DHANALAKSHMI SRINIVASAN INSTITUTE OF RESEARCH AND TECHNOLOGY

SIRUVACHUR, PERAMBALUR – 611 113

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CS6302-DATABASE MANAGEMENT SYSTEMS

QUESTION BANK/TWO MARKS

UNIT: 1- INTRODUCTION

1. Who is a DBA? What are the responsibilities of a DBA? April/May-2011

A database administrator (short form DBA) is a person responsible for the design, implementation, maintenance and repair of an organization's database. They are also known by the titles Database Coordinator or Database Programmer, and is closely related to the Database Analyst, Database Modeller, Programmer Analyst, and Systems Manager.

The role includes the development and design of database strategies, monitoring and improving database performance and capacity, and planning for future expansion requirements. They may also plan, co-ordinate and implement security measures to safeguard the database

2. What is a data model? List the types of data model used. April/May-2011

A **database model** is the theoretical foundation of a database and fundamentally determines in which manner data can be stored, organized, and manipulated in a database system. It thereby defines the infrastructure offered by a particular database system. The most popular example of a database model is the relational model.

Types of data model used

- ➢ Hierarchical model
- ➢ Network model
- ➢ Relational model
- ➢ Entity-relationship
- ➢ Object-relational model
- ➢ Object model

3. Define database management system?

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

4. 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

5. 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

6. 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 backup and recovery

7. Give the levels of data abstraction?

a) Physical level

b) Logical level

c) View level

8. 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.

9. 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.

10. Mention the actors on scene.

    ➢ Database administrator

    ➢ Database designer

    ➢ End users

11. What is conceptual schema?

The schemas at the view level are called subschema"s that describe different views of the database.

12. Define data model?

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

13. 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.

14. 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

15. 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 in to low level file system commands
  - c) Storing, retrieving and updating data in the database

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

The storage manager implements the following data structure
  - a) Data files
  - b) Data dictionary
  - c) Indices

17. What is a data dictionary?

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

18. What is an entity relationship model?

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

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. What are composite attributes?
Composite attributes can be divided in to sub parts.

25. 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.

26. 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.

27. What is meant by the degree of relationship set?
The degree of relationship type is the number of participating entity types.

28. 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.

29. 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.

30. 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

31. 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.**

32. What is meant by lossless-join decomposition? APRIL/MAY-2011
We claim the above decomposition is lossless. How can we decide whether decomposition is lossless?
1. Let R be a relation schema.
2. Let F be a set of functional dependencies on R.
3. Let R_1 and R_2 form a decomposition of R.
4. The decomposition is a lossless-join decomposition of R if at least one of the following functional dependencies are in :
a. $R_1 \cap R_2 \rightarrow R_1$
b. $R_1 \cap R_2 \rightarrow R_2$

33. List the disadvantages of relational database system
Repetition of data
Inability to represent certain information.

34. What is first normal form?
The domain of attribute must include only atomic (simple, indivisible) values.

35. What is meant by functional dependencies?
Consider a relation schema R and a C R and ß C R. The functional dependency a ß holds on relational schema R if in any legal relation r(R), for all pairs of tuples t1 and t2 in r such that t1 [a] =t1 [a], and also t1 [ß] =t2 [ß].

36. 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.

37. Explain trivial dependency?
Functional dependency of the form a ß is trivial if ß C a. Trivial functional dependencies are satisfied by all the relations.

38. What are axioms?
Axioms or rules of inference provide a simpler technique for reasoning about functional dependencies.

39. What is meant by computing the closure of a set of functional dependency?
+ The closure of F denoted b y F is the set of functional dependencies logically implied by F.

40. 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

UNIT:2-RELATIONAL MODEL

1. Define the terms i) DDL ii) 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

2. What is embedded SQL? What are its advantages? April/May-2011
Embedded SQL is a method of combining the computing power of a programming language and the database manipulation capabilities of SQL. Embedded SQL statements are SQL statements written in line with the program source code of the host language. The embedded SQL statements are parsed by an embedded SQL preprocessor and replaced by host-language calls to a code library. The output from the preprocessor is then compiled by the host compiler. This allows programmers to embed SQL statements in programs written in any number of languages such as: C/C++, COBOL and Fortran

3. 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.

4. Define tuple and attribute
• **Attributes:** column headers
• **Tuple** : Row

5. **Define the term relation**.
Relation is a subset of a Cartesian product of list domains.

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

7. Define the term Domain.
For each attribute there is a set of permitted values called the *domain* of that attribute.

8. What is a candidate key?
Minimal super keys are called *candidate keys.*

9. 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.

10. 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.

11. List the table modification commands in SQL?
  ➢ Deletion
  ➢ Insertion
  ➢ Updates
  ➢ Update of a view

UNIT:3-TRANSACTION PROCESSING AND CONCURRENCY CONTROL

1. What are the ACID properties? APRIL/MAY-2011
(*atomicity, consistency, isolation, durability*) is a set of properties that guarantee database transactions are processed reliably. In the context of databases, a single logical operation on the data is called a transaction. For example, a transfer of funds from one bank account to another, even though that might involve multiple changes (such as debiting one account and crediting another), is a single transaction.

2. What are two pitfalls (problem) of lock-based protocols? APRIL/MAY-2011
  ➢ Deadlock
  ➢ Starvation

3. What is transaction?
Collections of operations that form a single logical unit of work are called transactions.

4. What are the two statements regarding transaction?
The two statements regarding transaction of the form:
  ➢ Begin transaction
  ➢ End transaction

5. What are the properties of transaction?
The properties o f transactions are:
  ➢ Atomicity
  ➢ Consistency
  ➢ Isolation
  ➢ Durability

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

7. 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.

8. What are the states of transaction?

The states of transaction are

- ➢ Active
- ➢ Partially committed
- ➢ Failed
- ➢ Aborted
- ➢ Committed
- ➢ Terminated
- ➢

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

- ➢ Commit work
- ➢ Rollback work

10. 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.

11. 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.

12. 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.

13. What are the two types of serializability?

The two types of serializability is

- ➢ Conflict serializability
- ➢ View serializability

14. 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.

15. What are the different modes of lock?

The modes of lock are:

- ➢ Shared

➢ Exclusive

16. Define deadlock?

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

17. 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.

18. 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.

19. 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.

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

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

21. 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.

22. What are the two types of errors?

The two types of errors are:
➢ Logical error
➢ System error

23. What are the storage types?

The storage types are:
➢ Volatile storage
➢ Nonvolatile storage

24. Define blocks?

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

25. 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.

26. What is meant by buffer blocks?

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

27. What is meant by disk buffer?
The area of memory where blocks reside temporarily is called the disk buffer.

28. 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.

29. 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.

30. 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.

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

32. 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 p age table. Both the page tables are identical when the transaction starts. The current page table may b e changed when a transaction performs a write operation.

33. What are the drawbacks of shadow-paging technique?
> Commit Overhead
> Data fragmentation
> Garbage collection

34. 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.

35. 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.

36. 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.

37. 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.

UNIT:4-TRENDS IN DATABASE TECHNOLOGY

1. What are the advantages and disadvantages of indexed sequential file? APRIL/MAY-2011
The advantage of ordering records in a sequential file according to a key is that you can then search the file more quickly. If you know the key value that you want, you can use one of the relatively fast searches. The disadvantage is that when you insert, you need to rewrite at least everything after the insertion point, which makes inserts very expensive unless they are done at the end of the file. An indexed file approach keeps a (hopefully) small part of each row, and some kind of "pointer" to the row's location within the data file. This allows a search to use the index, which is ordered the index and (again hopefully) much smaller and therefore much faster than scanning the entire data file for the indexed data.

2. What is database tuning? APRIL/MAY-2011
Database tuning describes a group of activities used to optimize and homogenize the performance of a database. It usually overlaps with query tuning, but refers to design of the database files, selection of the database management system (DBMS), operating system and CPU the DBMS runs on.

3. Give the measures of quality of a disk.
Capacity
Access time
Seek time
Data transfer rate
Reliability
Rotational latency time.

4. 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

5. What are the types of storage devices?
> ➢ Primary storage
> ➢ Secondary storage
> ➢ Tertiary storage

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

7. 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.

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

9. 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.

10. 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.

11. 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.

12. What are called journaling file systems?
File systems that support log disks are called journaling file systems.

13. 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.

14. 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 fail ed disk is repaired.

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

16. What is meant by Multi-dimensional database?
A multi-dimensional database is a computer software system designed to allow for efficient and convenient storage and retrieval of large volumes of data that is (1) intimately related and (2) stored, viewed and analyzed from different perspectives and these perspectives are called dimensions.

17. What is meant by Spatial database?
A spatial database is a database that is optimized to store query data that represents objects defined in geometric space. Most spatial databases allow representing simple geometric objects such as points, lines and polygons. Some spatial databases handle more complex structures such as 3D objects, topological coverages,etc.

18. What is meant by Mobile database?
A mobile database is a database that resides on mobile device such as a PDA, a smart phone, or a laptop. Such devices are often limited in resources such as memory, computing power and battery power.

UNIT:5-ADVANCED TOPICS

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. What is the difference between Information Retrieval and DBMS.

S.No	Information Retrieval	DBMS
1	Imprecise semantics	Precise semantics
2	Keyword search	SQL
3	Unstructured data format	Structured data
4	Reads mostly. Adds document occasionally.	Expects reasonable number of updates.
5	Displays page through top k results.	Generates full answer.

7. List out the functionalities of Data warehouse.
> ➢ Data cleaning
> ➢ Data transformation
> ➢ Data integration
> ➢ Data loading &
> ➢ Periodic data refreshing

8. List the types of security mechanisms.
> ➢ Discretionary security mechanisms
> ➢ Mandatory security mechanisms

9. What are the database design issues?
> ➢ Legal and ethical issues
> ➢ Policy issues
> ➢ System related issues

10. What are the actions performed by DBA?
> ➢ Account creation
> ➢ Privilege granting
> ➢ Privilege revocation
> ➢ Security level assignment

11. What is the difference between Database and Data warehouse?

S.No	Database	Data warehouse
1.	Online transaction & query processing is done in database	Data analysis & decision making is done in data warehouse
2.	Online Transaction Processing (OLTP) is carried out in Database.	Online Analytical Processing (OLAP) is carried out in Database

12. 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

13. 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

14. What are the goals of data mining?

➢ Prediction
➢ Identification
➢ Classification
➢ Optimization

15. List out the types of Discovered knowledge.

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

16. What is meant by Association rule?

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

17. What is meant by Confidence rule?

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

18. 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.

19. 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.

20. 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.

21. What is meant by Classification?

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

22. List the applications of data mining.

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