



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Database Management systems
Third Year CSE(Sem:I)
2 marks Questions and Answers

UNIT: 1

1. Define database management system?

Database management system (DBMS) is a collection of interrelated data and a set of programs to access those data.

2. List any eight applications of DBMS.

- a) Banking
- b) Airlines
- c) Universities
- d) Credit card transactions
- e) Tele communication
- f) Finance
- g) Sales
- h) Manufacturing
- i) Human resources

3. What are the disadvantages of file processing system?

The disadvantages of file processing systems are

- a) Data redundancy and inconsistency
- b) Difficulty in accessing data
- c) Data isolation
- d) Integrity problems
- e) Atomicity problems
- f) Concurrent access anomalies

4. What are the advantages of using a DBMS?

The advantages of using a DBMS are

- a) Controlling redundancy
- b) Restricting unauthorized access

- c) Providing multiple user interfaces
- d) Enforcing integrity constraints.
- e) Providing back up and recovery

5. Give the levels of data abstraction?

- a) Physical level
- b) logical level
- c) view level

6. Define instance and schema?

Instance: Collection of data stored in the data base at a particular moment is called an Instance of the database.

Schema: The overall design of the data base is called the data base schema.

7. Define the terms 1) physical schema 2) logical schema.

Physical schema: The physical schema describes the database design at the physical level, which is the lowest level of abstraction describing how the data are actually stored.

Logical schema: The logical schema describes the database design at the logical level, which describes what data are stored in the database and what relationship exists among the data.

8. What is conceptual schema?

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

9. Define data model?

A data model is a collection of conceptual tools for describing data, data relationships, data semantics and consistency constraints.

10. What is storage manager?

A storage manager is a program module that provides the interface between the low level data stored in a database and the application programs and queries submitted to the system.

11. What are the components of storage manager?

The storage manager components include

- a) Authorization and integrity manager
- b) Transaction manager
- c) File manager

d) Buffer manager

12. What is the purpose of storage manager?

The storage manager is responsible for the following

- a) Interaction with the file manager
- b) Translation of DML commands into low level file system commands
- c) Storing, retrieving and updating data in the database

13. List the data structures implemented by the storage manager.

The storage manager implements the following data structures

- a) Data files
- b) Data dictionary
- c) indices

14. What is a data dictionary?

A data dictionary is a data structure which stores meta data about the structure of the database i.e. the schema of the database.

15. What is an entity relationship model?

The entity relationship model is a collection of basic objects called entities and relationships among those objects. An entity is a thing or object in the real world that is distinguishable from other objects.

16. What are attributes? Give examples.

An entity is represented by a set of attributes. Attributes are descriptive properties possessed by each member of an entity set.

Example: possible attributes of customer entity are customer name, customer id, customer street, customer city.

17. What is relationship? Give examples

A relationship is an association among several entities.

Example: A depositor relationship associates a customer with each account that he/she has.

18. Define the terms

- i) Entity set
- ii) Relationship set

Entity set: The set of all entities of the same type is termed as an entity set.

Relationship set: The set of all relationships of the same type is termed as a relationship set.

19. Define single valued and multivalued attributes.

Single valued attributes: attributes with a single value for a particular entity are called single valued attributes.

Multivalued attributes: Attributes with a set of value for a particular entity are called multivalued attributes.

20. What are stored and derived attributes?

Stored attributes: The attributes stored in a data base are called stored attributes.

Derived attributes: The attributes that are derived from the stored attributes are called derived attributes.

21. What are composite attributes?

Composite attributes can be divided in to sub parts.

22. Define null values.

In some cases a particular entity may not have an applicable value for an attribute or if we do not know the value of an attribute for a particular entity. In these cases null value is used.

23. Define the terms

i) Entity type

ii) Entity set

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

Entity set: The set of all entities of the same type is termed as an entity set.

24. What is meant by the degree of relationship set?

The degree of relationship type is the number of participating entity types.

25. Define the terms

i) Key attribute

ii) Value set

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

Value set: Each simple attribute of an entity type is associated with a value set that specifies the set of values that may be assigned to that attribute for each individual entity.

26. Define weak and strong entity sets?

Weak entity set: entity set that do not have key attribute of their own are called

weak entity sets.

Strong entity set: Entity set that has a primary key is termed a strong entity set.

27. What does the cardinality ratio specify?

Mapping cardinalities or cardinality ratios express the number of entities to which another entity can be associated. Mapping cardinalities must be one of the following:

- One to one
- One to many
- Many to one
- Many to many

28. Explain the two types of participation constraint.

- Total: The participation of an entity set E in a relationship set R is said to be total if every entity in E participates in at least one relationship in R.
- Partial: if only some entities in E participate in relationships in R, the participation of entity set E in relationship R is said to be partial.

29. Define the terms

- DDL
- DML

DDL: Data base schema is specified by a set of definitions expressed by a special language called a data definition language.

DML: A data manipulation language is a language that enables users to access or manipulate data as organized by the appropriate data model.

30. Write short notes on relational model

The relational model uses a collection of tables to represent both data and the relationships among those data. The relational model is an example of a record based model.

31. Define tuple and attribute

- Attributes: column headers
- Tuple: Row

32. Define the term relation.

Relation is a subset of a Cartesian product of list domains.

33. Define tuple variable

Tuple variable is a variable whose domain is the set of all tuples.

34. Define the term Domain.

For each attribute there is a set of permitted values called the domain of that attribute.

35. What is a candidate key?

Minimal super keys are called candidate keys.

36. What is a primary key?

Primary key is chosen by the database designer as the principal means of identifying an entity in the entity set.

37. What is a super key?

A super key is a set of one or more attributes that collectively allows us to identify uniquely an entity in the entity set.

38. Define- relational algebra.

The relational algebra is a procedural query language. It consists of a set of operations that take one or two relation as input and produce a new relation as output.

39. What is a SELECT operation?

The select operation selects tuples that satisfy a given predicate. We use the lowercase letter σ to denote selection.

40. What is a PROJECT operation?

The project operation is a unary operation that returns its argument relation with certain attributes left out. Projection is denoted by π (p).

41. Write short notes on tuple relational calculus.

The tuple relational calculation is a non procedural query language. It describes the desired information without giving a specific procedure for obtaining that information.

A query or expression can be expressed in tuple relational calculus as

$$\{t \mid P(t)\}$$

which means the set of all tuples 't' such that predicate P is true for 't'.

Notations used:

- $t[A]$ @ the value of tuple 't' on attribute, A
- $t \in r$ @ tuple 't' is in relation 'r'
- \exists @ there exists

Definition for 'there exists' (\exists):

$\exists t \in r(Q(t))$

which means there exists a tuple 't' in relation 'r' such that predicate Q(t) is true.

• "® for all

Definition for 'for all' ("):

$\forall t \in r(Q(t))$

which means Q(t) is true for all tuples 't' in relation 'r'.

• ® Implication

Definition for Implication (>):

P Q means if P is true then Q must be true.

42. Write short notes on domain relational calculus

The domain relational calculus uses domain variables that take on values from an attribute domain rather than values for entire tuple.

43. Define query language?

A query is a statement requesting the retrieval of information. The portion of DML that involves information retrieval is called a query language.

44. Write short notes on Schema diagram.

A database schema along with primary key and foreign key dependencies can be depicted pictorially by schema diagram. Each relation appears as a box with attributes listed inside it and the relation name above it.

45. What is foreign key?

A relation schema r1 derived from an ER schema may include among its attributes the primary key of another relation schema r2. This attribute is called a foreign key from r1 referencing r2.

UNIT: 2

1. What are the parts of SQL language?

The SQL language has several parts:

data - definition language

Data manipulation language

View definition

Transaction control

Embedded SQL

Integrity

Authorization

2. What are the categories of SQL command?

SQL commands are divided in to the following categories:

1. data - definition language
2. data manipulation language
3. Data Query language
4. data control language
5. data administration statements
6. transaction control statements

3. What are the three classes of SQL expression?

SQL expression consists of three clauses:

Select

From

where

4. Give the general form of SQL query?

Select A1, A2....., An

From R1, R2....., Rm

Where P

5. What is the use of rename operation?

Rename operation is used to rename both relations and a attributes.

It uses the as clause, taking the form:

Old-name as new-name

6. Define tuple variable?

Tuple variables are used for comparing two tuples in the same relation. The tuple variables are defined in the from clause by way of the as clause.

7. List the string operations supported by SQL?

- 1) Pattern matching Operation
- 2) Concatenation
- 3) Extracting character strings
- 4) Converting between uppercase and lower case letters.

8. List the set operations of SQL?

- 1) Union
- 2) Intersect operation
- 3) The except operation

9. What is the use of Union and intersection operation?

Union: The result of this operation includes all tuples that are either in r1 or in r2 or in both r1 and r2. Duplicate tuples are automatically eliminated.

Intersection: The result of this relation includes all tuples that are in both r1 and r2.

10. What are aggregate functions? And list the aggregate functions supported by SQL?

Aggregate functions are functions that take a collection of values as input and return a single value.

Aggregate functions supported by SQL are

Average: avg

Minimum: min

Maximum: max

Total: sum

Count: count

11. What is the use of group by clause?

Group by clause is used to apply aggregate functions to a set of tuples. The

attributes given in the group by clause are used to form groups. Tuples with the same value on all attributes in the group by clause are placed in one group.

12. What is the use of sub queries?

A sub query is a select-from-where expression that is nested with in another query. A common use of sub queries is to perform tests for set membership, make set comparisons, and determine set cardinality.

13. What is view in SQL? How is it defined?

Any relation that is not part of the logical model, but is made visible to a user as a virtual relation is called a view.

We define view in SQL by using the create view command. The form of the create view command is

Create view v as

14. What is the use of with clause in SQL?

The with clause provides a way of defining a temporary view whose definition is available only to the query in which the with clause occurs.

15. List the table modification commands in SQL?

Deletion

Insertion

Updates

Update of a view

16. List out the statements associated with a database transaction?

Commit work

Rollback work

17. What is transaction?

Transaction is a unit of program execution that accesses and possibly updated various data items.

18. List the SQL domain Types?

SQL supports the following domain types.

1) Char(n) 2) varchar(n) 3) int 4) numeric(p,d)

5) float(n) 6) date.

19. What is the use of integrity constraints?

Integrity constraints ensure that changes made to the database by authorized users do not result in a loss of data consistency. Thus integrity constraints guard against

accidental damage to the database.

20. Mention the 2 forms of integrity constraints in ER model?

Key declarations

Form of a relationship

21. What is trigger?

Triggers are statements that are executed automatically by the system as the side effect of a modification to the database.

22. What are domain constraints?

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.

23. What are referential integrity constraints?

A value that appears in one relation for a given set of attributes also appears for a certain set of attributes in another relation.

24. What is assertion? Mention the forms available.

An assertion is a predicate expressing a condition that we wish the database always to satisfy.

Domain integrity constraints.

Referential integrity constraints

25. Give the syntax of assertion?

Create assertion check

26. What is the need for triggers?

Triggers are useful mechanisms for alerting humans or for starting certain tasks automatically when certain conditions are met.

27. List the requirements needed to design a trigger.

The requirements are

Specifying when a trigger is to be executed.

Specify the actions to be taken when the trigger executes.

28. Give the forms of triggers?

The triggering event can be insert or delete.

For updated the trigger can specify columns.

The referencing old row as clause

The referencing new row as clause

The triggers can be initiated before the event or after the event.

29. What does database security refer to?

Database security refers to the protection from unauthorized access and malicious destruction or alteration.

30. List some security violations (or) name any forms of malicious access.

Unauthorized reading of data

Unauthorized modification of data

Unauthorized destruction of data.

31. List the types of authorization.

Read authorization

Write authorization

Update authorization

Drop authorization

32. What is authorization graph?

Passing of authorization from one user to another can be represented by an authorization graph.

33. List out various user authorization to modify the database schema.

Index authorization

Resource authorization

Alteration authorization

Drop authorization

34. What are audit trails?

An audit trail is a log of all changes to the database along with information such as which user performed the change and when the change was performed.

35. Mention the various levels in security measures.

Database system

Operating system

Network

Physical

human

36. Name the various privileges in SQL?

Delete

Select

Insert

update

37. Mention the various user privileges.

All privileges directly granted to the user or role.

All privileges granted to roles that have been granted to the user or role.

38. Give the limitations of SQL authorization.

The code for checking authorization becomes intermixed with the rest of the application code.

Implementing authorization through application code rather than specifying it declaratively in SQL makes it hard to ensure the absence of loopholes.

39. Give some encryption techniques?

DES

AES

Public key encryption

40. What does authentication refer?

Authentication refers to the task of verifying the identity of a person.

41. List some authentication techniques.

Challenge response scheme

Digital signatures

Nonrepudiation

42. Define Boyce codd normal form

A relation schema R is in BCNF with respect to a set F of functional dependencies if, for all functional dependencies in F

+

of the form. \rightarrow , where

43. List the disadvantages of relational database system

Repetition of data

Inability to represent certain information.

44. What is first normal form?

The domain of attribute must include only atomic (simple, indivisible) values.

45. What is meant by functional dependencies?

Consider a relation schema R and C R and C R. The functional dependency

holds on relational schema R if in any legal relation $r(R)$, for all pairs of

tuples t_1 and t_2 in r such that $t_1[A] = t_2[A]$, and also $t_1[B] = t_2[B]$.

46. What are the uses of functional dependencies?

To test relations to see whether they are legal under a given set of functional dependencies.

To specify constraints on the set of legal relations.

47. Explain trivial dependency?

Functional dependency of the form $A \rightarrow B$ is trivial if $A \supseteq B$. Trivial functional dependencies are satisfied by all the relations.

48. What are axioms?

Axioms or rules of inference provide a simpler technique for reasoning about functional dependencies.

49. What is meant by computing the closure of a set of functional dependency?

The closure of F denoted by F^+ is the set of functional dependencies logically implied by F .

50. What is meant by normalization of data?

It is a process of analyzing the given relation schemas based on their Functional Dependencies (FDs) and primary key to achieve the properties
Minimizing redundancy
Minimizing insertion, deletion and updating anomalies.

51. Define canonical cover?

A canonical cover F_c for F is a set of dependencies such that F logically implies all dependencies in F_c and F_c logically implies all dependencies in F . F_c must have the following properties.

52. List the properties of canonical cover.

F_c must have the following properties.

No functional dependency in F_c contains an extraneous attribute.

Each left side of a functional dependency in F_c is unique.

53. Explain the desirable properties of decomposition.

Lossless-join decomposition

Dependency preservation

Repetition of information

54. What is 2NF?

A relation schema R is in 2NF if it is in 1NF and every non-prime attribute A in R

is fully functionally dependent on primary key.

UNIT: 3

1. Give the measures of quality of a disk.

Capacity

Access time

Seek time

Data transfer rate

Reliability

Rotational latency time.

2. Compare sequential access devices versus random access devices with an example

sequential access devices random access devices

Must be accessed from the beginning It is possible to read data from any location

Eg:- tape storage Eg:-disk storage

Access to data is much slower Access to data is faster

Cheaper than disk Expensive when compared with disk

3. What are the types of storage devices?

Primary storage

Secondary storage

Tertiary storage

4. Draw the storage device hierarchy according to their speed and their cost.

5. What are called jukebox systems?

Jukebox systems contain a few drives and numerous disks that can be loaded into one of the drives automatically.

6. What is called remapping of bad sectors?

If the controller detects that a sector is damaged when the disk is initially formatted, or when an attempt is made to write the sector, it can logically map the sector to a different physical location.

7. Define access time.

Access time is the time from when a read or write request is issued to when data transfer begins.

8. Define seek time.

The time for repositioning the arm is called the seek time and it increases with the distance that the arm is called the seek time.

9. Define average seek time.

The average seek time is the average of the seek times, measured over a sequence of random requests.

Cache

Main memory

Flash memory

Magnetic disk

Optical disk

Magnetic tapes

10. Define rotational latency time.

The time spent waiting for the sector to be accessed to appear under the head is called the rotational latency time.

11. Define average latency time.

The average latency time of the disk is one-half the time for a full rotation of the disk.

12. What is meant by data-transfer rate?

The data-transfer rate is the rate at which data can be retrieved from or stored to the disk.

13. What is meant by mean time to failure?

The mean time to failure is the amount of time that the system could run continuously without failure.

14. What are a block and a block number?

A block is a contiguous sequence of sectors from a single track of one platter.

Each request specifies the address on the disk to be referenced. That address is in the form of a block number.

15. What are called journaling file systems?

File systems that support log disks are called journaling file systems.

16. What is the use of RAID?

A variety of disk-organization techniques, collectively called redundant arrays of independent disks are used to improve the performance and reliability.

17. Explain how reliability can be improved through redundancy?

The simplest approach to introducing redundancy is to duplicate every disk. This technique is called mirroring or shadowing. A logical disk then consists of two physical disks, and write is carried out on both the disk. If one of the disks fails the data can be read from the other. Data will be lost if the second disk fails before the first failed disk is repaired.

18. What is called mirroring?

The simplest approach to introducing redundancy is to duplicate every disk. This technique is called mirroring or shadowing.

19. What is called mean time to repair?

The mean time to failure is the time it takes to replace a failed disk and to restore the data on it.

20. What is called bit-level striping?

Data striping consists of splitting the bits of each byte across multiple disks. This is called bit-level striping.

21. What is called block-level striping?

Block level striping stripes blocks across multiple disks. It treats the array of disks as a large disk, and gives blocks logical numbers.

22. What are the two main goals of parallelism?

Load –balance multiple small accesses, so that the throughput of such accesses increases.

Parallelize large accesses so that the response time of large accesses is reduced

23. What are the factors to be taken into account when choosing a RAID level?

- o Monetary cost of extra disk storage requirements.
- o Performance requirements in terms of number of I/O operations
- o Performance when a disk has failed.
- o Performances during rebuild.

24. What is meant by software and hardware RAID systems?

RAID can be implemented with no change at the hardware level, using only

software modification. Such RAID implementations are called software RAID systems and the systems with special hardware support are called hardware RAID systems.

25. Define hot swapping?

Hot swapping permits the removal of faulty disks and replaces it by new ones without turning power off. Hot swapping reduces the mean time to repair.

26. Which level of RAID is best? Why?

RAID level 1 is the RAID level of choice for many applications with moderate storage requirements and high I/O requirements. RAID 1 follows mirroring and provides best write performance.

27. Distinguish between fixed length records and variable length records?

Fixed length records

Every record has the same fields and field lengths are fixed.

Variable length records

File records are of same type but one or more of the fields are of varying size.

28. What are the ways in which the variable-length records arise in database systems?

Storage of multiple record types in a file.

Record types that allow variable lengths for one or more fields.

Record types that allow repeating fields.

29. Explain the use of variable length records.

They are used for Storing of multiple record types in a file.

Used for storing records that has varying lengths for one or more fields.

Used for storing records that allow repeating fields

30. What is the use of a slotted-page structure and what is the information present in the header?

The slotted-page structure is used for organizing records within a single block.

The header contains the following information.

The number of record entries in the header.

The end of free space

An array whose entries contain the location and size of each record.

31. What are the two types of blocks in the fixed –length representation? Define them.

- Anchor block: Contains the first record of a chain.

- Overflow block: Contains the records other than those that are the first record of a chain.

32. What is known as heap file organization?

In the heap file organization, any record can be placed anywhere in the file where there is space for the record. There is no ordering of records. There is a single file for each relation.

33. What is known as sequential file organization?

In the sequential file organization, the records are stored in sequential order, according to the value of a “search key” of each record.

34. What is hashing file organization?

In the hashing file organization, a hash function is computed on some attribute of each record. The result of the hash function specifies in which block of the file the record should be placed.

35. What is known as clustering file organization?

In the clustering file organization, records of several different relations are stored in the same file.

36. What is an index?

An index is a structure that helps to locate desired records of a relation quickly, without examining all records.

37. What are the two types of ordered indices?

Primary index

Secondary index

38. What are the types of indices?

Ordered indices

Hash indices

39. What are the techniques to be evaluated for both ordered indexing and hashing?

Access types

Access time

Insertion time

Deletion time

Space overhead

40. What is known as a search key?

An attribute or set of attributes used to look up records in a file is called a search

key.

41. What is a primary index?

A primary index is an index whose search key also defines the sequential order of the file.

42. What are called index-sequential files?

The files that are ordered sequentially with a primary index on the search key are called index-sequential files.

43. What are the two types of indices?

Dense index

Sparse index

44. What are called multilevel indices?

Indices with two or more levels are called multilevel indices.

45. What are called secondary indices?

Indices whose search key specifies an order different from sequential order of the file are called secondary indices. The pointers in secondary index do not point directly to the file. Instead each points to a bucket that contains pointers to the file.

46. What are the disadvantages of index sequential files?

The main disadvantage of the index sequential file organization is that performance degrades as the file grows. This degradation is remedied by reorganization of the file.

47. What is a B+-Tree index?

A B+-Tree index takes the form of a balanced tree in which every path from the root of the root of the root of the tree to a leaf of the tree is of the same length.

Draw the structure of a B+ tree and explain briefly.

A node contains up to n-1 search key values and n pointers.

P1 K1 P2 K2 Pn-1 Kn-1 Pn

48. What is B-Tree?

A B-tree eliminates the redundant storage of search-key values .It allows search key values to appear only once.

49. What is hashing?

Hashing allows us to find the address of a data item directly by computing a hash function on the search key value of the desired record.

50. How do you create index in SQL?

We create index by the create index command.

Create index on ()

51. Distinguish between static hashing and dynamic hashing?

Static hashing

Static hashing uses a hash function in which the set of bucket addresses is fixed. Such hash functions cannot easily accommodate databases that grow larger over time.

Dynamic hashing

Dynamic hashing allows us to modify the hash function dynamically.

Dynamic hashing copes with changes in database size by splitting and coalescing buckets as the database grows and shrinks.

52. What is a hash index?

A hash index organizes the search keys, with their associated pointers, into a hash file structure.

53. What can be done to reduce the occurrences of bucket overflows in a hash file organization?

To reduce bucket overflow the number of buckets is chosen to be $(nr/fr) * (1+d)$.

We handle bucket overflow by using

- Overflow chaining (closed hashing)
- Open hashing

54. Differentiate open hashing and closed hashing (overflow chaining)

Closed hashing (overflow chaining)

If a record must be inserted into a bucket b , and b is already full, the system provides an overflow bucket for b , and inserts the record into the overflow bucket. If the overflow bucket is also full, the system provides another overflow bucket, and so on. All the overflow buckets of a given bucket are chained together in a linked list, overflow handling using linked list is known as closed hashing.

Open hashing

The set of buckets is fixed, and there are no overflow chains. Instead, if a bucket is full, the system inserts records into some other bucket in the initial set of buckets.

55. What is linear probing?

Linear probing is a type of open hashing. If a bucket is full the system inserts

records in to the next bucket that has space. This is known as linear probing.

56. What is called query processing?

Query processing refers to the range of activities involved in extracting data from a database.

57. What are the steps involved in query processing?

The basic steps are:

parsing and translation

optimization

evaluation

58. What is called an evaluation primitive?

A relational algebra operation annotated with instructions on how to evaluate is called an evaluation primitive.

59. What is called a query evaluation plan?

A sequence of primitive operations that can be used to evaluate a query is a query evaluation plan or a query execution plan.

60. What is called a query –execution engine?

The query execution engine takes a query evaluation plan, executes that plan, and returns the answers to the query.

61. How do you measure the cost of query evaluation?

The cost of a query evaluation is measured in terms of a number of different resources including disk accesses, CPU time to execute a query, and in a distributed database system the cost of communication

62. List out the operations involved in query processing

Selection operation

Join operations.

Sorting.

Projection

Set operations

Aggregation

63. What are called as index scans?

Search algorithms that use an index are referred to as index scans.

64. What is called as external sorting?

Sorting of relations that do not fit into memory is called as external sorting.

65. Explain nested loop join?

Nested loop join consists of a pair of nested for loops.

Example: $r \parallel s$

r is the outer relation and s is the inner relation.

66. What is meant by block nested loop join?

Block nested loop join is the variant of the nested loop join where every block of the inner relation is paired with every block of the outer relation. Within each pair of blocks every tuple in one block is paired with every tuple in the other blocks to generate all pairs of tuples.

67. What is meant by hash join?

In the hash join algorithm a hash function h is used to implement partition tuples of both relations.

68. What is called as recursive partitioning?

The system repeats the splitting of the input until each partition of the build input fits in the memory. Such partitioning is called recursive partitioning.

69. What is called as an N-way merge?

The merge operation is a generalization of the two-way merge used by the standard in-memory sort-merge algorithm. It merges N runs, so it is called an N-way merge.

70. What is known as fudge factor?

The number of partitions is increased by a small value called the fudge factor, which is usually 20 percent of the number of hash partitions computed.

71. Define query optimization.

Query optimization refers to the process of finding the lowest –cost method of evaluating a given query.

UNIT: 4

1. What is transaction?

Collections of operations that form a single logical unit of work are called transactions.

2. What are the two statements regarding transaction?

The two statements regarding transaction of the form:

Begin transaction

End transaction

3. What are the properties of transaction?

The properties of transactions are:

Atomicity

Consistency

Isolation

Durability

4. What is recovery management component?

Ensuring durability is the responsibility of a software component of the base system called the recovery management component.

5. When is a transaction rolled back?

Any changes that the aborted transaction made to the database must be undone.

Once the changes caused by an aborted transaction have been undone, then the transaction has been rolled back.

6. What are the states of transaction?

The states of transaction are

Active

Partially committed

Failed

Aborted

Committed

Terminated

7. What is a shadow copy scheme?

It is simple, but efficient, scheme called the shadow copy schemes. It is based on making copies of the database called shadow copies that one transaction is active at a time. The scheme also assumes that the database is simply a file on disk.

8. Give the reasons for allowing concurrency?

The reasons for allowing concurrency is if the transactions run serially, a short transaction may have to wait for a preceding long transaction to complete, which can lead to unpredictable delays in running a transaction.

So concurrent execution reduces the unpredictable delays in running transactions.

9. What is average response time?

The average response time is that the average time for a transaction to be completed after it has been submitted.

10. What are the two types of serializability?

The two types of serializability is

Conflict serializability

View serializability

11. Define lock?

Lock is the most common used to implement the requirement is to allow a transaction to access a data item only if it is currently holding a lock on that item.

12. What are the different modes of lock?

The modes of lock are:

Shared

Exclusive

13. Define deadlock?

Neither of the transaction can ever proceed with its normal execution. This situation is called deadlock.

14. Define the phases of two phase locking protocol

Growing phase: a transaction may obtain locks but not release any lock.

Shrinking phase: a transaction may release locks but may not obtain any new locks.

15. Define upgrade and downgrade?

It provides a mechanism for conversion from shared lock to exclusive lock is known as upgrade.

It provides a mechanism for conversion from exclusive lock to shared lock is known as downgrade.

16. What is a database graph?

The partial ordering implies that the set D may now be viewed as a directed acyclic graph, called a database graph.

17. What are the two methods for dealing deadlock problem?

The two methods for dealing deadlock problem is deadlock detection and deadlock recovery.

18. What is a recovery scheme?

An integral part of a database system is a recovery scheme that can restore the database to the consistent state that existed before the failure.

19. What are the two types of errors?

The two types of errors are:

Logical error

System error

20. What are the storage types?

The storage types are:

Volatile storage

Nonvolatile storage

21. Define blocks?

The database system resides permanently on nonvolatile storage, and is partitioned into fixed-length storage units called blocks.

22. What is meant by Physical blocks?

The input and output operations are done in block units. The blocks residing on the disk are referred to as physical blocks.

23. What is meant by buffer blocks?

The blocks residing temporarily in main memory are referred to as buffer blocks.

24. What is meant by disk buffer?

The area of memory where blocks reside temporarily is called the disk buffer.

25. What is meant by log-based recovery?

The most widely used structures for recording database modifications is the log.

The log is a sequence of log records, recording all the update activities in the database. There are several types of log records.

26. What are uncommitted modifications?

The immediate-modification technique allows database modifications to be output to the database while the transaction is still in the active state. Data modifications written by active transactions are called uncommitted modifications.

27. Define shadow paging.

An alternative to log-based crash recovery technique is shadow paging. This technique needs fewer disk accesses than do the log-based methods.

28. Define page.

The database is partitioned into some number of fixed-length blocks, which are referred to as pages.

29. Explain current page table and shadow page table.

The key idea behind the shadow paging technique is to maintain two page tables during the life of the transaction: the current page table and the shadow page table. Both the page tables are identical when the transaction starts. The current page table may be changed when a transaction performs a write operation.

30. What are the drawbacks of shadow-paging technique?

- Commit Overhead
- Data fragmentation
- Garbage collection

30. Define garbage collection.

Garbage may be created also as a side effect of crashes. Periodically, it is necessary to find all the garbage pages and to add them to the list of free pages. This process is called garbage collection.

32. Differentiate strict two phase locking protocol and rigorous two phase locking protocol.

In strict two phase locking protocol all exclusive mode locks taken by a transaction is held until that transaction commits.

Rigorous two phase locking protocol requires that all locks be held until the transaction commits.

33. How the time stamps are implemented

- Use the value of the system clock as the time stamp. That is a transaction's

time stamp is equal to the value of the clock when the transaction enters the system.

- Use a logical counter that is incremented after a new timestamp has been assigned; that is the time stamp is equal to the value of the counter.

34. What are the time stamps associated with each data item?

- W-timestamp (Q) denotes the largest time stamp if any transaction that executed WRITE (Q) successfully.
- R-timestamp (Q) denotes the largest time stamp if any transaction that executed READ (Q) successfully.

SCHEM

UNIT: 5

1. List the threats to databases.

- Loss of integrity
- Loss of availability
- Loss of confidentiality

2. List out the control measures.

- Access control
- Inference control
- Flow control
- Data encryption

3. What is meant by Data warehouse?

A data warehouse is a repository (archive) of information gathered from multiple sources, stored under a unified schema at a single site.

- Greatly simplifies querying, permits study of historical trends
- Shifts decision support query load away from transaction processing systems

4. Define Data mining.

Data mining - knowledge discovery in database. Data mining is the process of semi-automatically analyzing large databases to find useful patterns.

5. List out the operations done in data warehouse.

- Roll up
- Drill down
- Pivot

- Sorting
- Selection

6. List out the functionalities of Data warehouse.

- Data cleaning

- Data transformation
- Data integration
- Data loading &
- Periodic data refreshing

7. List the types of security mechanisms.

- Discretionary security mechanisms
- Mandatory security mechanisms

8. What are the database design issues?

- Legal and ethical issues
- Policy issues
- System related issues

9. What are the actions performed by DBA?

- Account creation
- Privilege granting
- Privilege revocation
- Security level assignment

10. What are the steps for designing a warehouse?

- Choose a *business process* to model
- Choose the *grain* of the business process
- Choose the *dimensions* that will apply to each fact table record
- Choose the *measures* that will populate each fact table record

11. What are the issues in data warehouse design?

- When and how to gather data
- What schema to use
- Data cleansing
- How to propagate updates
- What data to summarize

12. What are the goals of data mining?

- Prediction
- Identification
- Classification
- Optimization

13. List out the types of Discovered knowledge.

- Association rules
- Classification Hierarchies
- Sequential patterns
- Patterns within time series
- Clustering

14. What is meant by Association rule?

An association rule is of the form $X \rightarrow Y$, where $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_n\}$ are set of items with x_i and y_j being distinct items of all i and j . It must satisfy a minimum support and confidence.

15. What is meant by Confidence rule?

Given a rule of the form $A \rightarrow B$, rule confidence is the conditional probability that B is true when A is known to be true.

16. Define Apriori algorithm.

The Apriori algorithm was the first algorithm used to generate association rules. It uses the general algorithm for creating association rules together with downward closure and anti-monotonicity.

17. Define Sampling algorithm.

The Sampling algorithm selects samples from the database of transactions that individually fit into memory. Frequent itemsets are then formed for each sample. If the frequent itemsets form a superset of frequent itemsets for the entire database, then the real frequent itemsets can be obtained by scanning the remainder of the database.

18. What is meant by frequent pattern tree algorithm?

The Frequent pattern tree algorithm reduces the total number of candidate itemsets by producing a compressed version of the database in terms of an FP-tree. The FP-tree stores relevant information and allows for the efficient description of frequent itemsets. The algorithm consists of 2 steps:

1. Build FP-tree
2. Use the tree to find frequent itemsets.

19. What is meant by Classification?

Classification is the process of learning a model that is able to describe different classes of data.

20. List the applications of data mining.

- Marketing
- Finance
- Resource optimization
- Image Analysis
- Fraud detection