Nepal Institute of Engineering



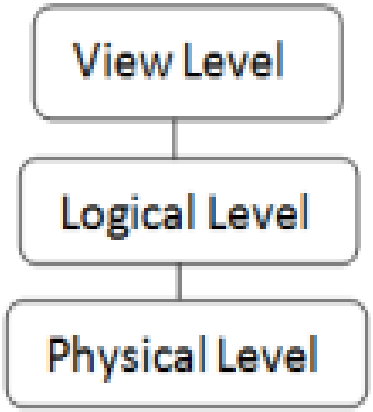
# Subject: Database Management System

By: Deepak Kumar Singh

# Introduction to data models

## Data Abstraction

- For the system to be usable, it must retrieve data efficiently.
- Need for efficiency has led designer to use complex data structures to represent data in the database.
- Developers hide complexity from users through several levels of abstraction



- **Physical Level**
  - Lowest level of abstraction describes how the data are actually stored.
- **Logical Level**
  - Describes what data are stored in the database and what relationship exist among those data.
- **View Level**
  - Describes only part of the entire database & helps to hide details from application program
  - The view level of abstraction exists to simplicity their interaction with the system.
  - The system may provide many view for the database.

- Which of the following isn't a level of abstraction?
  physical
  logical
  **user**
  view
- A level that describes how a record is stored.
  **physical**
  logical
  user
  view
- The _____level helps application programs hide the details of data types.
  physical
  logical
  user
  **view**

- The type data abstraction which allows the conceptual representation of data in database management system is considered as

  logical design model

  **data model**

  interface model

  user friendly model

- Considering abstraction concepts, the process of assigning the similar entities to similar entity types systematically is called

  **classification**

  instantiation

  identification

  exception abstract

# Data Independence

- defined as the capacity to change the schema at one level of a database system without having to change the schema at the next higher level
- Two types of data independence
  - Logical data independence
  - Physical data independence

**Q. An independent database is a system that permits _____ modifications on one level without affecting the _____ on the next level up.**

Instance

**Schema**

Both a and b

None

An independent conceptual schema can be changed without having to change an ____ schema.

Internal

**External**

Both A and B

None

Separating external data from ____ data is achieved using logical data independence.

Logical

**Conceptual**

Analytical

Physical

What is TRUE about Logical Data Independence?

We wouldn't impact the user view of the data if we changed the conceptual view of the data.

User interfaces are logically independent of data.

**Both A and B**

None of the above

It is possible to change the internal schema without affecting the conceptual schema to achieve ___ data independence.

Logical

**Physical**

Analytical

Conceptual

# Schema and Instance

- The skeleton that represents the logical view of the entire database.
- Defines how the data is organized and how the relations among them are associated.
- Two categories
  - Physical Database Schema − This schema pertains to the actual storage of data and its form of storage like files, indices, etc. It defines how the data will be stored in a secondary storage.
  - Logical Database Schema − This schema defines all the logical constraints that need to be applied on the data stored. It defines tables, views, and integrity constraints.
- Database Instance The data in the database at a particular moment in time is called a database state or snapshot.

- A logical structure of the database.
  **Schema**
  Attribute
  Parameter
  Instance
- The actual content in the database at a particular point.
  Schema
  Attribute
  Parameter
  **Instance**
- Map of concepts and their relationships is represented by _____ Schema.

  Logical

  Physical

  **Conceptual**

  None

- A map of entities and their attributes and relations is represented by ___

  **Logical**

  Physical

  Conceptual

  None

- The environment of database is said to be _____

  **Database Instance**

  Database Schema

- Overall design of the database is called as _____

  Database Instance

  **Database Schema**

  Database Abstraction

  None

# ER MODEL

- In this database model, relationships are created by dividing object of interest into entity and its characteristics into attributes.
- Different entities are related using relationships.
- E-R Models are defined to represent the relationships into pictorial form to make it easier for different stakeholders to understand.

- An _____ is a set of entities of the same type that share the same properties, or attributes.
  **Entity set**
  Attribute set
  Relation set
  Entity model
- Entity is a _____
  Object of relation
  Present working model
  **Thing in real world**
  Model of relation
- The descriptive property possessed by each entity set is _____
  Entity
  **Attribute**
  Relation
  Model
- The function that an entity plays in a relationship is called that entity's _____
  Participation
  Position
  **Role**
  Instance

# Strong and Weak Entity Set

| BASIS FOR COMPARISON | STRONG ENTITY | WEAK ENTITY |
|---|---|---|
| Basic | The Strong entity has a primary key. | The weak entity has a partial discriminator key. |
| Depends | The Strong entity is independent of any other entity in a schema. | Weak entity depends on the strong entity for its existence. |
| Denoted | Strong entity is denoted by a single rectangle. | Weak entity is denoted with the double rectangle. |
| Relation | The relation between two strong entities is denoted by a single diamond simply called relationship. | The relationship between a weak and a strong entity is denoted by Identifying Relationship denoted with double diamond. |
| Participation | Strong entity may or may not have total participation in the relationship. | Weak entity always has total participation in the identifying relationship shown by double line. |

- E-R model uses which symbol to represent weak entity set :

  Dotted rectangle

  Diamond

  **Doubly outlined rectangle**

  None of the above

- The weak entities and their identifying relationship in the ER diagrams are represented by

  oval shape with double line

  rectangle shape with double line

  square shape with double line

  **diamond shape with double line**

- The entities that have key attributes to be distinct are considered as

  weak entity

  **strong entity**

  single entity

  foreign entity

- The parent entity type or identifying entity type are considered as other names of

  **dominant entity type**

  non dominant entity type

  composite entity type

  non foreign entity type

- In entity-relationship, the entity type which has existence dependency constraint is classified as

  single entity

  foreign entity

  **weak entity**

  strong entity

- The set of key attributes that identify weak entities related to some owner entity is classified as

  structural key

  **partial key**

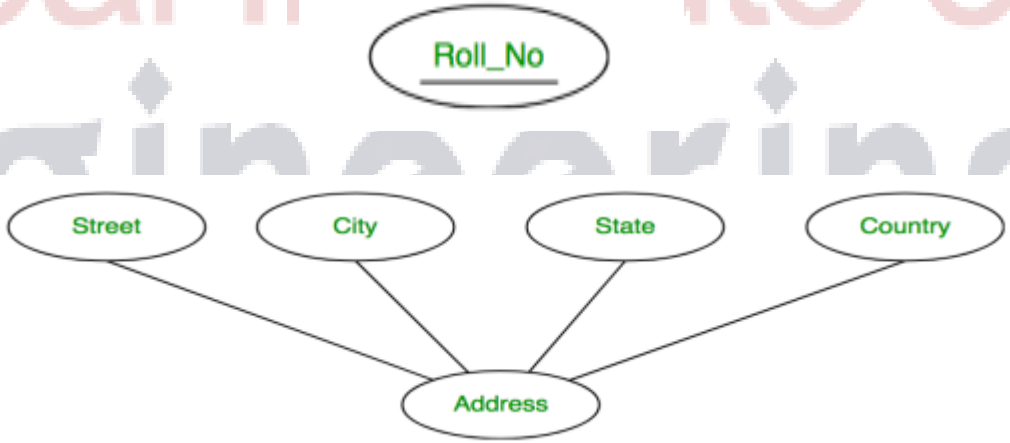
  string key

  foreign key

# Attributes And Keys

- Properties which define the entity type.
- Represented by an oval.
- Types

1. Key attribute




2. Composite attribute

3. Multivalued attribute


4. Derived attribute

- The descriptive property possessed by each entity set is _____
  a) Entity
  b) **Attribute**
  c) Relation
  d) Model
- The attribute *name* could be structured as an attribute consisting of first name, middle initial, and last name. This type of attribute is called
  a) Simple attribute
  b) **Composite attribute**
  c) Multivalued attribute
  d) Derived attribute
- The attribute AGE is calculated from DATE_OF_BIRTH. The attribute AGE is
  a) Single valued
  b) Multi valued
  c) Composite
  d) **Derived**
- Not applicable condition can be represented in relation entry as
  a) NA
  b) 0
  c) **NULL**
  d) Blank Space

- Which of the following can be a multivalued attribute?
  a) **Phone_number**
  b) Name
  c) Date_of_birth
  d) All of the mentioned
- Which of the following is a single valued attribute
  a) **Register_number**
  b) Address
  c) SUBJECT_TAKEN
  d) Reference
- In a relation between the entities the type and condition of the relation should be specified. That is called as_____attribute.
  a) **Descriptive**
  b) Derived
  c) Recursive
  d) Relative

# Database Keys

1. Primary Key

- Assigned in design phase of table
- Used to identify each record uniquely

2. Super Key

Assigned to set of attributes to identify each record uniquely

3. Candidate Key

Assigned to minimal set of attributes to identify each record uniquely

Create table student(id integer,

Name varchar(50),

Roll integer,

Primary key(id, roll));

- Which one of the following is a set of one or more attributes taken collectively to uniquely identify a record?
  a) Candidate key
  b) Sub key
  c) **Super key**
  d) Foreign key
- Consider attributes ID, CITY and NAME. Which one of this can be considered as a super key?
  a) NAME
  b**) ID**
  c) CITY
  d) CITY, ID
- The subset of a super key is a candidate key under what condition?
  a) **No proper subset is a super key**
  b) All subsets are super keys
  c) Subset is a super key
  d) Each subset is a super key

- A _____ is a property of the entire relation, rather than of the individual tuples in which each tuple is unique.
  a) Rows
  b) **Key**
  c) Attribute
  d) Fields
- Which one of the following attribute can be taken as a primary key?
  a) Name
  b) Street
  c) **Id**
  d) Department
- Which one of the following cannot be taken as a primary key?
  a) Id
  b) Register number
  c) Dept_id
  d) **Street**
- An attribute in a relation is a foreign key if the _____ key from one relation is used as an attribute in that relation.
  a) Candidate
  b) **Primary**
  c) Super
  d) Sub

# ER Diagram

Which of the following gives a logical structure of the database graphically?

a) **Entity-relationship diagram**
b) Entity diagram
c) Database diagram
d) Architectural representation

The entity relationship set is represented in E-R diagram as

a) Double diamonds
b) Undivided rectangles
c) Dashed lines
d) **Diamond**

The Rectangles divided into two parts represents

a) **Entity set**
b) Relationship set
c) Attributes of a relationship set
d) Primary key

- Consider a directed line(->) from the relationship set advisor to both entity sets instructor and student. This indicates _____ cardinality
  a) One to many
  b) **One to one**
  c) Many to many
  d) Many to one
- We indicate roles in E-R diagrams by labeling the lines that connect _____ to _____
  a) Diamond , diamond
  b) Rectangle, diamond
  c) Rectangle, rectangle
  d) **Diamond, rectangle**

# Different Normal Forms (1st, 2nd, 3rd and BCNF)

Normalization is used for mainly two purposes,

- Eliminating redundant(useless) data.
- Ensuring data dependencies make sense i.e. data is logically stored.

**First Normal Form**

For a table to be in the First Normal Form, it should follow the following 4 rules:

- It states that it should only have single(atomic) valued attributes/columns.
- Every values stored in a column should be of the same domain
- Every columns in a table should have unique names.
- And the order in which data is stored, does not matter.
- It ensures that each relation has a primary key.

**2nd Normal Forms**

relations must be in 1st normal form

there should not be partial dependency

**3rd normal form**

relations must be in 2nd normal form

there should not be transitive dependency

**BCNF**

relations must be in 3rd normal form

for each functional dependency, x->y, x should be super key

**4th normal form**

It should not contain multi valued dependency

**5th normal form**

Based on join dependency

relations must be in 4nf

It cant be further non loss decomposed

- A function that has no partial functional dependencies is in _____ form :

  A 3NF

  **B 2NF**

  C 4NF

  D BCNF

- 4NF is designed to cope with :

  **A** Transitive dependency

  **B** Join dependency

  **C Multi valued dependency**

  **D** None of these

- 5NF is designed to cope with :

  **A** Transitive dependency

  **B Join dependency**

  **C** Multi valued dependency

  **D** All of the above

- A BCNF is :

A loss less join and dependency preserving

**B loss less join but not dependency preserving**

C not loss less join but dependency preserving

D none of these

- Every Boyee-Codd normal form is in

A First normal form

B Second normal form

C Third normal form

**D All of the above**

- **By normalizing relations or sets of relations, one minimizes ___.**

A Data

B Fields

**B Redundancy**

D Database

# Functional Dependency

X → Y

Types

Trivial: ABC → B

Non Trivial: A→B

Multi-valued

Transitive

We can use the following three rules to find logically implied functional dependencies. This collection of rules is called

a) Axioms
b) **Armstrong's axioms**
c) Armstrong
d) Closure

Which of the following is not Armstrong's Axiom?

a) Reflexivity rule
b) Transitivity rule
c) **Pseudotransitivity rule**
d) Augmentation rule

- There are two functional dependencies with the same set of attributes on the left side of the arrow:
  A->BC
  A->B
  This can be combined as
  **A->BC**
  A->B
  B->C
  None of the mentioned
- **Functional dependencies are classified as ___ on the left.**
  Dependent
  Determined
  **Determinants**
  Database

# Integrity Constraints and Domain integrity constraints

Domain Integrity constraints

Tuple Integrity Constraints

Referential Integrity Constraints

Key Integrity Constraints

Entity Integrity Constraints



To include integrity constraint in an existing relation use :
a) Create table
b) Modify table
c) **Alter table**
d) Drop table

Which of the following is not an integrity constraint?
a) Not null
b) **Positive**
c) Unique
d) Check ‘predicate’

- Foreign key is the one in which the _____ of one relation is referenced in another relation.
  a) Foreign key
  b) **Primary key**
  c) References
  d) Check constraint
- Which of the following is used to delete the entries in the referenced table when the tuple is deleted in course table?
  a) Delete
  b) **Delete cascade**
  c) Set null
  d) All of the mentioned
- Domain constraints, functional dependency and referential integrity are special forms of _____
  a) Foreign key
  b) Primary key
  c) **Assertion**
  d) Referential constraint

- Which of the following is the right syntax for the assertion?
  a) **Create assertion ‘assertion-name’ check ‘predicate’;**
  b) Create assertion check ‘predicate’ ‘assertion-name’;
  c) Create assertions ‘predicates’;
  d) All of the mentioned
- Data integrity constraints are used to:
  a) Control who is allowed access to the data
  b) Ensure that duplicate records are not entered into the table
  c) **Improve the quality of data entered for a specific property (i.e., table column)**
  d) Prevent users from changing the values stored in the table
- To ensure that the value of budget is non-negative which of the following should be used?
  a) **Check(budget>0)**
  b) Check(budget<0)
  c) Alter(budget>0)
  d) Alter(budget<0)

## Queries Under DDL

1. CREATE (To create database and tables)

Syntax: Create database databasename;// Create database

Syntax: Create table tablename(columnname1 datatype1……columnnamen datatypen);

2. Alter: To change structure of tables

3. Drop: To drop structures in database or database too

4. Rename: To rename table, columns in database

5. TRUNCATE: To delete all records at-once

## DML(Data Manipulation Language)

SELECT: To visualize record

INSERT : To insert record in tables

UPDATE : To update columns values in tables

DELETE: To delete records

- **Commands that comes under DDL is/are –**

  CREATE

  DROP

  TRUNCATE

  **All of the above**

- **Which of the following is/are TRUE about DDL command?**

  Our data is stored in a table that is described by the schema, thus DDL commands deal with the schema.

  With the DDL commands, any structural changes can be made to the table, including creation, deletion, and alteration.

  **Both A. and B.**

  None of the above

**Select the correct statement.**

**With the DDL commands, any structural changes can be made to the table, including creation, deletion, and alteration.**

With the DML commands, any structural changes can be made to the table, including creation, deletion, and alteration.

With the DCL commands, any structural changes can be made to the table, including creation, deletion, and alteration.

With the TCL commands, any structural changes can be made to the table, including creation, deletion, and alteration.

## Assertions

- An assertion is a predicate expressing a condition we wish the database to always satisfy.
- Domain constraints, functional dependency and referential integrity are special forms of assertion.
- Where a constraint cannot be expressed in these forms, we use an assertion, e.g.
  - Ensuring the sum of loan amounts for each branch is less than the sum of all account balances at the branch.
  - Ensuring every loan customer keeps a minimum of $1000 in an account.
- An assertion in DQL-92 takes the form,

  **create assertion assertion-name check predicate**

## Triggers

- Another feature not present in the SQL standard is the trigger.
- Several existing systems have their own non-standard trigger features.
- A trigger is a statement that is automatically executed by the system as a side effect of a modification to the database.
- We need to
  - Specify the conditions under which the trigger is executed.
  - Specify the actions to be taken by the trigger.

## Views

Subset of tables

Simple view: Created using single table

Complex view: Created using multiple tables

Syntax:

Create view viewname as select (column list) from table1…table-n where[condition;]

- A _____ is a special kind of a store procedure that executes in response to certain action on the table like insertion, deletion or updation of data.
  a) Procedures
  **b) Triggers**
  c) Functions
  d) None of the mentioned
- The CREATE TRIGGER statement is used to create the trigger. THE _____ clause specifies the table name on which the trigger is to be attached. The _____ specifies that this is an AFTER INSERT trigger.
  a) for insert, on
  **b) On, for insert**
  c) For, insert
  d) None of the mentioned
- What are the after triggers?
  a) Triggers generated after a particular operation
  **b) These triggers run after an insert, update or delete on a table**
  c) These triggers run after an insert, views, update or delete on a table
  d) All of the mentioned

# Relational algebra

Relational Algebra is a procedural query language used to query the database tables to access data in different ways.

Relational Algebra consists of several groups of operations

Unary Relational Operations

- SELECT (symbol: σ (sigma))
- PROJECT (symbol: π (pi))
- RENAME (symbol: ρ (rho))

Relational Algebra Operations From Set Theory

- UNION (∪ ), INTERSECTION ( ), DIFFERENCE (or MINUS, – )
- CARTESIAN PRODUCT ( x )

Binary Relational Operations

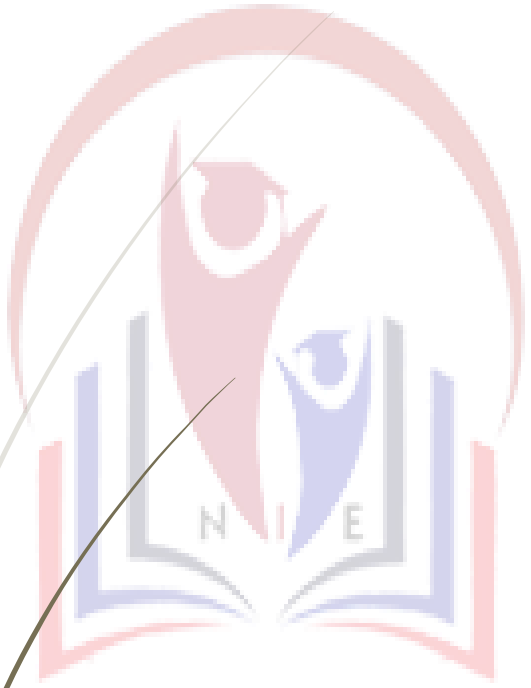
o JOIN (several variations of JOIN exist)

o DIVISION

Additional Relational Operations

o OUTER JOINS, OUTER UNION

o AGGREGATE FUNCTIONS (These compute summary of information: for example, SUM, COUNT, AVG, MIN, MAX)



Nepal Institute of Engineering

- Relational Algebra is a _____ query language that takes two relations as input and produces another relation as an output of the query.
  a) Relational
  b) Structural
  **c) Procedural**
  d) Fundamental

- Which of the following is a fundamental operation in relational algebra?
  a) Set intersection
  b) Natural join
  c) Assignment
  **d) None of the mentioned**

- Which of the following is used to denote the selection operation in relational algebra?
  a) Pi (Greek)
  **b) Sigma (Greek)**
  c) Lambda (Greek)
  d) Omega (Greek)

- For select operation the _____ appear in the subscript and the _____ argument appears in the paranthesis after the sigma.
  **a) Predicates, relation**
  b) Relation, Predicates
  c) Operation, Predicates
  d) Relation, Operation

# Query Processing

Query Cost Estimation

- Cost is generally measured as total time elapsed for answering query. To convert high level
- query to desired query we need some measurements.
- Basic measure for query cost are

  disk access

  CPU cycle

  Transit time in network
- Query Processing
- Evaluation of expression: Materialization and Pipelining
- Query Optimization: Cost based and Rule based
- Query Decomposition

- Which of the following are the process of selecting the data storage and data access characteristics of the database?

  Logical database design

  **Physical database design**

  Testing and performance tuning

  Evaluation and selecting

- **Query ___ is the activity performed in extracting data from the database.**

  Result

  Inhibition

  System

  **Processing**

- **Data is ___ from the database using various steps in query processing.**

  Extracted

  Added

  **Fetched**

  Deleted

- The process of finding a good strategy for processing a query is called

  **Query optimization**

  Query processing

  Query management

  Query cost

- The iterator operation of a demand driven pipeline provides the function of

  open()

  next()

  close()

  **All of the Above**

# Transaction Processing, Concurrency Control and Crash recovery

ACID Properties (Atomicity, Consistency, Isolation, Durability)

**Serializability**

- Serializability is a concept that helps us to check which schedules are serializable.
- A serializable schedule is the one that always leaves the database in consistent state.
- Two Types (Conflict and View Serializability)
  - Conflict Serializability : If any non-serial schedule is conflict equivalent to serial schedule, then it is called as conflict serializable schedule.

Test for conflict serializability

- To check conflict serializability, precedence graph is used.
- Let 'S' be a schedule, construct a directed graph known as precedence graph. Graph consists
- If graph is non cyclic, schedule is conflict serializable schedule.
- of a pair of G = (V , E)
- Where, V : a set of vertices(Transaction)

  E : a set of edges

## View Serializability

Two schedule S and S' are view equivalent if the following conditions are met :

Algorithm

Step 1 : For each data item Q, if Ti reads an item value of Q in schedule S, then Ti in S‘ must also reads an initial value of Q.

Step 2 : If Ti executes read Q in S and that value was produced by Tj (if any), then Ti in S' also read the value of Q that was produced by Tj

Step 3 : For each data item Q, the transaction that performs the final write (Q) operation in schedule S must also perform the final write (Q) in schedule S'.

- Collections of operations that form a single logical unit of work are called _____
  a) Views
  b) Networks
  c) Units
  **d) Transactions**
- The “all-or-none” property is commonly referred to as _____
  a) Isolation
  b) Durability
  **c) Atomicity**
  d) None of the mentioned
- Which of the following is a property of transactions?
  a) Atomicity
  b) Durability
  c) Isolation
  **d) All of the mentioned**
- Execution of translation in isolation preserves the _____ of a database
  a) Atomicity
  **b) Consistency**
  c) Durability
  d) All of the mentioned

- I and J are _____ if they are operations by different transactions on the same data item, and at least one of them is a write operation.
  **a) Conflicting**
  b) Overwriting
  c) Isolated
  d) Durable

- If a schedule S can be transformed into a schedule S' by a series of swaps of non-conflicting instructions, then S and S' are
  a) Non conflict equivalent
  b) Equal
  **c) Conflict equivalent**
  d) Isolation equivalent

- A schedule is _____ if it is conflict equivalent to a serial schedule.
  **a) Conflict serializable**
  b) Conflicting
  c) Non serializable
  d) None of the mentioned

- The set of _____ in a precedence graph consists of all the transactions participating in the schedule
  **a) Vertices**
  b) Edges
  c) Directions
  d) None of the mentioned

## Concurrent Execution

- It implies interleaving execution of operations of a transaction. Multiple transactions are allowed to run concurrently in the system.
- Benefits

  Helps in reducing waiting time

  Improved throughput and resource utilization

## Lock Based Protocol

- Shared Lock (Lock - S): To read data item only
- Exclusive Lock (Lock - X): To both read and write.

**Two Phase Locking (2PL)**

- It requires both Lock and Unlock being done in 2 phases.
- Phases

  Growing phase: New Locks on items can be acquired. When transaction acquire final item, phase reaches to point called Lock point from where shrinking phase starts.

  Shrinking phase: Existing locks are released but no new lock can be acquired. After lock point shrinking starts onwards.

- If a transaction has obtained a _____ lock, it can read but cannot write on the item
  **a) Shared mode**
  b) Exclusive mode
  c) Read only mode
  d) Write only mode
- If a transaction has obtained a _____ lock, it can both read and write on the item
  a) Shared mode
  **b) Exclusive mode**
  c) Read only mode
  d) Write only mode
- A transaction can proceed only after the concurrency control manager _____ the lock to the transaction
  **a) Grants**
  b) Requests
  c) Allocates
  d) None of the mentioned
- If a transaction can be granted a lock on an item immediately in spite of the presence of another mode, then the two modes are said to be _____
  a) Concurrent
  b) Equivalent
  **c) Compatible**
  d) Executable

- If a transaction may obtain locks but may not release any locks then it is in _____ phase

  **a) Growing phase**
  b) Shrinking phase
  c) Deadlock phase
  d) Starved phase

- If a transaction may release locks but may not obtain any locks, it is said to be in _____ phase

  a) Growing phase
  **b) Shrinking phase**
  c) Deadlock phase
  d) Starved phase

## Failure Classification

Transaction Failure (Logical and System error)

System Crash

Disk Failure

## Atomicity and Recovery, Log Based Recover

Log is a sequence of log records which maintain the records of actions performed by a transaction. It is important that the logs are written prior to the actual modification & stored on a stable storage media. (Log is the most commonly used structure for recording database modification.)

Updated log has the following fields :

- Transaction identifier
- Data item identifier
- Old value (Prior to write)
- New value (After write)

Example of log record

< T1 Start >

< T1, X2, 10, 15 >

< T1 Commit >

< T1 Abort >

- Which one of the following is a failure to a system
  a) Boot crash
  b) Read failure
  **c) Transaction failure**
  d) All of the mentioned
- Which of the following belongs to transaction failure
  a) Read error
  b) Boot error
  **c) Logical error**
  d) All of the mentioned
- The system has entered an undesirable state (for example, deadlock), as a result of which a transaction cannot continue with its normal execution. This is
  a) Read error
  b) Boot error
  **c) Logical error**
  d) System error
- Which kind of failure loses its data in head crash or failure during a transfer operation.
  a) Transaction failure
  b) System crash
  **c) Disk failure**
  d) All of the mentioned

The log is a sequence of _____ recording all the update activities in the database.

**a) Log records**

b) Records

c) Entries

d) Redo

If a transaction does not modify the database until it has committed, it is said to use the _____ technique.

**a) Deferred-modification**

b) Late-modification

c) Immediate-modification

d) Undo