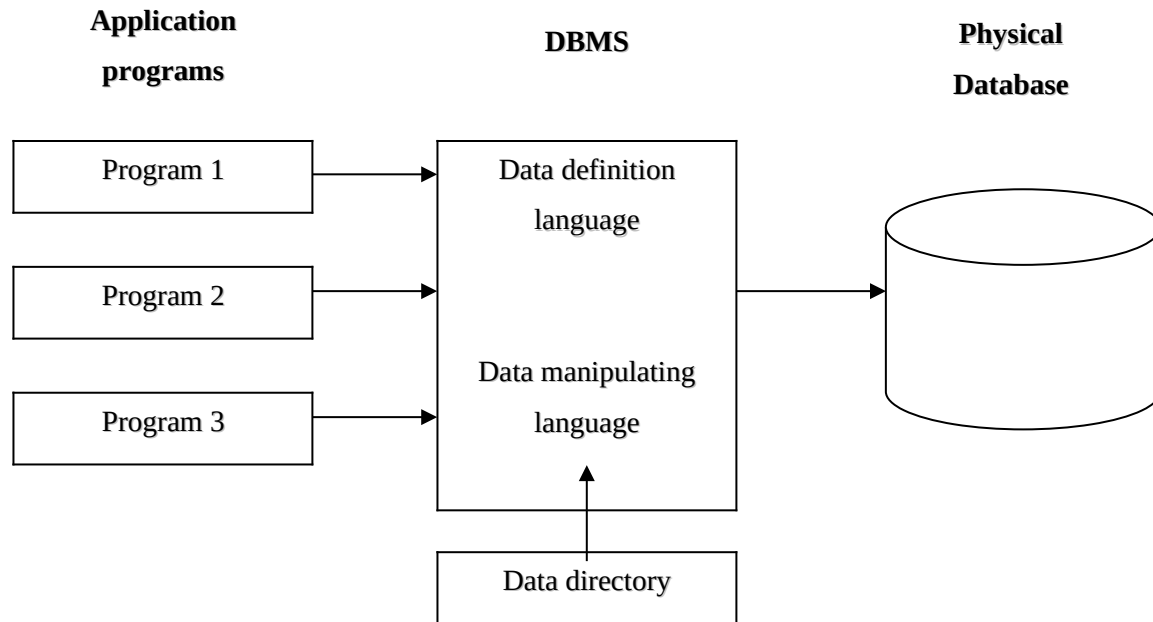
A database management system (DBMS) is simply the software that permits an organization, to centralize data, manage them efficiently, and provide access to the stored data by application programs. consider the following figure, that illustrates the database concept.

Figure5.4 : Contemporary Database Environment



programmer would have to define the data and then tell the computer when they are. The following figure illustrates the elements of a database management system.

Figure 5.5: Database Management System



A database management system has three components:

- A data definition language
 - A data manipulation language
 - A data directory
-
- **Data definition language:** the data definition language is the formal language used by programmers to structure of the database. The data definition language defines each data element as it appears in the database before that data element is translated into the forms required by the application programs.
 - **Data manipulation language:** most DBMs have a specialized language called a data manipulation language that is used in conjunction with some conventional third or fourth generation programming languages to manipulate the data in the database. This language contains commands that permit end users and programming specialists to extracts data

from the database to satisfy information requests and develop applications. The most prominent data manipulation language today is SQL, or structured Query language.

- **Data dictionary:** the third element of DBMs is a data dictionary. This is an automated or manual file that stores definitions of data elements and data characteristics such as usage, physical representation, ownership (who in the organization is responsible for maintaining the data), authorization, and security. Many data dictionaries can produce lists and reports of data utilization, groupings, program locations, and so on.

Advantages of Database Management System

The advantages of a DBMS are as follows:

- Complexity of the organization's information system environment can be reduced by central management of data, access, utilization, and security
- Data redundancy and inconsistency can be reduced by eliminating all of the isolated files in which the same data elements are repeated
- Data confusion can be eliminated by providing central control of data creation and definitions
- Program data dependence can be reduced by separating the logical view of data from its physical arrangement
- Program development and maintenance costs can be radically reduced
- Flexibility of information systems can be greatly enhanced/by permitting rapid and inexpensive ad hoc queries of very large pools of information
- Access and availability of information can be increased

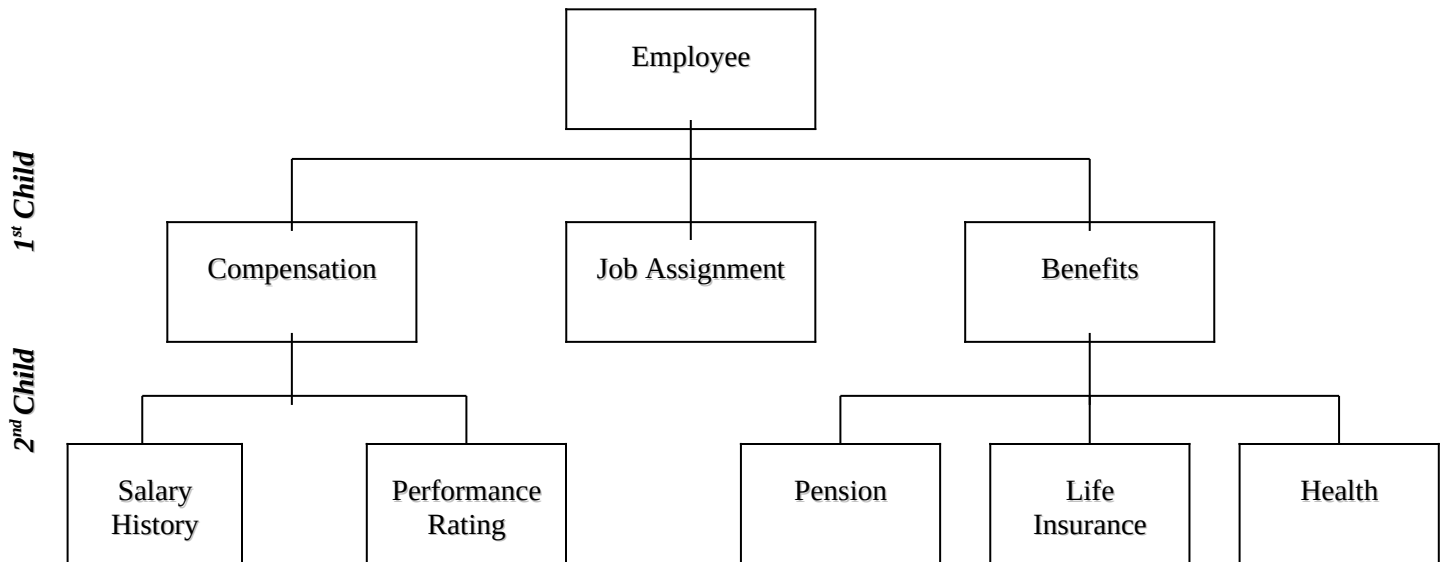
5.5.2 Designing Database

There are alternative ways of organizing data and representing relationship among data in a database. Conventional DBMS uses one of three principal logical database models for keeping track of entities, attributes, and relationships. The three principal logical database models are hierarchical, network, and relational. Each logical models has certain processing advantages and certain business disadvantages.

i) Hierarchical Data Model

The hierarchical data model presents data to users in a tree like structure. Within each record, data elements are organized into pieces of records called segments. To the user, each record looks like an organizations chart with one top level segment called the root. An upper segment is connected logically to a lower segment in a parent-child, but a child can have only one parent. Consider the following figure.

Figure 5.6 The Hierarchical Data Base



The above shows the hierarchical structure that might be used for a human resources database. The root segment is "employee" which contains basic employee information such as name, address, and identification number. Below it are three child segments: compensation (containing salary and promotion data), job assignment (containing data about job positions and departments), and benefits (containing data about beneficiaries and various benefits options). The compensation segments has two children below it: performance rating (containing data about employee's job performance evaluations) and salary history (containing historical data about employee's past salaries). Below the benefits segments are child segments for pension, life insurance, and health care, containing data about these various benefits plan.

Behind the logical view of data are a number of physical links and devices to tie the information together into a logical whole. In a hierarchical DBMS the data are physically linked to one another by a series of pointers that form chains of related data segments.

Pointers are data elements attached to the ends of record segments on the disk directing the system to related records.

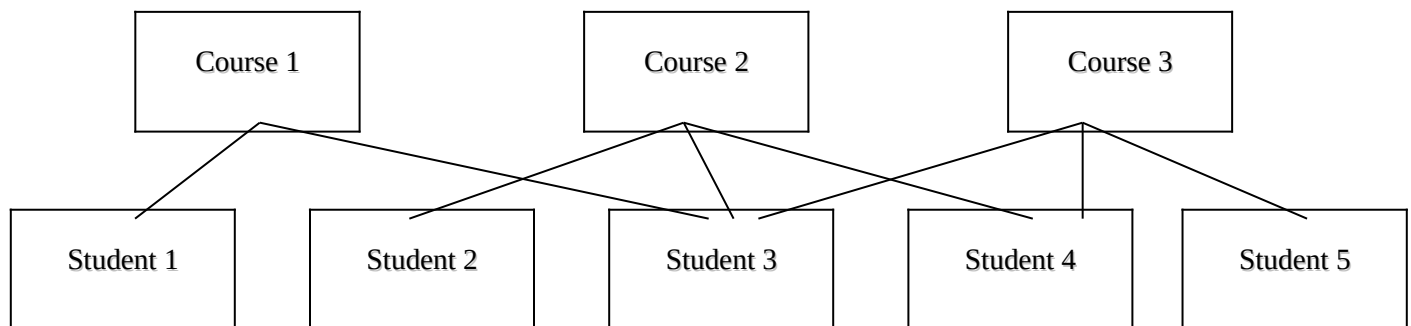
ii) Network Data Model

The network data model is a variation of the hierarchical data model. Indeed data bases can be translated from hierarchical to network and vice versa in order to optimise processing speed and convenience.

Where as, hierarchical structures, despite one to many relationships, network structures depict data logically as many to many relationships. In other words, parents have multiple "children" and a child can have more than one parent.

A typical many to many relationships in which network DBMS excels in performance is the student course relationship (See the following Figure)

Figure 5.7 Network Data Model



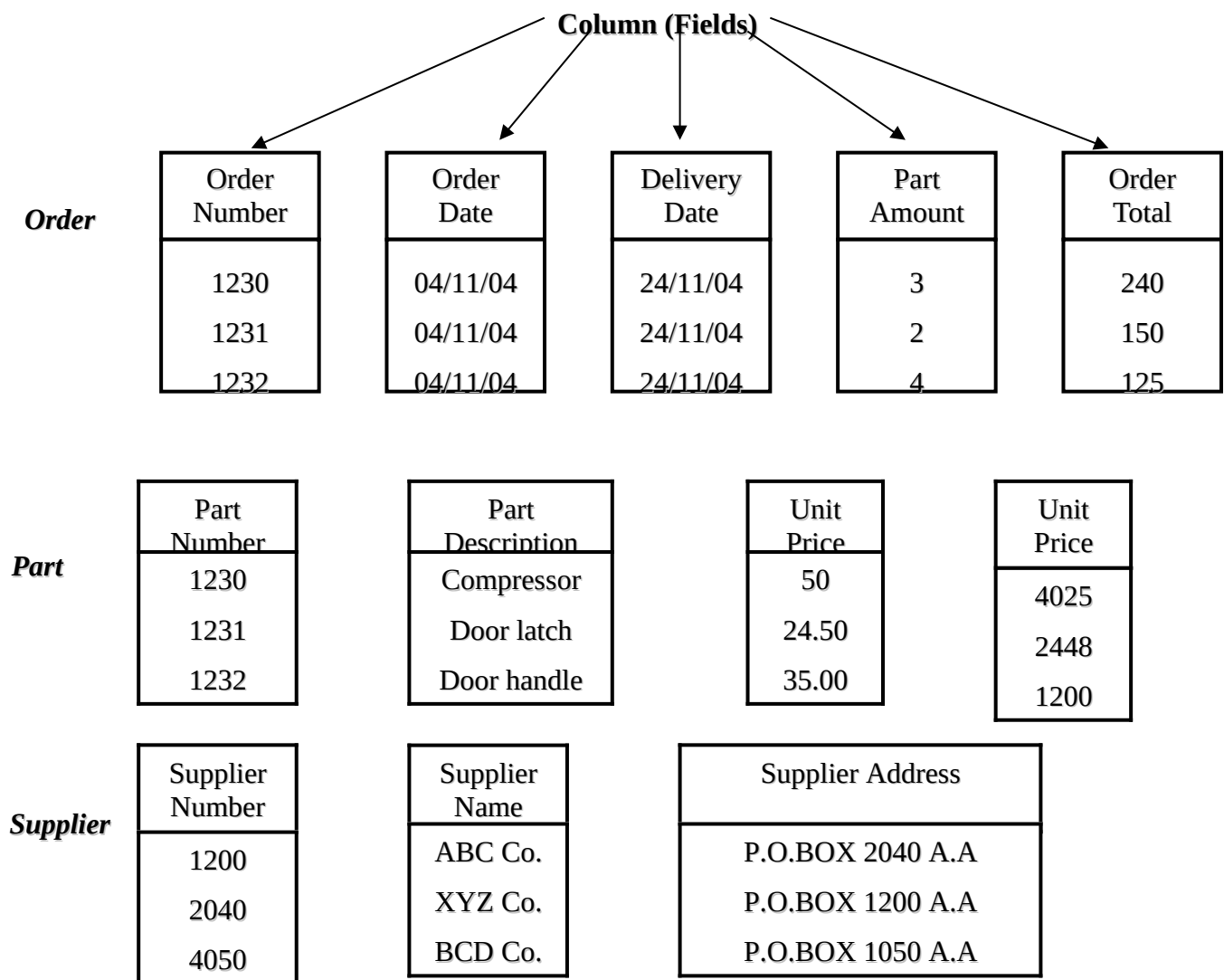
Network structures reduce redundancy and, in certain situations (where many to many relationships are involved), respond more quickly. However, there is a price for this reduction in redundancy and increased speed: the number of pointers in network structures rapidly increases, making maintenance and operation potentially more complicated.

iii) Related Data Model

The relational data model, the most recent of these three database models, overcomes some of the limitations of the other two models. The relational model represents all data in the database as simple two dimensional tables called relations. The tables appear similar to flat files, but the information is more than one file can be easily extracted and combined. Sometimes the tables are referred to as files. Consider the following figure.

Figure 5.8 Relational Data Model

Table (Relation)



The above figure shows a supplier table, a part table and an order table. In each table the rows are unique record and the columns are fields. Another term for a row or record in a relation is

a tuple. Often a user needs information from a number of relations to produce a report. Here is the strength of the relational model: it can relate data in any one file or table to data in another file or table as long as both tables shares a common data element.

Advantages And Disadvantage of The Models

The principal advantage of the hierarchical and network database models is processing efficiency. For instance, a hierarchical model is appropriate for airline reservation transactions processing systems, which must handle millions of structured routine requests each day for reservation information.

Hierarchical and network structures have several disadvantages. All of the access paths, directories, and indices must be specified in advance. Once specified, they are not easily changed without a major programming effort. Therefore, these designs have low flexibility.

Both hierarchical and network systems are programming – intensive, time consuming, difficult to install, and difficult to remedy if design error occurs. They do not support ad hoc, English languages – like inquiries for information.

The strengths of relational DBMS are great flexibility in regard to ad hoc queries, power to combine information from difficult sources, simplicity of design and maintenance, and the ability to add new data and records without disturbing existing programs and applications.

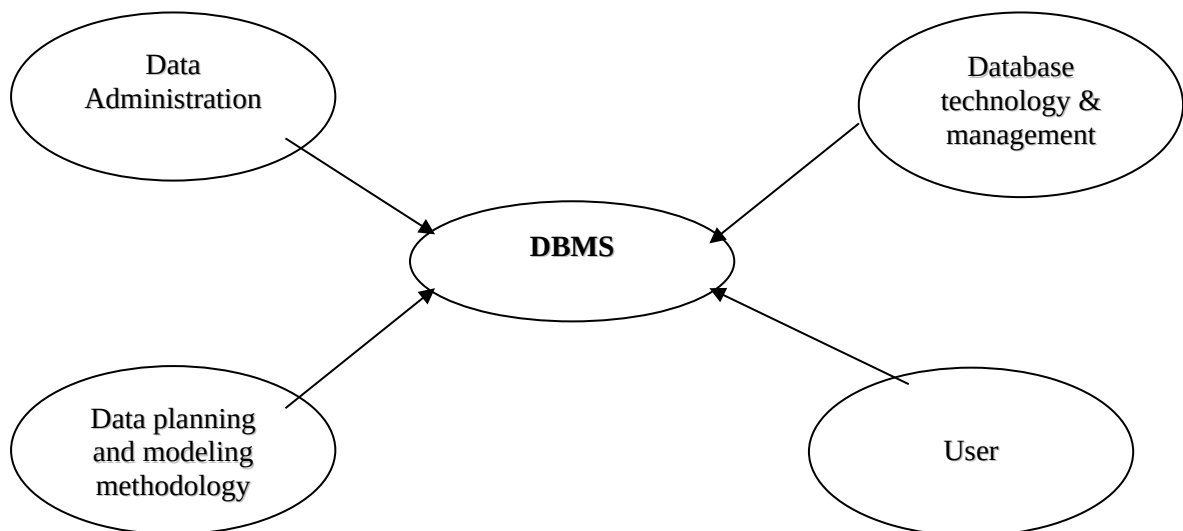
The weakness of relational DBMS are their relatively low processing efficiency. These systems are somewhat slower because they typically require many accesses to the data stored on disk to carry out the select, join, and project commands. Selecting one part number from among millions, one record at a time, can take a long time. Of course, the database can be indicated and "turned" to speed up pre-specified queries. Relational systems do not have the large number of pointers carried by hierarchical systems.

5.5.3 Requirements of Database System

Much more is required for the development of database systems than simply selecting a logical database model. Indeed, this selection may be among the last decision. The database is an organizational discipline, a method, rather than a tool or technology. It requires organizational and conceptual change.

Without management support and understanding, database efforts fail. The critical elements in a database environment are (1) data administration, (2) data planning and modeling methodology, (3) database technology and management, and (4) users. The environment is depicted in the following figure.

Figure 5.8 : Database Environment



1) Data Administration

Database systems require that the organization recognize the strategic role of information and begin actively to manage and plan for information as a corporate resource. This means that the organization must develop a data administration function with the power to define information requirements for the entire company and with direct access to senior management.

Data administration is responsible for the specific policies and procedures through which data can be managed as an organizational resource. These responsibilities include developing information policy, planning for data, overseeing logical database design and data dictionary

development, and monitoring the usage of data by information system specialists and end users groups.

An organization needs to formulate an information policy that specified its rules for sharing, disseminating, acquiring, standardizing, classifying, and inventorying information throughout the organization. Information policy lays out specific procedures and accountabilities, specifying which organizational units share information. While information can be distributed, and maintaining the information.

2) Data Planning And Modeling Methodology

Because the organizational interests served by the DBMS are much broader than those in the traditional file environment, the organization requires enterprise-wide planning for data. Enterprise analysis, which addresses the information requirements of the entire organization (as opposed to the requirements of individual applications), is needed to develop databases. The purpose of enterprise analysis is to identify the key entities, and relationships that constitute the organization's data.

3) Database Technology And Management

Databases require new software and a new staff specially trained in DBMS techniques as well as new management structures. Most corporations develop a database design and management group within the corporate information system division that is responsible for the more technical and operational aspects of managing data. The functions it performs are called database administration. This group does the following:

- Defines and organizes database structure and content
- Develops security procedures to safeguard the database
- Develop database documentation
- Maintains the database management software: in close cooperation with users, the design group establishes the physical database, the logical relations among elements, and the access rules and procedures.
- Users: a database serves a wider community of users than traditional systems. Relational systems with fourth-generation query languages permit employees who are not computer specialists to access large databases. In addition, users include trained computer specialists. In order to optimize access for non-specialists, more

resources must be devoted to training end users. Professional systems workers must be retained in the DBMS language, DBMS application development procedures, and new software practices.

Check Your Progress Exercise

1. Discuss in detail the file access methods?

2. Define database and database management systems.

3. List some of the benefits of DBMS.

4. Describe the three principal database models and their merits and demerits.

5. What are the four key elements of a database environment? Describe each briefly.

5.6 SUMMARY

In traditional file environment, data records are organized using either a sequential file organization or a direct or random file organization. Records in sequential file can be accessed sequentially or they can be accessed directly if the sequential file is on disk and uses an indexed sequential access method. Records on a file with direct file organization can be accessed directly without an index.

Database management systems (DBMS) are the software that permits centralization of data management. A DBMS includes a data definition language, a data manipulation language, and a data dictionary capability. The most important feature of the DBMS is its ability to separate the logical and physical view of data.

The DBMS retrieves information so that the user does not have to be concerned with its physical location. This feature separates programs from data and from the management of data.

There are three principal logical database models: hierarchical, network, and relational. Each has unique advantages and disadvantages. Hierarchical systems, which support one to many relationships are low in flexibility but high in processing speed and efficiency.

Network systems support many to many relationships. Relational systems are relatively slow but are very flexible for supporting ad hoc requests for information and for combining information from different sources.

Development of a database environment requires much more than selection of technology. It requires a change in the corporation's attitude toward information.

The organization must develop a data administration function and a data planning methodology. The database environment has developed more slowly than was originally anticipated. There is political resistance in organizations to many key database concepts, especially to sharing of information that has been controlled exclusively by one organizational group.

There are difficult cost/benefit questions database management. Often, to avoid raising difficult questions, database use begins and ends as a small effort isolated in the information systems department.

5.7 ANSWER TO CHECK YOUR PROGRESS EXERCISE

1. Refer Section 5.3
2. Refer Section 5.5
3. Refer Section 5.5
4. Refer Section 5.5.2
5. Refer Section 5.5.3