Z/K/TX

Reg. No: 22BCI0189

Final Assessment Test – November 2024

VIT Vellore Institute of Technology

Course: **BCSE302L - Database Systems**

Class NBR(s): **1511/1522/1528/1533/1539/1545/1549/ 1554/1560/1569/1581/1603/1616/1649/1666/1683/ 1691/1704/1716/1727/1736**

Slot: **A2+TA2**

Time: **Three Hours**

Max. Marks: **100**

- **KEEPING MOBILE PHONE/ANY ELECTRONIC GADGETS, EVEN IN 'OFF' POSITION IS TREATED AS EXAM MALPRACTICE**
- **DON'T WRITE ANYTHING ON THE QUESTION PAPER**

Answer ALL Questions

(10 X 10 = 100 Marks)

1. i. Explain the difference between a database schema and a database state with a suitable example. **[4]**

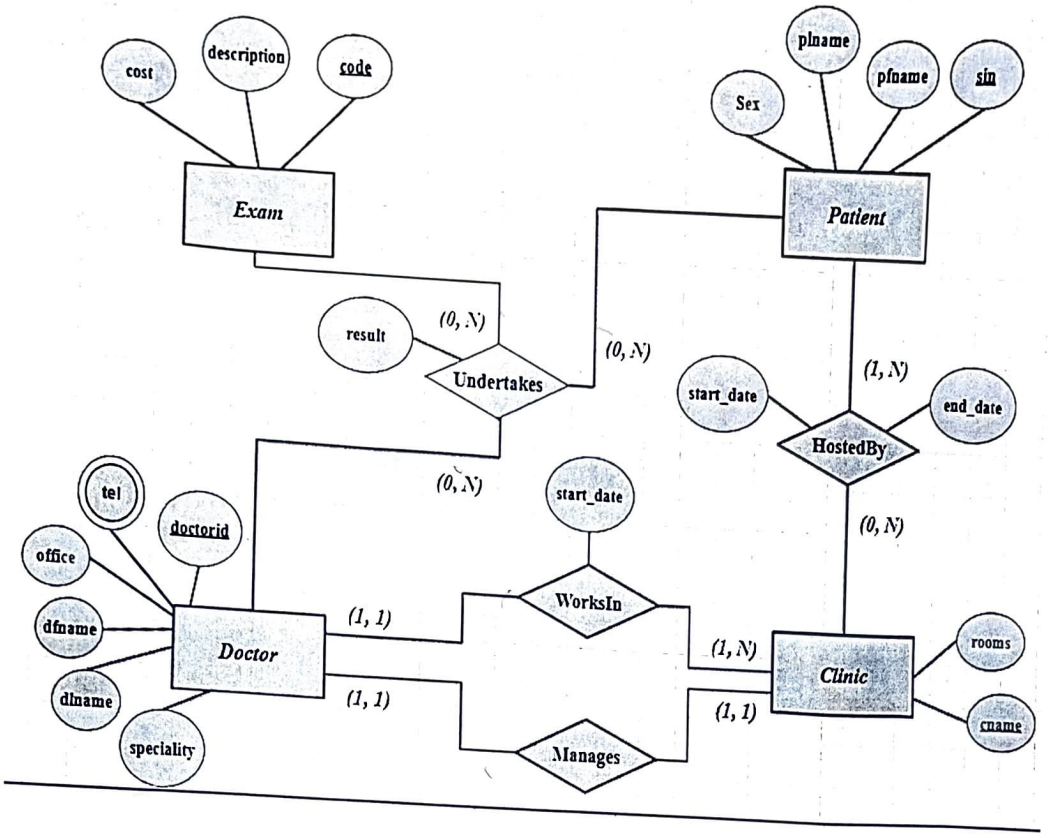
   ii. Differentiate between logical data independence and physical data independence, providing examples to illustrate their concepts. Finally, discuss the challenges and benefits associated with achieving these levels of data independence. **[6]**

2.a) Design an EER diagram to capture the requirements of an e-commerce platform that sells products online. The platform should allow users to browse products, add them to their shopping carts, checkout, and track their orders. Every product has a unique ID, name, description, price, quantity in stock, and category. Categories represent different product groups (e.g., electronics, clothing, books). Products can be categorized into different types (e.g., books, electronics, and clothing). Every customer has a unique ID, name, email, address, and shipping information. Each order has a unique ID, customer ID, order date, shipping address, and total amount. Each order can have multiple items. An order item specifies the product ID, quantity, and price. Each order has associated payment information, including payment method, transaction ID, and payment status. Customers can leave reviews for products, including a rating and a text comment. The platform can offer promotions and discounts on products.

State any assumptions you have that affects your design. Clearly indicate primary keys and the cardinalities with which an entity participates in a relationship with a pair of (minimum, maximum) value.

OR

2.b) Map the following ER diagram into relational schema(s) and specify all the relevant primary keys and foreign keys.



3. Suppose we have a database for an investment firm, consisting of the following attributes:

B – Broker,
O – Office of a broker
I – Investor
S – Stock
Q – Quantity of stock owned by an investor
D – Dividend paid by a stock.
Hence, the overall schema is R = (B, O, I, S, Q, D).

Assume that the following functional dependencies are required to hold on this database.

FD = {I → B, IS → Q, B → O, S → D}

i) List all the candidate keys for R.

ii) List the prime and non-prime attributes

iii) Classify to which normal form the present form of Relation R belongs to and justify the answer.

iv) Identify the FD's which violates the property of a particular normal form and perform lossless-join decomposition of R into subsequent normal forms (continue up to BCNF).

cify all the relevant

i. Consider the following relation R = {A, B, C, D, E} and the two sets of functional dependencies: [4]

F1 = { A → B, A → C, B → D, C → E }

F2 = { A → B, C → D, B → E }

Check whether F1 and F2 are equivalent.

ii. Suppose you are given a relation R(A,B,C,D). For each of the following sets of FDs, assuming the followings are the only dependencies that hold for R. Identify the candidate key(s) for R. Also state whether or not the proposed decomposition of R into smaller relations is a good decomposition (lossy or lossless) and briefly explain why or why not. [6]

a) B→C, D→ A; decompose into BC and AD.

b) A→BC, C→AD; decompose into ABC and AD.

c) A→B, B→C, C→D; decompose into AB and ACD.

5.a) Suppose we start with an empty B+ tree and keys arrive in the following order: 23, 7, 12, 20, 45, 30, 9, 14, 70, 25, 19, 8, 21, 17, 84. Construct a B+ tree of order 3. Delete 23, 9, 30, 20, 25 from the tree. Display the resultant tree after each deletion.

OR

5.b) i) Suppose that extendable hashing is being used on a database file that contains records with the following search key values: 8, 16, 4, 3, 14, 11, 12, 23, 42, 21. [5]

Construct the extendable hash structure for this file using the lowest-bits for the hash function (after converting the numbers into its binary form). That is, records in a bucket of local depth'd' agree on their rightmost 'n' bits. For example, the key 3 (0011) and 11 (1011) agree on their rightmost 3 bits (011). Each bucket can hold two records.

Answer the followings based on the constructed extendable hash structure.

1) What is the global depth of the resulting directory?

2) What is the local depth of the bucket contains key 14?

ii) Consider the following Publication Database [5]

Author

author_id	first_name	last_name
1	John	McCarthy
2	Dennis	Ritchie
3	Ken	Thompson
4	Claude	Shannon
5	Alan	Turing
6	Alonzo	Church
7	Perry	White
8	Moshe	Vardu
9	Roy	Batty

Author_Publication

author_id	pub_id	author_pos
1	1	1
2	2	1
3	2	2
4	3	1
5	4	1
5	5	1
6	6	1

Book

book_id	book_title	month	year	editor
1	CACM	April	1960	8
2	CACM	July	1974	8
3	BST	July	1948	2
4	LMS	November	1936	7
5	Mind	October	1950	NULL
6	AMS	Month	1941	NULL
7	AAAI	July	2012	9
8	NIPS	July	2012	9

Publication

pub-id	title	book_id
1	LISP	1
2	Unix	2
3	Info Theory	3
4	Turing Machines	4
5	Turing Test	5
6	Lambda Calculus	6

Based on the above database, answer the followings:

1) Write a relational algebra expression that returns the names of all authors who have at least one publication in the database.
2) Write a relational algebra expression that returns the names of all authors who are not book editors.
3) What question does the following expression answer?
   Author * (Author_Publication * ($\sigma_{month='July'}$ (Book) * Publication))
4) What question does the following relational algebra expression answer?
   $| \prod_{author\_id} (author) - \prod_{editor} (book) |$
5) How many tuples are returned by the following relational algebra expression?
   author $\bowtie_{author\ id=editor}$ book

6. i) For each of the following schedules, state whether they are serializable and conflict-serializable. For each schedule, draw the corresponding precedence graph. If the schedule is conflict-serializable, show all the conflict-equivalent serial schedules. If the schedule is not serializable, provide an initial DB state and some semantics (pseudo-code) for the transactions for which no serial schedule has the same net effect. **[6]**
   a) S_1 = r1(A) r2(A) w1(A) w2(A)
   b) S_2 = r1(A) r2(B) w3(A) r2(A) r1(B)
   c) S_3 = r1(A) w2(A) w1(A) r3(A)

   ii) Based on the scenarios given below, find out appropriate ACID property the scenarios are referring to and discuss various validation strategies to test those scenarios. **[4]**
   a) A bank account has an initial balance of $100. A transaction transfers $50 to another account. The database must ensure that the total balance of all accounts remains consistent before and after the transaction.
   b) The system experiences a power outage during a transaction. The transaction should be recoverable upon system restart.
   c) Multiple customers are trying to purchase the last available item of a product. The system should ensure that only one customer is able to successfully purchase the item, preventing race conditions.
   d) A customer places an order for multiple items. The transaction should either be fully committed (all items are added to the order and payment is processed) or fully rolled back (no changes are made to the database).

7. Define conflict serializable, recoverable, cascadeless, and strict schedules.
   For each of the following schedules, state which of the above mentioned properties hold. Justify your answer.
   a) S_1: R1(X), W2(X), W1(X), A2, C1
   b) S_2: R1(X), W2(X), W1(X), C2, C1
   c) S_3: R1(X), R2(X), W1(X), C1, W2(X), C2
   d) S_4: W1(X), R2(X), W1(X), C2, A1
   e) S_5: W1(X), R2(Y), R1(Y), R2(X), C1, C2
   f) S_6: R1(X), R2(X), C1, W2(X), C2

Analyze the trade-offs between different locking mechanisms and deadlock prevention protocols in terms of concurrency, performance, and overhead. Consider a scenario of a distributed database system, which is used to manage airline reservations, might encounter a deadlock situation. Explain how binary locks, exclusive/shared locks, and deadlock prevention protocols could be used to prevent or resolve the deadlock.

9. i. Explain why two-phase locking (2PL) is generally not used as a concurrency control method for indexes such as B+ trees. Provide a concrete example to illustrate the limitations of 2PL in this context. **[5]**

ii. Discuss the advantages and disadvantages of strict 2PL compared to other concurrency control mechanisms. Show that strict 2PL disallows the following schedule. **[5]**

$R1(X) \rightarrow R1(Y) \rightarrow R2(X) \rightarrow W2(X) \rightarrow R2(Y) \rightarrow W1(X) \rightarrow W1(Y) \rightarrow W2(Y) \rightarrow C1 \rightarrow C2$

10. Given a large-scale e-commerce application that requires high availability, low latency, and the ability to handle massive amounts of transactional data, which NoSQL database types would be most suitable? Justify your answer with specific scenarios to illustrate the key considerations. Discuss why other NoSQL types might not be suitable.

⇔⇔⇔ Z/K/TX ⇔⇔⇔