Unit IV

CHAPTER 11

DATABASE CONCEPTS

Learning Objectives

At the completion of this chapter, the student will be able to know

- the concept of a database and relational database.
- different components of the database.
- types of database models.
- types of relationship.
- the concepts of relational algebra.

Introduction

A database is an organized collection of data, generally stored and accessed electronically from a computer system. The term "database" is also used to refer to any of the DBMS, the database system or an application associated with the database. Because of the close relationship between them, the term "database" is often used casually to refer to both a database and the DBMS used to manipulate it. A school class register is a database where names are arranged alphabetically. Databases have been around since people started recording things. Here we tend to focus on electronic ones.

11.1 Data

Data are raw facts stored in a computer. A data may contain any character, text, word or a number.

Example : 600006, DPI Campus, SCERT, Chennai, College Road

11.2 Information

Information is formatted data, which allows to be utilized in a significant way.

Example

SCERT
College Road
DPI Campus
Chennai 600006

SYC51

As you can see from the example above, data appears as a set of words and numbers. However, when the data is processed, organized and formatted, it gives a meaningful information about the SCERT institution contact address.

11.3 Database

Database is a repository collection of related data organized in a way that data can be easily accessed, managed and updated. Database can be a software or hardware based, with one sole purpose of storing data.

11.4 DataBase Management System (DBMS)

A DBMS is a software that allows us to create, define and manipulate database, allowing users to store, process and analyze data easily. DBMS provides us with an interface or a tool, to perform various operations to create a database, storing of data and for updating data, etc. DBMS also provides protection and security to the

databases. It also maintains data consistency in case of multiple users.

Examples of DBMS softwares are : Foxpro, dbase.

11.4.1 Characteristics of Database Management System

1. Ability to manipulate data	DBMS provides the facility to manipulate data (store, modify and delete)in a data base.
2. Reduced Redundancy	In the modern world hard drives are very cheap, but earlier when hard drives were too expensive, unnecessary repetition of data in database was a big problem But DBMS follows Normalisation which divides the data in such a way that repetition is minimum.
3.Data Consistency	On live data, it is being continuously updated and added, maintaining the consistency of data can become a challenge. But DBMS handles it by itself.
4. Support Multiple user and Concurrent Access	DBMS allows multiple users to work on it(update, insert, delete data) at the same time and still manages to maintain the data consistency.
5.Query Language	DBMS provides users with a simple query language, using which data can be easily fetched, inserted, deleted and updated in a database.
6. Security	The DBMS also takes care of the security of data, protecting the data from unauthorized access. In a typical DBMS, we can create user accounts with different access permissions, using which we can easily secure our data by restricting user access.
7. DBMS Supports Transactions	It allows us to better handle and manage data integrity in real world applications where multi-threading is extensively used.

11.4.2 Advantages of DBMS

- Segregation of application program
- Minimal data duplication or Data Redundancy
- Easy retrieval of data using the Query Language
- Reduced development time and maintenance

11.4.3 Components of DBMS

The Database Management System can be divided into five major components as follows:

1.Hardware 2.Software

3.Data 4.Procedures/Methods

5.Database Access Languages

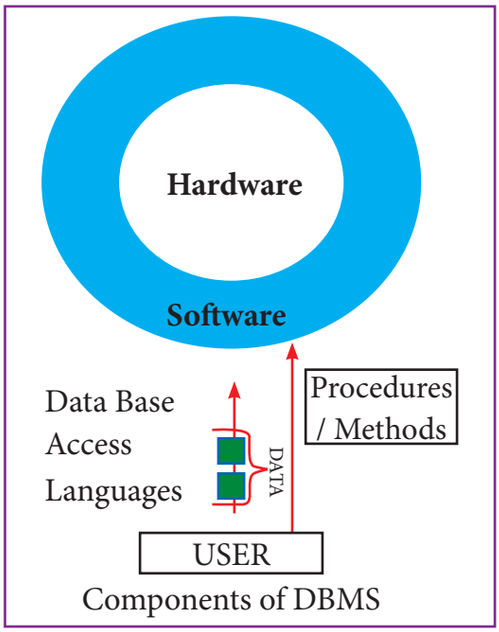


Figure 11.1

1. Hardware: The computer, hard disk, I/O channels for data, and any other physical component involved in storage of data

2. Software: This main component is a program that controls everything. The DBMS software is capable of understanding the Database Access Languages and interprets into database commands for execution.

3. Data: It is the resource for which DBMS is designed. DBMS creation is to store and utilize data.

4. Procedures/Methods: They are general instructions to use a database management system such as installation of DBMS, manage databases to take backups, report generation, etc.

5. DataBase Access Languages: They are the languages used to write commands to access, insert, update and delete data stored in any database.

Examples of popular DBMS: Dbase, FoxPro

11.5 Database Structure

Table is the entire collection of related data in one table, referred to as a File or Table where the data is organized as row and column.

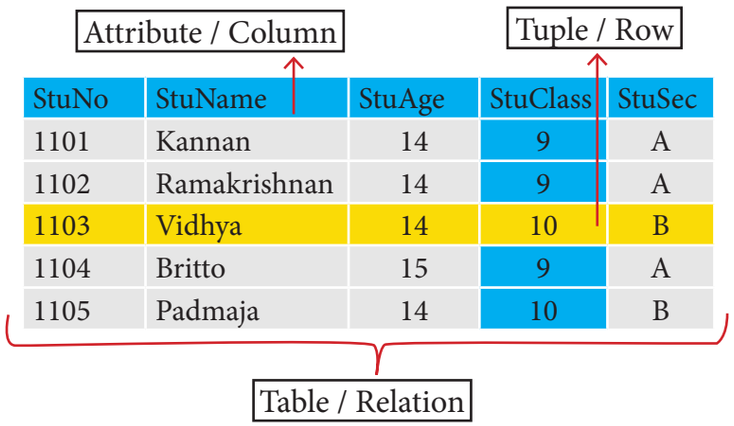
Each row in a table represents a record, which is a set of data for each database entry.

Each table column represents a Field, which groups each piece or item of data among the records into specific categories or types of data. Eg. StuNo., StuName, StuAge, StuClass, StuSec.

A Table is known as a RELATION

A Row is known as a TUPLE

A column is known as an ATTRIBUTE



StuNo	StuName	StuAge	StuClass	StuSec
1101	Kannan	14	9	A
1102	Ramakrishnan	14	9	A
1103	Vidhya	14	10	B
1104	Britto	15	9	A
1105	Padmaja	14	10	B

DataBase Structure

Fig 11.2

11.6 Data Model

- A data model describes how the data can be represented and accessed from a software after complete implementation
- It is a simple abstraction of complex real world data gathering environment.
- The main purpose of data model is to give an idea as how the final system or software will look like after development is completed.

11.6.1 Types of Data Model

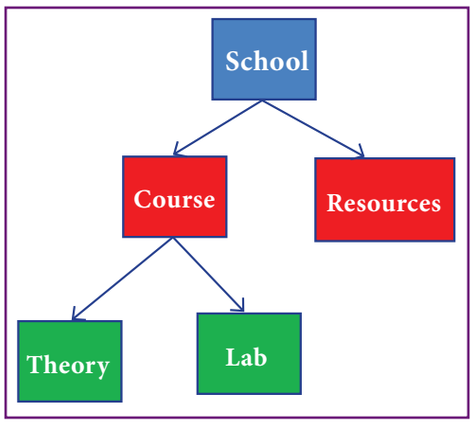
Following are the different types of a Data Model

- Hierarchical Model
- Relational Model
- Network Database Model
- Entity Relationship Model
- Object Model

1. Hierarchical Model

Hierarchical model was developed by IBM as Information Management System.

In Hierarchical model, data is represented as a simple tree like structure form. This model represents a one-to-many relationship i.e parent-child relationship. One child can have only one parent but one parent can have many children. This model is mainly used in IBM Main Frame computers.



Hierarchical Model Fig. 11.3

2. Relational Model

The Relational Database model was first proposed by E.F. Codd in 1970 . Nowadays, it is the most widespread data model used for database applications around the world.

The basic structure of data in relational model is tables (relations). All the information's related to a particular type is stored in rows of that table. Hence tables are also known as relations in a relational model. A relation key is an attribute which uniquely identifies a particular tuple (row in a relation (table)).

Stu_id	Name	Age
1	Malar	17
2	Suncar	16
3	Velu	16

Subj_id	Name	Teacher
1	C++	Kannan
2	Php	Ramakrishnan
3	Python	Vidhya

Stu_id	Subj_id	Marks
1	1	92
1	2	89
3	2	96

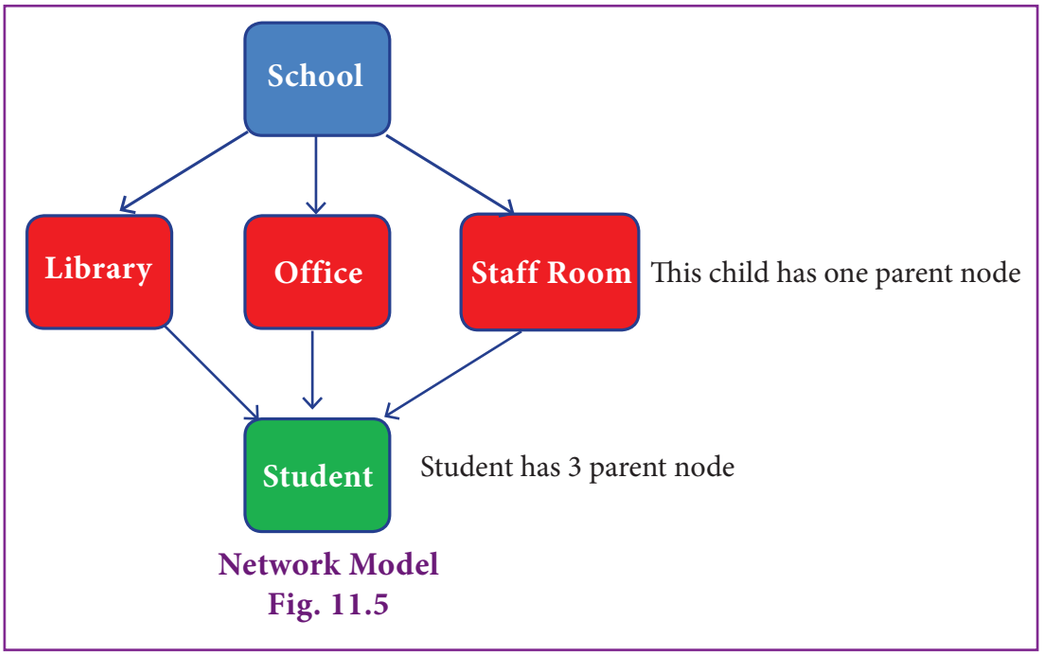
Relational Model
Fig. 11.4

3. Network Model

Network database model is an extended form of hierarchical data model. The difference between hierarchical and Network data model is :

- In hierarchical model, a child record has only one parent node,

- In a Network model, a child may have many parent nodes. It represents the data in many-to-many relationships.
- This model is easier and faster to access the data.



Network Model
Fig. 11.5

School represents the parent node

Library, Office and Staff room is a child to school (parent node)

Student is a child to library, office and staff room (one to many relationship)

4. Entity Relationship Model. (ER model)

In this database model, relationship are created by dividing the object into entity and its characteristics into attributes.

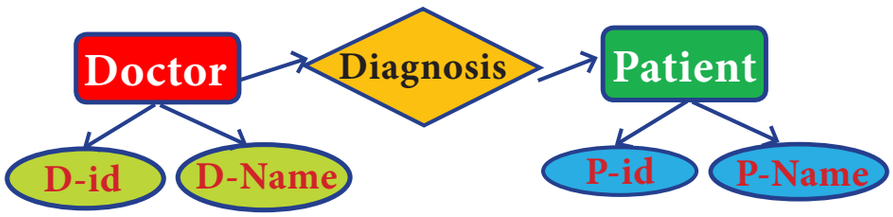
It was developed by Chen in 1976. This model is useful in developing a conceptual design for the database. It is very simple and easy to design logical view of data. The developer can easily understand the system by looking at ER model constructed.

Rectangle represents the entities. E.g. Doctor and Patient

Ellipse represents the attributes E.g. D-id, D-name, P-id, P-name. Attributes describes the characteristics and each entity becomes a major part of the data stored in the database.

Diamond represents the relationship in ER diagrams

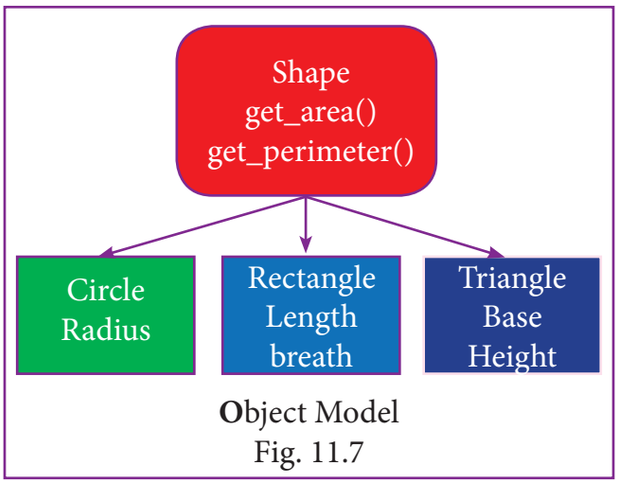
E.g. Doctor diagnosis the Patient



ER model
Fig. 11.6

5. Object Model

Object model stores the data in the form of objects, attributes and methods, classes and Inheritance. This model handles more complex applications, such as Geographic information System (GIS), scientific experiments, engineering design and manufacturing. It is used in file Management System. It represents real world objects, attributes and behaviors. It provides a clear modular structure. It is easy to maintain and modify the existing code.



Object Model
Fig. 11.7

An example of the Object model is **Shape, Circle, Rectangle** and **Triangle** are all objects in this model.

- **Circle** has the attribute **radius.**
- **Rectangle** has the attributes **length and breadth.**
- **Triangle** has the attributes **base and height** .
- The objects Circle, Rectangle and Triangle **inherit** from the object Shape.

11.6.2 Types of DBMS Users

Database Administrators

Database Administrator or DBA is the one who manages the complete database management system. DBA takes care of the security of the DBMS, managing the license keys, managing user accounts and access etc.

Application Programmers or Software Developers

This user group is involved in developing and designing the parts of DBMS.

End User

All modern applications, web or mobile, store user data. Applications are programmed in such a way that they collect user data and store the data on DBMS systems running on their server. End users are the one who store, retrieve, update and delete data.

Database designers: are responsible for identifying the data to be stored in the database for choosing appropriate structures to represent and store the data.

11.7 Difference between DBMS and RDBMS

Basis of Comparison	DBMS	RDBMS
Expansion	Database Management System	Relational DataBase Management System
Data storage	Navigational model ie data by linked records	Relational model (in tables). ie data in tables as row and column
Data redundancy	Exhibit	Not Present
Normalization	Not performed	RDBMS uses normalization to reduce redundancy
Data access	Consumes more time	Faster, compared to DBMS.
Keys and indexes	Does not use.	used to establish relationship. Keys are used in RDBMS.
Transaction management	Inefficient, Error prone and insecure.	Efficient and secure.
Distributed Databases	Not supported	Supported by RDBMS.
Example	Dbase, FoxPro.	SQL server, Oracle, mysql, MariaDB, SQLite.

Database normalization was first proposed by ***Dr. Edgar F Codd*** as an integral part of RDBMS in order to reduce data redundancy and improve data integrity. These rules are known as E F Codd Rules.

11.8 Types of Relationships

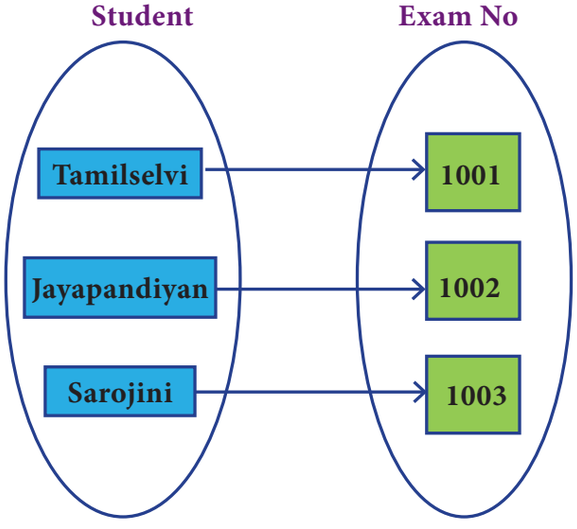
Following are the types of relationships used in a database.

1. One-to-One Relationship
2. One-to-Many Relationship
3. Many-to-One Relationship
4. Many-to-Many Relationship

1. One-to-One Relationship

In One-to-One Relationship, one entity is related with only one other entity. One row in a table is linked with only one row in another table and vice versa.

For example: A student can have only one exam number

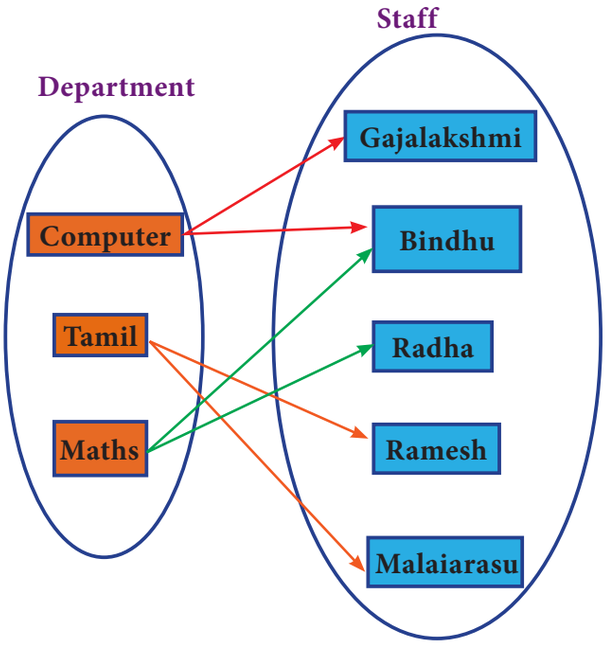


One to one Relationships
Fig 11.8

2. One-to-Many Relationship

In One-to-Many relationship, one entity is related to many other entities.

One row in a table A is linked to many rows in a table B, but one row in a table B is linked to only one row in table A. For example: One Department has many staff members.



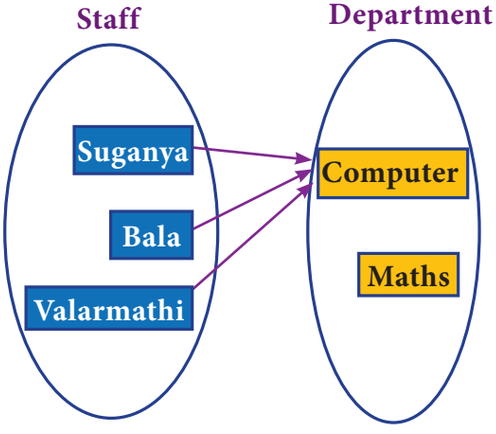
One to Many Mapping
Fig 11.9

3. Many-to-One Relationship

In Many-to-One Relationship, many entities can be related with only one in the other entity.

For example: A number of staff members working in one Department.

Multiple rows in staff members table is related with only one row in Department table.



Many to one Relationship
Fig 11.10

4. Many-to-Many Relationship

A many-to-many relationship occurs when multiple records in a table are associated with multiple records in another table.

Example 1: Customers and Product

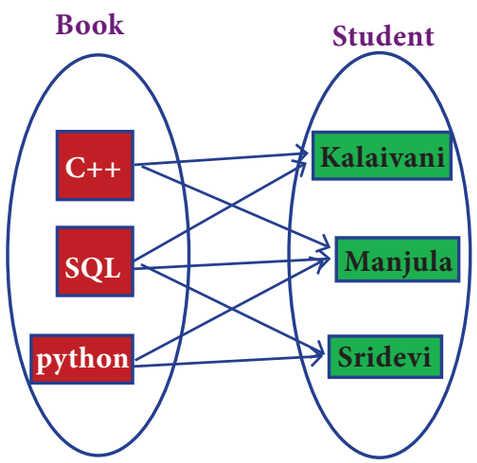
Customers can purchase various products and Products can be purchased by many customers

Example 2: Students and Courses

A student can register for many Courses and a Course may include many students

Example 3: Books and Student.

Many Books in a Library are issued to many students.



Many to Many Relationship
Fig 11.11

The relational model was invented by Edgar Frank Codd (Father of Relational DataBase) as a general model of data, and subsequently promoted by Chris Date and Hugh Darwen among others.

11.9 Relational Algebra in DBMS

What is Relational Algebra?

Relational Algebra, was first created by ***Edgar F Codd*** while at IBM. It was used for modeling the data stored in relational databases and defining queries on it.

Relational Algebra is a procedural query language used to query the database tables using SQL.

Relational algebra operations are performed recursively on a relation (table) to yield an output. The output of these operations is a new relation, which might be formed by one or more input relations.

Relational Algebra is divided into various groups

Unary Relational Operations

SELECT (symbol : σ)

PROJECT (symbol : $\prod$)

Relational Algebra Operations from Set Theory

- UNION ($\cup$)
- INTERSECTION ($\cap$)
- DIFFERENCE ($-$)
- CARTESIAN PRODUCT (X)

SELECT (symbol : σ)

General form σ_c (R) with a relation R and a condition C on the attributes of R.

The SELECT operation is used for selecting a subset with tuples according to a given condition.

Select filters out all tuples that do not satisfy C.

STUDENT

Studno	Name	Course	Year
cs1	Kannan	Big Data	II
cs2	Gowri Shankar	R language	I
cs3	Lenin	Big Data	I
cs4	Padmaja	Python Programming	I

Table 11.1

σ_{course} = "Big Data" **(STUDENT)**

Studno	Name	Course	Year
cs1	Kannan	Big Data	II
cs3	Lenin	Big Data	I

PROJECT (symbol : Π)

The projection eliminates all attributes of the input relation but those mentioned in the projection list. The projection method defines a relation that contains a vertical subset of Relation.

Example 1 using Table 11.1

Π_{course} (STUDENT)

Result

Course
Big Data
R language
Python Programming

Note

duplicate row is removed in the result

Example 2 (using Table 11.1)

$\Pi_{studno,\ course}$ (STUDENT)

Result

Studno	Course
cs1	Big Data
cs2	R language
cs3	Big Data
cs4	Python Programming

UNION (Symbol :∪)

It includes all tuples that are in tables A or in B. It also eliminates duplicates. Set A Union Set B would be expressed as A ∪ B

Example 3

Consider the following tables

Table A	
Studno	**Name**
cs1	Kannan
cs3	Lenin
cs4	Padmaja

Table B	
Studno	**Name**
cs1	Kannan
cs2	GowriShankarn
cs3	Lenin

Table 11.2

Result

Table A ∪ B	
Studno	**Name**
cs1	Kannan
cs2	GowriShankar
cs3	Lenin
cs4	Padmaja

SET DIFFERENCE (Symbol : -)

The result of A – B, is a relation which includes all tuples that are in A but not in B.

The attribute name of A has to match with the attribute name in B.

Example 4 (using Table 11.2)

Result

Table A – B	
cs4	Padmaja

INTERSECTION (symbol : ∩) A ∩ B

Defines a relation consisting of a set of all tuple that are in both in A and B. However, A and B must be union-compatible.

Example 5 (using Table 11.2)

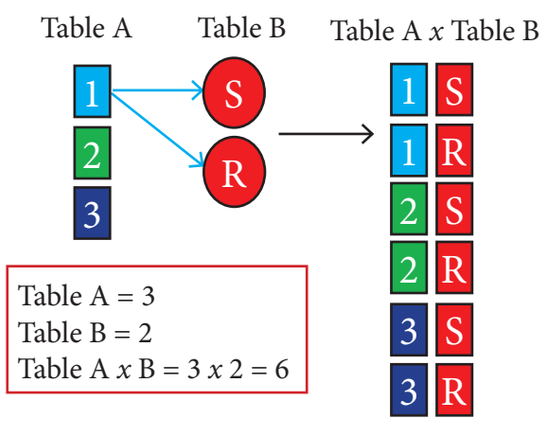
A ∩ B	
cs1	Kannan
cs3	Lenin

PRODUCT OR CARTESIAN PRODUCT (Symbol : X)

Cross product is a way of combining two relations. The resulting relation contains, both relations being combined.

A x B means A times B, where the relation A and B have different attributes.

This type of operation is helpful to merge columns from two relations.



Cartesian Product
Fig. 11.12

Table A		Table B	
studno	**name**	**course**	**subject**
cs1	Kannan	cs28	Big Data
cs2	Gowri Shankar	cs62	R language
cs4	Padmaja	cs25	python programming

Table 11.3

Cartesian product : Table A x Table B

studno	name	course	subject
cs1	Kannan	cs28	Big Data
cs1	Kannan	cs62	R language
cs1	Kannan	cs25	python rogramming
cs2	Gowri Shankar	cs28	Big Data
cs2	Gowri Shankar	cs62	R language
cs2	Gowri Shankar	cs25	python programming
cs4	Padmaja	cs28	Big Data
cs4	Padmaja	cs62	R language
cs4	Padmaja	cs25	python programming

Points to remember:

- DBMS is a computer based record keeping system
- Data is unprocessed data which contains any character, text, word or number has no meaning
- Information is processed data, organized and formatted.
- Examples of RDBMS are mysql, oracle, sql server, ibm db2
- Redundancy means duplication of data in a database.
- Data Consistency means that data values are the same at all instances of a database
- Data Integrity is security from unauthorized users
- Table is known a relation
- A row is called a tuple
- A column is known as an attribute
- Types of data model are Hierarchical, Relational, Network, ER and Object model.
- Hierarchical model is a simple tree like structure form with one-to-one relationship called parent-child relationship
- Relational Model represents data as relations or tables
- Network model is similar to Hierarchical model but it allows a record to have more than one parent
- ER model consists of entities, attributes and relationships

Points to remember:

- Object model stores data as objects, attributes, methods, classes and inheritance
- Normalization reduces data redundancy and improves data integrity
- Different types of Relationship are one-to-one, one-to-many, many-to-one and many-to-many relationships
- Database Normalization was proposed by Dr.Edgar F Codd
- Relational Algebra is used for modeling data stored in relational databases and for defining queries on it.

Evaluation

Part - A

Choose the best answer **(1 Marks)**

1. What is the acronym of DBMS?
   a) DataBase Management Symbol
   b) Database Managing System
   c) DataBase Management System
   d) DataBasic Management System

2 A table is known as
   a) tuple
   b) attribute
   c) relation
   d)entity

3 Which database model represents parent-child relationship?
   a) Relational
   b) Network
   c) Hierarchical
   d) Object

4 Relational database model was first proposed by
   a) E F Codd
   b) E E Codd
   c) E F Cadd
   d) E F Codder

5 What type of relationship does hierarchical model represents?
   a) one-to-one
   b) one-to-many
   c) many-to-one
   d) many-to-many

6. Who is called Father of Relational Database from the following?
   a) Chris Date
   b)Hugh Darween
   c) Edgar Frank Codd
   d) Edgar Frank Cadd

7. Which of the following is an RDBMS?

a) Dbase b) Foxpro

c) Mongo DB d) SQLite

8 What symbol is used for SELECT statement?

a) σ b) Π c) X d) Ω

9 A tuple is also known as

a) table b) row

c) attribute d) field

10. Who developed ER model?

a) Chen b) EF Codd

c) Chend d) Chand

Part - B

Answer the following questions (2 Marks)

1. Mention few examples of a DBMS.
2. List some examples of RDBMS.
3. What is data consistency?
4. What is the difference between Hierarchical and Network data model?
5. What is normalization?

Part -C

Answer the following questions (3 Marks)

1. What is the difference between Select and Project command?
2. What is the role of DBA?
3. Explain Cartesian Product with a suitable example.
4. Explain Object Model with example.
5. Write a note on different types of DBMS users.

Part -D

Answer the following questions (5 Marks)

1. Explain the different types of data model.
2. Explain the different types of relationship mapping.
3. Differentiate DBMS and RDBMS.
4. Explain the different operators in Relational algebra with suitable examples.
5. Explain the characteristics of DBMS.