1. The three main components of the relational model do NOT include:

A. Structural aspect (relations/table)

B. Manipulative aspect (relational algebra/SQL)

C. Integrity constraints

D. Indexing mechanisms

2. Which SQL command is used to remove a privilege that was previously granted?

A. DELETE

B. DROP

C. REVOKE

D. REMOVE

3. The primary goal of normalization in relational database design is to:

A. Increase the speed of query execution.

B. Reduce data redundancy and avoid insertion, deletion, and update anomalies.

C. Enhance data security.

D. Make application programs simpler.

4. A relation is in its First Normal Form (1NF) if:

A. It has no partial functional dependencies.

B. All non-prime attributes are fully functionally dependent on the primary key.

C. Every attribute is atomic and single-valued.

D. It has no transitive dependencies.

5. The durability property of a transaction ensures that:

A. A transaction is executed as a single, indivisible unit.

B. The database remains in a consistent state before and after the transaction.

C. The effects of a committed transaction survive any subsequent failures.

D. Concurrent transactions do not interfere with each other.

6. In an E-R diagram, a diamond shape is used to represent a(n):

A. Entity set

B. Attribute

C. Relationship set

D. Primary key

7. Which SQL command is used to undo the changes made by a transaction that has not been committed?

A. COMMIT

B. UNDO

C. ROLLBACK

D. REVOKE

8. The software package that enables users to create, maintain, and control access to the database is called the:

A. Operating System

B. Database Management System (DBMS)

C. Application Program

D. File System

9. Regarding SQL views, which of the following statements is TRUE?

A. A view is physically stored as a separate table.

B. Any operation on a view is translated into an operation on the underlying base tables.

C. Changes to the base table structure never affect existing views.

D. All views are updatable.

10. In relational algebra, the operation that selects a subset of tuples (rows) from a relation that satisfies a given predicate is called:

A. Projection

B. Selection

C. Join

D. Union

11. Which of the following is NOT a typical problem caused by uncontrolled concurrent transaction execution?

A. Lost Update

B. Dirty Read

C. Unrepeatable Read

D. Data Encryption Failure

12. If a transaction T holds an _____ lock on a data item Q, then T can both read and write Q, and no other transaction can acquire any lock on Q until T releases its lock.

A. Shared Lock (S-lock)

B. Exclusive Lock (X-lock)

C. Intention-Shared Lock (IS-lock)

D. Update Lock (U-lock)

13. The database component responsible for ensuring the Atomicity and Durability of transactions is the:

A. Concurrency-Control Manager

B. Recovery Manager

C. Security and Authorization Manager

D. Query Optimizer

14. In the database design process, the step where an E-R model is transformed into a collection of relation schemas is part of the:

A. Requirements Analysis stage.

B. Conceptual Design stage.

C. Logical Design stage.

D. Physical Design stage.

15. The SQL statement SELECT COUNT(DISTINCT Dept) FROM Employee; **returns:**

A. The total number of tuples in the Employee table.

B. The total number of departments listed in the Employee table.

C. The number of distinct department names in the Employee table.

D. The number of employees in each distinct department.

16. If a relation schema R is in 3NF, then:

A. It must also be in BCNF.

B. It must also be in 2NF.

C. It might not be in 2NF.

D. All transitive dependencies have been removed.

17. Data independence in a DBMS includes physical independence and:

A. Operational Independence

B. Logical Independence

C. Structural Independence

D. Program Independence

18. To allow a user user1 to grant a specific privilege to other users, the GRANT statement must include the clause:

A. WITH ADMIN OPTION

B. WITH GRANT OPTION

C. WITH PRIVILEGE OPTION

D. WITH PERMISSION OPTION

19. Concerning database indexes, which statement is FALSE?

A. Indexes can improve the performance of selection queries.

B. A unique index is automatically created for a primary key.

C. Creating more indexes on a table will always improve the performance of update operations.

D. Indexes require additional storage space.

20. The Two-Phase Locking (2PL) protocol requires that every transaction issues lock and unlock requests in two distinct phases:

A. Read Phase and Write Phase

B. Locking Phase and Unlocking Phase

C. Growing Phase and Shrinking Phase

D. Acquire Phase and Release Phase

21. In the context of query processing, which of the following is NOT typically considered a cost factor in the cost-based optimization of a join operation?

A. The number of disk I/O operations required.

B. The cardinality (number of tuples) of the input relations.

C. The specific business logic or meaning of the data being joined.

D. The availability of indexes on the join attributes.

22. According to the ACID properties, the consistency property primarily enforces:

A. That all operations in a transaction are completed or none are.

B. That transactions do not interfere with each other.

C. That only explicitly stated integrity constraints are maintained.

D. That the database's stated integrity constraints are not violated upon transaction commit.

23. In a B+-tree index structure used for a database file, the leaf nodes primarily contain:

A. Pointers to the root of the tree.

B. Pointers to other non-leaf nodes.

C. Search-key values and pointers to the actual records (or blocks) in the data file.

D. Only the search-key values for range queries.

24. When estimating the cost of a BNL between two relations R and S, where R has 500 pages and S has 1000

pages, and 3 buffer pages are available, which of the following is the most accurate description of the cost?

A. The cost is primarily determined by the number of tuples in R and S.

B. The cost is dominated by reading R once and reading S once for each record of R.

C. The cost is dominated by reading R once and reading S once for each page of R.

D. The cost is approximately the same as a merge-join.

25. The primary purpose of a "write-ahead log" (WAL) protocol is to ensure:

A. High availability of the database system.

B. That transactions are executed concurrently without conflict.

C. That no transaction is allowed to read a data item before it is written.

D. That the log records are flushed to stable storage before the corresponding data pages are updated.

26. A relation schema R is in Boyce-Codd Normal Form (BCNF) if, for every one of its functional dependencies X → Y:

A. Y is a prime attribute.

B. X is a superkey.

C. X is a candidate key.

D. Y is contained in a candidate key.

27. The "system catalog" or "data dictionary" in a DBMS is best described as:

A. A set of tables containing metadata about the database, such as schemas, constraints, and indexes.

B. A log of all transactions performed on the database.

C. The primary storage area for user data.

D. A cache for frequently accessed data pages.

28. In the context of functional dependencies, the attribute(s) on the left-hand side of a functional dependency is/are called the:

A. Determinant

B. Dependent

C. Closure

D. Superkey

29. In the context of database architecture, the "physical schema" describes:

A. The logical structure of the entire database as seen by the DBA.

B. The part of the database that a particular user group is interested in.

C. The storage details, such as file organization and indexes.

D. The high-level data model used for design.

30. When estimating the cost of a merge-join operation between two relations R and S, which of the following factors is LEAST likely to significantly impact the overall I/O cost?

A. The number of buffer pages available for the merge operation.

B. The initial sorting cost if the relations are not already sorted on the join attribute.

C. The specific values of the join attributes in the tuples of R and S.

D. The number of pages in relations R and S.

31. In a database system, the primary purpose of the recovery manager is to ensure:

A. That transactions are executed as quickly as possible.

B. That the database can be restored to a consistent state after a failure.

C. That users are authenticated before accessing the database.

D. That data is distributed evenly across storage devices.

32. The concept of a "dense index" in file structures is characterized by:
A. Having an index entry for every search-key value in the file.
B. Having an index entry for every data block in the file.
C. Having multiple levels of indexing.
D. Being used only for primary keys.

33. In the context of transaction schedules, a "conflict serializable" schedule is one that:
A. Is equivalent to some serial schedule when considering only conflicting operations.
B. Must execute transactions one after another without interleaving.
C. Is always recoverable and cascade-free.
D. Guarantees that no deadlocks will occur.

34. In the context of database concurrency control, a "wait-die" scheme is used for:
A. Deadlock prevention based on transaction timestamps.
B. Deadlock detection using a wait-for graph.
C. Ensuring strict two-phase locking.
D. Optimistic concurrency control.

----

1. **True/False:** In the context of the ACID properties, the 'I' (Isolation) ensures that the execution of transactions concurrently will result in a state that is equivalent to a state where these transactions were executed serially in some order.
2. **True/False:** A view in SQL is physically stored as a separate table derived from the base tables.
3. **True/False:** In a B+-tree index, all leaf nodes are at the same depth from the root.
4. **True/False:** The SQL COMMIT statement makes all the changes made by the current transaction permanent and visible to other transactions.
5. **True/False:** A schedule is conflict serializable if and only if its precedence graph is acyclic.
6. **True/False:** The cost of a block nested-loop join between two relations is independent of the number of available buffer pages.
7. **True/False:** In two-phase locking (2PL), a transaction must release all its shared locks before it can acquire any exclusive locks.
8. **True/False:** The database system catalog is considered a part of the database itself, storing metadata about the database schema.
9. **True/False:** A functional dependency $X \rightarrow Y$ is a constraint that specifies that the value of X is uniquely determined by the value of Y.
10. **True/False:** In an E-R diagram, total participation of an entity set in a relationship set is represented by a double line.
11. **True/False:** The main advantage of a dense index is that it requires less storage space than a sparse index.
12. **True/False:** A checkpoint record in the log is used to reduce the amount of work needed during transaction rollback.
13. **True/False:** A relation in Third Normal Form (3NF) is automatically also in Boyce-Codd Normal Form (BCNF).

Database Schema:
EMPLOYEE (emp_id, emp_name, dept_id, salary)
- emp_id: Employee ID (Primary Key)
- emp_name: Employee Name
- dept_id: Department ID (Foreign Key)

- salary: Salary

DEPARTMENT (dept_id, dept_name, location)
- dept_id: Department ID (Primary Key)
- dept_name: Department Name
- location: Location

PROJECT (proj_id, proj_name, budget, dept_id)
- proj_id: Project ID (Primary Key)
- proj_name: Project Name
- budget: Project Budget
- dept_id: Department ID (Foreign Key)

WORKS_ON (emp_id, proj_id, hours)
- emp_id: Employee ID (Part of Primary Key, Foreign Key)
- proj_id: Project ID (Part of Primary Key, Foreign Key)
- hours: Hours worked

Questions:

1. List the names and salaries of all employees who earn more than $50,000.
- **Relational Algebra:**
- **SQL Query:**

2. Find the names of all employees and the departments they work in.
- **Relational Algebra:**
- **SQL Query:**

3. For each project, find the project name and the total number of hours worked on it.
- **Relational Algebra:**
- **SQL Query:**

4. Find departments where the average salary of employees is more than $60,000. Show department name and average salary.
- **Relational Algebra:**
- **SQL Query:**

5. Find employees who work on projects with a budget over $100,000 and work more than 20 hours per week on those projects. Show employee name, project name, budget, and hours.
- **Relational Algebra:**
- **SQL Query:**

City General Hospital needs a database system to manage its operations. After requirements analysis, the following facts were identified:

1. The hospital has multiple **DEPARTMENTS**. Each department has a unique **department ID**, **department name**, and a **location** (e.g., "North Wing, 3rd Floor").
2. Each department employs several **DOCTORS**. Each doctor has a unique **doctor ID**, **name**, **specialization**, and **phone number**. A doctor works in exactly one department.
3. Doctors treat **PATIENTS**. Each patient has a unique **patient ID**, **name**, **date of birth**, and **phone number**.
4. A patient can be treated by multiple doctors, and a doctor can treat multiple patients. For each treatment, we need to record the **treatment date** and **diagnosis**.
5. Patients are admitted to **ROOMS**. Each room has a unique **room number**, **room type** (e.g., "Private", "Semi-Private", "General"), and **daily rate**.
6. A patient can be admitted to only one room at a time, but a room can accommodate multiple patients (depending on room type). We need to track the **admission date** and **discharge date** for each patient's room assignment.
7. The hospital stores **MEDICATIONS**. Each medication has a unique **medication ID**, **medication name**, **manufacturer**, and **unit price**.

8. The hospital wants to track **TREATMENT TEAMS**. A treatment team is formed for complex cases and consists of multiple doctors treating the same patient. Each team has a **team leader** (one of the treating doctors) and a **team name**. The same doctor-patient treatment relationship can be part of multiple teams over time.
9. **Treatment teams** (not individual doctors) prescribe medications to patients. A team can prescribe multiple medications, and the same medication can be prescribed by different teams. For each prescription, we need to record the **prescription date**, **dosage**, and **frequency**.

Questions:

Part 1: Design an ER diagram that captures all the requirements mentioned above.

Part 2: Convert your ER diagram into a set of relational schemas.

Part 3: Consider the following relation that was created in an initial design phase:

TREATMENT_TEAM_DETAILS (team_id, team_name, leader_doctor_id, leader_name, patient_id, patient_name, treatment_date, diagnosis, dept_id, dept_name)

    a) Identify all functional dependencies in this relation.

    b) What is the highest normal form (1NF, 2NF, 3NF, BCNF) that this relation satisfies? Justify your answer.

    c) If the relation is not in 3NF, decompose it into a set of 3NF relations. Show all steps and specify keys for the new relations.

Consider a relation SUPPLIER_ORDER_DETAILS that tracks procurement information in a company:

SUPPLIER_ORDER_DETAILS (order_id, order_date, supplier_id, supplier_name, supplier_city, product_id, product_name, category, quantity, unit_price, total_amount)

Where:

- order_id: Unique identifier for each order
- order_date: Date when the order was placed
- supplier_id: Unique identifier for each supplier
- supplier_name: Name of the supplier
- supplier_city: City where the supplier is located
- product_id: Unique identifier for each product
- product_name: Name of the product
- category: Product category
- quantity: Quantity ordered
- unit_price: Price per unit of the product
- total_amount: Total amount for the order line (quantity × unit_price)

Part 1: Identify all the functional dependencies (FDs) in the SUPPLIER_ORDER_DETAILS relation. Consider both trivial and non-trivial dependencies.

Part 2:
1. Find all candidate keys for this relation.
2. Determine the highest normal form (1NF, 2NF, 3NF, or BCNF) that this relation satisfies. Justify your answer by showing which normal form conditions are violated.

Part 3:

Let F be the set of functional dependencies you identified in Part 1.
1. Compute the closure of the attribute set {supplier_id, product_id} (denoted as {supplier_id, product_id}$^+$)
2. Using Armstrong's axioms, prove whether the following additional functional dependencies can be inferred from F:
      ○ order_id $\rightarrow$ supplier_city
      ○ product_id $\rightarrow$ category

o   {order_id, product_id} → total_amount

Part 4:

Decompose the SUPPLIER_ORDER_DETAILS relation into a set of 3NF relations that preserve all functional dependencies and is lossless-join.

1. Show each step of the decomposition process.
2. For each resulting relation, specify:
   - o   All attributes
   - o   Primary key
   - o   Foreign keys (if any)
   - o   Functional dependencies that apply to that relation

Part 5: BCNF Decomposition

Decompose the original relation (or your 3NF relations) into BCNF. Indicate whether this decomposition is dependency-preserving and explain why.

Query Processing and Optimization Problem: University Database

Database Schema:

STUDENT (s_id, s_name, dept_name, tot_cred)

- 10,000 tuples, 200 pages
- Primary Key: s_id
- Primary B+ tree index on dept_name (height: 3)

COURSE (c_id, title, dept_name, credits)

- 2,000 tuples, 50 pages
- Primary Key: c_id
- Secondary B+ tree index on dept_name (height: 2)

TAKES (s_id, c_id, semester, year, grade)

- 50,000 tuples, 1,000 pages
- Primary Key: (s_id, c_id, semester, year)
- Primary B+ tree index on s_id (height: 4)
- Secondary B+ tree index on c_id (height: 3)

Assume:

- Memory buffer size: 11 pages
- Cost metric: number of disk I/O operations
- All indexes are dense
- No data is cached initially

Part 1: Query Analysis

Consider the following SQL query:

sql

```
SELECT S.s_name, C.title, T.grade
FROM STUDENT S, TAKES T, COURSE C
WHERE S.s_id = T.s_id
   AND T.c_id = C.c_id
   AND S.dept_name = 'Comp. Sci.'
   AND C.credits > 3;
```

1. Draw the query tree for this query.
2. Identify the selection predicates and join predicates.
3. Estimate the size (number of tuples) of each relation after applying selections.

Selection Statistics:

- 10% of students are in 'Comp. Sci.' department, i.e., Selectivity of 'Comp. Sci.' in STUDENT: 0.1
- 20% of courses have credits > 3, i.e., Selectivity of credits > 3 in COURSE: 0.2

Part 2: Join Operation Cost Analysis

Calculate the I/O cost for the following join methods for joining **STUDENT** and **TAKES** on s_id:

1. **Block Nested-Loop Join** (STUDENT as outer, TAKES as inner)
2. **Indexed Nested-Loop Join** using the clustered B+ tree index on TAKES.s_id
3. **Merge Join** (assuming both relations are not sorted on s_id)

Show your calculations step by step.

Part 3: Query Evaluation Plans

Propose two different query evaluation plans for the entire query and estimate their total cost.

Plan A:
- Apply selections early
- Use indexed nested-loop join for STUDENT ⋈ TAKES
- Use merge join for the result with COURSE

Plan B:
- Use merge join for STUDENT ⋈ TAKES
- Apply selections during join processing
- Use block nested-loop join for the result with COURSE

Calculate the total I/O cost for each plan and recommend the better one.

Part 4: Optimization Techniques

1. Explain how **pipelining** could be used to improve the performance of this query.
2. Would creating any additional indexes help this query? If so, specify the index and explain why.
3. How would the presence of **sorted files** on the join attributes affect your cost calculations?

Part 5: Cost Estimation Refinement

The university database system maintains the following statistics:
- Number of distinct dept_name values in STUDENT: 20
- Number of distinct c_id values in TAKES: 1,500
- Number of distinct credits values in COURSE: 5

Refine your cost estimates using these statistics and explain how they affect your query optimization decisions.

Consider the following schedule S involving T1 and T2:

text
T1: read(A123)
T2: read(A123)
T1: write(A123)
T2: read(A456)
T1: write(A456)
T2: read(A456)
T1: write(A789)
T1: write(B123)
T1: commit
T2: commit

1. Draw the **precedence graph** for this schedule.
2. Is this schedule **conflict serializable**? Justify your answer.
3. Is this schedule **recoverable**? **Cascade-less**? Explain.