UNIT 16 DATABASE RESOURCE MANAGEMENT

Structure

16.1 INTRODUCTION

A university uses a database to keep student records, large libraries uses a database system to keep track of library and to provide various types of indexing to material by subject, title, author etc. and a bank uses database system to keep records of its customers, employees, loans etc. and many more. Businesses normally use a database to keep records of its employees, sales, production etc. All those who keep a collection of records for a common purpose, uses a database. Therefore, a database system is a collection of programs, which allows storage, modification, and manipulation of information from a database. It is a layer of software between the physical database and the user.

16.2 OBJECTIVES

After reading this unit, you should be able to:

- Describe the concept of a Database Management System (DBMS);
- Describe the importance of data in organizations;
- Explain the role of database administrator in DBMS; and
- Identify the factors responsible for the success of Data Ware Housing.

16.3 DATA AS ORGANIZATIONAL RESOURCE

Before the computer came into existence, there were many limitations associated with the physical handling of documents and human processing. Computers came into existence to speed up the data processing. Computers helped to manage data.

We are living in the age of information processing. The ability to acquire accurate and timely data, managing data efficiently, is all through which an organization succeed.

The development in hardware has catalysed rapid development in software and now we are having more sophisticated software to handle data as it was earlier.

The term data and information are often taken as to mean the same thing. Data are the details about factories, outlets, staff, competitors, customers and suppliers. Data is also kept for monitoring the activities and processes in business. Once businesses collect these details, information comes into picture. Information is the collection of meaningful facts that are derived from the data. Information is significant and relevant to the user unlike data, which has no meaning alone. Information leads to decision and thus an appropriate action. In fact, information consists of only the data that is useful, meaningful and needed. You come across many information systems in you daily life like banking information system, ticketing and reservations systems etc. Now what you get out of these information systems is pure information and these systems work out this information on the basis of data. In other words, data is raw and information is refined. The exact form of the refinement depends on the type of application one is dealing with.

The data must be accurate, timely and relevant. It is desired that any information system should have accurate data to work on. Inaccurate data howsoever well analysed will hardly be useful for any decision-making. Although collection of accurate data is costly and time consuming, every effort should be made to make it as accurate as possible. One may strike a balance between the cost of processing and value of accuracy. The data should also be timely. If the right data is not available at the right time then it is worth little use. For example if the updating of electoral lists cannot take place before the schedule of elections then it is of no use to the elections. The timeliness of the data is detrimental of success of any information system. The data should be relevant. For better understanding of the organizations and their information needs data has to be high on the relevance factor. There are many decision-support systems (DSS) and executive support systems that are present in today's scene, but the general feeling is that often the data is generating accurate and timely information, that is not very relevant. However, the future looks bright because of the emergence of high-end computing machines, sophisticated data capturing devices and other technology driven tools. Technology is making data available in larger quantities than ever before, due to lower cost storage, increased processing speeds, higher capacity communications and increased variety of information formats.

16.4 ORGANIZING DATA

In earlier times, data processing was done manually. Organizations appoint a large number of people called clerks. The information technology devices used at that time was forms, ledger books and basic mechanical adding machines. The results of such manual operations were obtained at a time when the information was almost out of date (e.g. census). Then some systems were invented in which processing was mostly mechanized e.g. Hollerith Tabulation System. In such systems data was recorded in the binary form of holes in cards, using a cardpunch. These stacks of cards could be sorted and tabulated. IBM and Remington Rand led the development of punched card technology.

Things have changed considerably with the advent of computers. There are a few terms that you need to know about the data organization. IT-specific encyclopedia of whatis.techtarget.com and searchdatabase.com defines theses terms as follows:

Bit: A bit (short for binary digit) is the smallest unit of data in a computer. A bit has a single binary value, either 0 or 1. Although computers usually provide instructions that can test and manipulate bits, they generally are designed to store data and execute

instructions in bit multiples called bytes. In most computer systems, there are eight bits in a byte.

Byte: In most computer systems, a byte is a unit of data that is eight binary digits long. A byte is the unit most computers use to represent a character such as a letter, number, or typographic symbol (for example, “g”, “5”, or “?”). A byte can also hold a string of bits that need to be used in some larger unit for application purposes.

Field: A field is an area in a fixed or known location in a unit of data such as a record, message header, or computer instruction that has a purpose and usually a fixed size. In some contexts, a field can be subdivided into smaller fields. In a database table, a field is a data structure for a single piece of data. Fields are organized into records, which contain all the information within the table relevant to a specific entity. For example, in a table called customer contact information, telephone number would likely be a field in a row that would also contain other fields such as street address and city. The records make up the table rows and the fields make up the columns.

Record: In a database, a record (sometimes called a row) is a group of fields within a table that are relevant to a specific entity. For example, in a table called customer contact information, a row would likely contain fields such as: ID number, name, street address, city, telephone number and so on.

File: In data processing, using an office metaphor, a file is a related collection of records. For example, you might put the records you have on each of your customers in a file. In turn, each record would consist of fields for individual data items, such as customer name, customer number, customer address, and so forth.

Database: A database is a collection of information that is organized so that it can easily be accessed, managed, and updated. In one view, databases can be classified according to types of content: bibliographic, full-text, numeric, and images.

In relation to database, an entity means a person, place, or thing that we wish to collect information on (e.g. customers). The word root is from the Latin, *ens*, or being, and makes a distinction between a thing’s existence and its qualities. Attribute is a characteristic of an entity (e.g. customer’s salary, customer’s address).
In a database management system (DBMS), an attribute may describe a component of the database, such as a table or a field, or may be used itself as another term for a field. Key Field is a special field that uniquely identifies a single record (*e.g.* customer’s registration number). It can be a collection of fields.

In a database management system (DBMS), files are either organised sequentially (one record after another, used for batch processing) or organised in an indexed sequential form (in sequence, but records can be directly accessed using an index) or in the form of direct or random form (records located using a key field generated by a mathematical formula.)

16.5 DATABASE MANAGEMENT SYSTEMS AND ITS COMPONENTS

We have discussed earlier that an organization must have accurate and reliable data for effective decision-making. For this, the organization maintains records by combining the data from different sources in an organization. Database is a collection of related information stored along with the details of interpretation of the data contained. For managing the data in the database we need a system called Database Management System. In other words, DBMS is a complex piece of software that

facilitates a flexible management of the data. Through DBMS we can access, monitor, store and modify the database. Through DBMS data can be made available to all users and redundant (duplicate) data can be minimized or completely eliminated. DBMS also makes possible for an organization to prevent important data access from unauthorized users by providing the security to the database at different levels. Some of the DBMS that are used are INGRES, ORACLE, SQL Server and SYBASE.

The DBMS allows users to access data from the database having no knowledge of how data is actually stored in it. The process is much the same as ordering a menu in the restaurant. A customer simply orders for the food to a waiter and waiter serves the specified order. A customer only checks the menu for the desired items and need not know how the items are arranged in a restaurant. Similarly, the database user need not know how the data is stored instead he needs to know only what he requires and the DBMS takes care of retrieving the required data on its own.

Lets look into the components of DBMS:

DML Precompiler (Data Manipulation Language Precompiler): As its name specifies, this is a compiler that converts DML statements (Statements that allows the users to manipulate the database) in an application program into normal procedure calls in the host language. This precompiler must interact with the query processor (discussed later) in order to generate the appropriate code.

DDL Compiler (Data Definition Language Compiler): The DDL compiler converts the data definition statements (statements that define the tables, database etc.) into a set of tables. These tables contain information in the form that can be used by other components of the DBMS.

File Manager: File manager manages the structure and space of the file on disk. This also locates the block in which the required record exists and requests from the disk manager for this block containing the required record and finally provides the required record to the data manager.

Disk Manager: As specified earlier, disk manager provides the block or page that the file manager asks for. The disk manager is a part of Operating System. Disk Manager does all the physical input and output operations.

Data Manager/Database Manager: It is also called database control system. It is a software component of the DBMS or we can say a program module that acts as an interface between the data stored in the database and the queries submitted to the system. Data Manager converts the queries from the user into the file system. It is actually responsible for the storage, retrieval and manipulation of the data in the database. Besides these responsibilities data managers also takes care of maintaining integrity. This means that the appropriate data should be stored in the database. For example, date of joining of an employee should not be less than the current date or the age of an employee for a particular post should lie between 25 and 40. If these constraints are specified in the database, then database manager takes care of them on its own. Database Manager also checks for an unauthorized user accessing the database. Database Manager does not allow any user to access the database without having the permission granted.

Query Processor: The Query Processor takes care of the queries by the database user. The database user when uses the Data Manipulation Language (DML) for retrieving the data, the Query Processor converts it into that form that could be sent to the Data Manager, so that Data Manager can execute it.

Data Dictionary: A data dictionary contains the information/ data about the data. A data dictionary includes all the database descriptions, entity/ table descriptions, attributes etc. Apart from these data dictionary includes the information about the users also, which tables are used by which programs, authorization of users. A data directory/ dictionary is just like a database and its cost depends on how complex it is? More complexity leads to increase in cost. Just like an index helps us to find the chapter easily, data dictionary helps DBA for designing, implementing and maintaining the database and searching for the desired record in the database. It also helps the managers and end users in their project planning.

16.6 MODELING DATA

As we have discussed earlier that any organization needs to store the information and this is done through database. Data models refer to the conceptual model of the data and the underlying relationships among them. DBMS abstract some generic structures to represent conceptually every possible file structure.

Data models can be classified in two classes viz; Record-based logical Models and Object-based logical models. Record-Based logical data models can be classified (Sadgopan, 1997) into the following categories:

Hierarchical Models: These are the early data models used in 1970's. Hierarchical models capture the intuitive hierarchy of the data elements. The early generation of large DBMS e.g. IMS belongs to the hierarchical data models. Even today some large databases are maintained on IMS platform.

Network Models: Since hierarchical models are unable to represent data items that existing at two different level of hierarchy, network models were proposed. The notable systems built using this model were ADABAS and DBMS-10 on DEC-10 machines.

Relational Models: Though network models were quite powerful, they lacked in elegance. The systems built on this data model were dBase, Xbase and ORACLE. Almost all commercial systems presently available like Oracle 8i, 9i, 11 etc., SQL Server, MySQL are built on the relational models. There are 12 rules that are required to be followed in a relational model.

Data modelling is achieved in two levels:

1) E-R modelling that builds the conceptual model of the data.

2) Normalization, which removes the redundancies.

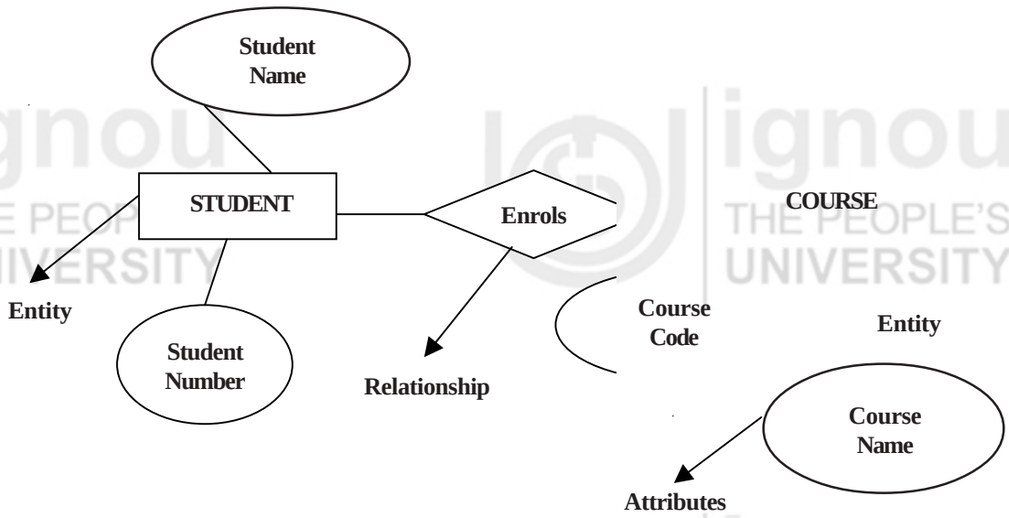
Object Oriented Logical Models can be of several types. One of them is Entity Relationship model. The Entity Relationship (ER) data model is the most common data model which is based on the perception of the real world.. This model allows us to represent the relationships among the objects called entities. It uses following three concepts to represent itself graphically. Although we have discussed them earlier, we revisit them here once again:

Entity: Any real-world object that has certain properties (attributes of its own) and this object are uniquely identified by the system on the basis of these. This is represented by a rectangle with the entity name specified in the centre.

For *e.g.*: student is an entity.

Attribute: Attribute is the properties of an entity, like a student entity can have attribute as student name, student roll no., student class etc. Attributes are placed inside the circles and attached to the entities and relationships.

Relationship: Relationship specifies the meaningful relation between two entities. This relationship can also have attributes. A rhombus represents these with a relation specified in it. For *e.g.*: if we have student and course as entities, then we can relate these entities as student enrols for a course.



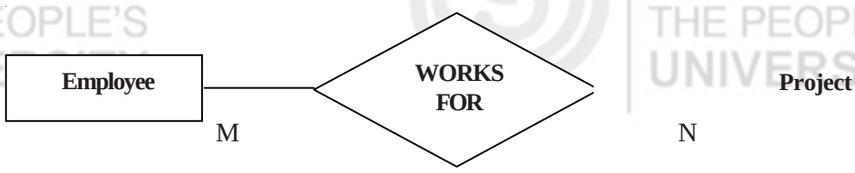
Degree of Relationship (DOR): This specifies the occurrence of entity with other entity.

One to One (1:1): In this, there can be almost one related occurrence for each entity. For example, A single manager manages one department and one department can be managed by a single manager. It is represented as follows.

One to Many (1:N): Here, for one occurrence of the first entity, there may exist many occurrences of the second entity and for every occurrence of the second entity, these exists only one occurrence of the first entity *e.g.*:

Here one manager supervises many employees and every employee reports to only one manager.

Many to Many (M: N): In this degree of relationship, for one occurrence of the first entity, these may exists many occurrences of the second entity and also for every occurrences of second entity, there exists many occurrences of the first entity *e.g.*:



One employee can work for many projects and many employees can handle one project.

Normalization is another concept in data modelling. Normalization is a process of converting complex data into simpler form without any loss of information. The Normalization technique ensures that there is no dependent and duplicate data when the E-R model has been made. Normalization first of all, converts the data into tables or relations. These tables are checked for redundancy. Finally the form is converted to a database definition. Why do we need normalization? Normalization improves the database design. Whenever the design is modified, we need to re-organize the data. Normalization reduces the need of reorganizing the data. Process of normalization removes all undesirable consequences that may occur due to inserting, updating and deleting values from the tables. Normalization reduces the unnecessary repetition of data (redundancy) that causes the problem with storage and retrieval of data. Poorly designed databases have several data management difficulties *e.g*.

Consider a relation

LIBRARY (MNAME, MADD, B_TITLE, ISS_DATE, DUE_DATE)

This relation has the following problems:

1) Here duplicate data exists because member address will be repeated if a member issues more than one book.
2) **Insertion Problems/ Anomalies:** If a member does not issues a book then we cannot record the address of a member. Here MNAME & B_TITLE are both necessary and hence we cannot leave any one of them as blank.
3) **Updating Problem:** If a member has issued two or more books, then address is necessary in all and if a member's address needs to be updated, then all the records needs to be updated.
4) **Delete Anomalies:** Here if we delete the issue details, then address would not be there.

But all the above problems can be eliminated by normalization. If this single relation is broken down in two relations:

MEMBER_DETAILS (MNAME, MADDR)

MEMBER_TRANS (MNAME, B_TITLE, ISS_DATE, DUE_DATE)

But this requires a join between these two relations, which is very expensive.
Now lets look into several normal forms with the help of an example:

Table 16.1: Unnormalized Data

Roll No	Name	Subject	Marks
1	Amit	English	65
		Hindi	72
		Maths	70
2	Seema	English	70
3	Anjali	English	54
		Hindi	60

To make this unnormalized data into 1 NF (First Normal Form), we have to make each cell containing one value. And all the repeated information should be removed.

Table 16.2: First Normal Form

Roll No.	Name	Subject	Marks
1	Amit	English	65
1	Amit	Hindi	72
1	Amit	Maths	70
2	Seema	English	70
3	Anjali	English	54
3	Anjali	Hindi	60

The data is still redundant in this form. Here combination of two keys (composite key), Roll No. and Subject, is a primary key (PK, that uniquely identifies a record). But the attributes of this table depend on the part of the PK. Roll No. and subject determines marks, Roll No. determines Name, Name has no dependency on the attribute subject. This may lead to following problems:

Insertion: the name of a subject cannot be recorded until a student gives any exam.

Updation: Roll No. and name is repeated several times. If name of the student is to be changed, then at every place the name needs to be updated otherwise may lead to inconsistencies.

Deletion: if a student has not given any of the three exams, then the name of that particular student will be deleted.

So, for overcoming these problems, this first normal form needs to be decomposed and converted into Second Normal Form (2NF) without any loss of information.

Table 16.3: Relation/Table in 2NF

Roll No.	Name	Roll No.	Name	Subject	Marks
1	Amit	1	Amit	English	65
2	Seema	1	Amit	Hindi	72
3	Anjali	1	Amit	Maths	70
		2	Seema	English	70
		3	Anjali	English	54
		3	Anjali	Hindi	60

A relation is said to be in 2NF, only if it is in 1NF and every attribute is dependent on the whole key and not just a part of it.

Other major normal forms are 3NF and 4NF. Third normal form or 3NF is used to prevent loss of information and dependencies preserving decomposition. Fourth normal form or 4NF is used to preserve multi-valued dependency, which is essentially a constraint. There are several other normal forms like project-join NF, domain-key NF. We are not describing them as it is not within the scope of this course.

Activity A

Prepare a comparative chart of the strengths and weaknesses of two competing DBMS product in the market e.g. dBase and Oracle.

.....

.....

.....

.....

.....

16.7 DATA TYPES

In database system, for storing the data, each column in the table have a certain type of data. For example when we create a table for employee's personal data, then we may have different fields like name, address, phone, age etc. In this if we take the field "name" then this would of character type and "age" and "phone" are of numeric type.

Sometimes an attribute of a table T takes the values from a set D that is known as a domain. A domain is defined in the cases when we want an attribute to take some exact set of constants. For instance, the domain of the attribute age for a particular post that a candidate can apply for, might be defined as consisting of all the integer numbers between 18 and 25. On the other hand, the domain of the marks column, might be defined as all the numbers between 0 and 100, with at most one non-zero digit after the decimal point (assuming the maximum marks to be 100). The significance of declaring a column of a table to have a particular type (or domain) is the ability to compare the values of two different column rests on this declaration. A particular column I a table must contain similar data, which is of a particular type. Besides storing character, numeric, we can also store data, binary data such as graphics, sound also. Whenever we have the bio data of employees or students or candidates, sometimes it is nice to store their photographs also, in that cases, we can use graphics as a data type for photograph column.

16.8 DEVELOPMENTS IN DATABASE TECHNOLOGY

DBMS consists of a collection of interrelated data that is called a database and a collection of programs to manage and access the data. An RDBMS is a collection of tables, each of which is assigned a unique name. Each table consists of a set of fields and stores a large set of records. A data warehouse is a repository of information collected from multiple sources, stored under a unified schema that resides at a single site. There have been a lot of developments in the field of database technology.

These developments include handling of spatial data such as maps; engineering design data such as design of buildings, components and circuits; multimedia data such as text, image, video and audio data; time related data such as historical records or stock exchange data and the World Wide Web data. Han & Kamber (2001) had described each of these advanced databases:

1. **Object Oriented Databases:** In object oriented databases each entity is considered as an object. These objects can be customers, suppliers, employees or items. Objects that share a common set of properties can be grouped into a class that represents properties common to a class.
2. **Object Relational Databases:** In object relational data model much is the same as the basic relational data model but for the fact that object relational databases also handles complex data types and complex object structures.
3. **Spatial Databases:** These databases contain geographic databases, clip design databases, medical and satellite image databases. These data are represented in raster format, consisting of n-dimensional bit maps or pixel maps as well as vector format, consisting of basic geometric shapes such as points, lines and polygon etc.
4. **Time Series & Temporal Database:** These databases contain time-related data. A temporal database usually stores relational data that include time-related attributes. A time series database stores sequences of values that change with time.
5. **Text & Multimedia Databases:** Text databases contain word description for objects (these could be long sentences or paragraphs). Multimedia databases store images, audio and video data. These databases are used in applications such as picture content-based retrieval, voice mail systems, video-on-demand systems & World Wide Web etc.
6. **Heterogeneous Databases:** Objects in one component database may differ greatly from objects in other component databases, making it difficult to assimilate their semantics into the overall heterogeneous database. There are special methods for analysis of such databases.
7. **World Wide Web Databases:** World Wide Web has data objects linked together to facilitate interactive access. Users seeking information of interest traverse from one object via links to another.

Activity B

Now that you know about various types of data, you can encounter in day-to-day life. Take the case of an organization of your choice and try to identify two examples of each of these data types. Also, comment on the usefulness of analyzing these data types for the organization

.....

.....

.....

.....

.....

.....

.....

16.9 DATABASE ADMINISTRATOR AND THE ROLE

Large and complex databases need a lot of maintenance to work in a proper manner. A database administrator (DBA) does this maintenance. DBA is the supervisor of the database and hence have the highest level of expertise. DBA centrally controls the database and administers the overall user community. DBA also takes care of the data and controls the database structures. DBA is responsible for accessing the database so that an unauthorized person should not access the database. DBA stores the permissions granted to the users created to verify that a particular user can only perform a given operation on the database. DBA can also revoke the permission if he founds that the user or user group needs no more access to that particular portion of the database.

DBA account is sometimes known as the system account that has the powerful capabilities that other regular database users don't have. Sometimes when more than one user uses the same resource such as the data, there may exists the conflict but the DBA seeks for the compromise and takes care of the conflict for the benefit of the organization. Hence, DBA also has the responsibility for maintaining the integrity of data. This means that only accurate data can be inserted into the database.

Duties of DBA includes the existence control, definition control, quality control, update control, access control etc. The DBA is also responsible for recovering the database from failures that may be due to human disasters, software and hardware causes or natural causes with the minimal loss of data. DBA changes the database according to the changes or growth in the organization. Hence, DBA is said to be a super user who is responsible for monitoring and improving the performance and operations to maintain the database activities. The database dictionary is one of the most important tools for the database administrator. It is used to maintain information related to various resources.

16.10 CRITICAL SUCCESS FACTOR OF DATA WARE HOUSING

In 1990, Bill Inmon first introduced the term Data Warehouse. According to him, "A warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision-making process'. You will read in detail about data warehousing in the next unit. A data warehouse is a central repository for significant pieces of the data that are collected by various business systems. Data from various online transaction processing applications and other sources is organized on the data warehouse for use decision-making process.

A data warehouse is used in applications like data mining, Web mining and a decision support system (DSS). Data Warehouse enables the executive, manager or the analyst to make better and faster decisions. It presents the right information in the right place at the right time with the right cost in order to support the right decision. According to Inmon, a data warehouse should firstly, have a common format meaning that data that goes into a warehouse should have a common data format. Since a data warehouse is made out of a combination of various organizational databases of the company, every care should be taken to maintain uniformity. Secondly, since different company functions such as Sales, Marketing, Finance, and Production have some variables common in a database.

A data warehouse should be able to use these data pertaining to different functions. Thirdly, since data often contain time as a dimension and is updated into a data warehouse periodically, a data warehouse should be efficient in computing all time dependent aggregations again and again.

Finally since data warehouses are never deleted or updated they should be able to sustain large chunks of data (may be in terabytes). A data warehouse grows all the time as new data keeps on adding. In practice, a data warehouse is often "resynchronized" to keep its volume manageable. Resynchronization means that data is kept in periods and as and when a new period is added the oldest one is deleted.

These are some very preliminary concepts about data warehousing. You will read about it in more detail in the next unit.

16.11 SUMMARY

Data consists of facts and details about things, activities, transactions etc. Data can be stored as a resource that can be drawn later on to produce information for the people and activities that need it. In this unit we have discussed about the prominent features of a modern database management system. Understanding a DBMS is very significant for a manager dealing with information systems. You have learnt about modelling data and different data types. You saw the developments in the field of database technology. This shows that almost any kind of data can be handled, be it a music database or a database of photographs stored on the computer. The unit closes with giving you a glimpse of success factors in data warehousing, which you would read in detail in the next unit.

We have limited our scope in the sense that we would not like to make you a programming expert but would be more interested in making you to capable of applying these concepts for deriving business advantage.

16.12 UNIT END EXERCISES

1) What do you understand by "data" and "information"?
Give three examples each to distinguish between data and information.

2) Explain the meaning of the terms database, database management system and database administrator in your own words.

3) Explain what do you understand by database management systems? Describe its components.

4) What are data models? How you would classify the data models? Differentiate between various data models with the help of an example.

5) What is normalization and why it is used? Explain with the help of an example how would you derive normal forms.

6) Discuss different types of data in database system. What are the developments in database technology with respect to data types?

16.13 REFERENCES AND SUGGESTED FURTHER READINGS

Combs, Martin R (1995), *Information Systems for Business Management*, Pitman Publishing, 128, Long Acre, London

Han Jiawei & Kamber Micheline (2001), *Data Mining: Concept and Techniques*, Morgan Kaufmann, Elsevier Science, India

Hicks Jr, James O. (1990), *Information Systems in Business: An Introduction, Second Edition*, West Publishing Company, St. Paul, MN

http://whatis.techtarget.com & http://searchdatabase.techtarget.com/ for definitions of terms. Site accessed on 19/05/2004.

Kroth Henry F, Silberschatz A(2001), *Database System Concepts*, McGraw Hill International Edition.

Sadgopan S. (1997), *Management Information Systems*, Prentice-Hall of India Private Limited, New Delhi, India