UNIT - I

1 Introduction to DBMS

Syllabus

Introduction : *Basic concepts, Advantages of DBMS over file processing systems, Data abstraction, Database languages, Data models, Data independence, Components of a DBMS, Overall structure of DBMS, Multi-user DBMS architecture, System catalogs, Data Modeling : Basic concepts, Entity, Attributes, Relationships, Constraints, Keys.*

Contents

Part I : Introduction

1.1 Basic Concepts

- **Definition :** A database management system (DBMS) is collection of **interrelated data** and various **programs** that are used to handle the data.
- The **primary goal** of DBMS is to provide a way to **store and retrieve** the required information from the database in convenient and efficient manner.
- For managing the data in the database two important **tasks** are conducted -
  - **Define the structure** for storage of information.
  - Provide mechanism for manipulation of information.
- In addition, the database systems must ensure the **safety of information** stored.

Database System Applications

There are wide range of applications that make use of database systems. Some of the applications are -

1) **Accounting :** Database systems are used in maintaining information employees, salaries, and payroll taxes.
2) **Manufacturing :** For management of supply chain and tracking production of items in factories database systems are maintained.
3) For maintaining customer, product and purchase information the databases are used.
4) **Banking :** In banking sector, for customer information, accounts and loan and for performing banking applications the DBMS is used.
5) For purchase on credit cards and generation of monthly statements database systems are useful.
6) **Universities :** The database systems are used in universities for maintaining student information, course registration, and accounting.
7) **Reservation systems :** In airline/railway reservation systems, the database is used to maintain the reservation and schedule information.
8) **Telecommunication :** In telecommunications for keeping records of the calls made, generating monthly bills, maintaining balances on prepaid calling cards, and storing information about communication networks the database systems are used.

1.1.1 Characteristics of Database Approach

Following are the characteristics of database system :

1) Representation of some aspects of **real world applications**.
2) Systematic **management of information**.
3) Representing the data by **multiple views**.
4) Efficient and easy implementation of various **operations** such as insertion, deletion and updation.
5) It maintains **data for** some specific **purpose**.
6) It represents **logical relationship** between records and data.

1.2 Advantages of DBMS over File Processing Systems

SPPU : May-18, Marks 5

1.2.1 Advantages of DBMS

Following are the advantages of DBMS :

1) DBMS **removes the data redundancy** that means there is no duplication of data in database.
2) DBMS allows to **retrieve the desired data** in required format.
3) **Data** can be **isolated** in separate tables for convenient and efficient use.
4) Data can be **accessed efficiently** using a simple query language.
5) The **data integrity** can be maintained. That means – the constraints can be applied on data and it should be in some specific range.
6) The **atomicity** of data can be maintained. That means, if some operation is performed on one particular table of the database, then the change must be reflected for the entire database.
7) The DBMS allows **concurrent access** to multiple users by using the synchronization technique.
8) The **security policies** can be applied to DBMS to allow the user to access only desired part of the database system.

1.2.2 Disadvantages of DBMS

1) **Complex design :** Database design is complex, **difficult and time consuming**.
2) **Hardware and software cost : Large amount of investment** is needed to setup the required hardware or to repair software failure.

3) **Damaged part :** If one part of database is corrupted or damaged, then **entire database** may get affected.

4) **Conversion cost :** If the current system is in conventional file system and if we need to convert it to database systems then large amount of cost is incurred in purchasing different tools, and adopting different techniques as per the requirement.

5) **Training :** For designing and maintaining the database systems, the people need to be trained.

1.2.3 File Processing System Vs. DBMS

- Earlier database systems are created in response to manage the commercial data. These data is typically stored in files. To allow users to manipulate these files, various programs are written for
  1) Addition of new data
  2) Updating the data
  3) Deleting the data.
- As per the need for addition of new data, separate application programs were required to write. Thus as the time goes by, the system acquires more files and more application programs.
- This typical **file processing system** is supported by conventional operating system. Thus the file processing system can be described as -
- "The system that stores the permanent records in files and it needs different application programs to extract or add the records".
- Before introducing database management system, this file processing system was in use. However, such a system has many drawbacks. Let us discuss them.

Disadvantages of Traditional File Processing System

The traditional file system has following disadvantages :

1) **Data redundancy :** Data redundancy means duplication of data at several places. Since different programmers create different files and these files might have different structures, there are chances that some information may appear repeatedly in some or more format at several places.

2) **Data inconsistency :** Data inconsistency occurs when various copies of same data may no longer get matched. For example changed address of an employee may be reflected in one department and may not be available (or old address present) for other department.

3) **Difficulty in accessing data :** The conventional file system does not allow to retrieve the desired data in efficient and convenient manner.

4) **Data isolation :** As the data is scattered over several files and files may be in different formats, it becomes difficult to retrieve the desired data from the file for writing the new application.

5) **Integrity problems :** Data integrity means data values entered in the database fall within a specified range and are of specific format. With the use of several files enforcing such constraint on the data becomes difficult.

6) **Atomicity problems :** An atomicity means particular operation must be carried out entirely or not at all with the database. It is difficult to ensure atomicity in conventional file processing system.

7) **Concurrent access anomalies :** For efficient execution, multiple users update data simultaneously, in such a case data need to be synchronized. As in traditional file systems, data is distributed over multiple files, one cannot access these files concurrently.

8) **Security problems :** Every user is not allowed to access all the data of database system. Since application program in file system are added in an ad hoc manner, enforcing such security constraints become difficult.

Database systems offer solutions to all the above mentioned problems.

Difference between Database System and Conventional File System

Sr. No.	Database systems	Conventional file systems
1.	Data redundancy is **less**.	Data redundancy is **more**.
2.	**Security** is **high**.	Security is very **low**.
3.	Database systems are used when security **constraints are high**.	Conventional file systems are used where there is **less demand** for **security constraints**.
4.	Database systems define the data in a **structured** manner. Also there is well defined co-relation among the data.	File systems define the data in **un-structured** manner. Data is usually in isolated form.
5.	Data **inconsistency** is **less** in database systems.	Data **inconsistency** is **more** in file systems.
6.	**User is unknown** to the physical address of the data used in database systems.	User locates the **physical address of file** to access the data in conventional file systems.
7.	We can retrieve the data in any **desired format** using database systems.	We cannot retrieve the data in any desired format using file systems.
8.	There is ability to access the data **concurrently** using database systems.	There is **no ability to concurrently** access the data using conventional file system.

Review Question

1. *State and explain the disadvantages of the File processing system.* **SPPU : May-18, (End Sem), Marks 5**

1.3 Data Abstraction

Definition of data abstraction : Data abstraction means retrieving only the required amount of information about the system and **hiding** background details.

There are several levels of abstraction that simplify user interactions with the system. These are :

1) Physical level :

- This is the **lowest level.**
- This level describes how the data are **stored.**
- The database administrators decide how to store data at the physical level.
- This level describes complex low-level data structures.

2) Logical level :

- This is the next **higher level,** which describes what data are stored in the database?.
- This level also describes the **relationship between the data.**
- The logical level thus describes the entire database in terms of a small number of relatively simple structures.
- The database administrators use a logical level of abstraction for deciding what information to keep in the database.

3) View level :

- This is the **highest level** of abstraction that describes only part of the entire database.
- The view level can provide access to only part of the database.
- This level helps in **simplifying the interaction** with the system.
- It can provide **multiple views** of the same system.
- For example – A Clerk at the reservation system can see only part of the database and access the passenger's required information.

Fig. 1.3.1 shows the relationship between the three levels of abstraction.

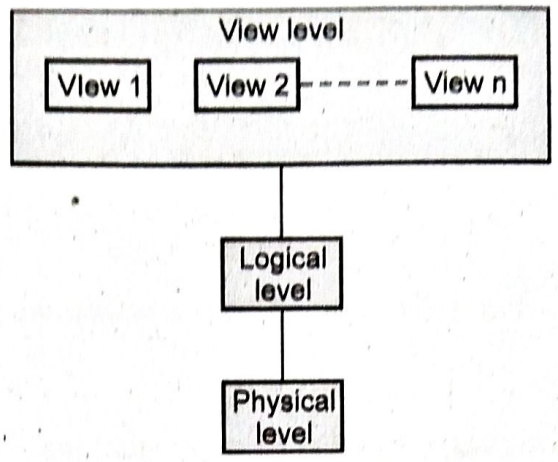


Fig. 1.3.1 : Levels of data abstraction

For example : Consider following record

```
type employee = record
                    empID : numeric(10)
                    empname : char(20)
                    dept_no : numeric(10)
                    salary : numeric(8,2)
                    end
```

This code defines a new record **employee** with four fields. Each field is associated with field name and its **type**. There are several other records such as **department** with fields dept_no, dept_name, building customer with fields cust_id,cust_name .

- **At the physical level**, the record - **customer, employee, department** can be described as block of **consecutive storage locations**. Many database systems hide lowest level storage details from database programmer.
- The type definition of the records is decided **at the logical level**. The **programmer** work of the record at this level, similarly **database administrators** also work at this level of abstraction.
- There is specific view of the record is allowed at the view level. For instance - customer can view the name of the employee, or id of the employee but cannot access employee's salary.

1.4 Database Languages

There are three types of languages supported by database systems.

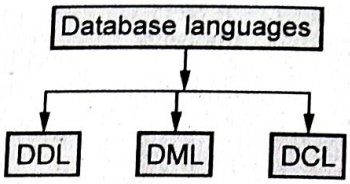


Fig. 1.4.1 Types of database languages

(1) DDL

- Data Definition Language (DDL) is a specialized language used to specify a database schema by a set of definitions.
- It is a language used for **creating** and **modifying** the structures of tables, views, indexes, etc.
- DDL is also used to specify additional properties of data.
- Some of the common commands used in DDL are -**CREATE, ALTER, DROP.**
- The **primary use** of CREATE command is to build a new table. Using ALTER command, the users can add up some additional column and drop existing columns. Using DROP command, the user can delete table or view.

(2) DML

- DML stands for Data Manipulation Language.
- This language enables users to access or manipulate data as organized by appropriate data model.
- The types of access are -
  - **Retrieval** of information stored in the database
  - **Insertion** of new information into the database.
  - **Deletion** of information from the database.
  - **Modification** of information stored in database.
- There are two types of DML -
  - **Procedural DML** - Require a user to specify what data are needed and how to get those data.
  - **Declarative DML** - Require a user to specify what data are needed without specifying how to get those data.
- **Query** is a statement used for requesting the retrieval of information. This retrieval of information using some specific language is called **query language.**

(3) DCL

- The Data Control Language (DCL) is used to control access to data stored in the database. This is also called as authorization.
- The typical command used in DCL are GRANT and REVOKE.
  - **GRANT :** This command is used to give access rights or privileges to the database.
  - **REVOKE :** The revoke command removes user access rights or privileges to the database objects

1.5 Data Models

SPPU : Nov.-19, Marks 5

- **Definition :** It is a collection of conceptual tools for describing data, relationships among data, semantics (meaning) of data and constraints.
- Data model is **a structure below the database.**
- Data model provides a way to describe the design of database at physical, logical and view level.
- There are various data models used in database systems and these are as follows -

(1) Relational model :

- The relation model consists of **collection of tables** which stores data and also represents the relationship among the data.
- The **table** is also known as **relation.**
- The table contains one or more **columns** and each column has unique name.
- Each table contains **record of particular type,** and each record type defines a fixed **number of fields or attributes.**
- **For example** – The following figure shows the relational model by showing the relationship between Student and Result database. For example – Student **Ram** lives in city **Chennai** and his marks are **78.** Thus the relationship between these two databases is maintained by the **SeatNo.** Column

Seat No	Name	City
101	Ram	Chennai
102	Shyam	Pune

SeatNo	Marks
101	78
102	95

Advantages :

(i) **Structural independence :** Structural independence is an ability that allows us to make changes in one database structure without affecting other. The relational model have structural independence. Hence making required changes in the database is convenient in relational database model.

(ii) **Conceptual simplicity :** The relational model allows the designer to simply focus on logical design and not on physical design. Hence relational models are conceptually simple to understand.

(iii) **Query capability :** Using simple query language (such as SQL) user can get information from the database or designer can manipulate the database structure.

(iv) **Easy design, maintenance and usage :** The relational models can be designed logically hence they are easy to maintain and use.

Disadvantages :

i) Relational model **requires powerful hardware** and large data storage devices.

ii) May lead to **slower processing time.**

iii) Poorly designed systems lead to **poor implementation** of database systems.

(2) Entity relationship model :

- As the name suggests the entity relationship model uses collection of basic objects called **entities** and **relationships.**
- The entity is a thing or object in the real world.
- The entity relationship model is widely used in database design.
- For example - Following is a representation of Entity Relationship model in which the relationship **works_for** is between entities **Employee** and **Department.**

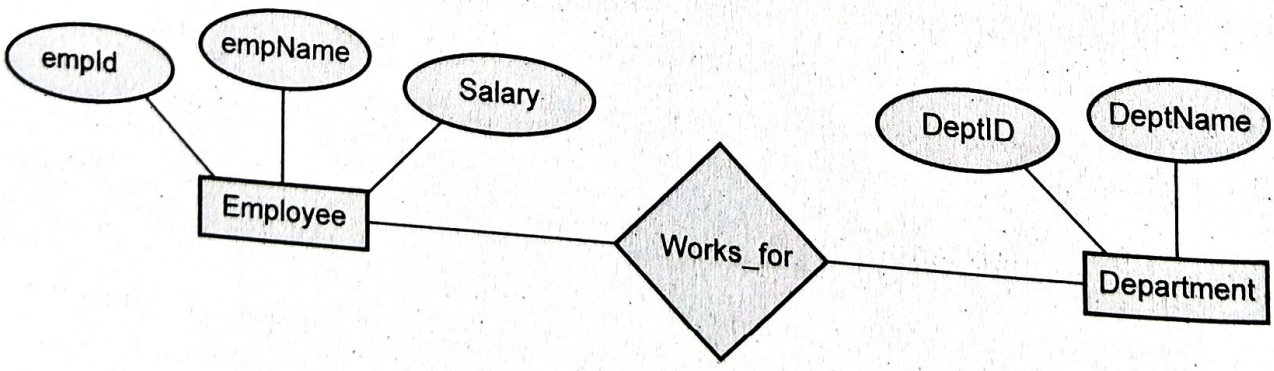


Fig. 1.5.1

Advantages :

i) **Simple :** It is simple to draw ER diagram when we know entities and relationships.

ii) **Easy to understand :** The design of ER diagram is very logical and hence they are easy to design and understand.

iii) **Effective :** It is effective communication tool.

iv) **Integrated :** The ER model can be easily integrated with Relational model.

v) **Easy conversion :** ER model can be converted easily into other type of models.

Disadvantages :

i) **Loss of information :** While drawing ER model some information can be hidden or lost.

ii) **Limited relationships :** The ER model can represent limited relationships as compared to other models.

iii) **No representation for data manipulation :** It is not possible to represent data manipulation in ER model.

iv) **No industry standard :** There is no industry standard for notations of ER diagram.

(3) Object Based Data Model :

- The **object oriented languages** like C++, Java, C# are becoming the dominant in software development.
- This led to object based data model.
- The object based data model combines **object oriented features** with relational data model.

Advantages :

i) **Enriched modelling :** The object based data model has capability of modelling the real world objects.

ii) **Reusability :** There are certain features of object oriented design such as inheritance, polymorphism which help in reusability.

iii) **Support for schema evolution :** There is a tight coupling between data and applications, hence there is strong support for schema evolution.

iv) **Improved performance :** Using object based data model there can be significant improvement in performance using object based data model.

Disadvantages :

i) **Lack of universal data model :** There is no universally agreed data model for an object based data model, and most models lack a theoretical foundation.

ii) **Lack of experience :** In comparison with relational database management the use of object based data model is limited. This model is more dependent on the skilled programmer.

iii) **Complex :** More functionalities present in object based data model make the design complex.

(4) Semi-structured data model :

- The semi-structured data model permits the **specification of data** where individual data items of same type may have **different sets of attributes.**
- The Extensible Markup Language **(XML)** is widely used to represent semi-structured data model.

Advantages

i) Data is not constrained by fixed schema.

ii) It is **flexible.**

iii) It is **portable.**

Disadvantage

i) **Queries** are **less efficient** than other types of data model.

(5) Hierarchical Model

- In this model each entity has only **one parent** but can have **several children.** At the top of hierarchy there is only one node called **root**. Refer Fig. 1.5.2.
- This model represents the relationship in 1 : N types. That means one university can have multiple courses. One course can have multiple projects and so on.

Advantage

1. This model groups the data into tables and defines the relationship between the tables.

Disadvantages

1. For searching any data, we have to start form the root and move downwards and visit each child node. Thus traversing through each node is required.
2. For addition of some information about child node, sometimes the parent information need to be modified.

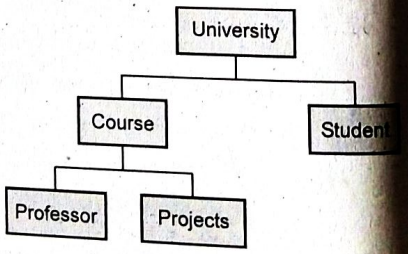


Fig. 1.5.2 Hierarchical model

3. It fails to handle many to many relationship (M : N) efficiently.
   It can cause duplication and data redundancy.

(6) Network Model

- This is enhanced version of hierarchical model. It overcomes the drawback of hierarchical model. It helps to address M : N relationship. That means, this model is not having single parent concept. Any child in this model can have multiple parents. Refer Fig. 1.5.3.
- The **main difference** between network model and hierarchical model is to allow many to many relationship.

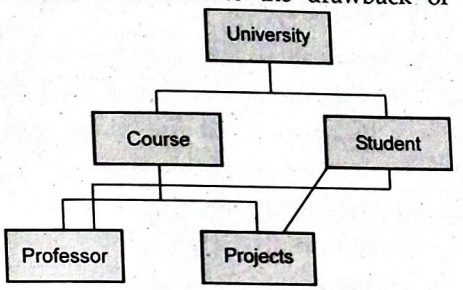


Fig. 1.5.3 Network model

Advantages

1. **Capability to handle more Relationships :** Since the network model allows many to many relationship, it helps in modeling the real life situations.
2. **Ease of data access :** The data access is easier and flexible than hierarchical model.
3. **Data Integrity :** In network model every member is associated with some other member in the model.
4. **Conformance to Standards :** The network model structure can be designed as per the standards.

Disadvantages

1. **Complex to implement :** For all the records the pointers need to be maintained, hence the database structure becomes complex.
2. **Complicated Operations :** The simple operations such as insertion, deletion and modification becomes complex due to adjustment of multiple pointer.
3. **Difficult to change structure :** The structural changes are difficult.

Review Question

1. *List out different Data Models. Explain any two.* **SPPU : Nov.-19, (End Sem), Marks 5**

1.6 Components of a DBMS

SPPU : Nov.-17,18, Marks 6

The functional components of database systems can be broadly classified into :

i) Storage manager and **ii) Query processor components.**

- Storage manager is to maintain the large storage space required for database.
- Query processor is helps to access the data of the database.

Let us first discuss the overall structure of DBMS.

1.6.1 Overall Structure of DBMS

Three Schema Architecture

- **Definition :** Database schema is a collection of database objects like tables, views, indexes and so on associated with one particular database username. This username is called the **schema owner**.
- For example Student Schema can be owner of STUDENT and MARKS tables. The Course schema can be the owner of SUBJECT table.
- The **goal** of three-schema architecture is to separate the user application from the physical database.
- The architecture of database is divided into **three levels** based on three types of schema - internal schema, conceptual schema or external schema.

1. Internal level :

- It contains **internal schema**.
- This schema represents the **physical storage structure of database**.
- This schema is maintained by the software and user is not allowed to modify it.
- This level is closest to the physical storage. It typically describes the record layout of the files and types of files, access paths etc.

2. Conceptual level :

- It contains **conceptual schema**.
- This schema **hides the details** of internal level.
- This level is also called as logical level as it contains the constructs used for designing the database.
- It contains information like table name, their columns, indexes and constraints, database operations.
- A representational data model is used to describe conceptual schema when a database system is implemented.

3. External level :

- It contains the external schema or user views.
- At this level, the user will get to see only the data stored in the database. Either they

will see whole data values or any specific records. They will not have any information about how they are stored in the database.

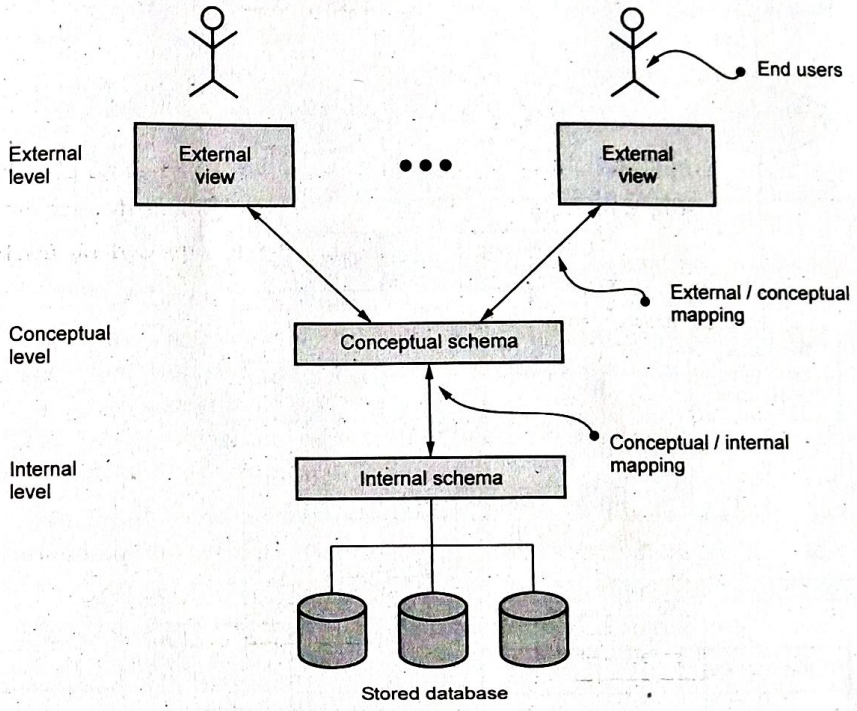


Fig. 1.6.1 Three schema architecture

- The processes of transforming requests and results between levels are called **mappings.**
- In the three schema architecture there are two mappings –

  1) External - Conceptual Mapping and

  2) Conceptual - Internal Mapping

1.6.2 Architecture of DBMS

- The typical structure of typical DBMS is based on relational data model as shown in Fig. 1.6.2. (Refer page 1-16).
- Consider the top part of Fig. 1.6.2. It shows **application interfaces** used by naïve users, application programs created by application programmers, query tools used by sophisticated users and administration tools used by database administrator
- The lowest part of the architecture is for **disk storage.**

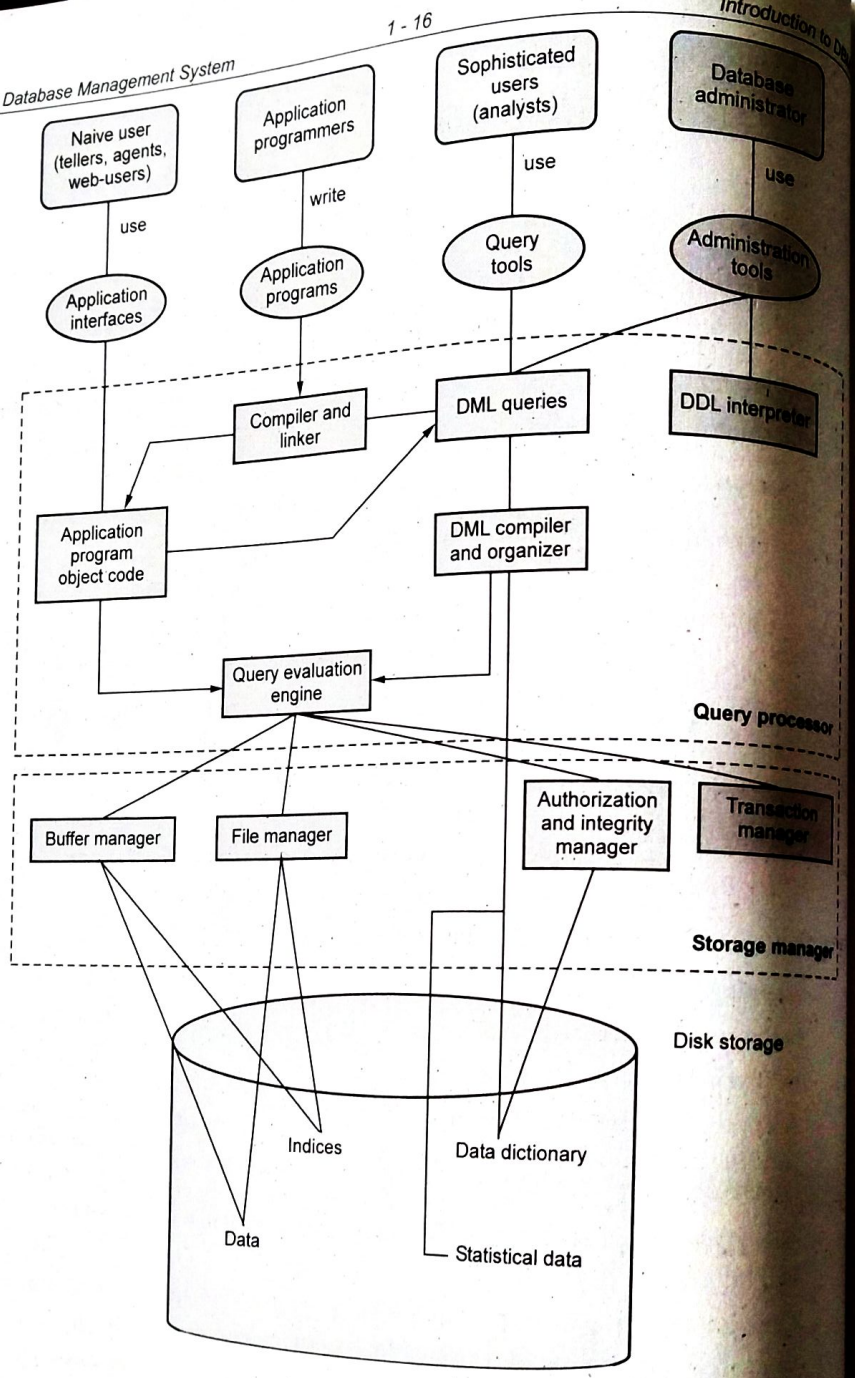


Fig. 1.6.2 Architecture of database

- The two important components of database architecture are - **Query processor and storage manager.**

Query processor :

- The interactive query processor helps the database system to simplify and facilitate access to data. It consists of DDL interpreter, DML compiler and query evaluation engine.
- With the following components of query processor, various functionalities are performed -
  - i) **DDL interpreter :** This is basically a translator which interprets the DDL statements in data dictionaries.
  - ii) **DML compiler :** It translates DML statements query language into an evaluation plan. This plan consists of the instructions which query evaluation engine understands.
  - iii) **Query evaluation engine :** It executes the low-level instructions generated by the DML compiler.
- When a user issues a query, the parsed query is presented to a query optimizer, which uses information about how the data is stored to produce an efficient execution plan for evaluating the query. An execution plan is a blueprint for evaluating a query. It is evaluated by **query evaluation engine**.

Storage Manager:

- Storage manager is the component of database system that provides interface between the low level data stored in the database and the application programs and queries submitted to the system.
- The storage manager is responsible for storing, retrieving, and updating data in the database. The storage manager components include -
  - i) **Authorization and integrity manager :** Validates the users who want to access the data and tests for integrity constraints.
  - ii) **Transaction manager :** Ensures that the database remains in consistent despite of system failures and concurrent transaction execution proceeds without conflicting.
  - iii) **File manager :** Manages allocation of space on disk storage and representation of the information on disk.
  - iv) **Buffer manager :** Manages the fetching of data from disk storage into main memory. The buffer manager also decides what data to cache in main memory. Buffer manager is a crucial part of database system.

- Storage manager implements several data structures such as -

i) **Data files :** Used for storing database itself.

ii) **Data dictionary :** Used for storing metadata, particularly schema of database.

iii) **Indices :** Indices are used to provide fast access to data items present in the database

1.6.3 Multi-user DBMS Architecture

Single Tier Architecture

- In this architecture application logic, presentation and data management all are combined in a single tier. For running the application, there was a use of mainframes. The application was accessible by **dumb terminals** that could perform only data input and display. It is as shown by following figure -

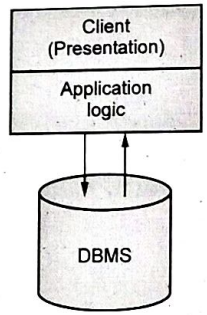


Fig. 1.6.3 Single tier architecture

- The advantage of this architecture is that it is easily maintained by central administrator.
- The main drawback of this architecture is that - it could support limited number of users.

Client Server Architecture

- The client server architecture is also called as **two tier architecture.**
- It consists of **Client computer** and **Server computer** which can communicate through well -defined protocol.

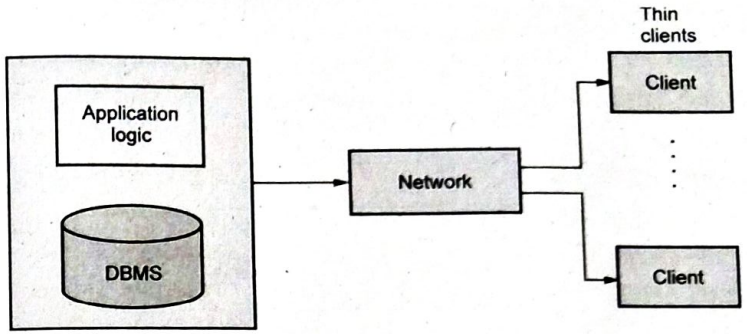


Fig. 1.6.4 Two server architecture

- In traditional client server architecture, the client computer implements simply the graphical user interface part. The Server computer implements application logic and data management part. In such architecture, the client is called as **thin client**.
- On the other hand there is some client server architecture, in which client implements graphical interface as well as business (application) logic part. Only data management part is taken care by server. Such client is called as **thick client.**
- The thick client model has several disadvantages as compared to thin client. Those are -
  - There is no central place to update and maintain the business logic as application program runs on the client computer only.
  - Client should run the business logic with reliability without affecting any security aspects.
  - The thick client architecture can handle only limited number of users.

Three Tier Architecture

The three tier architecture is made up of three tiers as –

(1) Presentation Tier : The presentation tier is comprised of graphical user interface. The user always expects the GUI which is easy to input. He/She expects the results in some organized format. The use web-based interfaces are getting popular.

(2) Middle Tier : This is a tier in which the application or business logic executes. The complex business processes get executed at this tier. The business logic can be implemented in some suitable programming language like C++ or Java.

(3) Data Management Tier : This tier takes care of data management activities. The database management (DBMS) systems are located in this architecture.

The typical three tier architecture is represented by Fig. 1.6.5.

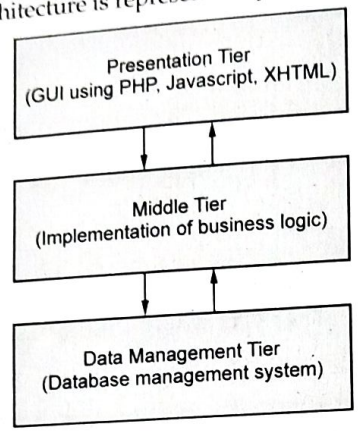


Fig. 1.6.5 Three tier architecture

Advantages of Three Tier Architecture

The Three tier architecture has following advantages –

(1) **Heterogeneous Systems :** Different platforms and different software components can be used at different tiers. Also it is easy to replace or modify some code present at any tier without affecting the other code.

(2) **Thin Clients :** Clients need enough computation power for presentation layer.

(3) **Integrated Data Access :** In many applications, data must be accessed from several sources. This is can be done transparently in three tier architecture by using the middle tier.

(4) **Scalability to Many Clients :** Multiple clients can access the system through middle tier.

(5) **Software Development Benefits :** By separating presentation, business logic and data management activities by means of separate tiers, it is easy to debug, maintain and modify the system as per the requirements. Interaction between tiers occur through well defined standardized APIs (Application Programming interfaces). Hence it is possible to create reusable components using three tier architecture.

Review Questions

1. *Draw and list different components of database system structure.* **SPPU : Nov.-17, (End Sem), Marks 4**
2. *Explain different database architectures.* **SPPU : Nov 17, (End Sem), Marks 6**
3. *Describe the three level architecture of DBMS. Explain how it is useful for achieving data independence.* **SPPU : Nov.-18, (End Sem), Marks 5**

1.7 Data Independence

SPPU : Nov.-19, Oct.-19, Marks 5

- **Definition :** Data independence is an ability by which one can change the data at one level without affecting the data at another level. Here level can be **physical, conceptual** or **external.**
- Data independence is one of the important characteristics of database management system.
- By this property, the structure of the database or the values stored in the database can be easily modified by without changing the application programs.
- There are two types of data independence

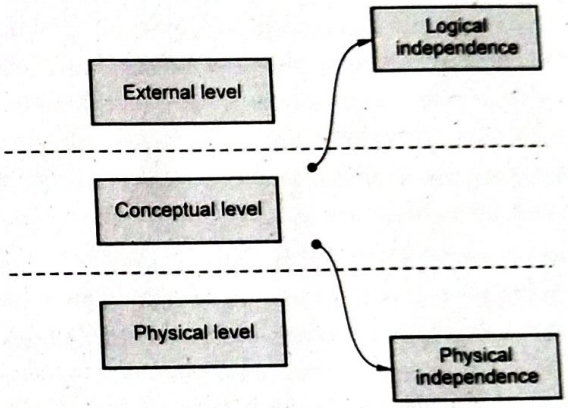


Fig. 1.7.1 Data independence

1. **Physical Independence :** This is a kind of data independence which allows the modification of physical schema without requiring any change to the conceptual schema. For example - if there is any change in memory size of database server then it will not affect the logical structure of any data object.
2. **Logical Independence :** This is a kind of data independence which allows the modification of conceptual schema without requiring any change to the external schema. For example - Any change in the table structure such as addition or deletion of some column does not affect user views.

- By these data independence the time and cost acquired by changes in any one level can be reduced and abstract view of data can be provided to the user.

Review Questions

1. *Describe the three level architecture of DBMS. Explain how is it useful for achieving data independence.* **SPPU : Nov.-19, (End Sem), Marks 5**
2. *Differentiate between logical and physical data independence.* **SPPU : Oct.-19, (In Sem), Marks 2**

SPPU : May 19, Nov.-18, Marks 5

1.8 Database Users

There **four different types** of database system **users** differentiated by the way they interact with the system. Different types of user interfaces for different types of users are -

i) **Naïve users :** This type of users interact with the system with the help of previously created program(known as application program). Typically a form interface is used by this type of user to interact with the system. For example - we may feel up the booking form for booking a ticket on an online system.

ii) **Application programmers :** These are computer professionals who write application programs. Normally Rapid Application Development(RAD) tools are used to quickly design forms and reports.

iii) **Sophisticated users :** These are the type of users who interact with the system without writing programs. These users may submit the database query to retrieve the desired information or tools from data analysis software. Database analysts fall in this category of database users.

iv) **Specialized users :** Specialized users are sophisticated users who write specialized database application that does not fit into the traditional data-processing framework. Among these applications are computer aided-design systems, knowledge-base and expert systems etc.

Responsibilities of DBA

Database Administrator (DBA) is a person who have a **central control** over both data and programs that access data in DBMS. The **functions of DBA** are -

i) **Schema definition :** DBA creates a database Schema using DDL statements.

ii) **Schema and physical organization modification :** In order to improve the overall performance of database management system DBA carried out changes in schema or physical organization.

iii) **Granting authorization for data access :** Granting authorization for data access means giving special permissions to the users for accessing the database. This task is done by DBA so that privacy of database can be maintained.

iv) **Maintenance :** Maintenance involve different types of tasks such as monitoring jobs and their performances, taking periodic back-ups, making free space available for normal operations or upgrading disk space as per the requirements. These all tasks are carried out by DBA.

Example 1.8.1 *Explain the problems that may arrive if the DBA does not discharge the responsibilities properly.* **SPPU : May 19, (End Sem), Marks 5**

Solution : Following are the problems that may arrive if DBA does not discharge the responsibilities properly -

1) The database can not perform without file manger interaction. If nothing is stored in the files then obviously we can not retrieve anything.
2) The consistency in database must operations must be maintained. If it is not, then it will create major problems. For instance – account balance may go below the minimum allowed, employees can earn too much overtime and so on.
3) If authorization for the authentic user is not done, then unauthorized users may access the database or users authorized to access part of the database may be able to access parts of the database for which they lack authority.
4) Data can be lost permanently.
5) Consistency constraints may be violated despite proper integrity enforcement in each transaction. For example, incorrect bank balances might be reflected due to simultaneous withdrawals and deposits, and so on.

Review Question

1. *List the responsibilities of DBA.* **SPPU : Nov.-18, (End Sem), Marks 5**

1.9 System Catalogs

- The system catalogue is a collection of tables and views that contain important information about a database or metadata(meta data means: data about data).
- For each scheme, following things must be at-least included –
  - The names of the tables in the database
  - The names of columns of each table
  - The data type of each column
  - The constraints present on the tables(primary key,NULL, NOT NULL, and some other types of constraints)
  - The access privileges for the elements of the databases.

- The storage structures and indexes

Such a database is often called as **system catalog**.

- Basically the system catalog defines the structure of database. It is sometimes referred as data dictionary.
- The system catalogue is basically a group of objects that contain information that defines other objects in the database.
- For example – Consider Student table as given below and corresponding system catalog.

Roll	Name	RegID	CourseID
101	Ankita	AA123	101
102	Aniket	AA227	105
103	Ashwini	SA823	101
104	Siddharth	HQ256	103
105	Nandan	YG491	105

System catlog (Meta data)

Table name	Column	Data type	Description
Student	Roll	int	Primary key of table
Student	Name	varchar(30)	Student's name
Student	RegID	varchar(10)	Registration ID
Student	CourseID	int	Course ID, Ref : Course table

Part II: Data Modelling

1.10 Basic Concepts

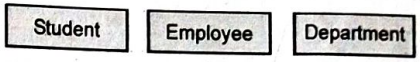
SPPU : May-18, Nov.-17,19, Marks 5

- Data Modelling in database management system is based on Entity Relationship modelling(ER Model)
- Entity Relational model is a model for identifying entities to be represented in the database and representation of how those entities are related.
- ER data model represents the overall **logical structure** of database.
- The E-R model is very useful in mapping the meanings and interactions of real-world entities onto a conceptual schema or database.
- The ER model consists of three basic concepts –

1. Entities 2. Relationships 3. Attributes

1.10.1 Entity and Entity Sets

- **Entity :** An entity is an object that exists and is distinguishable from other objects. For example - Student named "Poonam" is an entity and can be identified by her name. The entity can be concrete or abstract. The concrete entity can be - Person, Book, Bank. The abstract entity can be like - holiday, concept entity is represented as a box.



- **Entity set :** The entity set is a set of entities of the same types. For example - All students studying in class X of the School. The entity set need not be disjoint. Each entity in entity set have the same set of attributes and the set of attributes will distinguish it from other entity sets. No other entity set will have exactly the same set of attributes.

1.10.2 Relationship and Relationship Sets

- **Relationship:** Relationship is an association among two or more entities.
- **Relationship Set:** The **relationship set** is a collection of similar relationships. For example - Following Fig. 1.10.1 shows the relationship **works_for** for the two entities **Employee** and **Departments.**

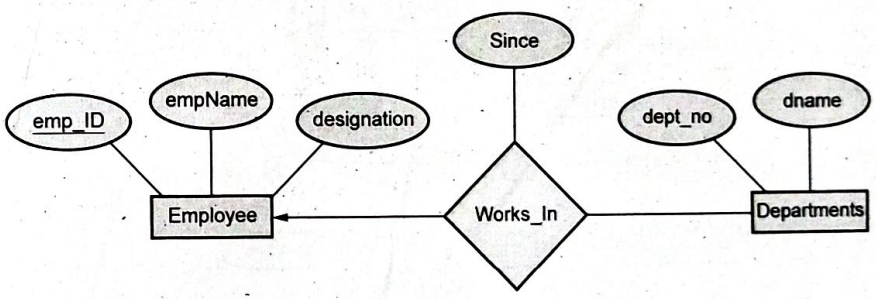


Fig. 1.10.1 : Relation set

- The association between entity sets is called as **participation**. that is, the entity sets E1, E2, . . . , En participate in relationship set R.
- The function that an entity plays in a relationship is called that entity's **role**.

1.10.3 Attributes

- Attributes define the properties of a data object of entity. For example: if student is an entity, his ID, name, address, date of birth, class are its attributes. The attributes help in determining the unique entity. Refer Fig. 1.10.2 for Student entity set with

attributes - ID, name, address. Note that entity is shown by rectangular box and attributes are shown in oval. The primary key is underlined.

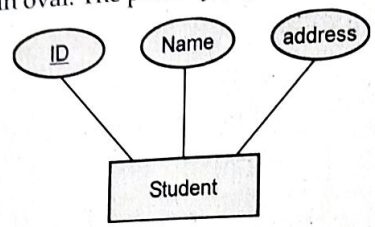


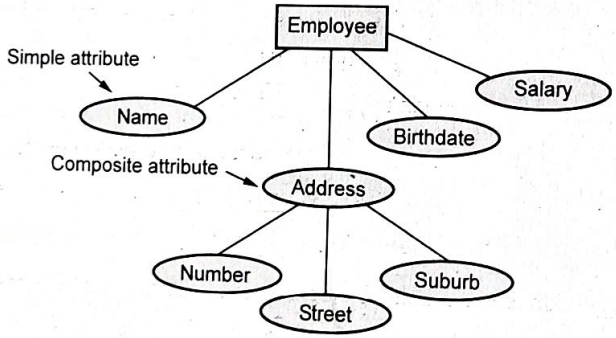
Fig. 1.10.2 : Student entity set with attributes

1.10.3.1 Types of Attributes

Following are the types of attributes -

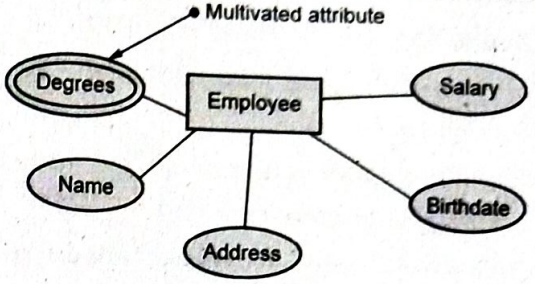
1) Simple and Composite Attributes :

1) Simple attributes are attributes that are drawn from the atomic value domains
   For example - Name = {Parth} ; Age = {23}
2) Composite attributes: Attributes that consist of a hierarchy of attributes
   For example – Address may consists of "Number", "Street" and "Suburb"
   Hence, **Address = {59 + 'JM Road' + 'ShivajiNagar'}**



2) Single valued and multivalued :

- There are some attributes that can be represented using a single value. For example - StudentID attribute for a Student is specific only one studentID.
- Multivalued attributes : Attributes that have a set of values for each entity. It is represented by concentric ovals
  For example - Degrees of a person: ' BSc' , 'MTech', 'PhD'



3) Derived attribute :

Derived attributes are the attributes that contain values that are calculated from other attributes. To represent derived attribute there is dotted ellipse inside the solid ellipse. For example – Age can be derived from attribute DateOfBirth. In this situation, DateOfBirth might be called Stored Attribute.

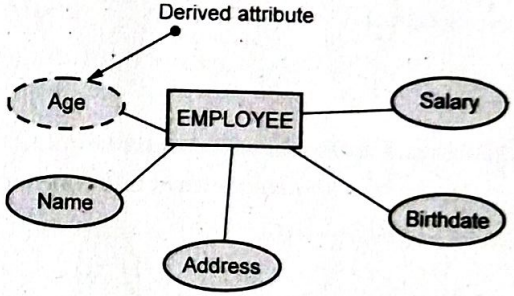


Fig. 1.10.3

Review Questions

1. *Define entity and entity-set. What is E-R model?* **SPPU : May-18. (End Sem). Marks 5**
2. *Explain different types of attributes of an entity with example.* **SPPU : Nov.-17. (End Sem). Marks 3; Nov.-19. (End Sem). Marks 4**

1.11 Constraints

- Relationship types have certain rules that limit the possible combination of entities that can take part in relationship. These rules or restrictions are called **structural constraints.**
- The common type of structural constraint is represented by the **cardinality ratio.**
- The **cardinality ratio** for a binary relationship specifies the **maximum** number of relationship instances that an entity can participate in.

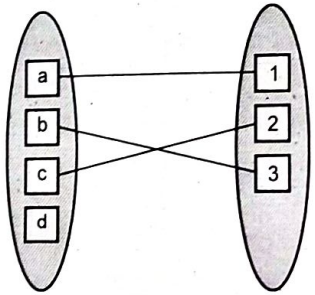
1.11.1 Types of Cardinality

Mapping Cardinality represents the number of entities to which another entity can be associated via a relationship set.

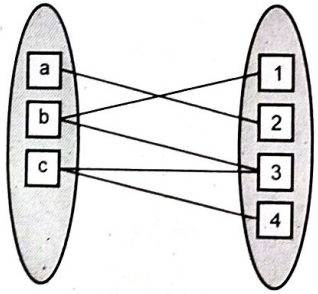
The mapping cardinalities are used in representing the binary relationship sets.

Various types of mapping cardinalities are -

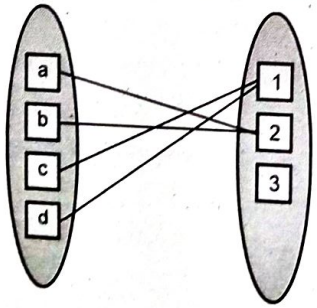
1. **One to One :** An entity A is associated with at least one entity on B and an entity B is associated with at one entity on A. This can be represented as



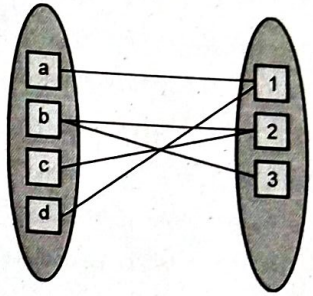
2. **One to Many :** An entity in A is associated with any number of entities in B. An entity in B, however, can be associated with at most one entity in A.



3. **Many to One :** An entity in A is associated with at most one entity in B. An entity in B, however, can be associated with any number of entities in A.



4. **Many to many :** An entity in A is associated with any number (zero or more) of entities in B, and an entity in B is associated with any number (zero or more) of entities in A.



1.12 Keys

- Keys are used to identify entities uniquely from the given entity set.
- A key can be a an attribute or a set of attributes that help us to identify the entity uniquely.
- Keys also help to identify relationships uniquely, and thus distinguish relationships from each other.
- The primary key of an entity set allows us to distinguish among the various entities of the set.
- For example – If a Student table contains the information about various students as given below –

RollNo	Name	City	Course
101	Ram	Pune	Computer
102	Sita	Pune	Electronics
103	Laxman	Chennai	Mechanical

In above table, **RollNo** is a primary key because, it uniquely identifies the student record.

1.13 Multiple Choice Questions

Q.1 A DBMS provides users with the conceptual representation of _____.

a register b data

c logical view d physical view

Q.2 DBMS helps to achieve _____.

[a] data Independence [b] centralized Control of Data
[c] neither a nor b [d] both a and b

Q.3 In view of total database content is _____.

[a] conceptual view [b] internal view
[c] external view [d] physical view

Q.4 The main purpose of DBMS is to provide _____ view of data to user.

[a] completer [b] abstract
[c] partial [d] none of these

Q.5 ___ means to hide certain details of how data is stored.

[a] Data Integrity [b] Data independence
[c] Data abstraction [d] Data separation

Q.6 How many levels of data abstraction are there ?

[a] One [b] Two
[c] Three [d] Four

Q.7 A _____view of data expresses the way a user thinks about data _____.

[a] logical view [b] physical view
[c] both [d] none

Q.8 A physical view of data refers to the way data is handled at a_____ its storage and retrieval.

[a] high level [b] low level
[c] medium level [d] all of these

Q.9 Architecture of the database can be viewed as _____.

[a] two levels [b] three levels
[c] four levels [d] one level

Q.10 In the architecture of a database system external level is the _____.

[a] physical level [b] logical level
[c] conceptual level [d] view level

Q.11 In hierarchical model records are organized as _____.

[a] lists [b] links
[c] tree [d] graph

Q.12 There are ___ levels of data independence

[a] one [b] two
[c] three [d] four

Q.13 The ability to modify the schema of database in one level without affecting the schema definition in higher level is called as _____.

[a] data isolation [b] data abstraction
[c] data hiding [d] data independence

Q.14 Which of the following is record based on logical model ?

[a] Network Model [b] Object Oriented Model
[c] E-R Model [d] None of these

Q.15 The DDL is used to specify the_____.

[a] conceptual schemas [b] internal schemas
[c] both [d] none

Q.16 DCL stands for _____.

[a] Data Control Language [b] Data Console Language
[c] Data Console Level [d] Data Control Level

Q.17 Which of the following is / are the DDL statements ?

[a] Create [b] Drop
[c] Alter [d] All of the above

Q.18 Which are the three levels of abstraction ?

[a] Physical [b] Logical
[c] External [d] All of these

Q.19 The statement in SQL which allows to change the definition of a table is _____.

[a] create [b] alter
[c] select [d] update

Q.20 Which of the following is NOT a basic element of all versions of the E - R model ?

[a] Entities [b] Relationships
[c] Attributes [d] Primary key

Q.21 Data independence means _____.

a data is defined separately and not included in programs

b programs are not dependent on the physical attributes of data.

c programs are not dependent on the logical attributes of data

d both (b) and (c)

Q.22 E-R model uses this symbol to represent weak entity set _____.

a dotted rectangle

b diamond

c doubly outlined rectangle

d none of these

Q.23 _____ express the number of entities to which another entity can be associated via a relationship set.

a Mapping cardinality

b Relational cardinality

c Participation constraints

d None of the mentioned

Q.24 In E-R diagram derived attribute is represented by _____.

a rectangle

b circle

c dashed Ellipse

d diamond

Q.25 DBA stands for _____.

a Data Building Administrator

b Database Access

c Database Authentication

d Database Administrator

Q.26 _____ represents the number of entities to which another entity can be associated

a Degree

b Cardinality

c Modality

d None of these

Q.27 Data Model is collection of conceptual tools for describing _____.

a Data

b Schema

c constraints

d All of the above

Q.28 Which of the following is example of Object based logical model ?

a Relational Model

b Hierarchical Model

c Network Model

d Entity Relationship Model

Q.29 Entity Relationship model consists of collection of basic objects called _____ and relationship among these objects.

a functions

b models

c entity

d all of these

Answer Keys for Multiple Choice Questions :

Q.1	b	Q.2	d	Q.3	a	Q.4	b
Q.5	c	Q.6	c	Q.7	a	Q.8	b
Q.9	b	Q.10	d	Q.11	c	Q.12	b
Q.13	d	Q.14	a	Q.15	a	Q.16	a
Q.17	d	Q.18	d	Q.19	b	Q.20	d
Q.21	d	Q.22	c	Q.23	a	Q.24	c
Q.25	d	Q.26	b	Q.27	d	Q.28	d
Q.29	c						

❑❑❑