DATABASE CONCEPTS

One Mark Questions:

1. What is data?

- Data is a collection of facts, numbers, letters or symbols that the computer process into meaningful information.

2. What is Information?

- Information is processed data, stored, or transmitted by a computer.

3. What is Database?

- A Database is a collection of logically related data organized in a way that data can be easily accessed, managed and updated.

4. What is a field?

- Each column is identified by a distinct header called attribute or filed.

5. What is a record?

- A single entry in a table is called a record or row. A record in a table represents set of related data. Records are also called the tuple.

6. What is an entity?

- An Entity can be any object, place, person or class. In E-R Diagram, an entity is represented using rectangles.

7. What is an instance?

- The collection of information stored in the database at a particular moment is called an instance of the database.

8. What is an attribute?

- It is defined as a named column of a relation. Ex: In STUDENT table, Regno, Name, Age, Class, Combination and Marks.

9. What is domain?

- It is defined as a set of allowed values for one or more attributes.

10. What is a relation?

- A relation is defined as a table with columns and rows. Data can be stored in the form of a two-dimensional table.

11. What is a table?

- A table is a collection of data elements organized in terms of rows and columns. Table is the simplest form of data storage.

12. What is normalization?

- Normalization is a step by step process of removing the different kinds of redundancy and anomaly one step at a time from the database.

13. What is a key?

- It is a column or columns which identifies the each row or tuple.

14. Give the symbol notation for project and select?

- SELECT – sigma (σ)
- PROJECT – Pi (Π)

15. What is data mining?

- Data mining is concerned with the analysis and picking out relevant information.

Two/Three Mark Questions:

1. Mention the applications of database.

- Banking: For customer information, accounts and loans, and banking transactions.
- Colleges: For student information, course registrations and grades.
- Credit card transactions: For purchases on credit cards and generation of monthly statements.
- Finance: For storing information about holdings, sales and purchases of financial instruments such as stocks and bonds.
- Sales: For customer, product, and purchase information.
- Telecommunication: For keeping records of call made, generating monthly bills, maintaining balance on prepaid calling cards, and storing information about the communication networks.
- Voter id/Aadhaar database: This is the biggest database in the world storing a data about 60 million people residing in India.

2. Explain database users.

To design, use and maintain the database, many peoples are involved. The people who work with the database include System Analysts, Application programmers, Database Administrators (DBA), End Users (Database Users)

- Database users are those who interact with the database in order to query and update the database, and generate reports.
- System Analysts: System analysts determine the requirement of end users; (especially naïve users), to create a solution for their business need and focus on non-technical and technical aspects.
- Application programmers: These are the computer professionals who implement the specifications given by the system analysts and develop the application programs.
- Database Administrators (DBA): DBA is a person who has central control over both data and application. Some of the responsibilities of DBA are authorization access, schema

definition and modification, new software installation and security enforcement and administration.

3. What is Data Independence? Mention the two types.

The capacity to change data at one layer does not affect the data at another layer is called data independence. Two types of data independence are:

- Physical Data Independence
- Logical Data Independence

4. Explain physical data independence.

It is the capacity to change the internal level without having to change either the schemas at the conceptual or external level. Changes to the internal schema may be needed because some physical files had to be reorganized. Physical data independence refers to the data insulation of an application from the physical storage structure only, it is easier to achieve than logical data independence.

The physical data independence are:

- File Organization
- Database Architecture
- Database Models

5. What is the difference between serial and direct access file organization.

Serial File Organization:

- Organization is continuous and simple.
- Data processing, which requires the use of all records, is best suited to use this method.

Direct Access File Organization

- The type of storage device used is comparatively expensive.
- It is less efficient in the usage of storage space compared to the sequential organization.

6. Explain ISAM with example.

The index sequential file organization is a combination of Sequential file organization and an Index file. It is also referred as ISAM (indexed sequential access method). Data is stored physically in adjacent storage locations and there exists a logical relationship among the data stored by using ordering field. An additional file called as Index file would be created, which contains n number of records. Each record of index file has two fields:

- The field is of the same data type as the ordering key field and
- The second field is a pointer to a disk block (a block address).

7. Give the advantages and disadvantages of ISAM.

Advantages

- Search time is less.

- There are fewer index entries than there are records in the data file.
- Quick access to the records even when the volume of records is high.

Disadvantages

- Additional file (index file) has to be created.
- Wastage of storage space by creating and maintaining the index file.
- Always indirect retrieval of data because first search begins in the index files then moves to the data file (No direct retrieval).

8. Classify the various types of keys used in database.

The different types of keys are:

Primary key:

- It is a field in a table which uniquely identifies each row/record in a database table. Primary keys must contain unique values. A primary key column cannot have NULL values. Ex: In Relation STUDENT, Regno serves as a primary key.

Candidate Key:

- When more than one or group of attributes serve as a unique identifier, they are each called as candidate key.

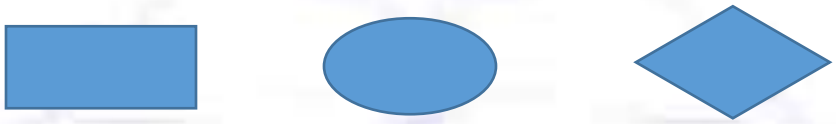
Alternate Key:

- The alternate key of any table are those candidate keys, which are not currently selected as the primary key. This is also known as secondary key.

Foreign key:

- A key used to link two tables together is called a foreign key, also called as referencing key. Foreign key is a field that matches the primary key column of another table.

9. Explain different notations of E-R diagram.



Entity Attribute Relationship

- Entity: An entity is represented using rectangles.
- Attribute: Attributes are represented by means of eclipses.
- Relationship: Relationship is represented using diamonds shaped box.

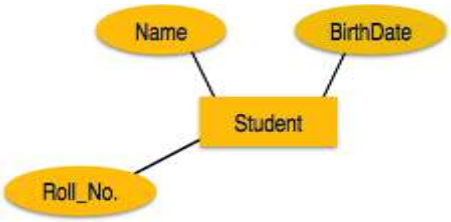
10. Explain any three components of E-R model.

ER-Diagram is a visual representation of data that describes how data is related to each other.

Entity:

- An Entity can be any object, place, person or class.

- In E-R Diagram, an entity is represented using rectangles.
- Rectangles are named with the entity set they represent.



Attribute:

- An Attribute describes a property or characteristic of an entity.
- Attributes are represented by means of eclipses.
- Every eclipse represents one attribute and is directly connected to its entity (rectangle).
- For example, Roll_No, Name and Birth date can be attributes of a student

Relationship:

- A relationship type is a meaningful association between entity types.
- Relationship is represented using diamond shaped box.
- Relationship types are represented on the E-R diagram by a series of lines.

11. What is a Relationship? Classify and give example.

A Relationship describes relations between entities. Relationship is represented using diamonds shaped box. There are three types of relationship that exist between entities.

- Binary Relationship
- Recursive Relationship
- Ternary Relationship

Binary Relationship: It means relation between two entities. This is further divided into three types.

1. One to One:

- This type of relationship is rarely seen in real world.
- The above example describes that one student can enroll only for one course and a course will have only one Student. This is not what you will usually see in relationship.

2. One to Many:

- It reflects business rule that one entity is associated with many number of same entity.
- For example, Student enrolls for only one Course but a Course can have many Students.

3. Many to Many:

- It reflects business rule that many entity are associated with many number of same entity.
- The above diagram represents that many students can enroll for more than one course.

12. What is generalization?

In generalization, a number of entities are brought together into one generalized entity based on their similar characteristics. For example, pigeon, house sparrow, crow and dove can all be generalized as Birds.

13. What is specification?

Specialization is the opposite of generalization. In specialization, a group of entities is divided into sub-groups based on their characteristics. Take a group 'Person' for example. A person has name, date of birth, gender, etc. Similarly, in a school database, persons can be specialized as teacher, student, or a staff, based on what role they play in school as entities.

14. **What is Relation algebra?**

Relational algebra is a procedural query language that consists of a set of operations that take one or more relations as input and result into a new relation as an output. The relational algebraic operations can be divided into:

- Basic set-oriented operations: Union, Set different, Cartesian product
- Relational-oriented operations: Selection, Projection, Division, Joins

15. **What is Cartesian product?**

It is a binary operation, and it is denoted by the symbol x. The Cartesian product of two relations R and S, denoted by R x S, defines a new relation, which is the concatenation of each tuple of relation R with each tuple of relation S.

17. **What is Data warehouse?**

A data warehouse is a repository of an organization's electronically stored data. Data warehouse are designed to facilitate reporting and supporting data analysis. The concept of data warehouses was introduced in late 1980's.

18. **List the components of data warehouse.**

The components of data warehouse are:

- Data Source
- Data Transformation
- Reporting
- Metadata

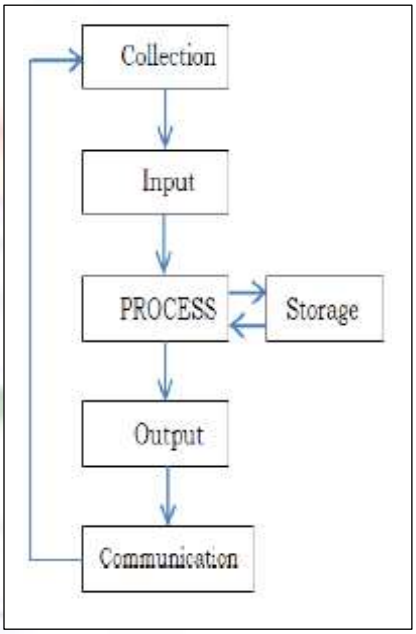
Additional components are Dependent data marts, Logical Data marts, Operational Data store.

Five Mark Questions:

1. Give the difference between Manual and Computerized data processing.

	Manual Data Processing	Computerized Data Processing
1	The volume of data, which can be processed, is limited.	The volume of data, which can be processed is large
2	Requires large quantity of paper	Requires less quantity of paper
3	Speed and accuracy is executed is limited	Faster and Accurate
4	Labour cost is high	Labour cost is low
5	Storage medium is paper.	Storage medium is Hard disk etc.

2. Explain Data processing cycle.



- **Data Collection:** It is the process of systematic gathering of data from various sources that has been systematically observed, recorded and organized.
- **Data Input:** The raw data is put into the computer using a keyboard, mouse or other devices such as the scanner, microphone and the digital camera.
- **Data Processing:** Processing is the series of actions or operations on the input data to generate outputs.
- **Data storage:** Data and information should be stored in memory so that it can be accessed later.
- **Output:** The result obtained after processing the data must be presented to the user in user understandable form. The output can be generated in the form of report as hard copy or soft copy.
- **Communication:** Computers now-a-days have communication ability which increases their power. With wired or wireless communication connections, data may be input from a far place, processed in a remote area and stored in several different places and then transmitted by modem as an e-mail or posted to the website where the online services are rendered.

3. Explain the features or advantages of Database.

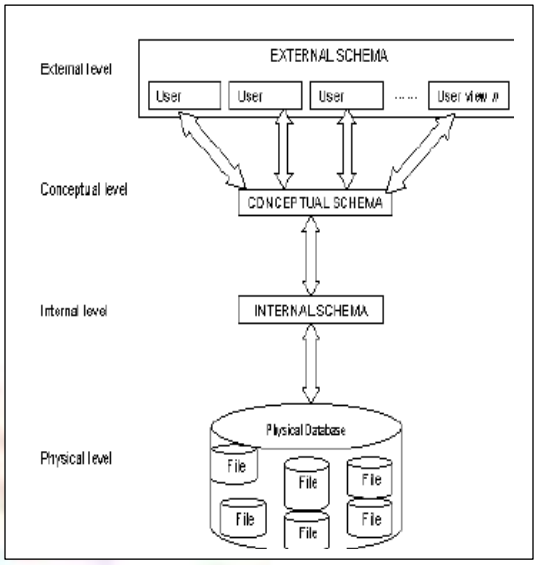
- **Redundancy can be minimized or controlled**: In DBMS environment if redundancy is present, then it can be controlled by propagating updates in all the places where ever redundant data is present.
- **Data Integrity**: Data Integrity refers to the correctness of the data in the database. In other words, the data available in the database is reliable data.
- **Data Sharing**: In DBMS, data is stored in the centralized database and all the permitted users can access the same piece of information required at the same time.
- **Database Security**: DBMS provides a variety of security mechanisms for the user to protect his or her data stored in the database.

- **Supports Concurrent access**: DBMS supports concurrent access to the same data stored in the database by applying locking and time stamp mechanisms.

4. Explain the concept of data abstraction.

A major purpose of a database system is to provide users with an abstract view of the data. That is the system hides certain details of how the data are stored and maintained.

There are three level of data abstraction.



- Physical Level(Internal level)
- Conceptual Level (Logical level)
- View Level(External level)

Physical Level:

- It is the lowest level of abstraction that describes how the data are actually stored.
- The physical level describes complex low-level data structures in detail.
- It contains the definition of stored record and method of representing the data fields and access aid used.

Conceptual Level:

- It is the next higher level of abstraction that describes what data are stored in the database and what relationships exist among those data.
- It also contains the method of deriving the objects in the conceptual view from the objects in the internal view.

View Level:

- It is the highest level of abstraction that describes only part of the entire database.
- It also contains the method of deriving the objects in the external view from the objects in the conceptual view.

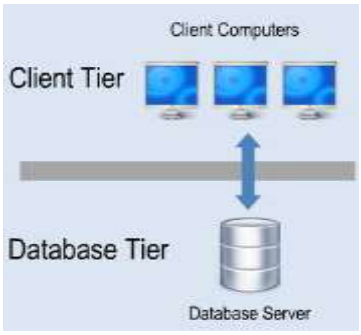
5. Explain DBMS Architecture.

- The design of Database Management System highly depends on its architecture.
- It can be centralized or decentralized or hierarchical.
- Database architecture is logically divided into three types.
    - Logical one-tier in 1-tier Architecture
    - Logical two-tier Client/Server Architecture.
    - Logical three-tier Client/Server Architecture.

Logical one-tier in 1-tier Architecture:

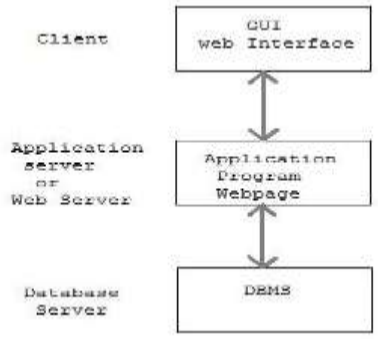
- DBMS is the only entity where user directly sits on DBMS and uses it. Any changes done here will directly be on DBMS itself. It does not provide handy tools for end users and preferably database designers and programmers use single tier architecture.

Two-tier Client / Server Architecture:



- Two-tier Client / Server architecture is used for User Interface program and Application Programs that runs on client side. An interface called ODBC (Open Database Connectivity) provides an API that allows client side program to call the DBMS.
- Most DBMS vendors provide ODBC drivers. A client program may connect to several DBMS's. In this architecture some variation of client is also possible for example in some DBMS's more functionality is transferred to the client including data dictionary, optimization etc.

Three-tier Client / Server Architecture:



- Three-tier Client / Server database architecture is commonly used architecture for web applications. Intermediate layer called **Application server** or Web Server stores the web connectivity software and the business logic (constraints) part of application used to access the right amount of data from the database server.
- This layer acts like medium for sending partially processed data between the database server and the client.

6. Explain Database Model.

Data model is a collection of conceptual tools for describing data, data relationship, data semantics and constraints. A data model generally consists of:

- **Data model theory**, which is a formal description of how data may be structured and used.
- **Data model instance,** which is a practical data model designed for a particular application.

The process of applying model theory to create a data model instance is known as **data modeling**. In history of database design, three models have been in use.

- Hierarchical Model
- Network Model
- Relational Model

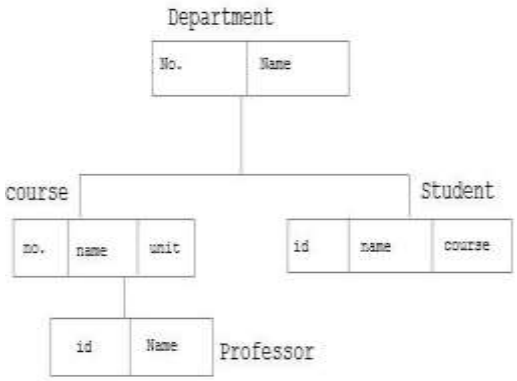
7. Explain Hierarchical data model.

The Hierarchical data model organizes data in a tree structure. In this data model, data is represented by a collection of **records** and the relationships are represented by **links**. In this model, each entity has only one parent but can have several children. At the top of hierarchy, there is only one entity, which is called **Root node**.

Advantages:

- **Simplicity**: The relationship between the various layers is logically simple.
- **Data Security**: The data security is provided by the DBMS.

- **Data Integrity**: There is always link between the parent segment and the child segment under it.
- **Efficiency**: It is very efficient because when the database contains a large number of one to many relationships and when the user requires large number of transaction.



Disadvantages:

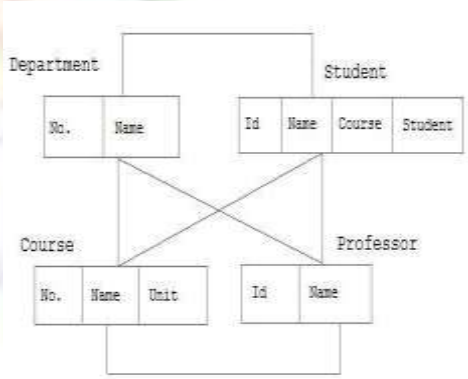
- Implementation complexity
- Database management problem
- Lack of structural Independence.
- Operational Anomalies

8. Explain Network data model.

In 1971, the Conference on Data Systems Languages (CODASYL) formally defined the network models. In this model, data is represented by a collection of records and the relationships are represented by links. Each record is collection of fields, which contains only one data value. A link is an association between two records. In the network model, entities are organized in a graph, in which some entities can be accessed through several paths.

Advantages:

- It is simple and easy to implement.
- It can handle many relationships within the organization.
- It has better data independence compared to hierarchical model.

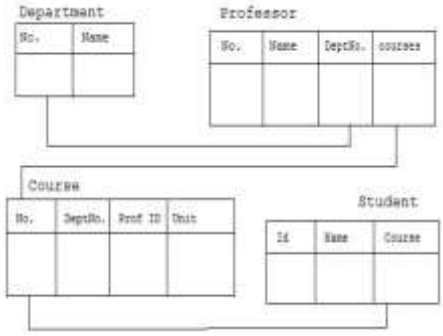


Disadvantages:

- More complex system of database structure
- Lack of structural dependence.

9. Explain Relation Data Model.

E.F Codd developed the relation data model in 1970. Unlike, hierarchical and network model, there are no physical links. All data is maintained in the form of tables consisting of rows and columns. Each row (record) represents an entity and a column (field) represents an attribute of the entity. In this model, data is organized in two-dimensional tables called **relations**. The tables or relations are related to each other.



10. Explain any 5 Codd's rule for database management.

E.F Codd was a Computer Scientist who invented **Relational model** for Database management. Based on relational model, **Relation database** was created.

Rule zero

- This rule states that for a system to qualify as an **RDBMS**, it must be able to manage database entirely through the relational capabilities.

Rule 1: Information rule

- All information (including meta-deta) is to be represented as stored data in cells of tables. The rows and columns have to be strictly unordered.

Rule 2: Guaranteed Access

- Each unique piece of data (atomic value) should be accessible by:
- **Table Name + primary key (Row) + Attribute (column)**.
- **NOTE:** Ability to directly access via POINTER is a violation of this rule.

Rule 3: Systematic treatment of NULL

- **Null** has several meanings; it can mean missing data, not applicable or no value. It should be handled consistently. Primary key must not be null. Expression on **NULL** must give null.

Rule 4: Active Online Catalog

- This rule states that the structure description of whole database must be stored in an online catalog i.e. data dictionary, which can be accessed by the authorized users.

Rule 5: Powerful language

- One well-defined language must be there to provide all manners of access to data.
- Example: **SQL**. If a file supporting table can be accessed by any manner except SQL interface, then its a violation to this rule.

Rule 6: View Updation rule

- All view that is theoretically updatable should be updatable by the system.

11. Write comparing RA and SQL.

	RA	**SQL**
1	Relation Algebra	Structured Query Language
2	Is closed (the result of every expression is a relation)	Is a superset of relation algebra
3	Simple semantics	Complicated Semantics
4	It is used for reasoning, query, optimization etc	It is an end-user language.
5	Relationally Complete	Relationally Complete

CHAPTER 13 – DATABASE CONCEPTS BLUE PRINT				
VSA (1 marks)	SA (2 marks)	LA (3 Marks)	Essay (5 Marks)	Total
01 Question	01 Question	01 Question	01 Question	04 Question
Question No 06	Question No 16	Question No 24	Question No 35	11 Marks

Important Questions One Marks Questions:

1. Define Primary key [March 2015]
2. What is a database? [June 2015, June 2016]
3. Define Data Mining. [March 2016]
4. Define an Entity. [March 2017, June 2017]

Two Marks questions:

1. What is data independence? Mention the types of data independence? [March 2015]
2. Write any two advantages of database system. [June 2015]
3. What are the advantages and disadvantages of ISAM? [March 2016]
4. Define Primary key and Secondary keys. [June 2016]
5. Write the difference between data and information. [March 2017]
6. Mention the database users. [March 2017, June 2017]

Three Marks Questions:

1. Briefly explain one-tier database architecture. [March 2015]
2. Write the different symbols used in E-R diagram with their significance. [June 2015]
3. Explain relational data model with example. [March 2016]
4. Define hierarchical data model. Give one advantage and disadvantage. [June 2016]
5. Mention any three advantages of random/direct access file organization. [June 2017]
6. Give the different notations for E-R diagram.
7. What is an entity-relationship diagram? Explain its components Entity and Attribute?
8. Define DBMS. Write any one feature of it.
9. What is a relationship? List different types of relationships.
10. Give the advantages and disadvantages of indexed sequential file organization.

Five Marks Questions:

1. What is data warehouse? Briefly explain its components. [March 2015]
2. Define the following database terms:

a. Data Model b. Tuple c. Domain d. Primary key e. Foreign key [June 2016]

3. Write the difference between manual and electronic data processing. [March 2016]
4. Explain any five applications of database. [June 2016]
5. Briefly explain the data processing cycle. [March 2017]
6. Write the difference between Hierarchical data model and network data model. [June 2017]
7. What is normalization? Explain second normal form with an example.
8. What is database model? Explain Hierarchical model.
9. Define any 5 Codd's rule.
10. Explain 3-level DBMS architecture.