01CT0407 - Database Management System

Unit - 1
Introductory concepts of DBMS

Prof. Harikesh Chauhan
Information Communication and Technology

Outline

- Introduction of DBMS
- Applications of DBMS
- Advantages of DBMS
- Basic terms
- Data models
- Three levels ANSI SPARC database system
- Data Abstraction in DBMS
- Mappings and data independence
- Database users and DBA
- Database system architecture

Introduction of DBMS

What is Database Management System (DBMS)?

- Data - Fact that can be recorded or stored
  - e.g. Person Name, Age, Gender and Weight etc.
- Database - Collection of logically related data
  - e.g. Books Database in Library, Student Database in University etc.
- Management - Manipulation, Searching and Security of data
  - e.g. Viewing result in GTU website, Searching exam papers in GTU website etc.
- System - Programs or tools used to manage database
  - e.g. SQL Server Studio Express, Oracle etc.
- DBMS - A Database Management System is a software for creating and managing databases.
- Database Management System (DBMS) is a software designed to define, manipulate, retrieve and manage data in a database.
  - e.g. MS SQL Server, Oracle, My SQL, SQLite, MongoDB etc.

Applications of DBMS

Applications of DBMS

- DBMS is a computerized record-keeping system.
- DBMS is required where ever data need to be stored.
  - E-Commerce (Flipkart, Amazon, Shopclues, eBay etc...)
  - Online Television Streaming (Hotstar, Amazon Prime etc...)
  - Social Media (WhatsApp, Facebook, Twitter, LinkedIn etc...)
  - Banking & Insurance
  - Airline & Railway
  - Universities and Colleges/Schools
  - Library Management System
  - Human Resource Department
  - Hospitals and Medical Stores
  - Government Organizations

Exercise Write down any five applications of DBMS other than above.

Advantages of DBMS

Reduce data redundancy (duplication)

Computer

Emp_Name	Address	Mobile	Subject
Prof. Amit Shah	Rajkot	1234	C lang.

Civil

Emp_Name	Address	Mobile	Subject
Prof. Amit Shah	Rajkot	1234	C lang.

Database management system can remove such data redundancy by storing data centrally.

Same data is stored at four different places.

Emp_Name	Address	Mobile	Subject
Prof. Amit Shah	Rajkot	1234	C lang.

Mechanical

Emp_Name	Address	Mobile	Subject
Prof. Amit Shah	Rajkot	1234	C lang.

Electrical

Remove data inconsistency

Computer			
Emp_Name	**Address**	**Mobile**	**Subject**
Prof. Amit Shah	Rajkot	6789	C lang.

Civil			
Emp_Name	**Address**	**Mobile**	**Subject**
Prof. Amit Shah	Rajkot	1234	C lang.

Database management system can keep data in consistent state.

Same data having different state (values)

Mobile no is changed

Mechanical			
Emp_Name	**Address**	**Mobile**	**Subject**
Prof. Amit Shah	Rajkot	6789	C lang.

Electrical			
Emp_Name	**Address**	**Mobile**	**Subject**
Prof. Amit Shah	Rajkot	1234	C lang.

Data isolation

- Data are scattered in various files.
- Files may be in different formats.
- Difficult to retrieve the appropriate data.

DBMS allow us to access (retrieve) appropriate data easily.

Data isolation is a property that controls when and how changes made by one operation become visible to other concurrent users and systems.
This issue occurs in a concurrency situation.

File - 1

Emp_Name	Address	Mobile	Subject
Prof. Amit Shah	Rajkot	1234	C lang.

File - 2

Emp_Name	Post	Salary	Load
Prof. Amit Shah	Professor	50000	16

File - 3

Emp_Name	Teaching	Knowledge	Rating
Prof. Amit Shah	Good	Excellent	9

Guaranteed atomicity

Atomicity: Either transaction execute 0% or 100%.

Sum of both account before transfer is 3000

Person A
Account A
Bal : 2000

Transfer 500

Person B
Account B
Bal : 1000

Step 1 : Debit 500 from Account A
Step 2 : Credit 500 into Account B

Transaction is failed

Sum of both account after transfer is 3000

Sum of both account is 2500 so inconsistent

Allow to implement integrity constraints

Personal

Student_Name	Address	Mobile_No	Subject
Amit Shah	Rajkot	9825098251	C lang.

Should contain exact 10 digits

Academic

Student_Name	Branch	Backlog	SPI
Amit Shah	Professor	0	9.5

Should be between 0 to 10

DBMS allows us to implement such business rules in our database.

Sharing of data among multiple users
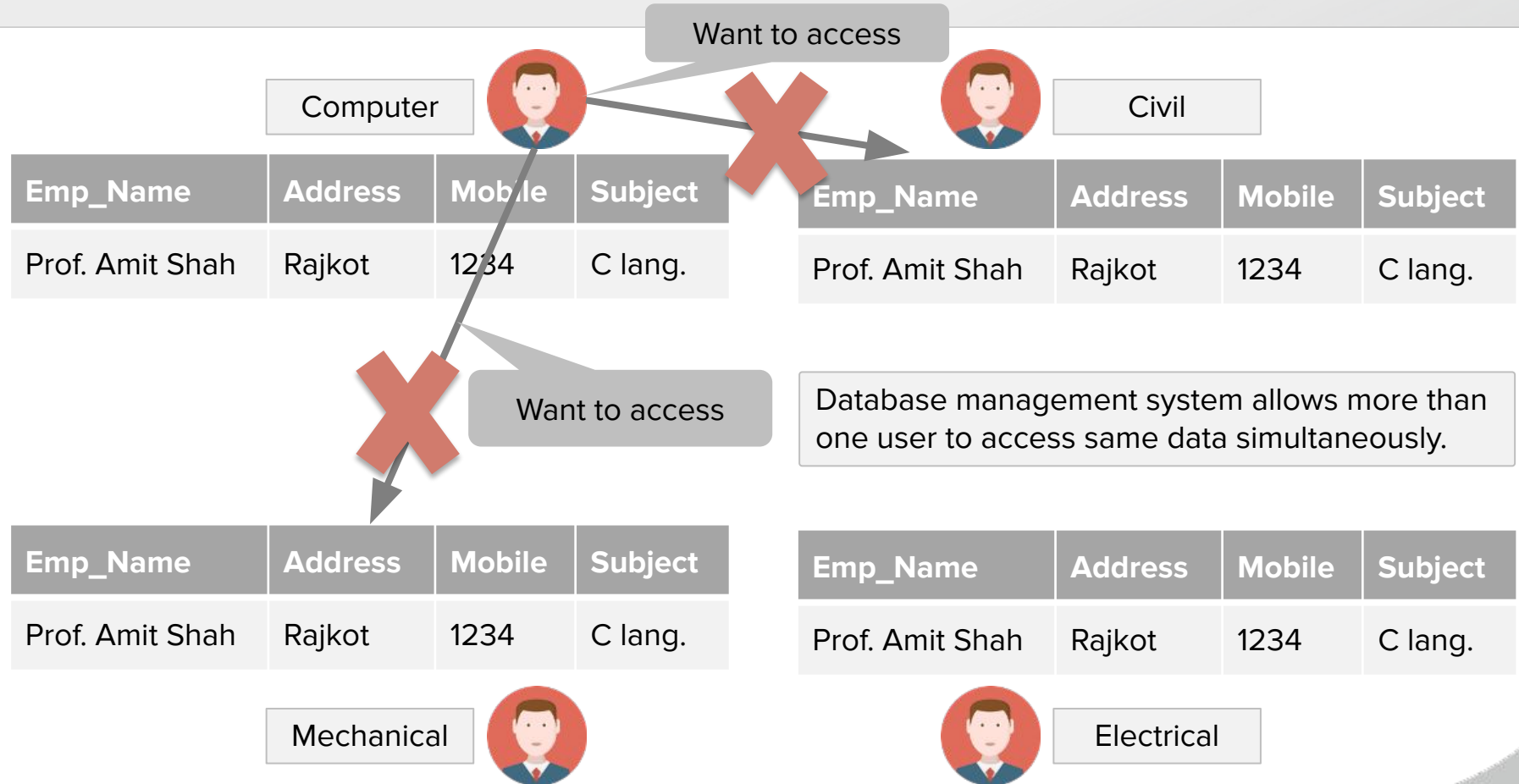
Want to access

Computer			

Emp_Name	Address	Mobile	Subject
Prof. Amit Shah	Rajkot	1234	C lang.

Civil			

Emp_Name	Address	Mobile	Subject
Prof. Amit Shah	Rajkot	1234	C lang.

Want to access

Database management system allows more than one user to access same data simultaneously.

Emp_Name	Address	Mobile	Subject
Prof. Amit Shah	Rajkot	1234	C lang.

Emp_Name	Address	Mobile	Subject
Prof. Amit Shah	Rajkot	1234	C lang.

Mechanical			

Electrical			

Restricting unauthorized access to data

Marwadi University

File - 1

Emp_Name	Address	Mobile	Subject
Prof. Amit Shah	Rajkot	1234	C lang.

File - 2

Emp_Name	Post	Salary	Load
Prof. Amit Shah	Professor	50000	16

✔️

File - 3

Emp_Name	Teaching	Knowledge	Rating
Prof. Amit Shah	Good	Excellent	9

❌

Wants to access ❌ Faculty of other college

Wants to access Marwadi Faculty

DBMS prevents unauthorized user to access data.

Providing backup and recovery services

Backup

Database

Backup Disk
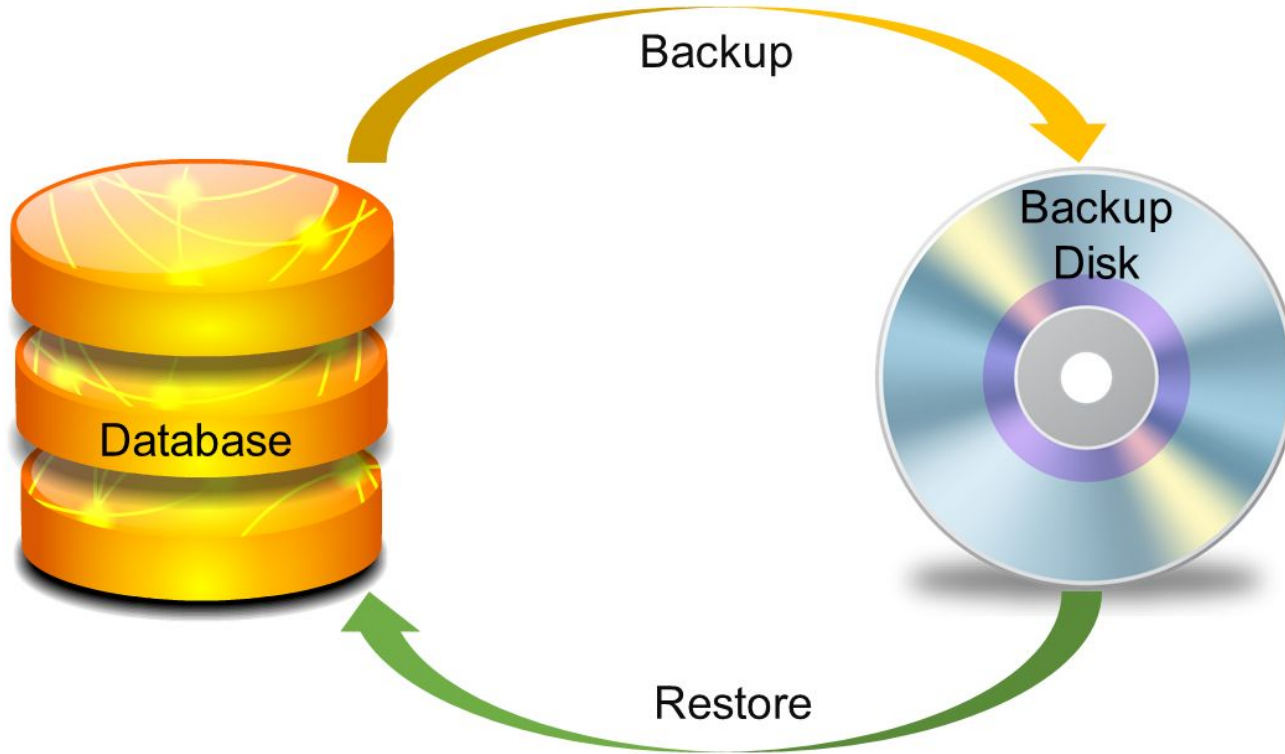
Restore

Provides facilities to backup and restore the database in case of failure.

Advantages of DBMS (Summary)

- Reduce data redundancy (duplication)
  - Avoids unnecessary duplication of data by storing data centrally.
- Remove data inconsistency
  - By eliminating redundancy, data inconsistency can be removed.
- Data isolation
  - A user can easily retrieve proper data as per his/her requirement.
- Guaranteed atomicity
  - Either transaction executes 0% or 100%.

Advantages of DBMS (Summary)

- Allow implementing integrity constraints
  - Business rules can be implemented such as do not allow to store amount less than Rs. 0 in balance.
- Sharing of data among multiple users
  - More than one users can access same data at the same time.
- Restricting unauthorized access to data
  - A user can only access data which is authorized to him/her.
- Providing backup and recovery services
  - Can take a regular auto or manual backup and use it to restore the database if it corrupts.

Basic Terms

- Data
    - Data is raw, unorganized facts that need to be processed.
    - Example: Marks of students
    - Student_1 = 50/100, Student_2 = 25/100.

- Information
    - When data is processed, organized, structured or presented in a given context so as to make it useful, it is called information.
    - Example: Result of students (Pass or Fail)
    - Student_1 = Pass, Student_2 = Fail.

Basic terms

- Metadata
  - Metadata is data about data.
  - Data such as table name, column name, data type, authorized user and user access privileges for any table is called metadata for that table.

Faculty			
Emp_Name	**Address**	**Mobile**	**Subject**
Prof. Amit Shah	Rajkot	1234	C lang.

  - Metadata of above table is:
    - Table name such as Faculty
    - Column name such as Emp_Name, Address, Mobile_No, Subject
    - Datatype such as Varchar, Decimal
    - Access privileges such as Read, Write (Update)

Basic terms

- Data dictionary
  - A data dictionary is an information repository which contains metadata.

  > Table Name – Faculty
  > Column Name – Emp_Name, Address, Mob, Subject, Salary
  > Datatype – Varchar, Decimal
  > Access Privileges – Read, Write (Update)

- Data warehouse
  - A data warehouse is an information repository which stores data.

Faculty			
Emp_Name	**Address**	**Mobile**	**Subject**
Prof. Amit Shah	Rajkot	1234	C lang.
Prof. Ajay Patel	Rajkot	9876	C++

Exercise   Why data dictionary and data warehouse are stored in the different places?

Basic terms

- Field
  - A field is a character or group of characters that have a specific meaning.
  - E.g, the value of Emp_Name, Address, Mobile_No etc are all fields of Faculty table.

Faculty			
Emp_Name	**Address**	**Mobile**	**Subject**
Prof. Amit Shah	Rajkot	1234	C lang.

Fields

Prof. Ajay Shah	Rajkot	1234

- Record / Tuple
  - A record is a collection of logically related fields.
  - E.g, the collection of fields (Emp_Name, Address, Mobile_No, Subject) forms a record for the Faculty.

Prof. Amit Shah	Rajkot	1234	C lang.
Prof. Ajay Patel	Rajkot	9876	C++

Record / Tuple

- Instance
  - The data which is stored in the database at a particular moment of time is called an instance of the database.

Faculty			
Emp_Name	**Address**	**Mobile**	**Subject**
Prof. Amit Shah	Rajkot	1234	C lang.
Prof. Jay Mehta	Surat	9879	C++
Prof. Jeet Oza	Baroda	9876	JAVA

  - Let's say a table faculty in our database whose name is MU, suppose the table has 2 records so the instance of the database has 2 records.
  - Now we are going to add another one records so the instance have total 3 records.
  - The instances can be changed by certain CRUD operations as like addition, deletion of data.

- Schema
  - The overall design of a database is called schema.
  - The basic structure of how the data will be stored in the database is called schema.

Faculty			
Emp_Name	**Address**	**Mobile**	**Subject**
Prof. Amit Shah	Rajkot	1234	C lang.
Prof. Jay Mehta	Surat	9879	C++

  - Let's say a table faculty in our database name DU, the faculty table require the emp_name, address, mobile, subject in their table so we design a structure as :
    - faculty table
    - emp_name : string
    - address: string
    - mobile : string
    - subject : string

> A schema contains schema objects like table, foreign key, primary key, views, columns, data types, stored procedure, etc.

  - Above given is the schema of the table faculty.

3 Levels
ANSI SPARC
Database System

3 Levels ANSI SPARC Database System

3 Levels ANSI SPARC Database System

- Internal level (Physical level)
  - It describes how a data is stored on the storage device.
  - Deals with physical storage of data.
  - Structure of records on disk - files, pages, blocks and indexes and ordering of records
  - Internal view is described by the internal schema.
- Conceptual level (Logical level)
  - What data are stored and what relationships exist among those data?
  - It hides low level complexities of physical storage.
  - For Example, STUDENT database may contain STUDENT and COURSE tables which will be visible to users but users are unaware about their storage.
  - Database administrator works at this level to determine what data to keep in the database.
- External level (View level)
  - It describes only part of the entire database that an end user concern or how data are viewed by each user.
  - Different user needs different views of the database, so there can be many views in a view level abstraction of the database. Used by end users and application programmers.
  - End users need to access only part of the database rather than the entire database.
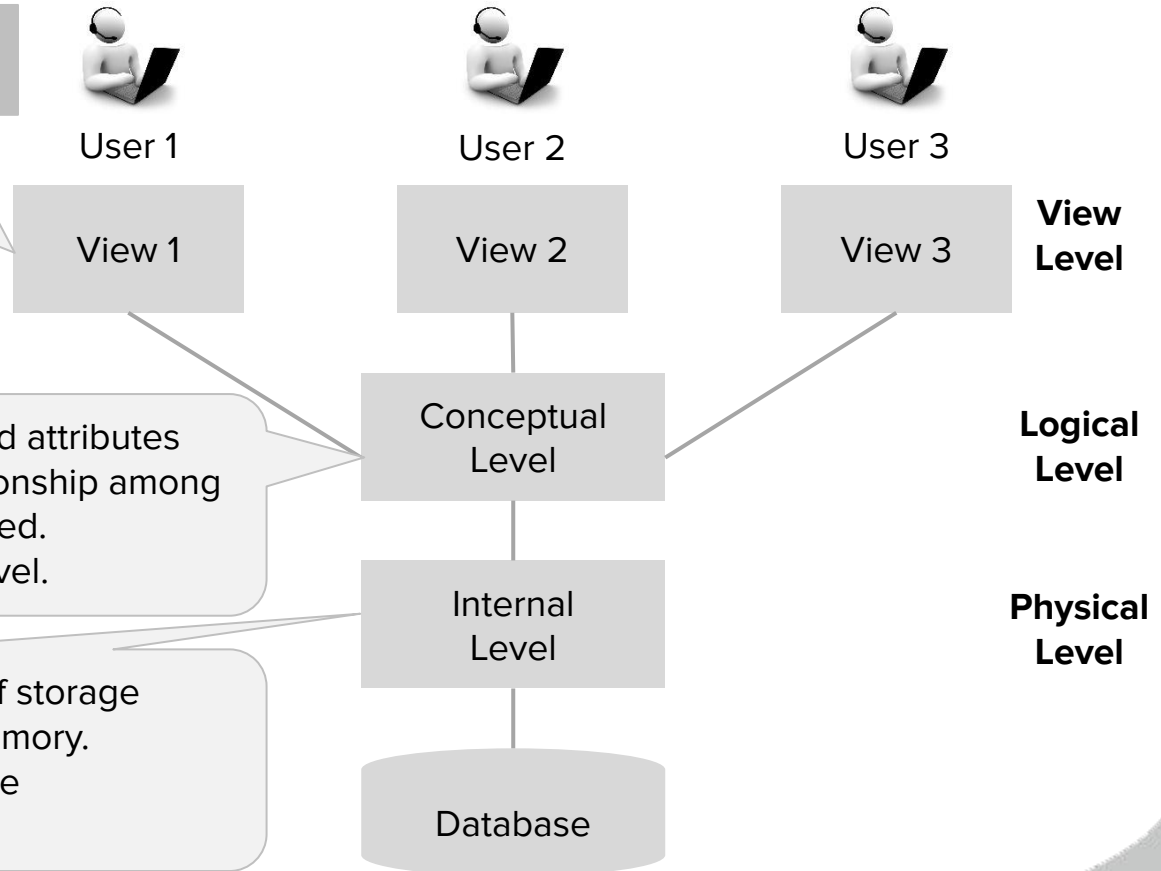
3 Levels ANSI SPARC Database System

We are storing student information in a student table.

User just interact with system with the help of GUI.
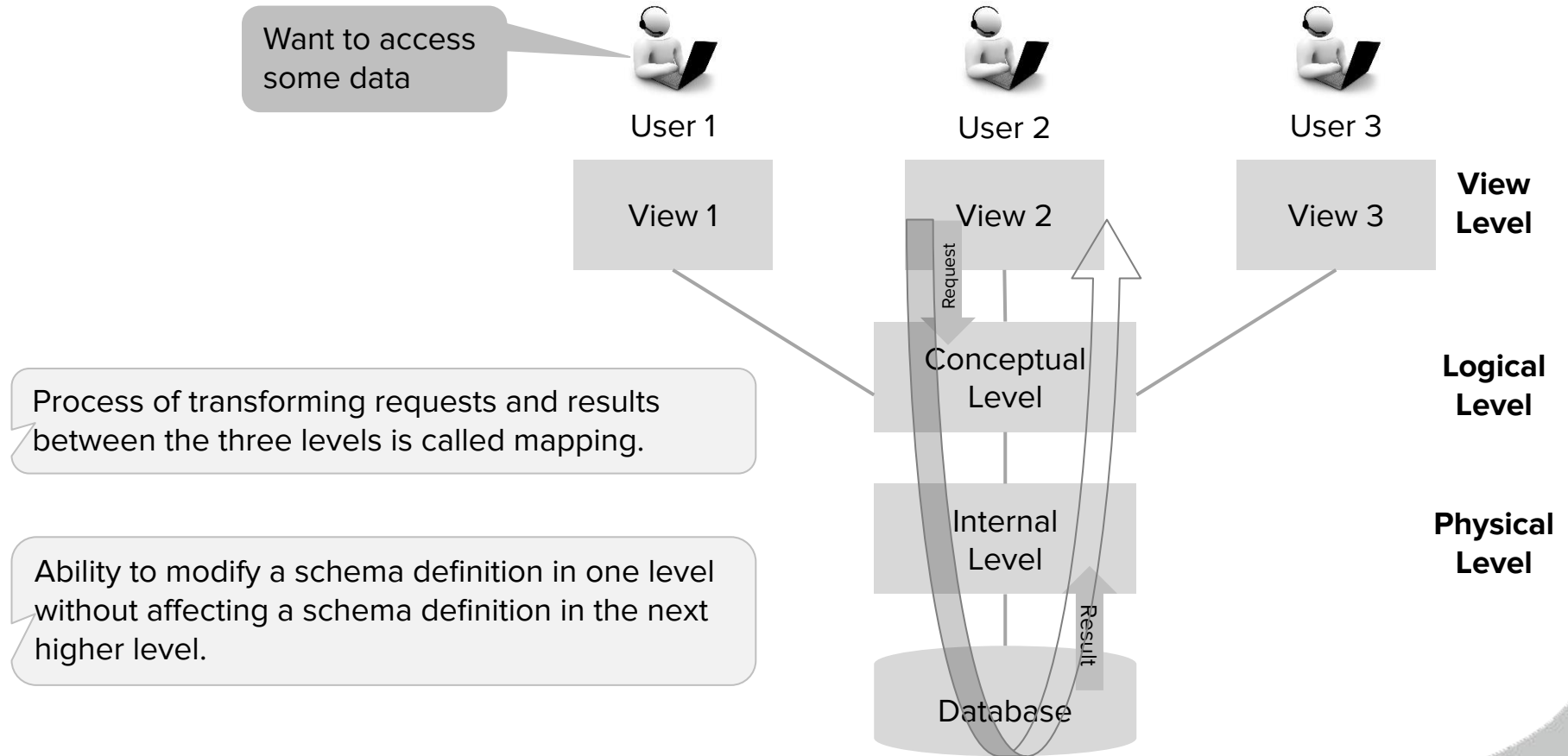Users are not aware of how and what the data is stored.

Records can be described as fields and attributes along with their data types, their relationship among each other can be logically implemented.
Programmers generally work at this level.

Records can be described as blocks of storage (bytes, gigabytes, terabytes etc.) in memory.
These details are often hidden from the programmers.

User 1   User 2   User 3

View 1   View 2   View 3   **View Level**

Conceptual Level   **Logical Level**

Internal Level   **Physical Level**

Database

Data Abstraction in DBMS

- Database systems are made-up of complex data structures.
- To ease the user interaction with database, the developers hide internal irrelevant details from users.
- This process of hiding irrelevant details from user is called data abstraction.

Mapping and Data Independence

Process of transforming requests and results between the three levels is called mapping.

Ability to modify a schema definition in one level without affecting a schema definition in the next higher level.

Types of Data Independence

- Physical Data Independence
  - Physical Data Independence is the ability to modify the physical schema without requiring any change in logical (conceptual) schema and application programs.
  - Modifications at the internal levels are occasionally necessary to improve performance.
  - Possible modifications at internal levels are changes in file structures, compression techniques, hashing algorithms, storage devices, etc.
- Logical Data Independence
  - Logical data independence is the ability to modify the conceptual schema without requiring any change in application programs.
  - Modification at the logical levels is necessary whenever the logical structure of the database is changed.
  - Application programs are heavily dependent on logical structures of the data they access. So any change in logical structure also requires programs to change.

Types of Database Users

Types of Database Users

- Naive Users (End Users)
  - Unsophisticated users who have zero knowledge of database system
  - End user interacts to database via sophisticated software or tools
  - e.g. Clerk in bank
- Application Programmers
  - Programmers who write software using tools such as Java, .Net, PHP etc…
  - e.g. Software developers
- Sophisticated Users
  - Interact with database system without using an application program
  - Use query tools like SQL
  - e.g. Analyst
- Specialized Users (DBA)
  - User write specialized database applications program
  - Use administration tools
  - e.g. Database Administrator

Role of DBA
(Database Administrator)

Role of DBA (Database Administrator)

- Schema Definition
    - DBA defines the logical schema of the database.
- Storage Structure and Access Method Definition
    - DBA decides how the data is to be represented in the database & how to access it.
- Defining Security and Integrity Constraints
    - DBA decides on various security and integrity constraints.
- Granting of Authorization for Data Access
    - DBA determines which user needs access to which part of the database.
- Liaison with Users
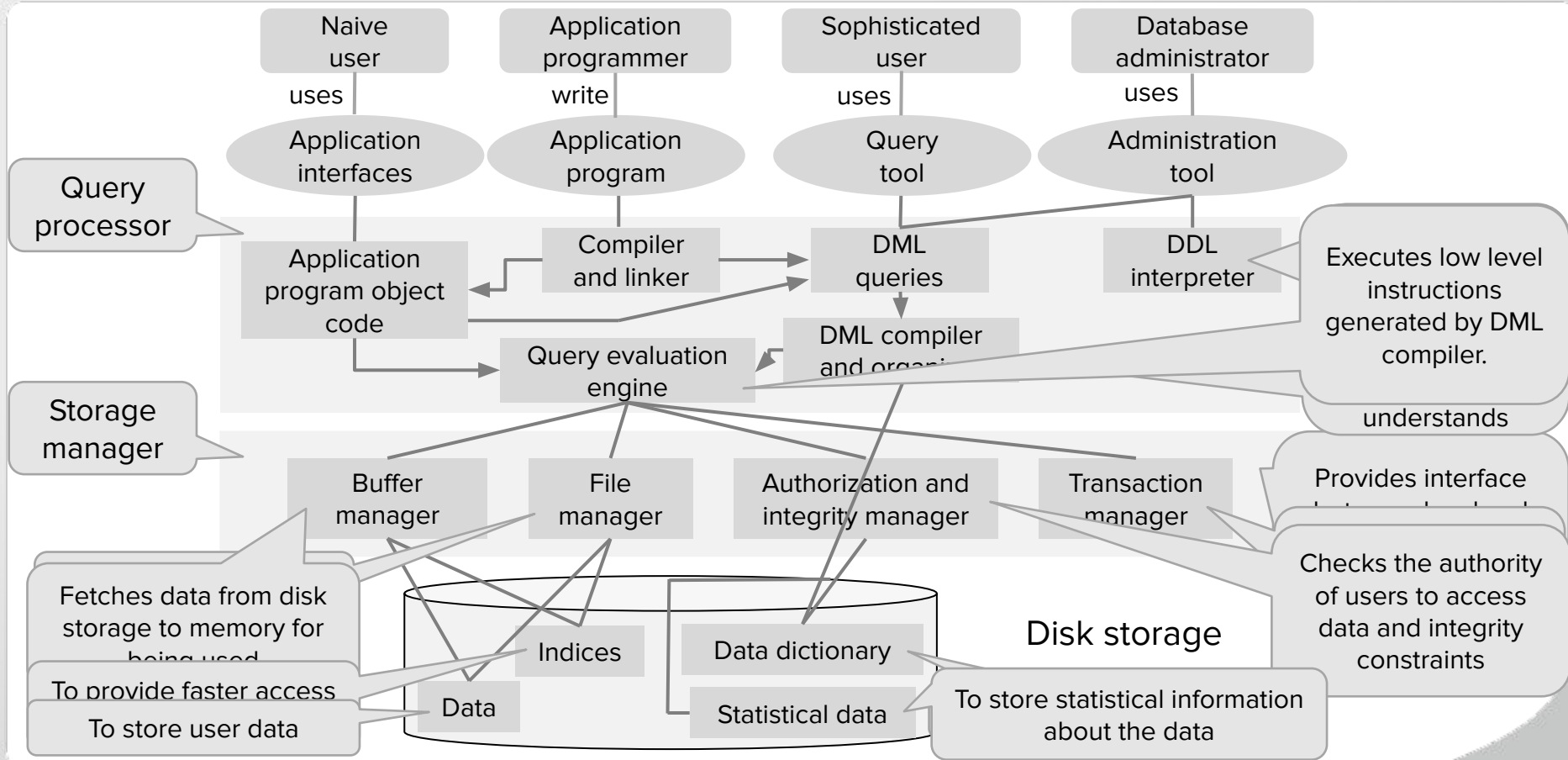   - DBA provide necessary data to the user.

Role of DBA (Database Administrator)

- Assisting Application Programmer
  - DBA provides assistance to application programmers to develop application programs.
- Monitoring Performance
  - DBA ensures that better performance is maintained by making a change in the physical or logical schema if required.
- Backup and Recovery
  - DBA backing up the database on some storage devices such as DVD, CD or magnetic tape or remote servers and recover the system in case of failures, such as flood or virus attack from this backup.
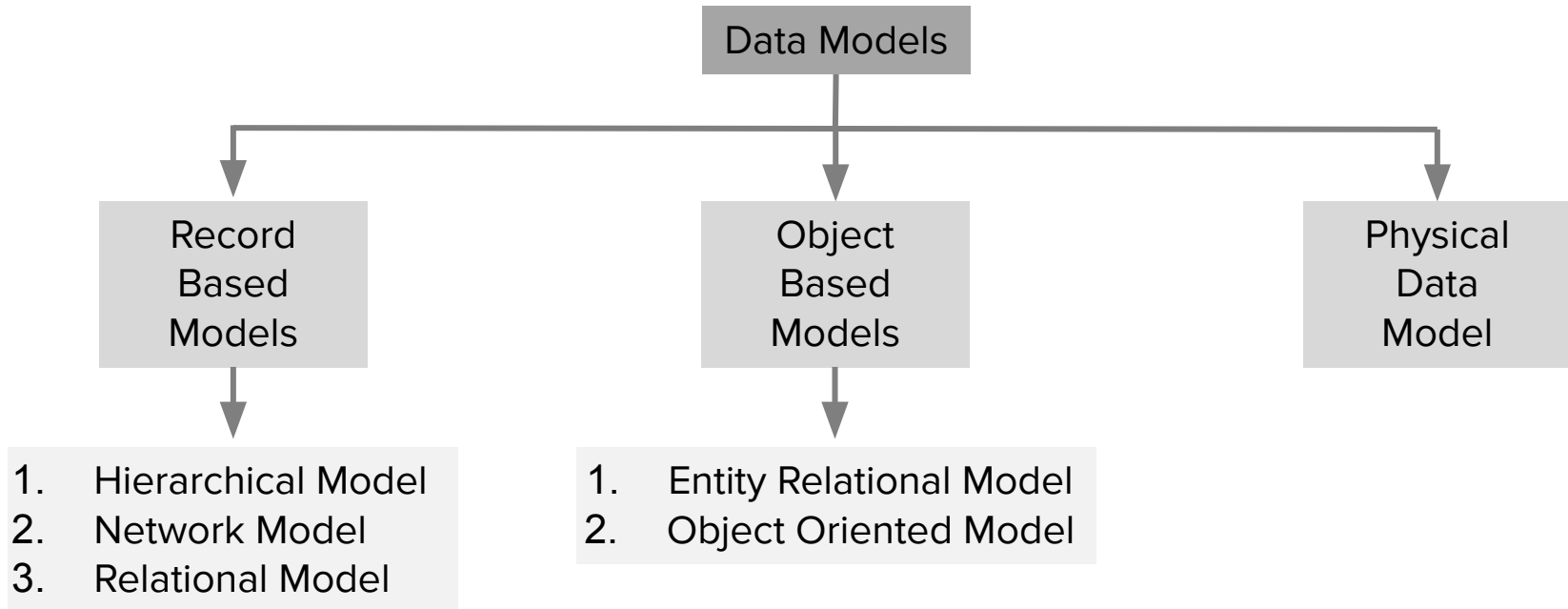
Database System Architecture

Database System Architecture

Naive user — uses → Application interfaces
Application programmer — write → Application program
Sophisticated user — uses → Query tool
Database administrator — uses → Administration tool

Query processor

Application program object code
Compiler and linker
DML queries
DDL interpreter
Query evaluation engine
DML compiler and organi...

Storage manager

Buffer manager
File manager
Authorization and integrity manager
Transaction manager

Disk storage
Indices
Data dictionary
Data
Statistical data

Executes low level instructions generated by DML compiler.

understands

Provides interface...

Checks the authority of users to access data and integrity constraints

Fetches data from disk storage to memory for being used

To provide faster access

To store user data

To store statistical information about the data

Data Models

Data Models

- Data models define how the logical structure of a database is modeled.
- It defines how data can be stored, accessed and updated in a database.
- It also define how data is connected to each other and how they are processed and stored inside the system.

```
                          Data Models
        ┌──────────────────────┼──────────────────────┐
        ▼                      ▼                      ▼
  Record               Object               Physical
  Based                Based                Data
  Models               Models               Model
        │                      │
        ▼                      ▼
```

1. Hierarchical Model
2. Network Model
3. Relational Model

1. Entity Relational Model
2. Object Oriented Model

Hierarchical Model

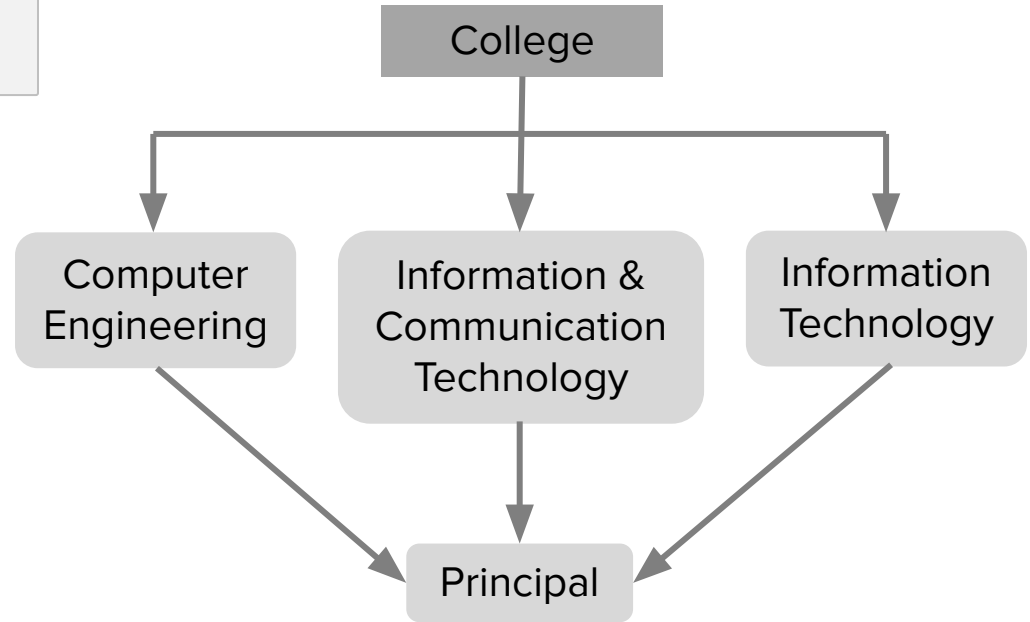- Here tree concept is used to represent data and relationship among data.

- Here each child record can have only one parent record.
- Each parent record can have zero or more than one child record.

- Records are represented by rectangular box and relationships between records are represented by arcs.
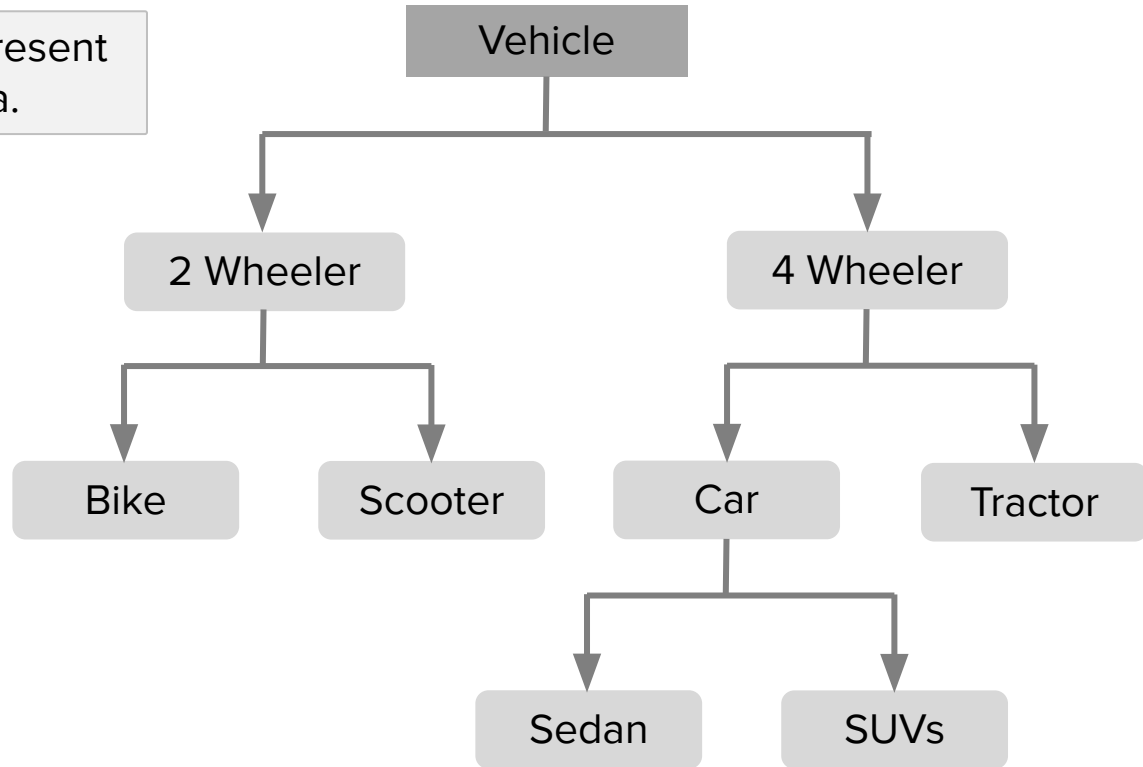
- This model is an extension of the hierarchical model.

- This model is the same as the hierarchical model, the only difference is that a record can have more than one parent.

- It replaces the hierarchical tree with a graph in which object types are the nodes and relationships are the edges.

College → Computer Engineering, Information & Communication Technology, Information Technology

Computer Engineering → Principal
Information & Communication Technology → Principal
Information Technology → Principal

Rational Model

- Relational Model is the most widely used model.
- In this model, the database is represented as a collection of tables in the form of rows and columns of a two-dimensional table.
- In this model, the data is maintained in the form of a two-dimensional table. All the information is stored in the form of row and columns.
- The basic structure of a relational model is tables. So, the tables are also called relations in the relational model.
- Each row is known as a tuple (a tuple contains all the data for an individual record) while each column represents an attribute.

Faculty

| Emp_Name	Address	Mobile	Subject
Prof. Amit Shah	Rajkot	1234	C lang.
Prof. Jay Mehta	Surat	9879	C++
Prof. Raj Patel	Baroda	9825	Java

Entity Relational Model

- In this model, we represent the real-world problem in the pictorial form to make it easy for the stakeholders to understand.
- An ER model is the logical representation of data as objects and relationships among them. These objects are known as entities, and relationship is an association among these entities.

- Entities: Entity is a real-world thing. It can be a person, place, or even a concept.
    Example: Faculty, Students, Course, Department etc are some of the entities of a   College Management System.
- Attributes: An entity contains a real-world property called attribute. This is the characteristics of that attribute.
    Example: The entity Faculty has the property like Name, Salary, Age, etc.
- Relationship: Relationship tells how two attributes are related.
    Example: Faculty teaches to the Students.

| Entity | Relationship | Entity |

- This model is based on collection of objects.
    We have two objects Employee and Department.
- An object body consists of data as well as methods.
    The attributes like Name, Job_title of the employee and the methods which will be performed by that object are stored as a single object.
- The objects that contain same type of data and same type of function are group together as a class.
- The two objects are connected through a common attribute i.e the Dept_id and the communication between these two will be done with the help of this common id.

Physical Data Model
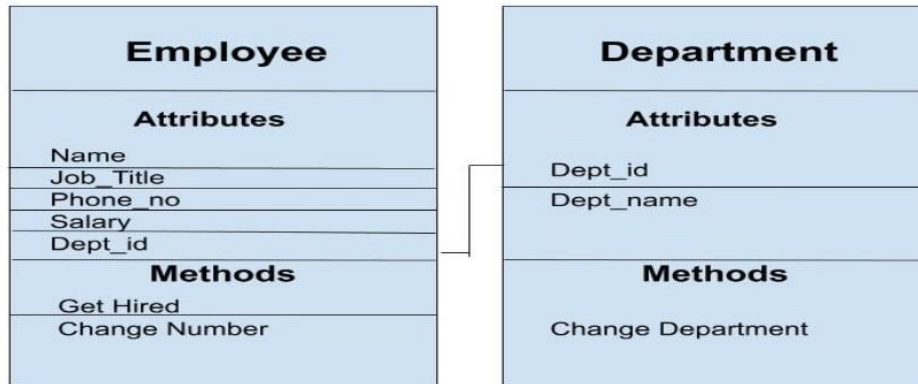
- Physical data models are used for a higher level description of storage structure and access mechanism.
- They describe how data is stored in database.