I - 12481

Register No. | | | | | | | | |

BTech Degree Examination May 2022

Fourth Semester

Information Technology

20ITT42 – DATABASE MANAGEMENT SYSTEMS

(Regulations 2020)

Time: Three hours Maximum: 100 marks

Answer all Questions

Part – A (10 × 2 = 20 marks)

1. Mention some of the major responsibilities of a database administrator [CO1,K1]
2. Consider the following three relations in a relational database. [CO1,K3]
   Employee(eId, Name), Brand(bId, bName), Own(eId, bId)
   State the relational algebra expressions return the set of elds who own all the brands?
3. Express the term entity relationship model with one example [CO2,K2]
4. Consider the following two tables and write query to perform Right Outer Join. [CO2,K3]
   Book (isbn, bname), Stock (isbn, copies)
5. In this schema R(A, B, C, D) and functional dependencies A→B and C →D. Then what will be the properties of the decomposition of R into R1 (A, B) and R2(C, D) ? [CO3,K3]
6. How does of records are organized in files? [CO3,K1]
7. What are the causes of bucket overflow in a hash file organization? [CO4,K1]
8. List the ACID properties of Transaction. [CO4,K1]
9. Mention the two techniques of deadlock prevention [CO5,K1]
10. Classify the failures in transactions. [CO5,K1]

Part – B (5 × 16 = 80 marks)

11. a. i) Illustrate the overall architecture of the data base system in detail. (8) [CO1,K2]

    ii) Determine atleast five superkeys, possible candidate keys and suitable primary key for the following relation. Also, calculate the number of super keys. (8) [CO1,K3]
    Student(Sid, Sname, address, phone#, mailed, CGPA, year, class, section)

    (OR)

    b. i) Consider the following relational database (8) [CO1,K3]
    Emp(empname, street, city)
    Works(empname, name, salary)
    Company(bankname, city)
    Manages(empname, managername)
    Give an expression in the relational algebra to express each of the following queries.
    1. Find the names and cities of residence of all employes who work for 'State Bank of India'.
    2. Find the names, street, cities of residence of all employees who work for 'IOB' and earn more than Rs. 40,000.
    3. Find the names of all employees who live in the same city as the bank for which they work.
    4. Find the names of all employees who line in the same city as do their managers.

ii) Explain about the database applications are handled in two and three lies architecture. (8) [CO1,K2]

12. a. i) Design a database for a world-wide package delivery company (eg.DHL or fedEx). The database must be able to keep track of customers (who ship items) and customers (who receive items); some customers may do both. Each package must be identifiable and trackable, so the database must be able to store the location of the package and its history of locations. Locations include trucks, planes, airports and warehouses. Draw an E-R diagram. (8) [CO2,K3]

ii) Convert the above ER diagram into Relational database with list of constraints including primary key and foreign key constrainsts. (8) [CO2,K3]

(OR)

b. EMPLOYEE (ESSN, ENAME, DEPT_NO, SALARY) (16) [CO2,K3]

DEPENDENT (ESSN, DEPEND_NAME, RELATION, DOB)

DEPARTMENT (DEPT_NO, DEPT_NAME, MANAGER)

(i) Write queries to create tables using SQL for above schema.

Write SQL queries for following statements.

(ii) Find details of dependents for employee having name AJAY.

(iii) Find the name of the manager of the department in which employee with ESSN Code 5078 works.

(iv) Find the name of all employees whose age is less than 58 years.

(v) Find the DOB of the son of the employee having employee code ESSN 5078.

13. a. i) What is Normalization? Explain First normal form, second normal form and third normal form with an example (8) [CO3,K2]

ii) Apply 4NF and 5NF for the given student database. Also explain about multivalued dependencies. (8) [CO3,K3]

Student ID	Areas of interest	Hobbies
100	OS, DBMS	Singing, Painting
101	JAVA	Reading

(OR)

b. i) State about RAID technologies and it's levels. (8) [CO3,K2]

ii) Find all keys for R. Consider a relation R with attributes ABCDEFGH and FDS S as follows: (8) [CO3,K3]

$$S = \{A \rightarrow CD, ACF \rightarrow G, AD \rightarrow BEF, BCG \rightarrow D, CF \rightarrow AH, CH \rightarrow G, D \rightarrow B, H \rightarrow DEG\}$$

14. a. i) Construct B+ tree to insert the following key values (the order of the tree is three) 32,11,15,13,7,22,15,44,67,4. (8) [CO4,K3]

ii) Apply extendable hashing on a file that contains records with the following search key values: 2, 3, 5, 7, 11, 17, 19, 23, 29 and 31. The hash function is h(x) = x and 8, and buckets can hold three records. (8) [CO4,K3]

(OR)

b. Consider the following schedules. The actions are listed in the order they are scheduled and prefixed with the transaction name: (16) [CO4,K3]

S1: T1:R(X),T2:R(X),T1:W(Y),T2:W(Y),T1:R(Y),T2:R(Y)

S2: T3:W(X),T1:R(X),T1:W(Y),T2:W(Z),T3:R(Z)

For each of the schedules answer the following questions:

(i) What is the precedence graph for the schedule?

(ii) Is the schedule conflict serializable? If so what are all the conflict equivalent serial schedules?

(iii) Is the schedule view serializable? If so what are all the view equivalent serial schedule?

15. a. Elucidate the functionalities of lock based protocol with your own examples. (16) [CO5,K2]

(OR)

b. Illustrate the working principle of recovery algorithm with relevant example. (16) [CO5,K2]

Bloom's Taxonomy Level	Remembering (K1)	Understanding (K2)	Applying (K3)	Analysing (K4)	Evaluating (K5)	Creating (K6)
Percentage	7	36	57	-	-	-

KONGU ENGINEERING COLLEGE, PERUNDURAI, ERODE- 638 060
B.Tech Degree Examinations May 2022
Fourth Semester
Information Technology
20ITT42 – Database Management Systems
Answer Key
(Regulation 2020)

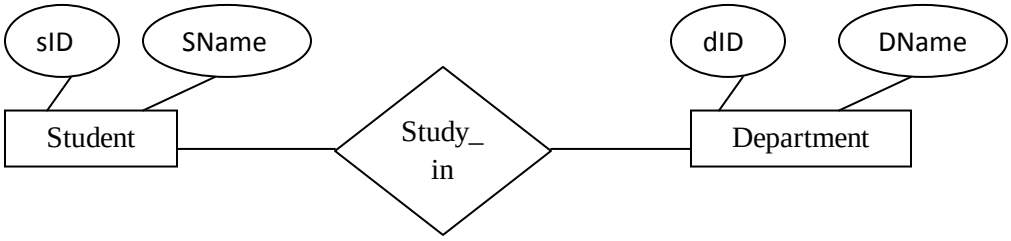
PART - A

1. **Any two** of the following responsibilities,
   - Schema definition
   - Periodically backing up the database
   - Monitoring jobs running on the database
   - Storage structure and access-method definition
   - Schema and physical-organization modification
   - Granting of authorization for data access

2. T1 ← $\prod_{bId}(Brand)$

   T2 ← Own ÷T1

3. Entity Relationship model (ER model) was developed to facilitate database design.
   The ER model can express the overall logical structure of a database graphically.
   Any Example similar to below ER-diagram:



4. Any SQL query similar to the following query,
   Select * from Book right Join Stock on Book.isbn = Stock.isbn

5. The given decomposition is a lossy decomposition.

6. Records are organized using any of the following methods, (**any two** of the following methods)
   - Sequential File Organization
   - B^+ Tree File Organization
   - Heap File Organization
   - Hash File Organization
   - Multitable clustering file organization

7. The cause of bucket overflow in hash file organization are,
   - Insufficient buckets.
   - Skew in distribution of records. This can occur due to two reasons:

- multiple records have same search-key value
- chosen hash function produces non-uniform distribution of key values

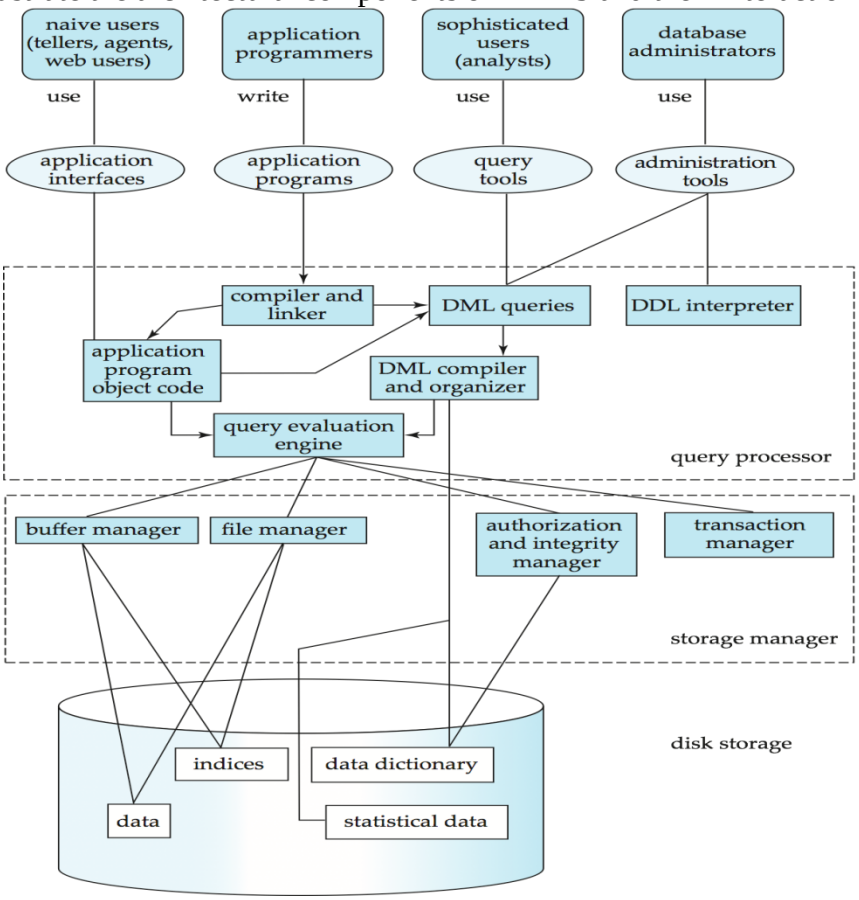
8. ACID properties are,
   - Atomicity
   - Consistency
   - Isolation
   - Durability

9. The two techniques in deadlock prevention are,
   - Wait-Die
   - Wound-Wait

10. The Transaction failures are classified as,
    - Logical errors
    - System errors
    - System crash
    - Disk failure

PART - B

11. a) i) Illustrate the architectural components of DBMS and their interaction within each other.



a) ii) Given the relation
Student(Sid, Sname, address, phone#, mailed, CGPA, year, class, section)
Possible candidate keys are {{Sid}, {mailed}}

Since the primary key is chosen from candidate key, the primary key can be any one of the following, {Sid}, {mailed}

Super key is a set of those keys that identify a row or a tuple uniquely. Following are some of the super keys of the Student relation, **(any set of attributes containing a candidate key is acceptable)**

- {Sid, Sname}
- {Sid, year, class}
- {mailed, year, class, section}
- {mailed, address}

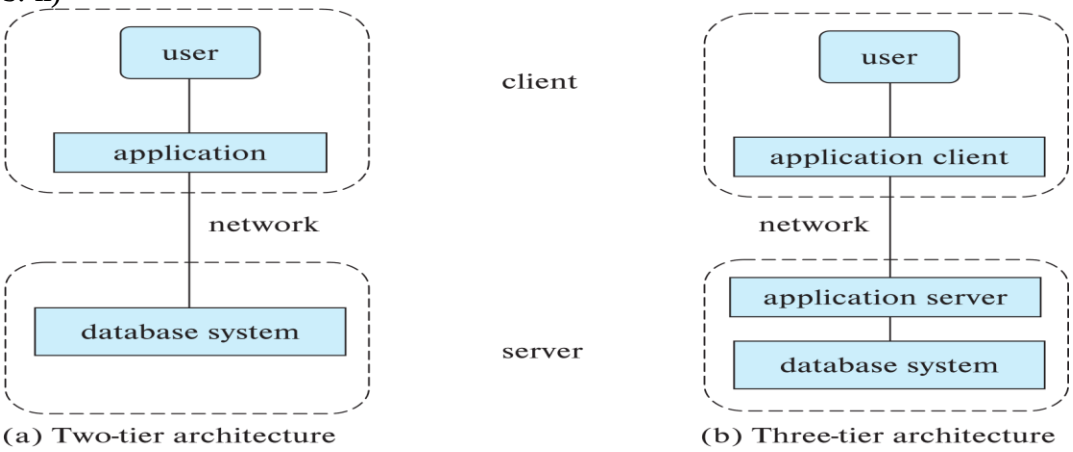
The number super keys for the relation Student is = Maximum Super keys = $2^n - 1$ = $\mathbf{2^9-1 = 511}$

b) i) Relational Algebra expressions

1. $\Pi_{empname,\ city}(Emp \bowtie \sigma_{(bankname\ =\ \text{“State bank of India”}}(works)))$
2. $\Pi_{empname,street,\ city}(\sigma_{(bankname\ =\ \text{“IOB”}\ \wedge\ salary\ >\ 40000)}\ works \bowtie employee)$
3. $\Pi_{empname}(employee \bowtie works \bowtie company)$

4. $\Pi_{empname}(Emp \bowtie Manages) \bowtie (managername=Emp2.empname \wedge Emp.city = Emp2.city)(\rho_{Emp2}(Emp)))$

b. ii)



(a) Two-tier architecture (b) Three-tier architecture

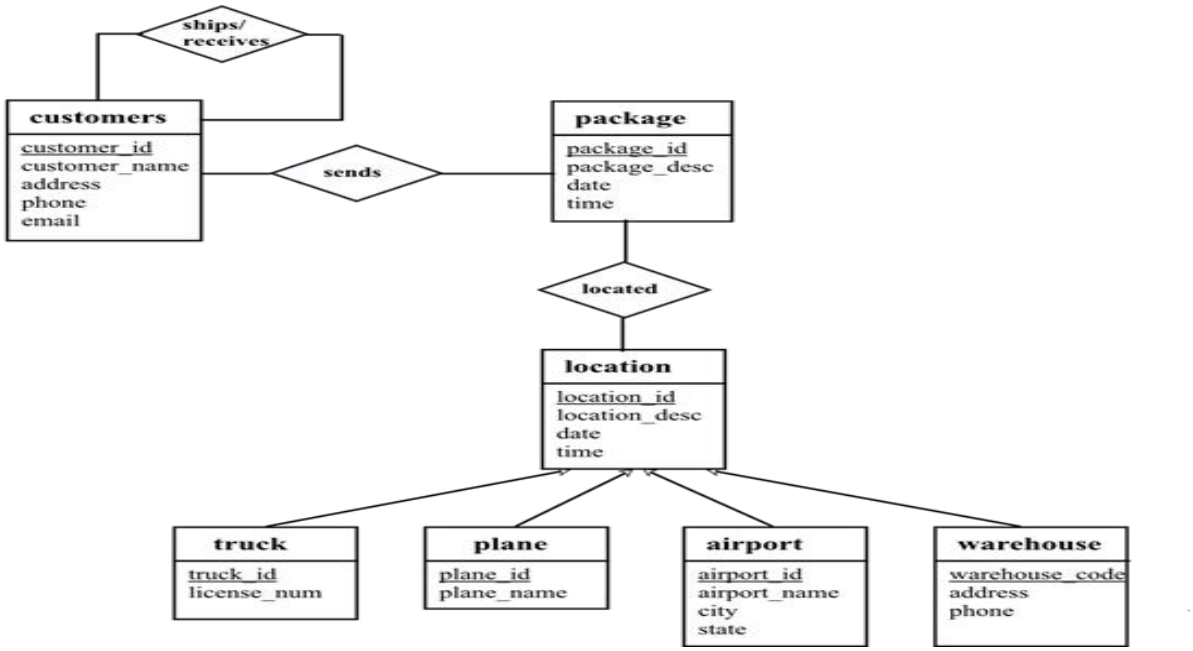
Two tier architecture

- The two-tier is based on Client Server architecture. The two-tier architecture is like client Server application.
- The direct communication takes place between client and server.
- There is no intermediate between client and server. Because of tight coupling a 2 tiered application will run faster.

Three tier architecture

- Three-tier architecture typically comprises a presentation tier, a business or data access tier, and a data tier.
- The communication between the client and the database is controlled by the middle ware or a data access layer

12. a. i) ER-diagram similar to the following is acceptable,



ii) customers (customer_id, customer_name, address, phone, email)
package(package_id, package_desc, date, time, customer_id, location_id)
location(location_id, location_desc, date, time)

b) i) Queries similar to following are acceptable,
create table EMPLOYEE (ESSN int primary key, ENAME varchar(50), DEPT_NO int foreign key references DEPARTMENT (DEPT_NO), SALARY int);

create table DEPENDENT (ESSN int primary key, DEPEND_NAME varchar(50), RELATION varchar(20), DOB date);

create table DEPARTMENT (DEPT_NO int primary key, DEPT_NAME varchar(20), MANAGER int foreign key references EMPLOYEE(ESSN));

ii) select DEPEND_NAME, RELATION, DOB from DEPENDENT where ESSN IN (select ESSN from EMPLOYEE where ENAME='AJAY');

iii) select ENAME from EMPLOYEE where ESSN = (select MANAGER from DEPARTMENT where DEPT_NO = (select DEPT_NO from EMPLOYEE where ESSN=5078))

iv) Assume DOB attribute is in EMPLOYEE. select name from employee where age< 58;

v) select DOB from DEPENDENT where ESSN=5078 and RELATION='son';

13. a) i) **Normalization** is the process of minimizing redundancy from a relation or set of relations. Redundancy in relation may cause insertion, deletion, and update anomalies.
First Normal form: A relational schema R is in **first normal form** if the domains of all attributes of R are atomic. Example: Gender attribute values Male, female are atomic.

Second Normal form: Partial dependency is not allowed. (i.e. partial key determining non-prime attributes are not allowed). Example: R(A,B,C) FD's AB -> C, A -> D. candidate key for R is {AB}. In FD A-> D LHS A is a partial key and the RHS D is a non-prime attribute.

Third Normal form: Transitive dependency is not allowed. (i.e. non-prime attribute determining non-prime attribute is not allowed). Example: R(A,B,C,D) FD's AB -> C, C -> D. candidate key for R is {AB}. In FD C -> D both LHS and RHS are non-prime attributes.

ii) **Multivalued Dependency**: For a dependency A → B, if for a single value of A, multiple values of B exists, then the relation will be a multi-valued dependency.

4NF: a relation should be in Boyce-Codd Normal Form and may not contain more than one multi-valued attribute.
In the given relation StudentID ->> Area_of_interest and StudentID ->> Hobbies
Multi-valued dependency (MVD) exists in the given relation. MVD can be eliminated by decomposing the relation into R1(StudentID, Area_of_interest) and R2(StudentID, Hobbies).

5NF: It is also known as project-join normal form (PJNF). A relation is in Fifth Normal Form (5NF), if it is in 4NF, and does not have lossless join decomposition.

b) i) RAID (Redundant array of independent disks) is a technology to connect multiple secondary storage devices and use them as a single storage media.

RAID Level 0: Block striping; non-redundant.
RAID Level 1: Mirrored disks with block striping
RAID Level 2: Memory-Style Error-Correcting-Codes (ECC) with bit striping.
RAID Level 3: Bit-Interleaved Parity
RAID Level 4: Block-Interleaved Parity; uses block-level striping, and keeps a parity block on a separate disk for corresponding blocks from N other disks
RAID Level 5: Block-Interleaved Distributed Parity; partitions data and parity among all N + 1 disks.
RAID Level 6: P+Q Redundancy scheme

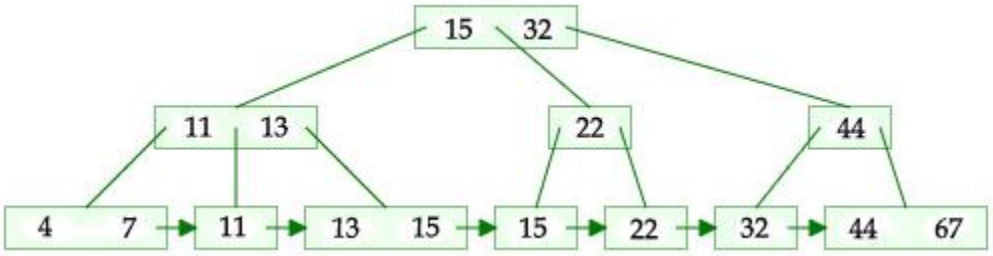
ii)

LHS	Result	Decision
$\mathbf{A}^+$	= ACD from A → CD = ACDBEF from AD → BEF = ACDBEFG from ACF → G = ACDBEFGH from CF → AH	Result includes all the attributes of relation R. That is, if we know A, then all the attributes of R could be uniquely determined. Hence, **A is one candidate key**.
ACF^+	No need to find the closure of (ACF) because the subset A is already a candidate key.	ACF is a super key but not candidate key.
AD^+	No need to find the closure of (AD) because the subset A is already a candidate key.	AD is a super key but not candidate key.
BCG^+	= BCGD from BCG → D	Result does not include all the attribute of relation R. Hence, (BCG) cannot be a candidate key.
$\mathbf{CF}^+$	= CFAH from CF → AH Further, as we know A now, then we can conclude that CF will uniquely determine all the other attributes of A.	Result includes all the attributes of relation R. Hence, **(CF) is one candidate key**.

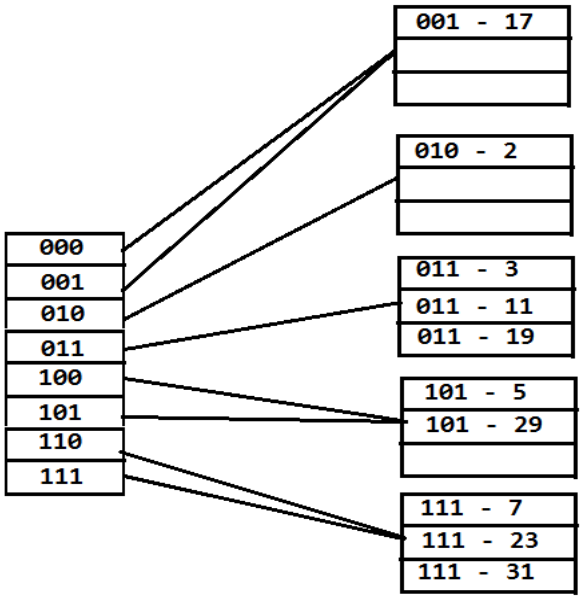
D+	= DB from D → B	Result does not include all R. Hence, D cannot be a key.
H+	= HDEG from H → DEG = HDEGB from D → B	Result does not include all R. Hence, H cannot be a key.

From the above table, it is clear that only **A and CF are the candidate keys**.

14. a) i) B+ Tree elements to be inserted are, 32, 11, 15, 13, 7, 22, 15, 44, 67, 4.



ii) Number of records per Bucket = 3
Hash function H(x) = x mod 8
Search keys: 2, 3, 5, 7, 11, 17, 19, 23, 29, 31



b) **Schedule S1**

T1	T2
R(X)	
	R(X)
W(Y)	
	W(Y)
R(Y)	
	R(Y)

i) Schedule S1's precedence graph:

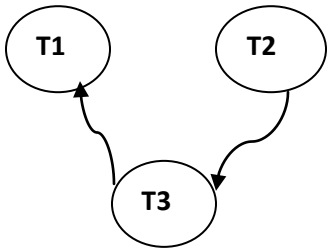
ii) The given schedule S1 is **not conflict serializable**

iii) The given schedule S2 is **view serializable.** The view equivalent serial schedule is,
T1 → T2

Schedule S2

T1	T2	T3
		W(X)
R(X)		
W(Y)		
	W(Z)	
		R(Z)

i) Schedule S2's precedence graph:



ii) The given schedule S2 is **conflict serializable.** The conflict equivalent serial schedule is,
T2 → T3 → T1

iii) The given schedule S2 is **view serializable.** (all conflict serializable schedules are view serailizable). The view equivalent serial schedule is,
T2 → T3 → T1

15. a) Lock based Protocol: A lock is a mechanism to control concurrent access to a data item
Data items can be locked in two modes :
    - exclusive (X) mode. Data item can be both read as well as written. X-lock is requested using lock-X instruction.
    - shared (S) mode. Data item can only be read. S-lock is requested using lock-S instruction. Lock requests are made to concurrency-control manager
- A locking protocol is a set of rules followed by all transactions while requesting and releasing locks.
- Example: (any example similar to the following)

T_1	T_2	concurrency-control manager
lock-X(B)		
		grant-X(B, T_1)
read(B)		
$B := B - 50$		
write(B)		
unlock(B)		
	lock-S(A)	
		grant-S(A, T_2)
	read(A)	
	unlock(A)	
	lock-S(B)	
		grant-S(B, T_2)
	read(B)	
	unlock(B)	
	display($A + B$)	
lock-X(A)		
		grant-X(A, T_1)
read(A)		
$A := A + 50$		
write(A)		
unlock(A)		

b) Recovery Algorithm:

- Database systems, like any other computer system, are subject to failures but the data stored in it must be available as and when required.
- Recovery techniques are heavily dependent upon the existence of a special file known as a system log.
- The log keeps track of all transaction operations that affect the values of database items.
- **Recovery from failure**: Two phases
  - **Redo phase**: replay updates of **all** transactions, whether they committed, aborted, or are incomplete
  - **Undo phase**: undo all incomplete transactions
  - Examples similar to below are accepted