BACHELOR OF COMPUTER APPLICATIONS

SEMESTER 3

DCA2102

DATABASE MANAGEMENT SYSTEM

Unit 1

Database Management System Concepts

Table of Contents

1. INTRODUCTION

In this unit, we will introduce the basic concepts of DBMS. Database technology development evolved rapidly in the three decades since the rise and eventual dominance of relational database systems. While many specialized database systems (spatial, object-oriented, multimedia, etc.) have found substantial user communities in the science and engineering sections, relational systems remain the preferred database technology for business enterprises.

A database is a collection of related information stored so that it is available to several users for several different purposes. The content of a database is obtained by combining data from all the different sources in an organization so that data are available to all users and replicated data can be minimized or eliminated. A computer database gives us an electronic filing system, which has a large number of ways of cross-referencing and this allows the user several different ways in which to retrieve and reorganize data.

1.1 Objectives:

By the end of Unit 1, the learners should be able to:

- *Understand the definition of database and database management system*
- *Study the applications of database system*
- *Understand modeling for a database*
- *Understand the advantages and disadvantages of DBMS*
- *Differentiate between DBMS and RDBMS*

2. SIGNIFICANCE OF DATABASE

The **database** is a collection of related data. A data item is the smallest identified unit of data that has value in the real world – for example, last name, first name, street address, ID number, or political party – and is the fundamental component of a file in a file system. A group of related data items considered as a single unit by an application is called a *record*. Examples of types of records are salesperson, customer, order, product,and department. A *file* is a collection of several records of a single type.

A database has the following properties:

- A database represents an aspect of the real world and is sometimes called the universe of discourse (UoD) or mini-world. The database keeps track of changes to the mini-world.
- A database is a logically organised collection of data that has some meaning..
- A database is designed, built, and populated with relevant data for a specific purpose.

A *database* is a more complex object; it is a collection of interrelated stored data that facilitates the requirements of several users within one or more organizations, that is, interrelated collections of many different types oftables. The encouragement for using databases rather than files includesless redundancy of data, greater availability to a diverse set of users, and integration of data for easier access to and updating of complextransactions.

The basic definitions of some database concepts are:

Data: Data is a collection of known facts that may be recorded and has an underlying significance.

Database: It is a collection of related data.

Database System: It is the DBMS software together with the data itself.Sometimes, the applications are also included.

Database Management System (DBMS): It is a software package/systemto facilitate the creation and maintenance of a computerized database.

Examples of database management systems software are ORACLE, SQL Server, MS Access, DB2, SYBASE, INFORMIX, etc.

A database can handle accounting and filing, and business inventory, and use the information in its files to prepare summaries, estimates, and other reports. There can be a database that stores books, newspaper articles, magazines, and comics. On practically every issue, there is already a well-defined market for specialised knowledge for a small number of customers.

The management of data in a database system is handled by a database management system, which is a general-purpose software programme. The database management system is the major software component of a database system. Therefore a database management system is a combination of hardware and software that can be used to set up and monitor a database and can manage the retrieval andupdating the database that has been stored in it. The majority of database management systems have the following capabilities:

- Creating of a file, addition to data, modification of data, deletion of data,creation, addition, and deletion of files.
- Retrieving data selectively or collectively.
- The data can be sorted or indexed at the user's discretion and direction.
- Several reports can be generated from the system. These may be eitherstandardized reports or specifically generated according to specific user requirements.
- Mathematical functions can be performed on data stored in the database and can be manipulated using these functions to perform the desired calculations.

Self-Assessment Questions – 1

1. The basic component of a file in a file system is a_____.
2. UoD stands for_____.
3. Database Management System is a_____system to facilitates thecreation and maintenance of a computerized database.
4. The database management system is the major software component of a _____.

3. DATABASE SYSTEM APPLICATIONS

Databases are widely used. Here are some of the representative applications:

- **Universities:** For student information, course registrations, exam, and grades.
- **Credit card transactions:** For purchases on credit cards andgeneration of monthly statements, and payments.
- **Banking:** For accounts, and loans, customer information, and bankingtransactions online/offline.
- **Finance:** For storing information regarding financial instrument holdings, sales, and purchases, such as stocks and bonds.
- **Sales:** For the customer, product, and purchase information.
- **Telecommunication:** To keep track of incoming and outgoing calls, monitor prepaid calling card balances, generate monthly bills, and store information about communication networks..
- **Manufacturing:** For inventories of items in warehouses/ stores,management of supply chain, and tracking the production of items in factories, and orders for different items.
- **Human resources:** For information about employees, payroll taxes and benefits, salaries, and for generation of paychecks.
- **Airlines:** For reservations, cancellations, and schedule information. Airlines were among the first to use databases in a geographically distributed manner terminals situated across the world accessed the central database system through some data networks.

So we can say that databases form an essential part of almost all enterprises today. During the last four decades of the twentieth century, the useof databases grew in all enterprises. Few people engaged directly with database systems in the early days, but they did so indirectly — through printed reports such as credit card statements, or through agents such as bank tellers and airline reservation agents – without recognising it. Then automated teller machines came along and let users interact directly with databases. Phone interfaces to computers also facilitated users to deal with databases directly. For example – a caller could dial a number, and press phone keys to enter information or selectalternative options, to find flight arrival/departure times.

During the late 1990s, the internet revolution sharply increased direct user access to databases. Organizations converted many of their phone interfaces to databases into Web interfaces and facilitates a different kinds of services and information available online. For example, when you access a Web site, information about you may be retrieved from a database, to selectwhich advertisements should be shown to you.

When you explore a book or music collection in an online retailer, you are accessing data stored in a database.

When you place an online order, it is saved in a database.

When you go to a bank's website to get your account balance and transaction history, the data is pulled from the bank's database system. Moreover, data related to Web access may be stored in a database. Thus, although user interfaces hide details of access to a database, and most people are unaware they are dealing with a database, accessingdatabases forms an essential part of almost everyone's life these days. The importance of database systems can be judged in another way – today, database system vendors like Oracle are among the largest software companies in the world, and database systems form an essential part of the product line of more diversified companies like IBM and Microsoft.

Self-Assessment Questions – 2

5. _____ were among the first to use databases in ageographically distributed manner.
6. Web accesses may be stored in a_____.
7. The_____machines came along and let users interactdirectly with databases.
8. In_____database is used for keeping records of calls made,generating monthly bills, maintaining balances on prepaid calling cards,and storing information about the communication networks.

4. DATA INDEPENDENCE

We can define data independence as the capability to modify the schema definition at one level without disturbing a schema definition at the next higher level. It is usually understood from two points of view: **physical data independence** and **logical data independence**. Physical data independence permits changes in the physical storage devices or organization of the files to be made without causing changes in the conceptual view or any of the external views and hence in the application programs using the database. Thus, the files may migrate from one physicalmedia to another or the file structure may change without requiring any changes in the application programs.

Logical data independence refers that application programs need not be changed if fields are added/deleted to/from an existing record. Logical data independence indicates that the conceptual schema can be changed without affecting the existing external schemas. In a database environment, data independence is advantageous, as it allows for changes at one level of the database, without requiring any change in other levels. The mappings between the layers absorb these changes.. Since application programs are heavily dependent on the logical structure of the data they access, so it is more typical to achieve logical data independence than physical independence.

In many respects, the concept of data independence is similar to the conceptof abstract data type in programming languages like C++. Both hide implementation details from the users. This facilitates users to concentrate on the general structure rather than low-level complex implementation details.

Self-Assessment Questions – 3

9. Data independence is usually considered from_____points of view.
10. _____data independence allows changes in the physicalstorage devices.
11. _____ data independence implies that application programs need not be changed if fields are added to an existingrecord.
12. Logical data independence is more _____ to achieve than physical independence.

5. DATA MODELING FOR A DATABASE

The **data model** is considered one part of the conceptual design process. The other is the **function model**. The data model emphasizes what data should be stored in the database whereas the function model deals with how the data is processed. In the context of the relational database, the data model is used to design the relational tables, whereas the functional model is used to design the queries that will access and perform operations on those tables.

Data modeling is preceded by planning and analysis. The effort done to this stage is proportional to the scope of the database. The planning and analysis of a database help to serve the needs of an enterprise and requiresmore effort than one intended to serve a small workgroup.

5.1 Entities and Their Attributes

The entity-relationship (E-R) data model is based on a consideration of the real-world that consists of a set of basic objects called ***entities***, and of ***relationships*** among these objects. It was developed to facilitate database design by considering the specification of an ***enterprise schema***, which represents the overall logical structure of a database. The E-R data model is one of numerous semantic data models; the semantic aspect of the model is based on the attempt to reflect the data's meaning.. The E-R model is very much useful in mapping and interactions of real-world enterprises onto a conceptual schema. Due to this utility, many database-design tools draw on concepts from the E-R model. More about the E-R model is explained in Unit 3 (Entity-Relationship Model).

Entities

These are the main data objects about which information is to be collected; they usually denote a person, place, thing, or event of informational interest.A particular occurrence of an entity is said to be an ***entity instance*** or sometimes an ***entity occurrence***. Suppose a Company database; here department, division, project, skill, employee, and location are all examples of entities. In Unit 3, you'll obtain more examples of entities (Entity- Relationship Model)

Attributes

These are characteristics of entities that provide details about them.

An attribute value is a unique instance of an attribute within an entity or relationship.

Employees may have attributes such as emp-id, emp-name, phone-no, fax-no, job-title, emp-address, and so on.

The attribute has a link to the entity it describes.

There are two types of attributes: **identifiers** and **descriptors**. An identifier (or key) is used to uniquely identify an instance of an entity also known as a **key attribute**; a descriptor (or monkey attribute) is used to specify a non-unique characteristic of a particular entity instance. Both identifiers and descriptors may consist of single or composite attribute.

For example, an identifier or key attribute of an Employee is emp-id, and a descriptor of an Employee is emp-name or job-title. Strong entities have internal identifiers that uniquely determine the existence of entity instances, but weak entities derive their identity from the identifying attributes of one or more "parent" entities. Weak entities are often displayed with a double-bordered rectangle, which denotes that all occurrences of that entity dependon an associated (strong) entity for their existence in the database. More about attributes is explained in Unit 3 (Entity-Relationship Model).

5.2 Relationships and Relationship Types

Relationships

Relationships represent associations among one or more entities and have no physical or conceptual existence, other than that which depends upon their entity associations. Degree, connectedness, and existence are all concepts used to characterise relationships.. The most common meaning associated with the term *relationship* is indicated by the connectivity between entity occurrences: one-to-one, one-to-many, and many-to-many.

The relationship construct is a diamond box that connects the associated entities. The relationship name can be written inside or just outside the diamond box.

A role is the name of one end of a relationship when each end needs a distinct name for clarity of the relationship. The individual duties of each entity in the relationship are clearly defined

by the entity names paired with the relationship name. However, in some cases, role names should be used to clarify ambiguities. Role names are generally nouns. More information about relationships and their types is given in Unit 3 (Entity-Relationship Model).

Relationships Types

The contents of a database must conform to specific limitations defined by an E-R enterprise schema.. Cardinality ratios or mapping cardinalities shows the number of entities to which another entity can be associated through a relationship set.

For a binary relationship set R between two entity sets A and B, the mapping cardinality must be one of the following:

- One-to-one (1:1): Each entity in A is linked to only one entity in B, and each entity in B is linked to only one entity in A.
- A one-to-many (1:N) relationship exists between an entity in A and any number of entities in B.
  However, an entity in B can only be linked to one entity in A.
- Many-to-one (N:1) means that an entity in A is linked to only one entity in B.
  However, one entity in B can be linked to any number of entities in A.
- Many-to-many (M: N"): An entity in A is linked to an unlimited number of entities in B. A single entity in B can be linked to any number of entities in A

Self-Assessment Questions – 4

13. The_____is one part of the conceptual design process.
14. Data modeling is preceded by_____and analysis.
15. The E-R data model is based on a perception of real-world that consistsof a set of basic objects called_____.
16. A particular occurrence of an entity is called an *entity*_____.
17. An entity in A is associated with at most one entity in B, and an entity Bis associated with at most one entity in A, this type of relationship is _____ .
18. In one-to-many relationships, an entity in A is associated with any number of entities in B. An entity in B, however, can be associated with most _____entities in A.

6. ADVANTAGES AND DISADVANTAGES OF DATABASE MANAGEMENT SYSTEM

Out of several advantages, one of the main advantages of using a database system is that the organization can exert, through the DBA, centralized management and control over the data.

The centralised control is focused on the database administrator.

Any application that necessitates a change in the structure of a data record must make prior arrangements with the DBA, who will make the appropriate changes.

The following are some of the most significant advantages of DBMS:

6.1 Advantages

Sharing Data

A database permits the sharing of data under its control by any number of users or application programs.

Data Redundancy Reduction

Centralized control of data by the DBA avoids undesirable duplication of data and effectively reduces the total amount of data storage required.

It also reduces the need for further processing to find the relevant data in a huge amount of data.

Further

Another benefit of minimising duplication is that it eliminates discrepancies that are common in redundant data files.

Any DBMS redundancies are managed, and the system assures that numerous copies of the same data are consistent.

Data Integrity

Centralized control can also ensure that sufficient checks are incorporated in the DBMS to facilitate data integrity. When we talk about data integrity, we're referring to the fact that the data is accurate.

The database's information is both accurate and consistent.

As a result, data values entering for storage could be double-checked to ensure that they are within a certain range and are in the correct format. For example, consider the value for the age of an employee may be in the range of 16 and 75. Another integrity check that should be incorporated into the database is to ensure that if there is a reference to a certain object, that objectmust exist. In the case of an automatic teller machine, for example, a user isnot allowed to transfer funds from a nonexistent saving account to anexisting one.

Data Security

Data is of vital importance for any organization and may be confidential. Such confidential data must be protected from unauthorized persons. The DBA, as the owner of the data in the DBMS, can ensure that suitable access protocols are followed, such as proper authentication schemas for DBMS access and additional checks before granting access to sensitive data.

For different types of data and procedures, multiple levels of security could be established. The enforcement of security could be data value-dependent (e.g., a manager has access to the salary details of employees in his or her department only), as well as data-type dependent (e.g. the manager cannot access the medical history ofany employees, including those in his/her department).

Conflict Resolution

Since the database is under the control of the DBA, she/ he should resolve the conflicting requirements of various users and applications. In other words, the DBA selects the appropriate file format and access mechanism for response-critical applications while enabling less important apps to continue to use the database with a delayed response time.

Disadvantages

The expense of the DBMS system is a big disadvantage.. In addition to the cost of purchasing or developing the software, the hardware has to be upgraded to facilitate the extensive programs and the workspaces required for their storage and execution. The processing overheadintroduced by the DBMS to implement security, integrity and sharing of the data results in a degradation of the response and throughput times. Furthermore, an additional

cost is that of migration from a traditionally separate application environment to an integrated one.

While centralization reduces duplication, the lack of duplication requires thatthe database be sufficiently backed up, so that, the data can be recoveredin the case of failure. In a DBMS setting, backup and recovery operations are fairly difficult, and this is worsened in a concurrent multi - user database system.Moreover, a database system needs a certain amount of controlled redundancies and duplication to allow access to related data items.

Due to centralization, the data is accessible from a single source namely the database.

As a result, the severity of security breaches and disruptions to the organization's operations due to downtime and failures is increased.

Many of the problems caused by downtime and failures are reduced when a centralised database is replaced with a federation of independent and cooperating distributed databases.

Self-Assessment Questions – 5

19. The database administrator is the focus of the_____control.
20. Any redundancies that exist in the DBMS are controlled and the systemensures that these multiple copies are_____.
21. _____ means that the data contained in the database is bothaccurate and consistent.
22. Data is of_____importance to an organization and may beconfidential.
23. A significant disadvantage of the DBMS system is_____.

7. DBMS Vs RDBMS

Currently, the market-leading DBMS products are all SQL DBMS products. They were originally based on the relational database management model - but did not implement the model completely accurately. First of all, the SQL DBMS permits the data to be queried based on any column in any table. Relational/SQL data is easier to query than CODASYL, hierarchical, or some other model.

Secondly, since the relational model is based on set theory its usefulness and accuracy have a basis in mathematics. A basis in mathematics thatindeed is centuries old and proven.

Thirdly, a relational database describes data in terms of its natural structure only – which means, it excludes all details having to do with machinerepresentation. The comparison between DBMS and RDBMS is shown in table 1.1.

Table 1.1: DBMS Vs RDBMS

Sr. No.	DBMS	RDBMS
1.	It is Introduced in the 1960s.	It is Introduced in the 1970s.
2.	It takes more time to fetch data from a complex and large amount ofdata set.	It is comparatively faster because of its relational model.
3.	It is used for applications thatuse a small amount of data.	It is used for huge applications which use a complex and large amount of data.
4.	Managing DBMS is difficult due tohigher data redundancy.	Managing RDBMS is easier by avoiding data redundancy.
5.	Some examples are dBase,Microsoft Acces, and FoxPro.	Some example systems are SQLServer, Oracle, and MySQL.

Self-Assessment Questions – 6

24. Relational/SQL data is_to query than hierarchical,CODASYL, or some other model.
25. The relational model is based on _____ its accuracy and usefulness and has a basis in mathematics.

8. SUMMARY

A database system is a collection of related files along with information about their definition, interpretation, maintenance, and manipulation. A DBMS is a vital software component of the database system. It consists of the collection of interrelated data and programs to access that data. The main goal of a DBMS is to provide an environment that is both efficient and convenient to use in retrieving desired information from and storing information in the database.

The DBMS not only makes the integrated collection of reliable and accurate data available to multiple applications and users but also prohibits unauthorized users to access the data. The DBMS has its advantages and disadvantages. DBMS is different from RDBMS.

9. TERMINAL QUESTIONS

1. List out the database implicit properties.
2. What are the representative applications of Databases? List them.
3. Differentiate between physical data independence and logical dataindependence.
4. What are entities and attributes? Give one example.
5. What are relationships? Explain the relationship types.
6. Explain the Advantages and Disadvantages of the DBMS.
7. Compare DBMS with RDBMS.

10. ANSWERS

Self Assessment Questions

1. Data item
2. Universe of Discourse
3. Software
4. Database system
5. Airlines
6. Database
7. Automated teller
8. Telecommunication
9. Two
10. Physical
11. Logical
12. Difficult
13. data model
14. Planning
15. *Entities*
16. Instance
17. One-to-one
18. One
19. Centralized
20. Consistent
21. Data integrity
22. Vital
23. Cost
24. Easier
25. set theory

Terminal Questions

1. Implicit properties of a database are: (i) A database represents some aspect of the real world, sometimes called the mini-world or universe of discourse (UoD). (ii) A database is a logically coherent collection of data with some inherent meaning and so on. (Refer to section 1.2 for detail)
2. List of representative applications of Databases are Banking, Airlines, Universities, Credit card transactions, Telecommunication, etc. (Refer to section 1.3 for detail.)
3. Physical data independence allows changes in the physical storage devices or organization of the files to be made without requiring changesin the conceptual view whereas Logical data independence implies that application programs need not be changed if fields are added to an existing record. (Refer to section 1.4 for detail)
4. Entities are the principal data objects about which information is to be collected. Attributes are characteristics of entities that provide descriptive details about them. (Refer to section 1.5 for detail)
5. Relationships represent real-world associations among one or more entities. A relationship is indicated by the connectivity between entityoccurrences: one-to-one, one-to-many, and many-to-many. (Refer to section 1.5.2 for detail)
6. The advantages of using a database system are that the organization can exert, via the DBA, centralized management and control over the data apart from Reduction of Redundancies, Data integrity, Data security, and Conflict resolution. (Refer to section 1.6.1 for detail)
7. Relational/SQL data is easier to query than hierarchical, CODASYL, or some other model. SQL DBMS products were originally based on the relational database management model – but did not implement the model fully or completely accurately. (Refer to section 1.7)