UNIT 4 DATABASES MANAGEMENT SYSTEMS (DBMS)

Structure

4.0 OBJECTIVES

After reading this unit, you will be able to :

- understand the concept of database management system;
- know the structure and types of databases; and
- comprehend the database organization.

4.1 INTRODUCTION

The non-technical dictionary meaning of database is "a store of a large amount of information, especially in a form that can be handled by a computer". A database is a collection or set of related data arranged-logically in a structured form designed to meet the information requirement on non redundant operational data which are sharable between different application systems. The advantage with a database is that the data remains independent of the application programs that use them. Further, the data is accessible to any programme with a legitimate need for them, regardless of where the data is physically located. It is also accessible to any programme regardless of the language in which the programme is written. Data are not duplicated in different locations. A database basically comprise data elements or fields each of which contains a data value about an attribute of a particular entity. A set of similarly constructed records constitute a file which contains data records about an entity type. Ultimately a set of related files stored together in a logical fashion comprise a database. Data are raw facts or intangible ideas about something and include numbers, words, symbols, ideas, concepts and oral verbalisation.

The purpose of information systems is to collect, process and store large quantity of data to obtain the information for effective decision making, planning and control.

Moreover, in most organizations for reasons of volume, complexity, timing and computational demand, this collected data must be organised in a manner to serve a variety of user's information request.

Data that stored in a database generally fall into a data hierarchy made up of categories: These categories in the hierarchical order are : Fields, Records, Files and Database

Database

↓

Files

↓

Records

↓

Fields

At the lowest level data is organized into fields. A field represents a subject data item, e.g., in case of bibliographic data for books, we have Author Field, Title Field, etc. Fields are made up of individual letters, numbers or symbols. Once characters are joined into a field, then the field is treated as a Unit.

e.g. Ashok Kumar Singh = Author Field

Record is made up of a set of related fields. A record can contain either one field or a number of fields, for example a record for a book may contain- Author Field, Title Field, Imprint Field, Edition Field, Collation Field etc. So for each book we may have a separate record.

File is a set of related records. A file may contain as few as one or as many as millions of records.

Database is the highest in the hierarchy. A database is a collection or set of related files.

Historically, files came before database. Though database represents a step forward from files, they are in fact constructed from files.

Due to intrinsic advantages of database as mentioned earlier, the database organisers are willing to pay for the cost of creation and maintenance of database. An organization that uses database rather than files can save time and money in developing the application program. It can exploit the data more efficiently, since they are easier to get it. In a database, the data is organized into Record (logical) and Record (physical). A logical record consists of more than one data elements or fields among which some logical connection exists. The contents of the field represent a set of qualities or attributes about the particular real world entity represented by the record and the logical connection among the data elements or field is maintained by their physical representation or the data structure within the storage medium. The physical record is a basic unit of data which is read from or written on the storage medium by a single input/output command to the computer. One physical record often contains multiple logical records or segments. The data file is a collection of one type of stored records or interrelated data that are treated as unit and kept on a secondary computer storage device. A set of similarly constructed records comprise a file. Physical storage area containing data could be program or group of data or records managed as a single unit by the operating system of a computer. A given physical file can be accessed in a wide variety of ways.

The records can be located in a storage medium randomly or directly, i.e., independent of the location of any other record in the file. This type of organization is called direct physical file organization or random access file. When a file is maintained on a sequential or random storage device in sequential access mode then it is called a sequential file. The searching of data or data records is generally through an index. The input to the index file from each record is very important. The operation of creating an index file from a master file is referred to as inversion and files with indexes are often called inverted files.

A database concept has traditionally evolved from file-processing or management system. It is a collection of related and cross referred files designed and created to minimise the repetition of data. These integrated files are part of the overall database system including the specialized software called the database management system (DBMS), which allow data records to be created, accessed, updated, deleted and retrieved. With the evolution of the database system and concepts, the same physical data could be viewed in different logical ways by different applications. The database management system bridges the gap between the logical file description and the physical organisation of database.

4.2 CONCEPTS AND DEFINITION OF DBMS

Database management system (DBMS) is a programme, that makes it possible to create, use and maintain a database. When DBMS is used, the users need not be concerned about the way data are organised or retrieved from the direct access storage device or the particular file from which the data are obtained.

The database management system (DBMS) can be defined as an integrated set of programmes that serve as interface between application programmes and a set of coordinated integrated physical files called a database. It provides access to the organized group of data or data records for purposes of searching, manipulating and reporting. The management of a database involves both definition of structure for the storage of data and the provision of mechanisation for the manipulation of data. Therefore any database management system uses two languages a data definition language (DDL) and a data manipulation language (DML)

The main role of data definition language is to :

— Describe schema and subschema

— Describe the fields in each record

— Indicate key words

— Provide for logical and physical data independence

— Provide means for associating record.

The function of data manipulation language is to provide the machines the information for data manipulation, such as creation, updating, deletion, replacement, retrieval and sorting of data or records. It also enables the user and the application programmes to process the data using logically meaningful data names rather than physical storage location. It also allows the user of the application programmes to be independent of physical data storage and maintenance. The DML also provides for the use of logical relationship among the items. Besides DDL and DML, there is a query language for the users of database. In brief, the DBMS will comprise the following modules to operate on the central database :

a) Data Definition Language (DDL), a mechanism for organising and structuring the database and defining the data elements.

b) Data Manipulation Language (DML), used by programmers to manipulate data, often used in conjuction with a high level language.

c) Query Language, for users of the database who need answers to their question

d) a report writer for producing printed customised report.

e) a data input system.

f) a communication system to support a number of on-line areas.

The database approach to library automation can be explained with the help of a diagram:

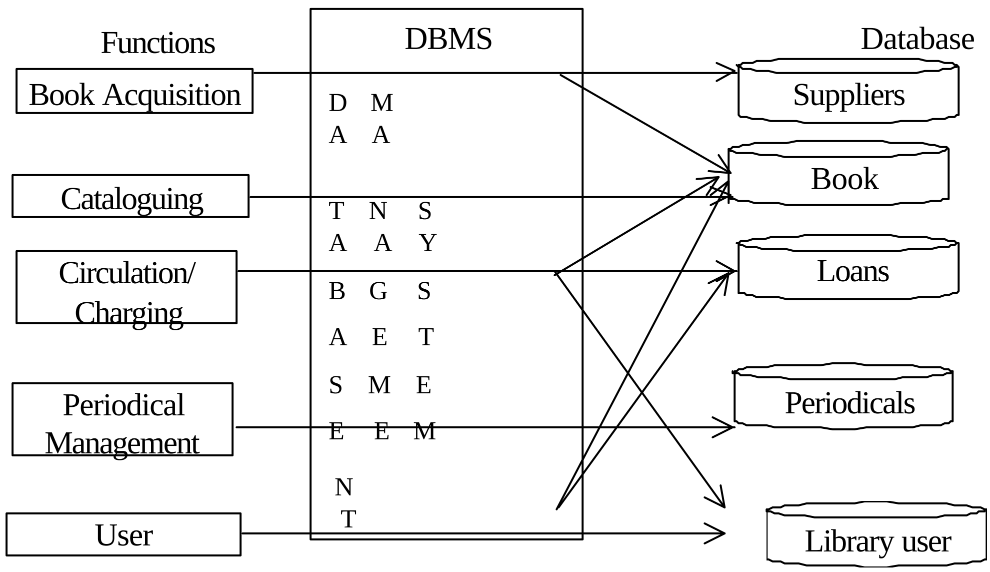


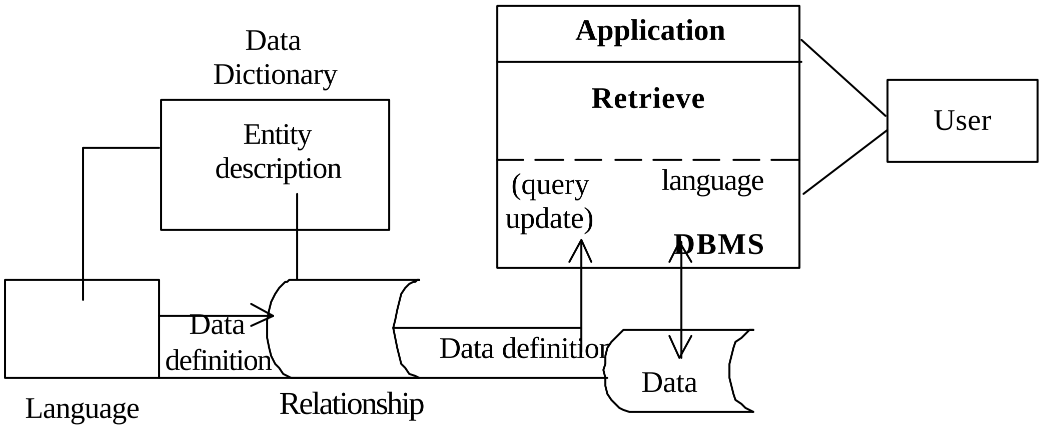
Fig. 1: Database Approach to Library Automation

Advantage of a DBMS

A complete database management system separates the definition of data from the programme that accesses it. This concept of data independence is one of the key advantage of a database management system. The lack of data independence from traditional approaches to progarmming also creates a significant maintenance problem. As programmes are changed to reflect changing conditions or request from users, all the programmes in a system that access the files have to be altered. At a minimum, the record description in the programmes will have to be changed. It may also be necessary to make modifications in the programme themselves to process added data. With a database management system only programmes that access the actual fields altered are generally affected by a change. Programmes that do not use the altered fields do not usually have to be changed. As a result, we have gained some independence between the data and the programmes that access those data. This kind of data independence is the essence of the database concept.

The main functions and the design objectives and benefits of the database management system are: (i) Integration, (ii) data independence, (iii) data retrieval, analysis, modification and storage, (iv) privacy, (v) integrity controls and recovery methods, (vi) compatibility, (vii) concurrency support, (viii) support of complex file structure and access paths.

The overview of the database management system is given below :



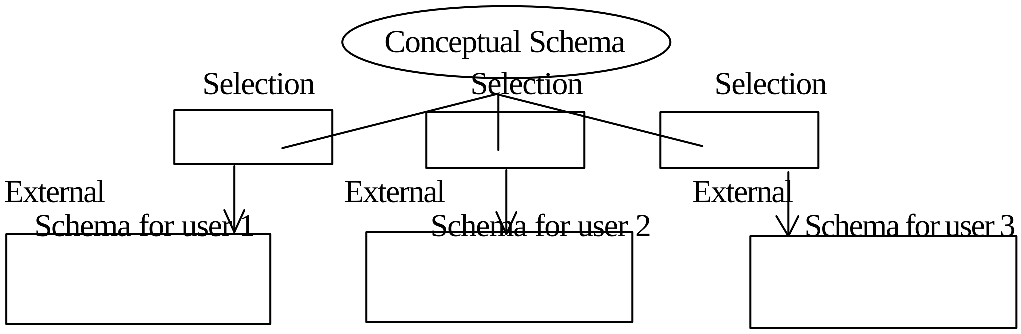
4.3 STRUCTURE OF DBMS

The conceptualisation of the database management system starts with the definition of the schema. A schema is a formulation of the schema in the language interface offered by a DBMS. A schema comprises two parts :

i) a logical schema and

ii) a physical schema.

The logical schema is concerned with exploiting the data structure offered by a DBMS in order to make the schema understandable to a computer. The physical schema, on the other hand, deals with the manner in which the conceptual database shall get represented in the computer as a stored database. It can be said that there is no straight forward step by step procedure to be followed in arriving at the conceptual schema. It is not surprising, therefore, that two different people looking at the same universe of discourse may end up with different conceptual schema or even the same person looking at a given universe at two different times may produce two different schema.



The conceptual schema should have the following properties :

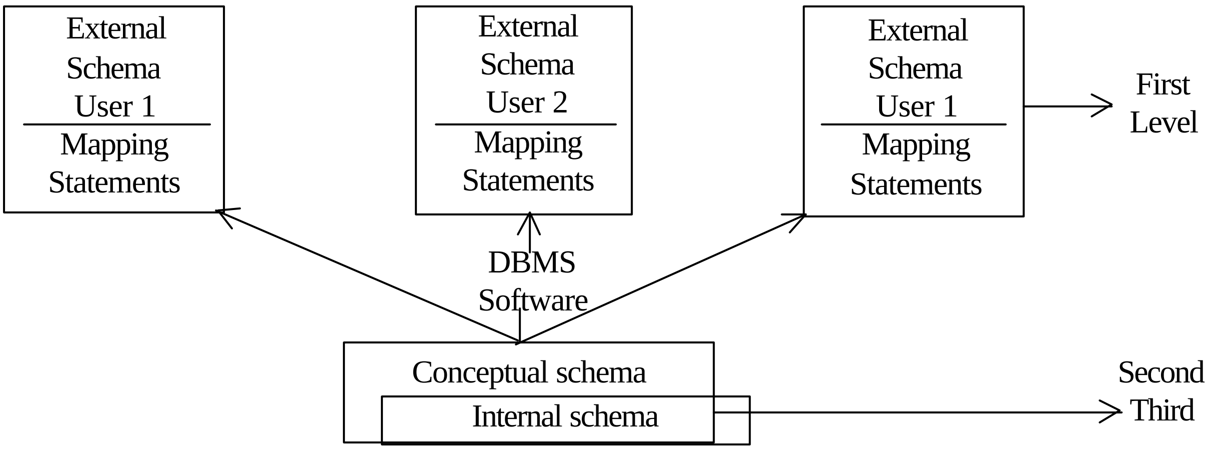
— It should correspond well with the real world of which it is a model.

— It should capture as precisely as possible the information content of the real world.

— It should be complete.

— It showed be possible for the conceptual schema to support different external schema as per the requirements of the users of the information system.

— The representation of conceptual schema should be stable in order to make the specification of both, the logical schema and the physical .



Schema easy for the DBA. In terms of schema, the DBMS has a three-level architecture. The first level structure of DBMS is the external level which refers to the way the users view the data. The external level is also called sub-schema. A user may be interested in a small portion of the database which will form his external view. The architecture of a DBMS is illustrated in figure 2.

The second level conceptual schema represents the total information content of the database. It gives a global or integrated view of the database. The third level of the architecture is the internal schema. The Internal view describes how the data is actually stored and managed on the storage media. It specifies what indexes exist, how fields are represented and the physical sequence of the stored records. The internal schema is built up using the internal DDL. The conceptual/internal mapping statements ensure physical data interdependence. The mapping components between internal schema and secondary storage device, i.e., direct access storage device (DASD) is called access method.

Thus designing of efficient and effective database structure has led to the identification of three distinct planes or levels of data abstraction and description. These three levels are :

i) External or users view level.

ii) Conceptual schema level.

iii) Internal or physical level.

The conceptual schema is machine and application software independent description of total database. The term schema is used to mean an overall chart of the entire database. The conceptual schema might be regarded as an overall logical database description. The model of conceptual schema or model base has to be as stable as possible.

Data Models

In a database the data is usually logically and physically organised according to some data model. A data model is a collection of conceptual tools for describing data, data relationship, data semantics and data constraints.

The data base generally structure their data on the basis of one of the following four data models :

i) Relational data model

ii) Network data model

iii) Hierarchical data model

iv) or a combination of these three or some subsets of these three models.

The hierarchical or network data models have been in use since the early 1960s, whereas the relational model was proposed as a modeling structure in the early 1970s. The difference between these three data models is the way they represent the relationship of entities or their attributes. These models make use of three database constraints-simple sequence relationship, hierarchical relationship and network relationship.

Internal Schema

The physical organization and layout of the database on the storage device is called the internal view. The internal view is represented by means of the internal schema, which not only defines the physical structure of the stored database, but also specifies the methods that may be used to locate the logically related data records, insert new record and delete records.

4.4 TYPES OF DBMS

The DBMS can be classified from different perspectives. But the DBMS is classified according to the way they present data to the user. But it is important to stress that the way data are presented to the user does not bear any resemblance to the way the data are actually physically stored. On the basis of the data structure models, the DBMS are generally of three types :

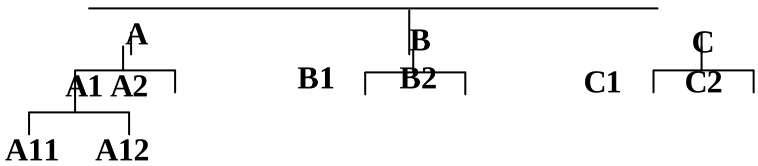
i) Hierarchical DBMS

ii) Network DBMS

iii) Relational DBMS

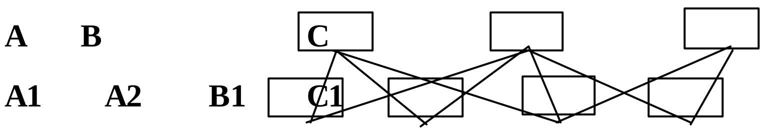
The hierarchical DBMS is the oldest of the three types. In this the data is structured in such a manner that each element is subordinate to another in a strict hierarchical manner. The hierarchical relationship between data elements or records is similar to genus-species, parent-child relationships. The relationship between the two data elements or records in a hierarchical relationship is therefore one of super and subordinate relationship. The advantage of such DBMS is simplicity.

Entities



ii) **Network DBMS**

In a network DBMS, the network support n:m relationship. A network can be a hierarchy but hierarchy cannot be a network.. Defining a network style of database requires considerable skill and planning and changing such system is difficult.



The network structure

Further if the original designer who is familiar with the relationship is not available, then maintenance and change can consume a considerable number of man-hours .

iii) **RDBMS**

Relational DBMS have been evolving since the relational data model was proposed in 1970 by Edgar F Codd of IBM. The relational model is a radically different approach to the problem in that it concentrates on simplifying the view of the database into one made up of flat tables. The data are organised in two dimensional tables as given below. Such tables are easy for a user to develop and understand. The name is derived from the fact that each table represents a relation.

Name	City	Profession	Income (monthly)
Ashok	Delhi	Medical	40000
Neelam	Varanasi	Legal	15000
Mohan	Calcutta	Scientists	20000
Deepak	Mumbai	Engineer	30000

Since different users see different sets of data and different relationships among them, it is necessary to extract subsets of the table columns for same users and to join tables together to form larger tables for others. The mathematics provides the basis for extracting some columns from the tables and for joining various columns. This capability to manipulate relations provides a flexibility not normally available in a hierarchical or network structure.

The relational database management system has many advantages. Most of the DBMS are based on relational model because it is relatively easier for the user to understand.

4.5 DATABASE ORGANIZATION AND DEVELOPMENT

The database design is a step by step procedure. The database design is done usually by a system analyst or database administrator. The overall process of database development and its organization are :

i) Planning – The overall scope and objectives of the database is defined. The hardware and software requirements and constraints must be identified.

ii) Survey of Users Needs – The survey/study must be conducted to identify the needs of potential users. An attempt should be made to ascertain present and future needs.

iii) Conceptual Design – The conceptual design of the database is a two-stage process:

   a) The design of a conceptual schema which is an abstraction of the real world situation under consideration, and

   b) Design of a logical structure, representing this schema.

   The designing of a conceptual schema implies the representation of that part of the 'real model' that the database is about. The entities must be identified about which information or data is to be included in the database.

iv) Physical Database Design – At this stage the conceptual model is translated into physical database model that conforms to the features of particular DBMS. Every DBMS has its own database structure to store information.

v) Design Implementation – The physical model is implemented with the aid of the database management system in establishing information storage and retrieval systems of the organization.

vi) Testing and Evaluation – The implemented design is tested to ensure whether the system is working as expected. The system must be evaluated/monitored to maintain optimum level of efficiency.

4.6 RELATIONAL DATABASE MANAGEMENT SYSTEM

The database management system based on the relational model was proposed by E.F. Codd of the IBM in 1970 and since then it has become an important concept in DBMS technology. The DBMS for micro-computers are based on relational model because it is relatively easy to understand from the user point of view. The basic concept in the relational model is that of a relation which is a set of theoretic notion.

From the pragmatic points of view, one can look at a relation as a table. The relations expressed in the form of a table has the following characteristics :

i) Columns in a table are always homogenous. In other words the column of a table always contain items of the same kind whereas items in different columns may not be of the same kind.

ii) Each item is a simple number or character string.

iii) All rows of a table are distinct i.e. a table does not contain two row which are identical in every column.

iv) The ordering of rows within a table is immaterial.

v) The column of a table are assigned distinct names and the ordering of these columns is immaterial.

Hence each column and rows of a table can be uniqely identified by its content. Consequently one can use content addressibility as a means for accessing individual rows of the table.

So a relational database is a collection of data belonging to a system spread over a number of relations.

Relational Algebra

Let us consider, the relational model formally. This is called relational as relational algebra shall be concerned with the relation as a data structure and with the operations that can be performed on it. The keys (data items) play an important role in relational data base management system. The keys serve two basic purposes :

— Identification of tuples

— Establishing relationship between tables

The keys can be made up of a single attribute called single key attribute or multiple attributes called multi-key attributes. Keys formed by multiple attributes are also called composite or concatenated key.

The Cartesian product of two set A and B, denoted by $A \times B$ is

$$A \times B = \{(a,b) : a \in A \text{ and } b \in B\}$$

The expanded Cartesian product X of n-set A1, A2, … An is defined by

$$X\ (A_1, A_2 \dots An) = \{a_1, a_2, \dots a_n) .\}$$

The element (a_1, a_2 …. a_n) is called an n-tuple. Let there be two tuples d and e which are (d_1, d_2….. d_m) and (e_1, e_2…. e_n) respectively. One can define the operation of concatenation (^) as:

$$a\wedge e = (d_1, d_2 \dots dm, e_1, e_2 \dots e_n)$$

That is the concatenation operator applied to two tuples of degree m and n respectively yields a (m+n) tuple.

The unit on which a relation is constructed is known as a domain. They are two kinds of domain –simple and compound. The simple domain is defined as a set of which each element is an integer or each element of which is a character string. A compound domain is the expanded Cartesian product of a finite number , N($N \geq 1$) of simple domains. N is called the degree of the compound domain.

Given domain D_1, D_2, …. D_n, we define a relation R as a subset of the expanded Cartesian product of these domains as follows :

$$R\ (D_1, D_2 \dots D_n)\ C\ X\ (D_1, D_2 \dots D_n)$$

In general, we say that a relation defined over n domain has a degree n or is in n-ary. The element of this set are n-tuples. The definition of a relation gives a name to the relation and specifies the components over which it is defined. These components are referred to as relation attributes or attributes for short. A relational schema is defined to be a collection of relation definitions. It should be obvious that the relational schema can be used to represent the data base schema.

Relational Table for books

Order no	Author	Title	Supplier	Date
215	William, J.	Information Technology	Student Friends	31.3.1999
216	Naveen Prakash	Database Management	Vishwa Vidyalaya Prakashan	1.4.1999
217	Singh, R.K.	Child Psychology	Current Books	2.4.1999
218	Mishra, L.	Organic Chemistry	Popular Book House	3.4.1999
219	T. Lal	Algebra	Vikas	5.4.99

The above table contains information about books: order no, supplier, title, date of order. While constructing the tables of the relational model one has to fulfil the following requirements:

— Each table must be given a unique heading

— Each attribute must be given a name, referred to as keys

— Each entity record or tuple corresponding to a row in the table must have an attribute or combination of attributes which serves as unique identifier referred to as a primary key.

Within one table, we consider a set of attributes which defines the entity type or a particular facet of the entity type with which we may be concerned in a given context. Each row or tuple in the table represents data for a particular order. An attribute B is said to be functionally dependent on the attribute A if the value of attribute A always determines the value of the attribute B.

A⟶B i.e., A is determinant of B

A functional dependency may involve more than two attributes.

Normalisation

The normalisation technique is concerned with translating a conceptual design into a set of well designed relational table. The normalisation is a major task in designing a relational database. The process of normalisation ensures that there will be no problem in updating the database and that operations on the various relations will not lead to inconsistent and incorrect data. During the normalisation process, the designer first looks to be sure that the relations are in first normal form, next he or she checks for second normal form and finally for third. The first normal form requires that all occurrences of a record type contain the same number of fields. It may be noted that normalisation is primarily aimed at preventing or reducing data maintenance problems rather than improving retrieval efficiency. Normalisation of relations removes anomalies in the database.

The second and third normal forms require the designer to examine the relationship between key fields and other fields in the record. To conform to second and third normal forms, each non-key field must give us information about the entire key and nothing but the key.

e.g. suppose that one has a relationship as follows :

Order No.	Author	Title	Supplier	Data

If author and title forms a composite key, this relationship is not in second normal form. Note that its author and title would be repeated in each record that stores information about a part in another. If the author changed then every record of a order number would have to be updated. What would happen if there were no order number relating to author or book. Then it is possible that database would like to keep track of the author, since there would be no record having its author. The relations can be made to conform to second normal form by splitting it into two relations.

Order No.	Author	Title	Supplier	Data
	Author	Title	Supplier	Data

Ordernumber and Author could be the combined key for the relation and author can be the key for the second.

Thus normalisation is a systematic process of transforming initial conceptual design first into a set of relational table in First normal form (1NF) by assigning a unique key to each entity type table and removing repeating group from the tables by splitting each into two or more new entity types or relation tables. The set of relational tables in 1NF may be changed into set of tables in second normal form (2NF) by removing partial functional dependencies within tables. It is accomplished by splitting of these dependencies into new separate tables. The set of tables in 2NF may be converted into a new set of tables in third normal form (3NF) by eliminating transitive dependencies.

First normal form (1NF), Second normal form (2NF), Third normal form (3NF), Fourth normal form (4NF), Fifth normal form (5NF), and the highest normal form is called domain/key normal form (DK/NF).

Example of Normalisation Process

Employee

Employee Name Place of Work Child

Child name Date of Birth Sex

Employee name	Place of work	Child		
		Child name	**Date of birth**	**Sex**
Ashok	Varanasi	Suraj	12-7-1985	M
Vinay	New Delhi	Arpita	14-7-1986	F
Surendra	Allahabad	Ashish	15-10-1988	M

Representation of employee

Employee name	Place of work	Child name	Date of birth	Sex
Ashok	Varanasi	Suraj	12-7-1985	M
Vinay	New Delhi	Arpita	14-7-1986	F
Surendra	Allahabad	Ashish	15-10-1988	M

Fig 2(a) Employee in 2 NF

It can be split into two representation employee and child.

Employee
name **Place** of work

Employee Child
Employee name **Child** name **Date** of birth **Sex**

Employee name	Place of work
Ashok	Varanasi
Vinay	New Delhi
Surendra	Allahabad

Employee name	Child name	Date of birth	Sex
Ashok	Suraj	12-7-1985	M
Vinay	Arpita	14-7-1986	F
Surendra	Ashish	15-10-1988	M

Fig 2 (b) Employee in 1NF

Functional Dependencies

Dependency refers to the relationship among attributes. These attributes may belong to the same relation or different relations. Dependencies can be of various types, viz. functional dependencies, transitive dependencies, multi-level dependencies etc.

Functional dependence is a relationship that exists between any two fields or attributes. We may say that a field A of an n-set O is functionally dependent upon field B of O (written $B \rightarrow A$) if each value of B has precisely one value of A corresponding to it at every instant of time.

It can be represented as

i) $J \rightarrow K$

ii) $J \rightarrow L$

iii) $K \rightarrow J$

iv) $L \rightarrow K$

Transitive dependency is a form of intermediate dependency. For example if we have attributes or groups of attributes A, B and C such that A determines B and B determines C, i.e.

$A \rightarrow B$

$B \rightarrow C$

In the above example, we can say that there exist transitive dependency between A and O.

The multi-valued dependency refers to m:n (many to many) relationships. The multivalued dependency exist between two data items when one value of the first data item gives a collection of values of the second data item i.e., it multi-determines the second data items.

Self Check Exercise

1) Discuss the definition and functions of DBMS.

2) Explain the important components of DBMS.

3) What is relational database management system ?

Note: i) Write your answers in the space given below.

ii) Check your answer with the answers given at the end of this unit.

....

....

....

....

....

4.7 SUMMARY

The use of database management system represents one of the most significant trends in the field of computer-based information systems. A database management system (DBMS) is a collection of software for processing a collection of interrelated data known as 'database'. The objective of a database management system is to facilitate the creation of data structure and relieve the programmer of the problem of setting up complicated files. A complete DBMS separates the definition of data from the programmes that access it. With a DBMS, it is possible to design file structures much more easily and to set up a database that can be used by a number of different application programmes. In modern times RDBMS is being developed and used more and more.

4.8 ANSWERS TO SELF CHECK EXERCISES

1) The database management system (DBMS) can be defined as an integrated set of programmes that serves as interface between application programmes and a set of coordinated integrated physical files called a database. A database management system provides the capabilities for creating maintaining and changing a database as well as retrieving information from it. In other words DBMS is a series of programmes that allows the user to easily create and modify database. The database is a collection of related sets of data or data files that are structured in various ways to meet multiple processing and retrieval needs of users.

   The main functions of DBMS are:

   i) Integration

   ii) Data independence – logical and physical

   iii) Data retrieval, analysis, modification and storage

   iv) Primary

   v) Integrity controls and recovery methods

   vi) Compatibility

   vii) Concurrency support

   viii) Support of complex file structures and access paths.

2) The important components of DBMS are :

   i) Data Definition Language (DDL) It provides the mechanism for organizing and structuring the databases and defining the data elements.

   ii) Data Manipulation Language (DML). It provides the mechanism for data manipulation such as creation, updating, deletion, replacement, retrieval and storage of data or records.

   iii) Query Language – This provides the mechanism to provide answers to the user queries.

   iv) Data input system – for inputting of data

   v) Communication sub-system - to support a number of on-line users.

3) The relational database management system is a DBMS based on the relational model of data organization. It was first propounded by EF Coded of the IBM in 1970. In the relational DBMS the relations between data or entities are expressed in the form of tables.

4.9 KEYWORDS

Data : Collection of facts made up of numbers, characters and symbol stored on a computer in such a way that it can be processed by the computer.

Database : An integrated collections of files of data stored in a structured form in a large memory, which can be accessed by one or more users at different terminals.

DBMS : Database management system is a series of programmes that allow the user to easily create and modify databases.

4.10 REFERENCES AND FURTHER READING

Codd, E F(1982). Relational Database: A Practical Foundation for Productivity. *Comm ACM.* 25, 2; 109-17.

Date, C.J. (1975). An Introduction to Database System. Reading, Mass: Addison Wesley.

Dickson, Gray W and Wetherbe, James C (1985). The Management of Information System. NY : McGraw Hill.

Dodd, G. (1969). Elements of Database Management System. *Computing Survey*. vol. 1 no 7; 117-133.

IGNOU (1997). Design and Management of Databases. MLIS-06 (course material), Unit 3. New Delhi : IGNOU.

Kashyap, MM (1993). Database System: Design and Development, New Delhi: Sterling publishes.

Korth H and Silber Schatz A(1986). Database System Concepts. NY : McGraw Hill.

Lucas, Henry C (1990). Information Systems Concept for Management. NY: McGraw Hill.

Martin, James (1976). Principles of Database Management System. N.Y.: Prentice Hall.

Naveen Prakash (1984). Understanding Database Management. New Delhi : Tata McGraw Hill.