

XCS 235 – DATABASE MANAGEMENT SYSTEMS

UNIT-1

Questions And Answers----Two Marks

1. Define database management systems?

A database management system is a collection of programs that enables users to create, store modify maintain, and extract information from a database. A database management system is a piece of software that provides services for accessing a database, while maintaining all the required features of the data.

2. State the various processes of database?

- a) Defining
- b) Constructing
- c) Manipulating

3. What are the characteristics of database?

- a) self describing nature of the database
- b) insulation between programs and data and data abstraction.
- c) Support for multiple views of data
- d) Sharing of data and multi-user transaction processing
- e) security- Data should be protected from unauthorized users.

4. Define-metadata?

The information stored in the catalog is called meta-data and it describes the structure of the primary database.

5. Define data abstraction?

The characteristics that allows program data independence and program operation independence are called data abstraction.

6. What are the advantages of DBMS?

- a) Controlling redundancy
- b) Restricting unauthorized access
- c) Providing persistent storage for program objects and data structures
- d) Permitting inference and actions using rules.

7. What is data model?

A data model is a collection of concepts that can be used to describe the structure of the database, which are data types, relationships and constraints that should hold on the data.

8. What are the categories of data model?

- a) High level or conceptual data model
- b) Low level or physical data model
- c) Representational or implementational data model.

9. Define entity, attributes and relationships?

Entity represent a real world object
Attribute represent property that describes an entity.
Relationship represents an interaction among the entities.

10. Define access path?

An access path is a structure that makes the search for particular database records efficient.

11. Define database schema?

The description of a database is called the database schema

12. Define database state or snapshot?

The data in the database at a particular moment in time is called database state or snapshot.

13. Distinguish intension and extension?

The schema is called intension and a database state is called extension of the schema.

14. What are the types of attributes in the ER model?

- a) Simple versus composite
- b) Single valued versus multi-valued
- c) Stored versus derived

15. What are composite and simple attributes?

Composite attributes can be divided into smaller subparts. Simple or atomic attributes are not divisible.

16. What are stored and derived attribute?

The age attribute is called derived attribute and is said to be derivable from the birth data attribute which is called stored attribute.

17. How are complex attributes represented?

Grouping components of a composite attribute between paranthesis and separating the components with commas and by displaying multivalued attributes between the braces {}. Such attributes are called complex attributes.

18. What is an entity type?

An entity type defines a collection of entities that have the same attributes.

19. What is an entity set?

The collection of all entities of a particular entity type in the database at any point in time is called an entity set.

20. What is a key attribute?

An entity type usually has an attribute whose values are distinct for each individual entity in the collection. Such an attribute is called a key attribute

unit 2

1. Define weak entity type?
An entity type may have no key; it is called weak entity type.
2. Define valuesets of attributes?
It specifies the set of values that may be assigned to that attribute for each individual entity.
3. Define degree of relationship type?
The degree of relationship type is the number of participating entity types.
4. What is a role name?
The role name specifies the role that a participating entity from the entity type plays in each relationship.
5. What are the types of relationship constraints?
 - a) Cardinality ratio
 - b) Participation
6. Define cardinality ratio of binary relationship?
The cardinality ratio for a binary relationship specifies the number of relationship instances that an entity can participate in.
7. Define participation constraint?
The participation constraint specifies whether the existence of an entity depends on its being related to another entity via the relationship type.
8. What are the types of participation constraints?
 - a) Total participation
 - b) Partial participation
9. Define a partial key?
A weak entity type has a partial key which is a set of attributes that can uniquely identify weak entities that are related to the same owner entity.
10. Define specialization?
Specialization is the process of defining a set of subclasses of an entity type. This entity type is called the superclass of the specialization. specialization is a top-down process. An example of specialization is the creation of three subclasses for the EMPLOYEE superclass.
11. What are the main categories of storage hierarchy?
 - a) Primary storage
 - b) Secondary storage
12. Define track?
Each concentric circle in a disk surface is called a track.

13. What is the hardware address of a block?
The hardware address of a block is the combination of a surface number, track number and block number.
14. Define seek time?
To transfer a disk block, given its address, the disk controller must first position the read write head on the correct track. The time required to do this is called seek time.
15. Define rotational delay or latency?
The beginning of the desired block rotates into position under the read write head is called rotational delay or latency.
16. Define block transfer time?
Time needed to transfer the data is called block transfer time.
17. Define bulk transfer rate?
The time required to transfer consecutive blocks is called bulk transfer rate.
18. What is data striping?
Data striping distributes the data transparently over multiple disks to make them appear as a single large fast disk.
19. What is mirroring or shadowing?
Data is written redundantly to two identical physical disks that are treated as one logical disk. When data is read it can be retrieved from the disk, with short queuing, seek time and rotational delay.
20. What is bit level data striping?
Bit level data striping consists of splitting a byte of data and writing bit j to the j th disk.
21. What are the records-at a time operations?
Reset, find, read, findnext, delete, modify, insert
22. What are the set at a time operations?
Findall, Findordered, reorganize.
23. Define hashing function?
It provides a function h called the hash function that is applied to the hash field value of a record and yields the address of the disk block in which the record is stored.
24. What is folding?
Folding is a hashing technique that involves applying an arithmetic function such as addition or logical function such as exclusive or(XOR) to different portions of the hash field value to calculate the hash address.

unit 3

1. When does collision occur?

A collision occurs when a hash field value of a record that is being inserted hashes to an address that already contains a different record. In this situation the new record must be inserted to some other position because its hash address is already occupied.

2. What is collision resolution?

The process of finding new location to insert record is called collision resolution.

3. State the methods of collision resolution?

- a) Open addressing
- b) Chaining
- c) Multiple hashing

4. What is external hashing?

Hashing for disk files is called external hashing. In the case of disks, the target address space is made of buckets. Each bucket contains multiple records. A bucket is either one disk block or a cluster. The hashing function maps a key into a relative bucket number, rather than assign an absolute block address to the bucket.

5. What is local depth in extendible hashing?

Several directory locations with the same first 'l' bits for their hash values may contain the same bucket address if all the records that hash to these locations fit in a single bucket. In this case 'l' is called the local depth. The local depth stored with each bucket specifies the number of bits on which the bucket contents are based.

6. What is file load factor?

The file load factor α can be defined as $\alpha = r / (bfr * N)$

7. Define relation schema?

A relation schema red denoted by $R(A_1, A_2, \dots, A_n)$ is made up of a relation name Red and a list of attributes A_1, A_2, \dots, A_n .

8. Define degree of relation?

The degree of relation is the number of attributes n of its relation schema.

9. Define relation state?

A relation red of the relation schema $R(A_1, A_2, \dots, A_n)$ is a set of tuples $r = \{t_1, t_2, \dots, t_n\}$.

10. State the characteristics of relations?

- a) ordering of tuples in a relation
- b) ordering of values within a tuple
- c) values in the tuples
- d) interpretation of a relation.

11. Define super key?

A super key specifies a uniqueness constraint that no distinct tuples in a state r of Red can have the same value for superkey.

12. Define a candidate key?

A relation schema may have more than one key is called a candidate key.

13. Define relational database schema?

A relational database schema S is a set of relation schemas $S=\{R_1,R_2,\dots,R_n\}$ and a set of integrity constraints.

14. Define relational database state?

A relational database state of S is a set of relation states $Db=\{r_1,r_2,\dots,r_m\}$ such that each r_i is a state of R_i and such that the r_i relation state satisfy the integrity constraints.

15. What are the basic update operations on relations?

- a) insert
- b) delete
- c) modify

16. Define mapping or select from where clause?

```
SELECT <ATTRIBUTE LIST> FROM <TABLE LIST>  
WHERE <CONDITION>
```

18. Define functional dependency?

A functional dependency is a constraint between two sets of attributes from the database.

19. State the inference rules of functional dependency?

- a) reflexive rule
- b) augmentation rule
- c) transitive rule
- d) decomposition rule
- e) union or additive rule
- f) pseudo transitive rule.

20. What is normalization?

Normalization of data is a process of analyzing the given relation schemas based on their FDs and primary keys to reduce redundancy and to minimize insertion, deletion and updation.

21. What is a prime attribute?

An attribute of relation schema R is called prime attribute of R if it is a member of some candidate key of R.

unit 4

1. Define full functional dependency?
If the removal of any attribute A from X means that the dependency does not hold any more.
2. What are the two techniques of implementing query optimization?
 - a. Heuristic rules for ordering the operations in a query execution strategy
 - b. Systematically estimating the cost of different execution strategies
3. Define external sorting?
It refers to sorting algorithms that are suitable for large files of records stored on disk that do not fit entirely in main memory such as most database files.
4. What are the phases of external sorting?
 - a. Sorting phase
 - b. Merging phase
5. Define degree of merging?
It is the number of runs that can be merged together in each pass.
6. Give the search methods for simple select operation?
 - a. linear search
 - b. Binary Search
 - c. Using a primary index
 - d. Using a primary index to retrieve multiple records
 - e. Using a clustering index to retrieve multiple records
 - f. Using secondary index on equality comparison
7. What are the methods for implementing joins?
 - a. Nested loop join
 - b. Single loop join
 - c. Sort-merge join
8. What is a query tree?
A query tree is a data structure that corresponds to a relational algebra expression. It represents the input relations of the query as leaf nodes of the tree, and represents the relational algebra operations as internal nodes.
9. What are the cost components for query execution?
 - a. Access cost to secondary storage
 - b. Storage cost
 - c. Computation cost
 - d. Memory usage cost
 - e. Communication cost
10. Define selectivity?
Selectivity is the fraction of records satisfying an equality condition on the attribute.

11. Define selection cardinality?

It is the average number of records that will satisfy an equality selection condition on that attribute.

12. Define transaction?

A transaction is a logical unit of database processing that includes one or more database access operations that include insertion, deletion, modification or retrieval operations.

13. What are the problems that occur in transaction if they run concurrently?

- a. Lost Update problem
- b. Temporary update problem
- c. Incorrect summary problem

14. What are the types of failures?

- a. Computer failure(system crash)
- b. Transaction or system Error
- c. Local errors
- d. Concurrency control enforcement
- e. Disk failures
- f. Physical problems and catastrophes

15. What are the transaction operations?

- a. Begin transaction
- b. Read or write
- c. End transaction
- d. Commit transaction
- e. Rollback or abort

16. When does a transaction reach its commit point?

A transaction T reaches its commit point when all its operations that access the database have been executed successfully and the effect of all the transaction operations on the database have been recorded in the log;

17. What are the properties of transaction?

- a. Atomicity
- b. Consistency reservation
- c. Isolation
- d. Durability or permanency

18. Define consistent state.

A consistent state of the database satisfies the constraints specified in the schema as well as any other constraints that should hold on the database.

19. When is a schedule said to be serial?

A schedule S is said to be serial if, for every transaction T participating in the schedule all the operations of T is executed consecutively in the schedule otherwise it is non serial.

20. When is a schedule serializable?

A schedule S of transactions is serializable if it is equivalent to some serial schedule of the same n transactions.

21. Define precedence graph or serialization graph?

It is a directed graph $G = (n, E)$ that consists of a set of nodes $N = \{T_1, T_2, \dots, T_n\}$ and a set of directed edges $E = \{e_1, e_2, \dots, e_n\}$

unit 5

1. What is a lock?

A lock is a variable associated with a data item that describes the status of the item with respect to possible operations that can be applied to it.

2. What is a binary lock?

A binary lock can have two states. Locked and Unlocked.

3. What are the fields present in a lock?

- a. Data item name
- b. LOCK
- c. Locking transaction

4. What are the locking operations of a shared/exclusive lock?

- a. Read lock(X)
- b. Write lock(X)
- c. Unlock(X)

5. What are the phases in a locking transaction?

- a. Expanding phase
- b. Shrinking Phase

6. Define deadlock?

Deadlock occurs when each transaction T in a set of two or more transactions is waiting for some item that is locked by some other transaction T in the set.

7. Define a Timestamp?

A time stamp is a unique identifier created by the DBMS to identify a transaction.

8. What are the values present in a timestamp algorithm?

- a. Read TS(X)
- b. Write TS(X)

9. What are the phases in concurrency control protocol?

- a. Read phase
- b. Validation phase
- c. Write phase

10. What are the conditions that hold in a validation phase?

- a. Transaction T_j completes its write phase before T_i starts its read phase
- b. T_i starts write phase after T_j completes its write phase and the read set of T_i has no items in common with the write set of T_j
- c. Both the read set and the write set of T_i have no items in common with the write set of T_j and T_i completes its read phase before T_i completes its read phase

11. Define granularity?

The size of data item is called data item granularity. A database consists of many data items on which lock is acquired. The data item as large as the entire database or it can be as small as a field value of a record. When the size of the data item is small we say the granularity is fine and when it is large we say it is coarse.

12. What are types of intension locks?

- a. Intension shared
- b. Intension Exclusive
- c. Shared-Intension Exclusive

13. What are the techniques for recovery from non catastrophic failures?

- a. Deferred update
- b. Immediate update

14. Define directory?

A directory for the cache is used to keep track of which database items are in the buffers.

15. What is a dirty bit?

It indicates whether or not the buffer has been modified.

Sixteen Marks Questions and Answers

1. Explain: Data models, Schemas, and Instances

a) Categories of data models

A database management system is a collection of programs that enables users to create and maintain a database.

b) Schemas, Instances, and Database state

c) self describing nature of the database

d) insulation between programs and data and data abstraction.

e) Support for multiple views of data

f) Sharing of data and multi-user transaction processing

2. Explain: DBMS architecture and data independence

a) The three-schema architecture

A data model is a collection of concepts that can be used to describe the structure of the database, which are data types, relationships and constraints that should hold on the data.

b) Data Independence

c) High level or conceptual data model

d) Low level or physical data model

e) Representational or implementational data model.

3. Write about Entity types, Entity sets, Attributes, and keys.

a) Entities and attributes

Entity represent a real world object

Attribute represent property that describes an entity.

Relationship represents an interaction among the entities.

b) Entity types, Entity sets, Keys, and Value sets

c) Simple versus composite

d) Single valued versus multi-valued

e) Stored versus derived

4. Explain the constraints and characteristics of generalization and specialization.

a) constraints on specialization/generalization

b) specialization/generalization hierarchies and lattices

Specialization is the process of defining a set of subclasses of an entity type. This entity type is called the superclass of the specialization.

c) utilizing specialization and generalization in conceptual data modeling.

5. Explain about secondary storage devices

a) Hardware description of disk devices

Each concentric circle in a disk surface is called a track.

The hardware address of a block is the combination of a surface number, track number and block number.

To transfer a disk block, given its address, the disk controller must first position the read write head on the correct track. The time required to do this is called

seek time.

b) Magnetic tape storage devices

6. Explain the concept of parallelizing disk access using RAID technology.

a) Improving reliability with RAID

Data striping distributes the data transparently over multiple disks to make them appear as a single large fast disk.

b) Improving performance with RAID

Data is written redundantly to two identical physical disks that are treated as one logical disk. When data is read it can be retrieved from the disk, with short queuing, seek time and rotational delay.

c) RAID organization and levels

7. Write about various operations that can be performed on files.

a) Open

b) Reset

c) Find(or Locate)

d) Read(or Get)

e) FindNext

f) Delete

g) Modify

h) Insert

i) Close

j) Scan

k) FindAll

l) FindOrdered

m) Reorganize

8. Explain: Hashing Techniques

It provides a function h called the hash function that is applied to the hash field value of a record and yields the address of the disk block in which the record is stored.

a) Internal hashing

b) External hashing for disk files

It involves applying an arithmetic function such as addition or logical function such as exclusive or to different portions of the hash field value to calculate the hash address.

c) Hashing techniques that allow dynamic file expansion

A collision occurs when a hash field value of a record that is being inserted hashes to an address that already contains a different record.

9. Explain the different types of single-level ordered indexes

a) Primary indexes

Ordering fields based on primary key.

b) Clustering indexes

Ordering fields based on secondary key.

c) Secondary indexes

Non-ordering file

10. Explain the basic relational algebra operations.

a) The SELECT operation

b) The PROJECT operation

c) Sequences of operations and the RENAME operation

d) Set Theoretic operation

- e) The JOIN operation
 - f) The DIVISION operation
11. Write about Structured Query Language(SQL) commands.
- a) The CREATE table command
 - b) The DROP TABLE commands
 - c) The ALTER TABLE command
 - d) The SELECT-FROM-WHERE structure of SQL queries.
 - e) Nested queries
 - f) The EXISTS and UNIQUE functions in SQL.
 - g) Aggregate functions and grouping
 - h) Insert, Delete, and Update statements in SQL.
 - i) Views(Virtual Tables) in SQL.
12. Explain: The Tuple Relational Calculus
- a) Tuple variables and Range Relations
 - b) Expressions and Formulas in Tuple relational calculus.
 - c) The Existential and Universal Quantifiers
 - d) Safe Expressions
 - e) Quantifiers in SQL.
13. Explain: Functional Dependency
- a) Definition of Functional Dependency
 - b) Inference Rules for Functional Dependencies
 - c) Equivalence of sets of functional dependencies
 - d) Minimal sets of Functional Dependencies
14. Explain the process of Normalization
- a) First Normal form(1NF)
 - b) Second Normal Form(2NF)
 - c) Third Normal Form(3NF)
 - d) Boyce-Codd Normal Form(BCNF)
 - e) Fourth Normal Form(4NF)
 - f) Fifth Normal Form(5NF)
15. Explain the process of database tuning
- a) Tuning Indexes
The indexes can be rebuild to improve performance and to reduce response time.
 - b) Tuning the Database Design
The queries are altered and modified in a way to improve performance.
 - c) Tuning Queries
16. Explain the basic algorithms for executing query operations
- a) External Sorting
 - b) Implementing the SELECT operation
 - c) Implementing the JOIN operation
 - d) Implementing PROJECT and Set operations
 - e) Implementing Aggregate operations
 - f) implementing Outer Join
17. Explain: Serializability of Schedules
- a) Serial, Nonserial, and Conflict-Serializable Schedules
 - b) Testing for conflict Serializability of a Schedule
 - c) Uses of Serializability
 - d) View Equivalence and View Serializability

18. Explain: Locking techniques for Concurrency Control

- a) Types of Locks and System Lock Tables
- b) Guaranteeing Serializability by Two-Phase Locking

A schedule S is serial if, for every transaction T participating in the schedule all the operations of T is executed consecutively in the schedule otherwise it is non serial

- c) Dealing with Deadlock and Starvation

19. Explain the process of concurrency control based on timestamp ordering.

- a) Timestamps

A timestamp is a unique identifier created by the DBMS to identify a transaction.

- b) The Timestamp ordering algorithm

The values present in the timestamp algorithm are Read TS(x) and write Ts(x).

Algorithm for timestamp ordering.

20. Explain: Database security and Authorization.

- a) Types of security
- b) Database Security and the DBA.
- c) Access protection, User accounts, and Database Audits.
- d) Types of Discretionary Privileges
- e) Specifying privileges using Views
- f) Revoking Privileges
- g) Propagation of privileges using the GRANT OPTION

