

BABU MADHAV INSTITUTE OF INFORMATION TECHNOLOGY, UTU Integrated M.Sc.(IT)

Semester-I

060010110 | CC2 Database Management Systems | Question Bank-Unit: 02

Unit-2: Database Management System

Short Questions [1 Mark]

1. What is a database?

Ans:

4 Database:

- ➤ A database is a collection of information that is organized so that it can easily be accessed, managed and updated.
- ➤ The collected information could be in any formats like (electronic, printed, graphic, audio, video).

Examples:

- Phone book
- Address book
- Shopping list



2. Explain role of Database Administrator.

Ans:

- ➤ Database Administrator is the person who is responsible to manage the database.
- ➤ The DBA is responsible for...
 - 1. Authorizing access to the database
 - 2. Monitoring the use of database
 - 3. Maintain security
 - 4. Maintain proper performance

3. Define Database Management System.

Ans:

A database Management System is a collection of database and programs that enable you to store, modify, and extract information from a database.

➤ For Example:

- Oracle (Oracle Corporation)
- SQL Server (Microsoft)
- DB2 (IBM)

4. What is the goal of Database Management System?

Ans: The main goal of DBMS is to provide efficiency and convenience in storing and retrieving database information and also provides safety of the information against system crashes as well as an unauthorized access.

5. What is data dictionary?

Ans: Data dictionary stores metadata. Metadata means data about data. In data dictionary various information about the database are stored like attribute names, data type, length etc.

- ➤ Data dictionary is set of read-only tables that provide information about the database. A data dictionary contains metadata. Metadata means data about data.
- ➤ It stores...
 - 1. The definitions of all tables and views (attribute name, data type, size etc)
 - 2. Default values for columns
 - 3. Integrity constraint information

6. Define metadata.

Ans: > The data about data is called metadata. In databases, metadata describes the structure and definition of all objects (like tables, views, index, procedure, trigger etc).

- Metadata is used by the system for managing the database and query processing.
- For Example:
 - Metadata about a table may include name of attributes, data types, size etc.

7. What is the difference between database and table?

Ans: A database is defined as a collection of logically related tables stored together that are designed to meet the information needs of an organization.

➤ A table is a collection of columns and rows where columns are the attributes and rows contain the values for those attributes.

8. Differentiate Physical Schema and Logical Schema.

Ans: Physical schema describes the database design at the physical level, while the logical schema describes the database design at logical level.

9. Which are the components of Storage Manager?

Ans: The components of the storage manager are:

- 1. Authorization and integrity manager
- 2. Transaction manager
- 3. File manager
- 4. Buffer manager

10. What is Domain Constraints?

Ans: A domain is a set of values that may be assigned to an attribute. All values that appear in a column of a relation must be taken from the same domain. Domain constraint allows to define the domain of an attribute.

11. List the levels of data abstraction.

Ans: There are three levels of data abstraction:

- 1. Physical level
- 2. Logical level
- 3. View level

12. Define Data Model.

Ans: \(\rightarrow \) Underlying structure of the database is called as data models.

- A data model is a collection of conceptual tools for describing data, data relationships, data semantics and consistency constraints.
- ➤ It is a way to describe the design of the database at physical, logical and view level.

13. Differentiate Schema and Instances.

or

State the major difference between the term's relation and relation schema.

Ans:

Relation Schema:

The schema of a relation refers to its logical design.

Relation Instance:

An instance of the relation refers to data (records) stored into relation.

Schema is analogous to variable declaration while instance is analogous to the value of a variable.

For Example:

STUDENT(EN, NAME, COURSE) is relation schema that represent the logical design of table.

A record (201, AJAY, BCA-H) stored in STUDENT table is an instance. It represents one student.

14. What is conceptual schema?

The schemas at the view level are called subschemas that describe different views of the Ans: database.

15. What is an attribute?

An attribute is a property or characteristic (field) of an entity. Ans:

16. Define relation in terms of RDBMS.

In the relational model the term relation is used to refer to a table. Ans:

Short Questions [2 Marks]

Differentiate Data and Information. 1.

Ans:

	<u>DATA</u>		<u>INFORMATION</u>
1.	Data is a raw fact or material	1.	This is processed from the data
2.	It may not be meaningful	2.	It is always meaningful
3.	It is used for analysis and reasoning	3.	The result of analysis and reasoning
4.	It is input processing unit	4.	It is output processing unit
5.	It may be accurate but not necessarily passes other characteristics of information	5.	It is relevant, timely, accurate, reliable, concise, and completes
6.	Ex:	6.	Ex:
	Marks of the student 1 account number		NO. BRANCH STUDENT
			a. 1 CE 60
			b. 2 ME 50
			Information of student, branch and no
			of student study in college.

2. List four applications you have used that most likely employed a database system to store persistent data.

or

List any four applications of DBMS.

- **Ans:** 1. Banking: For account information, transfer of funds, banking transactions.
 - 2. Universities: For student information, online assignment submissions, course registrations, and grades.
 - 3. Airlines: For reservation of tickets, and schedule information.
 - 4. Online news sites: For updating new, maintenance of archives.

5. Online-trade: For product data, availability and pricing information, order-tracking facilities, and generating recommendation lists.

3. What are the disadvantages of file processing system?

Ans:

- > The disadvantages of file processing systems are
 - 1. Data redundancy and inconsistency
 - 2. Difficulty in accessing data
 - 3. Data isolation
 - 4. Integrity problems
 - 5. Atomicity problems
 - 6. Concurrent access anomalies

4. List five responsibilities of a database management system.

Ans:

- The five responsibilities of a database management system are as follows:
 - 1. Interaction with the file manager
 - 2. Integrity enforcement
 - 3. Security enforcement
 - 4. Backup and recovery
 - 5. Concurrency control

5. Define data abstraction.

Ans:

It means to hide certain details which are not required. Abstraction allows the programmer to concentrate on the relevant information.

6. Define: 1) DDL 2) DML

Ans:

DDL:

- > DDL stands for Data Definition Language.
- ➤ DDL specifies the database schema and some additional properties to data.
- For Example:

Create, Alter, Rename and Drop statements.

♣ DML:

- > DML stands for Data Manipulation Language.
- > DML statements help users to access and manipulate data.
- ➤ For Example:

Insert, Update, Select and Delete statements.

7. Write at least four differences between database schema and instance.

Ans:				Schema					Instance
	1.	The	overall	logical	design	of	the	1.	The collection of information stored in
		databa	ase.						the database at a particular moment.
	2.	It incl	ludes tab	ole names	s, columi	n nai	nes,	2.	It includes actual data or information
		data-t	ypes and	l size of	columns	, var	ious		stored in tables in form of different
		constr	raints at l	logical le	vel.				records/rows.
	3.	It doe	s not cha	inge freq	uently.			3.	It changes frequently.
	4.	It cha	anges w	hen inse	ertion/del	etior	n of	4.	It changes when insert, delete or
		tables	or colu	mns; cha	nge in d	ata-t	ype,		update operation on data stored in
		size o	r constra	int on an	y columr	occ	urs.		database occurs.
	5.	It is a	nalogous	to varial	ole decla	ratio	n.	5.	It is analogous to value of the variable.

8. What are the categories of data model?

Ans: Data model is divided into two categories.

- 1. <u>Logical Data Model:</u>
 - For Example: E-R Model, Relational Model, Object Oriented Model, etc.
- 2. Physical Data Model:
 - Example: Semantic Model, Frame Model.

9. Write the purpose of storage manager.

- Ans: > The storage manager translates the various DML statements into low-level file system commands.
 - > Thus, the storage manager is responsible for storing, retrieving, and updating data in the database.
 - ➤ The storage manager is responsible for the interaction with the file manager

10. Differentiate Naive users and Sophisticated users.

Ans: > Naive users:

These users interact with the system by invoking one of the application programs that have been written previously (typical users of form interface).

> Sophisticated users:

These types of users interact with the system without writing programs; they form their requests in a database query language.

11. What are five main functions of a database administrator?

- **Ans:** > Following are the functions of DBA:
 - 1. To backup data.
 - 2. In some cases, to create the schema definition.
 - 3. To define the storage structure and access methods.
 - 4. To modify the schema and/or physical organization when necessary.
 - 5. To grant authorization for data access.

12. Differentiate Two-tier and Three-tier architecture. Which is better suited for web applications? Why?

Ans:

- In two tier architecture, the application is partitioned into a component that resides at the client machine and invokes database functionality at the server machine through query language.
- > In three tier architecture, the client machines act as a front end and do not contain any direct database calls.
- Three tier applications are more appropriate for applications that run in the World Wide Web because business logic of application server is embedded in application server, instead of being distributed across multiple clients.

13. Explain the concept of physical data independence, and it's importance in database systems.

Ans:

- > Physical data independence is the ability to modify the physical scheme without making it necessary to rewrite application programs.
- > Such modifications include changing from unblocked to blocked record storage, or from sequential to random access files.
- > Such a modification might be adding a field to a record; an application program's view hides this change from the program.

14. How Physical Data Independence is easy to achieve as compared to Logical Data **Independence?**

- > Physical data independence exists in most databases and file environments where physical details such as the exact location of data on disk, and hardware details of storage encoding, placement, compression, splitting, merging of records, and so on are hidden from the user. Applications remain unaware of these details.
- > On the other hand, logical data independence is harder to achieve because it allows structural and constraint changes without affecting application programs—a much stricter requirement.

15. What is referential integrity? Give relevant example.

Ans: A value that appears in one relation for a given set of attributes also appears for a certain set of attributes in another relation. This condition is called referential integrity.

➤ Referential integrity is a relational database concept in which multiple tables share a relationship based on the data stored in the tables, and that relationship must remain consistent.

For Example:

artist_id	artist_name
1	Bono
2	Cher
3	Nuno Bettencourt

 Link Broken

 artist_id
 album_id
 album_name

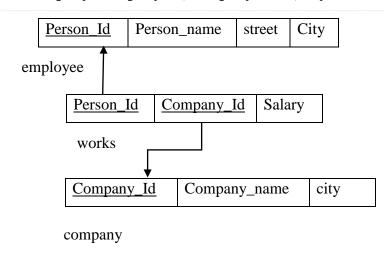
 3
 1
 Schizophonic

 4
 2
 Eat the rich

 3
 3
 Crave (single)

16. Consider the following relational database. Identify the primary key and foreign key in each relation.

employee (person_Id, person name, street, city)
works (person_Id, company_id, salary)
company (company_id, company name, city)



Primary Key: person_Id in employee table and Company_Id in company table are primary keys.

Foreign key: person_Id and Company_Id in works table are foreign keys.

Scenario based Questions [5 Marks]

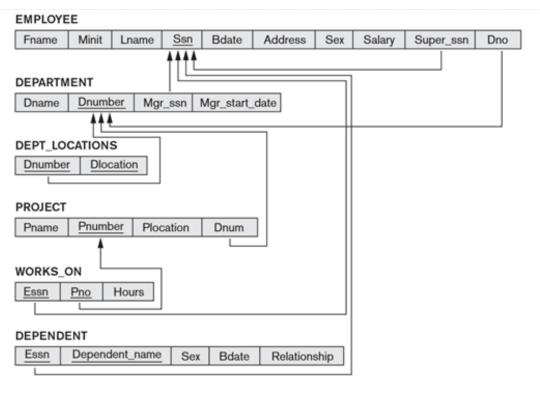
1. List six major steps that you would take in setting up a database for a particular enterprise.

Ans: Six major steps in setting up a database for a particular enterprise are:

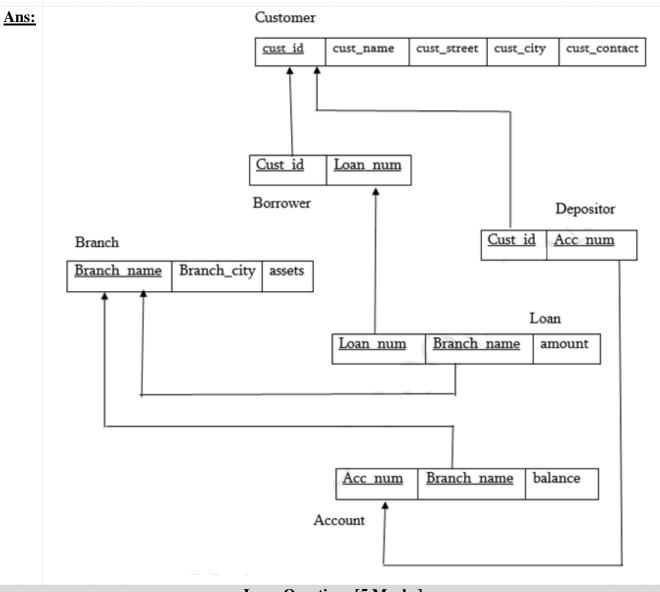
- 1. Define the high-level requirements of the enterprise (this step generates a document known as the system requirements specification.)
- 2. Define a model containing all appropriate types of data and data relationships.
- 3. Define the integrity constraints on the data.
- 4. Define the physical level.
- 5. For each known problem to be solved on a regular basis (e.g., tasks to be carried out by clerks or Web users) define a user interface to carry out the task, and write the necessary application programs to implement the user interface.
- 6. Create/initialize the database.
- 2. Suppose you want to build a video site similar to YouTube. Consider each of the disadvantages of keeping data in a file-processing system. Discuss the relevance of each of these points to the storage of actual video data and to metadata about the video, such as title, the user who uploaded it, tags, and which users viewed it.
- Ans: 1. Data redundancy and inconsistency: This would be relevant to metadata to some extent, although not to the actual video data, which is not updated. There are very few relationships here, and none of them can lead to redundancy.
 - 2. **Difficulty in accessing data:** If an organization needs to find video data based on specific search conditions (beyond simple keyword queries) if metadata were stored in files it would be hard to find relevant data without writing application programs. Using a database would be important for the task of finding data.
 - 3. **Data isolation:** Since data is not usually updated, but instead newly created, data isolation is not a major issue. Even the task of keeping track of who has viewed what videos is (conceptually) appended only, again making isolation not a major issue. However, if authorization is added, there may be some issues of concurrent updates to authorization

information.

- 4. **Integrity problems:** It seems unlikely there are significant integrity constraints in this application, except for primary keys. If the data is distributed, there may be issues in enforcing primary key constraints.
- 5. **Atomicity problems:** When a video is uploaded, metadata about the video and the video should be added atomically, otherwise there would be an inconsistency in the data. An underlying recovery mechanism would be required to ensure atomicity in the event of failures.
- 6. **Concurrent-access anomalies**: Since data is not updated, concurrent access anomalies would be unlikely to occur.
- 7. Security problems: These would be an issue if the system supported authorization.
- 3. Assume the database system of Company. Write relational database schema for Employee, Department, Department Locations, Project and Dependent with appropriate attributes. Also show the primary key and its referential keys.



- 4. Consider the bank database:
 - branch (branch name, branch city, assets)
 - customer (cust_id, cust_name, cust_street, cust_city, cust_contact)
 - loan (loan number, branch name, amount)
 - borrower (cust_id, loan number)
 - account (account number, branch name, balance)
 - depositor (cust_id, account number)
 - a. What are the appropriate primary keys?
 - b. Given your choice of primary keys, identify appropriate foreign keys.



Long Questions [5 Marks]

1. Write the drawbacks of conventional File Processing Systems.

or

What are the disadvantages of File Processing Systems?

Ans: The Drawbacks of Conventional File Processing System are:

1) Data redundancy and inconsistency:

- Since the files and application programs are created by different programmers over a long period of time, the files have different formats and the programs may be written in several programming language. The same piece of information may be duplicated in several files.
- <u>For Example:</u> The address and phone number of particular customer may appear in a file that consists of personal information and in saving account records file also.
- This redundancy leads to data inconsistency that is, the various copies of the same data may no longer agree.
- <u>For Example:</u> A changed customer address may be reflected in personal information file, but not in saving account records file.

2) Difficulty in accessing data:

- Conventional file processing environments do not allow needed data to be retrieved in a convenient and efficient manner.
- For Example: Suppose that bank officer needs to find out the names of all customers who live within the city's 411027 zip code. The bank officer has now two choices: Either get the list of customers and extract the needed information manually, or ask the data processing department to have a system programmer write the necessary application program. Both alternatives are unsatisfactory.

3) Data isolation:

• Since, data is scattered in various files, and files may be in different formats, it is difficult to write new application programs to retrieve appropriate data.

4) Concurrent access anomalies:

- In order to improve the overall performance of the system and obtain a faster response time many systems allow multiple users to update the data simultaneously.
- In such environment, interaction of concurrent updates may results in inconsistent data.

• For Example: Consider bank account A, with \$500.If two customers with draw funds (say \$50 and \$100 respectively) from account A at the same time, the result of the concurrent executions \$400, rather than \$350. In order to guard against this possibility, some form of supervision must be maintained in the system.

5) Atomicity Problem:

- System failure will lead to atomicity problem.
- <u>For Example:</u> Failure during transfer of fund from system A to A. It will be debited from A but not credited to B leading to wrong transaction.

6) Concurrent Access Anomalies:

- In order to improve the overall performance of the system and obtain a faster response time many systems allow multiple users to update the data simultaneously.
- In such environment, interaction of concurrent updates may result in inconsistent data.
- For Example: Consider bank account A, containing \$500. If two customers withdraw funds say \$50 and \$100 respectively) from account A at about the same time, the result of the concurrent executions may leave the account in an incorrect (or inconsistent) state. Balance will be \$400 instead of \$350. To protect against this possibility, the system must maintain some form of supervision.

7) Security problems:

- Not every user of the database system should be able to access all the data. System should be protected using proper security.
- For Example: In a banking system, pay roll personnel should be only given authority to see the part of the database that has information about the various bank employees. They do not need access to information about customer accounts.
- Since application programs added to the system in an ad-hoc manner, it is difficult to enforce such security constraints.

8) Integrity problems:

- The data values stored in the database must satisfy certain types of consistency constrains.
- <u>For Example:</u> The balance of a bank account may never fall below a prescribed amount (say \$100). These constraints are enforced in the system by adding appropriate code in the various application programs.

2. Explain the merits of using Database Management System.

Ans: 4 The merits of DBMS:

1. <u>Centralized Management and Control</u>:

One of the main advantages of using a database system is that the organization can exert, via the DBA, centralized management and control over the data.

2. Reduction of Redundancies and Inconsistencies:

Centralized control avoids unnecessary duplication of data and effectively reduces the total amount of data storage required. Removing redundancy eliminates inconsistencies.

3. Data Sharing:

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

4. **Data Integrity:**

Data integrity means that the data contained in the database is both accurate and consistent. Centralized control can also ensure that adequate checks are incorporated in the DBMS to provide data integrity.

5. Data Security:

Data is of vital importance to an organization and may be confidential. Such confidential data must not be accessed by unauthorized persons. The DBA who has the ultimate responsibility for the data in the DBMS can ensure that proper access procedures are followed. Different levels of security could be implemented for various types of data and operations.

6. Data Independence:

Data independence is the capacity to change the schema at one level of a database system without having to change the schema at the next level.

It is usually considered from two points of view:

- 1. Physical data independence and
- 2. Logical data independence.

Physical data independence is the capacity to change the internal schema without having to change conceptual schema.

Logical data independence is the capacity to change the conceptual schema without having to change external schemas or application programs.

3. Explain the storage manager and query processor in the structure of DBMS.

Ans: Storage Manager:

- A storage manager is a program module that provides the interface between the lowlevel data stored in the database and the application programs and queries submitted to the system.
- The storage manager is responsible for the interaction with the file manager. The storage manager translates the various DML statements into low-level file system commands.

Components of the storage manager are:

- 1. **Authorization and integrity manager:** It test for satisfaction of various integrity constraints and checks the authority of users accessing the data.
- 2. **Transaction manager:** It ensures that the database remains in a consistent state despite system failures, and concurrent executions proceed without conflicting.
- 3. **File manager:** It manages the allocation of space on disk storage and the data structures used to represent information stored on disk.
- 4. **Buffer manager:** It is responsible for fetching data from disk storage into main memory and to decide what data to cache in main memory. It enables the database to handle data sizes that are much larger than the size of the main memory.

The storage manager implements several data structures as part of physical system implementation.

Query Processor:

• The query processor is an important part of the database system. It helps the database system to simplify and facilitate access to data.

Components of the query processor are:

- **1. DDL interpreter**, which interprets DDL statements and records the definitions in the data dictionary.
- **2. DML compiler**, which translates DML statements in a query language into an evaluation plan consisting of low-level instructions that the query evaluation engine understands. A query can be translates into any number of evaluations plans that all give the same result. The DML compiler also performs query optimization, that is, it picks up the lowest cost evaluation plan from among the alternatives.
- **3. Query evaluation engine**, which executes low-level instructions generated by the DML compiler.

4. List four significant differences between a file-processing system and a DBMS.

Ans: Few main differences between file-processing system and DBMS are:

- 1. Both systems contain a collection of data and a set of programs which access that data.

 A database management system coordinates both the physical and the logical access to the data, whereas a file-processing system coordinates only the physical access.
- 2. A database management system reduces the amount of data duplication by ensuring that a physical piece of data is available to all programs authorized to have access to it, whereas data written by one program in a file-processing system may not be readable by another program.
- 3. A database management system is designed to allow flexible access to data (i.e., queries), whereas a file-processing system is designed to allow pre-determined access to data (i.e., compiled programs).
- 4. A database management system is designed to coordinate multiple users accessing the same data at the same time. A file-processing system is usually designed to allow one or more programs to access different data files at the same time. In a file-processing system, a file can be accessed by two programs concurrently only if both programs have read-only access to the file.

5. Differentiate between logical database design and physical database design. Show how this separation leads to data independence.

Ans:

Basis	Logical Database Design	Physical Database Design
Task	Maps or transforms the conceptual schema (or an ER schema) from the high-level data model into a relational database schema.	The specifications for the stored database in terms of physical storage structures, record placement, and indexes are designed.
Choice of criteria	The mapping can proceed in two stages: System-independent mapping but data model-dependent Tailoring the schemas to a specific DBMS	The following criteria are often used to guide the choice of physical database design options: Response Time Space Utilization Transaction Throughput
Result	DDL statements in the language of the chosen DBMS that specify the conceptual and external level schemas of the database system. But if the DDL statements include some physical design parameters, a complete DDL specification must wait until after the physical database design phase is completed.	An initial determination of storage structures and the access paths for the database files. This corresponds to defining the internal schema in terms of Data Storage Definition Language.

- The database design is divided into several phases.
 - 1. The Logical database design and
 - 2. The Physical database designs.
- This separation is generally based on the concept of three-level architecture of DBMS, which provides the data independence.
- Therefore, we can say that this separation leads to data independence because the output of the logical database design is the conceptual and external level schemas of the database system which is independent from the output of the physical database design that is internal schema.

6. Define Data Independence. Explain the types of Data Independence.

Ans:

- ➤ The ability to modify a scheme definition in one level without affecting a scheme definition in a higher level is called **data independence**.
- ➤ There are two types of Data Independence:
 - 1. Physical data independence:

The ability to modify the physical scheme without causing application programs to be rewritten - Modifications at this level are usually to improve performance.

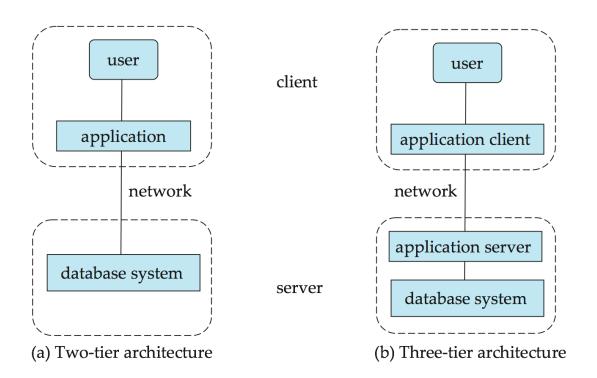
2. <u>Logical data independence:</u>

The ability to modify the conceptual scheme without causing application programs to be rewritten - Usually done when logical structure of database is altered.

Logical data independence is harder to achieve as the application programs are usually heavily dependent on the logical structure of the data. An analogy is made to abstract data types in programming languages.

7. Explain two-tier architecture and three-tier architecture in detail.

- ➤ In **two tier architecture**, the application is partitioned into a component that resides at the client machine and invokes database functionality at the server machine through query language.
- Application program interface standards like ODBC and JDBC are used for interaction between the client and the server.

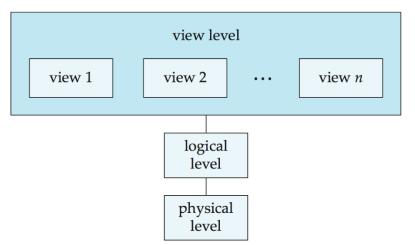


- In three tier architecture, the client machines act as a front end and do not contain any direct database calls.
- ➤ The client end communicates with the application servers through interface.
- ➤ The application server interacts with database system to access data.
- The business logic of application says what actions to be carried out under what condition.
- Three tiers are more appropriate for large applications.

8. Explain Data abstraction in brief.

Ans:

- ➤ The Complexity is hidden from the users through several level of abstraction.
- > There are three levels of data abstraction:



1. Physical level:

- The lowest level of abstraction describes how the data are actually stored.
- The physical level describes complex low-level data structures in detail.

2. Logical level:

- The next-higher level of abstraction describes what data are stored in the database, and what relationships exist among those data.
- The logical level thus describes the entire database in terms of a small number of relatively simple structures.
- Although implementation of the simple structures at the logical level may involve complex physical-level structures, the user of the logical level does not need to be aware of this complexity.
- This is referred to as physical data independence.
- Database administrators, who must decide what information to keep in the database, use the logical level of abstraction.

3. View level:

- The highest level of abstraction describes only part of the entire database.
- Even though the logical level uses simpler structures, complexity remains because of the variety of information stored in a large database.
- Many users of the database system do not need all this information; instead, they need to access only a part of the database.

- The view level of abstraction exists to simplify their interaction with the system.
- The system may provide many views for the same database.

9. What is Data model? Explain various types of data models.

Ans:

- ➤ Data Models is an underlying structure of the database.
- ➤ It is a collection of conceptual tools for describing data, data relationships, data semantics, and consistency constraints.
- It is a way to describe the design of the database at physical, logical and view level.
- The data models can be classified into four different categories:
 - 1. Relational model
 - 2. Entity-Relationship data model
 - 3. Object-based data models
 - 4. Semistructured data model (XML)

1. Relational Model:

- The relational model uses a collection of tables to represent both data and the relationship among those data. Each table has multiple columns and each column has a unique name.
- Software such as Oracle, Microsoft SQL Server and Sybase are based on the relational model.
- E.g. Record Based model. It is based on fixed format records of several types.

2. Entity relationship model:

- It is based on a collection of real world things or objects called entities and the relationship among these objects.
- The Entity relationship model is widely used in database design.

3. Object-Based Data model:

- The object- oriented model is an extension of E-R model.
- The object- oriented model is based on a collection of objects.
- An object contains values stored in instance variables within the object.
- An object also contains bodies of code that operate in the object these bodies of code are called methods.
- Objects that contain the same types of values and methods are grouped together into classes.

4. Semi Structured Data Model:

- This data model allows the individual data items of same type to have different sets of attributes.
- Other data model allows a particular type of data item to have same set of attributes.
- Extensible Markup Language (XML) is used to represent structured data.

10. Draw the labeled diagram of database architecture.

