Questions

qs-UGC_NET-2018, Check Referential Integrity

Let $R_1(a, b, c)$ and $R_2(x, y, z)$ be 2 relations in which 'a' is a foreign key in R_1 that refers to the primary key (unspecified) of R_2. Consider 4 options:

1. Insert into R_1
2. Insert into R_2
3. Delete from R_1
4. Delete from R_2

Which of theses options is true?

- Option 'a' & 'c' will cause violation.
- Option 'b' & 'c' will cause violation.
- Option 'c' & 'd' will cause violation.
- **Option 'd' & 'a' will cause violation.**

Basics

Structured data

- Data that needs to be stored in a structured manner.
- Data stored in a RDB, ie a Relational Database, in the form of a relation/table. It is managed by a RDBMS.
- Users can operate on the data using DBMS/RDBMS. Possible operations: Insert, Add, Delete, Update.
- Examples: SQL Server, Oracle, MySQL, etc.

Unstructured data:

- Data that does not need to be stored in a particular pre-defined structure.
- Example: Websites, photos, videos, etc.

File System vs DBMS

- Filesystems were used before DBMS existed.
- In 1970's, the user used to access data in their own system.
- Advantages of using databases & database management systems (`-` : filesystem, `+` : database):

```
- Data is stored in 1 user's computer, and is only accessible by that p
+ We have the data stored in a centralized server, and multiple users c

- Users need to download the full file even if they only want some speci
+ Users can send a query to the server to access only the data they nee

- Users need to know the location of the file, to fetch it.
+ Users don't need to know details about the data ie where it is stored

- Concurrency: In modern times, a lot of people access/update data from
+ There are protocols in place to handle inconsistency & conflict during

- Security: Since filesystem is controlled by OS, there is no level-by-l
+ Role-based (hierarchical) security protocols are available, so differe

- Redundancy: Multiple files with same content but different names, can b
+ DBMS has various contraints to ensure we don't waste storage space by
```

2-Tier Architecture

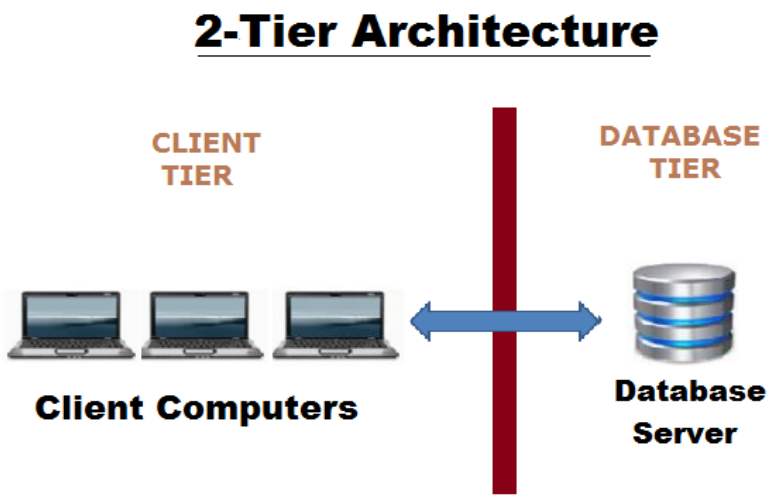


Image taken from here

- Consists of multiple clients, directly interacting with a single database server.
- Clients have an interface which help them interact with the database server.
- Client connects to the database, sends the query, and fetches the information. Information can also be modified or deleted from the database server as needed.

```
+ Simple architecture.
+ Easy to maintain.
- Difficult to scale.
- Implementing security is difficult,because client is directly is interactin
```

3-Tier Architecture:

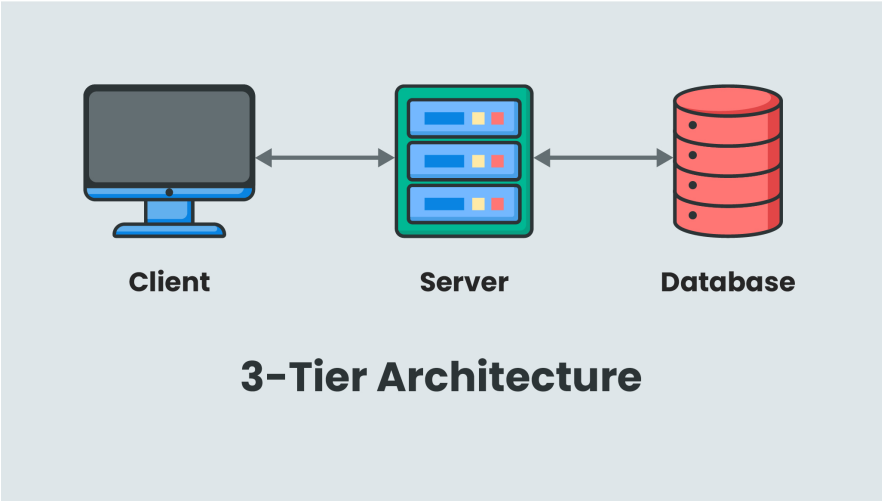


Image taken from here

- 3 Layers: Client (Application) Layer > Application Server / Business Layer >

Database Layer

- Clients have an interface which allows them to interact with the middleware application server. They don't directly connect to the database server.
- Database server does not get overwhelmed since the middleware server handles all the queries, so load on it is reduced.

```
+ Can be scaled.
+ Secure, since client is not directly interacting with the database server
- More complex and difficult to maintain compared to 2-Tier architecture.
```

Schema & 3-Schema Architecture

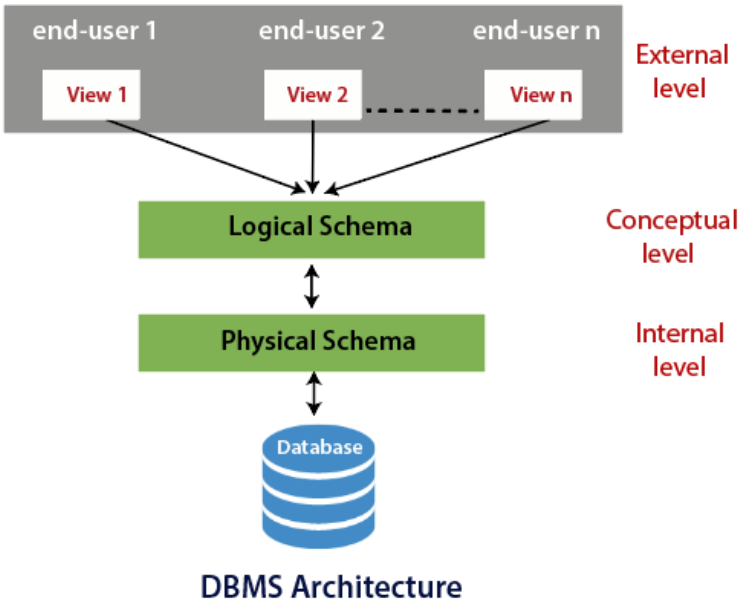


Image taken from here

- Schema means the logical representation of data in the database. It tells us about the structure in which the data will be stored. We also get information about the elements in the structure, their size, etc.
- We're not interested in the physical representation of the data, ie how exactly they're stored in the database. We're only interested in a legible representation of the data.
- A 3-Schema architecture consists of 3 levels: View | Logical | Physical
- We see data in the form of a table, but it is actually stored in the form of files.

1: External Schema / View Level:

- Controlled by the Front-end Developer.
- Different types of users (student, faculty, etc.) have different views.
- Within a particular type, different users are shown different data, relevant to them, within the particular view.

2: Conceptual Schema / Logical Level:

- Controlled by the Database Designer.
- Can be called a blueprint of the structure.
- Stores information about the tables, and the data in the tables.
- Stores Relationships between the tables ie how they are connected.

3: Physical Schema / Physical Level:

- Controlled by the Database Administrator.
- Stores information about where the data is stored, how it is stored.
- Decides if data is centralized or de-centralized.

Data Independence

- The actual data is shown to the user, but information about how it is stored, where it is stored, etc. is hidden.
- We don’t show all the tables to the user, we only show the relevant information the user wants to see.
- The application program is not changed, since the user directly interacts with it and it used to the particular view.
- The objective is to make changes in the Conceptual and Physical Schema in such a way that the user (View Level) is not affected.
- **Logical Data Independence**:
    - Changes made in the Conceptual Schema don’t affect the External Schema.
    - Example: The WebUI shown to an user is not affected even after making some changes to the Conceptual Schema of a database system.
- **Physical Data Independence**:

- Changes made in the Physical Schema don't affect the Conceptual Schema or the External Schema.
- Example: Changing file structure, physical storage, Indexing, etc don't affect the Conceptual Schema.

Keys

- Any piece of data within a tuple in a database, can be called a key.
- Keys are of many types: Primary Key, Candidate Key, etc.

Primary & Candidate Key

- Any key that can be used to uniquely identify data tuples (or rows) in a database, is called a primary key.
- The primary key must be unique. Also, it must not be wrong or have any default value.
- It must not be null.
- There must only be 1 primary key in a database table (in case of RDB).
- Usually, if there's a possibility that the user cannot provide the data for a primary key at this time, it is auto-generated and assigned to the user. For example, registration number, roll number, etc.
- The set of all possible keys that can be used as a primary key, is called Candidate Key.
- We can have keys who value is null, in the set of Candidate Keys. However, the `Unique` property must be maintained.
- Example:

Roll Number	Registration Number	SName	City	Age	Phone Number
101	REG2022001	Alice	Paris	20	555-1234-5678

Roll Number	Registration Number	SName	City	Age	Phone Number
102	REG2022002	Bob	London	22	555-1234-5678
103	REG2022003	Charlie	Berlin	21	555-3456-7890
104	REG2022004	David	Tokyo	23	null
105	REG2022005	Emily	Rome	20	555-5678-9012

    - Here, the candidate keys are: <Roll No, Registration Number, Phone Number>
    - Among these, Phone number can be null, so it can't be used as a primary key. Any 1 of the other 2 can be used for the purpose.

Foreign Key

- An attribute or set of attributes that references to `Primary Key` of the same table or another table.
- It maintains referential integrity.
- If 2 tables are related, there will be atleast 1 common column between them. However, it's not necessary for this column to have the same name in both the tables.
- Example: **Table 1**: Student Information

Roll No	Name	Address
101	Alice	Paris
102	Bob	London
103	Charlie	Berlin
104	David	Tokyo

Table 2: Course Information

Course ID	Course Name	Roll Number

Course ID	Course Name	Roll Number
201	Math	101
202	Science	102
203	English	103
204	History	104

- Here, `Roll Number` of Table 2 is a foreign key, referencing `Roll No` in Table 1.
- The names of the columns may or may not be unique.
- SQL query for setting a foreign key during table creation:

```
CREATE TABLE StudentInformation (
    RollNo INT PRIMARY KEY,
    Name VARCHAR(20),
    Address VARCHAR(20)
);

CREATE TABLE CourseInformation (
    CourseID INT PRIMARY KEY,
    CourseName VARCHAR(20),
    RollNumber INT,
    FOREIGN KEY (RollNumber) REFERENCES StudentInformation(RollNo)
);
```

- SQL query for setting a foreign key after table creation (CONSTRAINT name can be anything as it's the name of the constraint):

```
ALTER TABLE CourseInformation
    ADD CONSTRAINT FK_StudentInfo_RollNo
    FOREIGN KEY (RollNumber) REFERENCES StudentInformation(RollNo);
```

Referential Integrity

Integrity means consistency of the data across multiple tables. By preserving referential integrity, we aim to ensure consistency

among the data across both the referenced (base) & referencing table.

- Now, let's check if INSERT, DELETE & UPDATE operations on either of the tables disturbs the Referential Integrity of the data.
- Example: **Referenced (Base) Table 1**: Student Information

Roll No	Name	Address
101	Alice	Paris
102	Bob	London
103	Charlie	Berlin
104	David	Tokyo

  - Primary Key: `Roll No`

  Referencing Table 2: Course Information

Course ID	Course Name	Roll Number
201	Math	101
202	Science	102
203	English	103
204	History	104

  - Foreign Key: `Roll Number` REFERENCES `Roll No` .
- Operations performed on `Referenced (Base) Table 1` :
  - **INSERT**: No violation
  - **UPDATE**: Possible violation.
    - Updating the `Roll No` on `Table 1` can introduce inconsistency, because the tuple in `Table 2` that was referencing to it, now links to nothing.
    - Solution:
      i. Update manually: Recommended.
      ii. $onUpdateCascade()$: When `Roll No` is updated, delete the

corresponding row on `Table 2`. This can result in loss of data.

iii. $onUpdateSetNull()$: When `Roll No` is updated, set the `Roll Number` of the corresponding row on `Table 2`, to null. However, this may cause issues if the foreign key is also a primary key of `Table 2`.

- **DELETE**: Possible violation.
  - We may accidentally delete a tuple whose `Roll No` has been referenced by `Table 2`.
  - Solution:
    i. No action: Do nothing, this is not recommended for obvious reasons.
    ii. $onDeleteCascade()$: When the tuple is deleted, delete the corresponding row on `Table 2`. This can result in loss of data.
    iii. $onDeleteSetNull()$: When the tuple is deleted, set the `Roll Number` of the corresponding row on `Table 2`, to `null`.

- Operations performed on `Referencing Table 2`:
  - **INSERT**: Possible violation.
    - We cannot insert a tuple in this table with a `Roll Number` that doesn't reference to any tuple in `Table 1`.
    - Solution: Ensure that the correnssponding tuple exists in `Table 1`.
  - **UPDATE**: Possible violation.
    - Updating the foreign key, `Roll Number` on `Table 2` can introduce inconsistency, while it's fine to update other columns.
    - Solution:
      i. Before updating the foreign key, ensure that it's not referencing to any primary key in `Table 1`.
  - **DELETE**: No violation.

Super Key

- It is a combination of all possible attributes which can uniquely identify 2 tuples in a table. We pair the candidate key with any other key, to uniquely identify the tuple.
- A super-set of any candidate key is called a super key.
- Example 0: IF $R(A_1, A_2, A_3, ...A_n)$:
  - If A_1 is a candidate key, then how many super keys are possible?

        - Usually, per key, we have 2 choices. We can either include it in our set, or exclude it.
            > Total number of possible choices in general: 2^n
        - But, here A_1 is said to be the candidate key, and must be included. For the rest, they may or may not be included.
            > Total number of possible choices: 2^{n-1}
    - If A_1 & A_2 are candidate keys:
        - If we just take A_1, we have 2^{n-1} possible choices.
        - If we just take A_2, we have 2^{n-1} possible choices.
        - **However**, both these super sets will have some elements in common. For A_1 set, we will have elements which contain A_2 in them, and vice versa. So, we need to exclude them. We intend to exclude elements which contain both A_1 & A_2 in them, which is 2^{n-2}.
            > Number of possible sets: $2^{n-1} + 2^{n-1} - 2^{n-2}$
    - If A_1A_2 & A_3A_4 are candidate keys:
        > Similarly, the number of possible sets: $2^{n-2} + 2^{n-2} - 2^{n-4}$

Data Models

- Before implementing a database, we need to draw a logical design for the end-product. This design is called a data model.
-

Entity Relationship Model

- Entity: Object which has some attributes.
- Relationship: Relationships among the entities.

Attributes

Example: Student

1. Single: Attribute which has only 1 value.

> Registration number.

2. Multivalued: has more than 1 values.
> Mobile Number

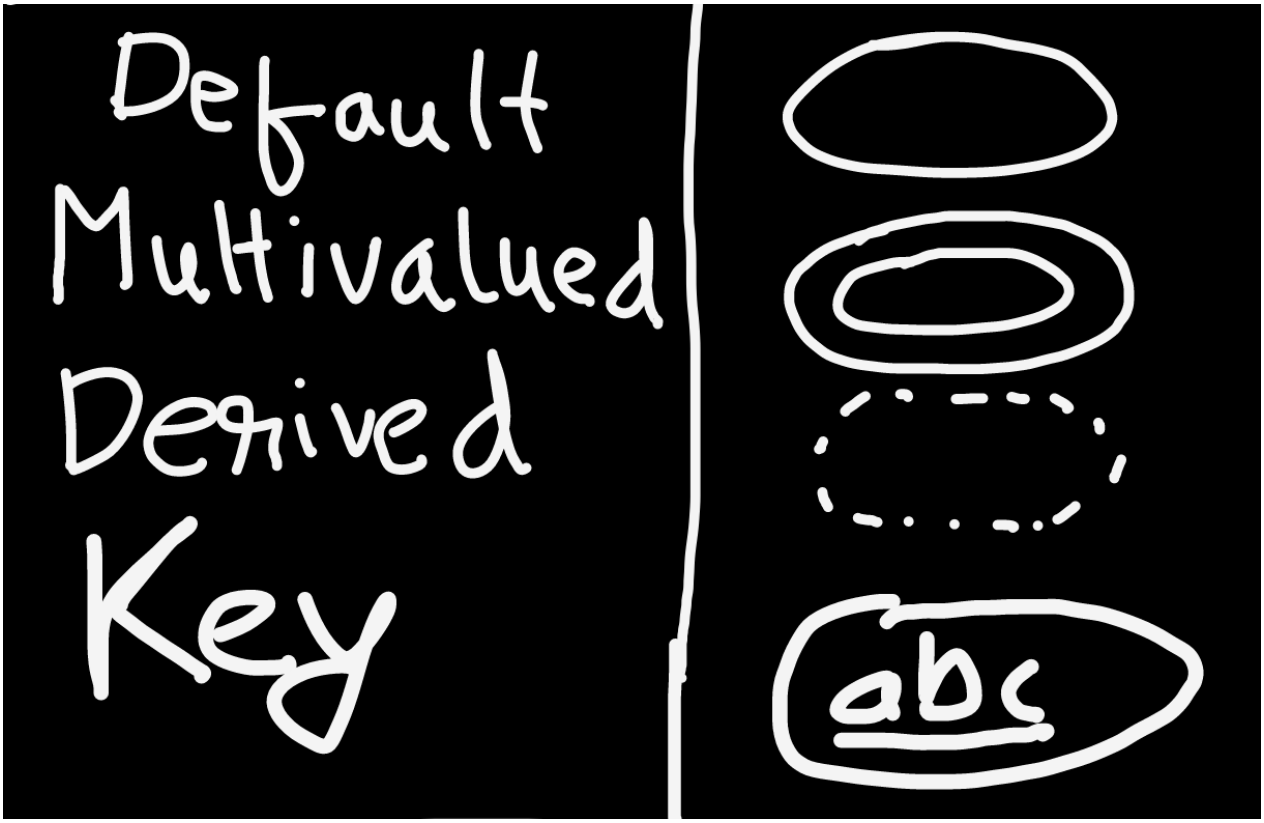
1. Simple: cannot be broken further
> Age

2. Composite: can be broken further
> Name, cab be broken down into first name, middle name, last name
> Date of Birth, can be broken down into DD,MM,YY
> Address

3. Complex: Composite + Multivalued
> Address, it is composite, and a single student can have multiple addresses too.

1. Stored: directly stored.
> Date of Birth

2. Derived: derived from stored attribute.
> Age, derived from Date of Birth.



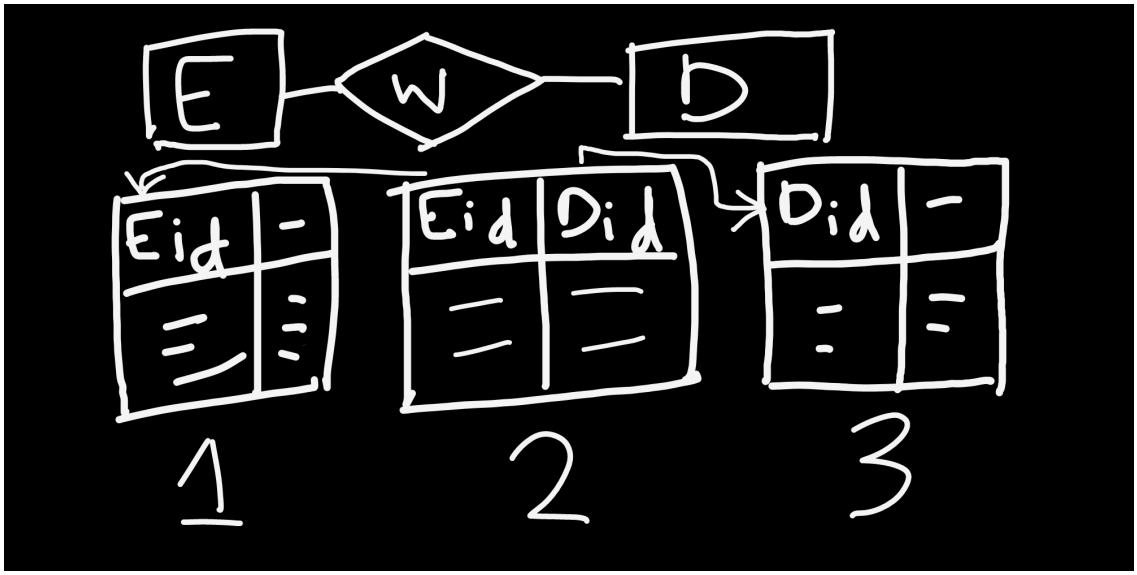
1. Key: can be used as the Primary key in a database.
   > Registration number
2. Non-key: All keys other than the key attribute.
   > Name

1. Required: mandatorily needs to have a value.
   > Username, usually.
2. Optional: may not have a value.
   > Mobile number, usually.

Relationships

One to One

- Representation: 1-1 | One-One
- Example:
  - 3 tables: `Employee` , `Work` , `Department`



- In `Employee` : E_id is the primary key
- In `Department` : D_id is the primary key
- In the Relationship table, `Work` :
  - Foreign keys:
    - E_id: referencing to E_id of `Employee` .
    - D_id: referencing to D_id of `Department` .
  - Relationship (1-1): An employee only works in a single Department.
    - E_id: Value is unique
    - D_id: Value is unique
  - Primary Key: Either of the primary keys of the LHS or RHS table.

    > Any of `E_id` or `D_id`

- We **can reduce** (merge) `Employee` & `Work` into a single `Employee` table. In fact, we can merge the Relationship table with either the LHS or the RHS table.

One to Many

- Representation: 1-M | One-Many
- Example:
  - 3 tables: `Customer` , `Places` , `Order`
  - In `Customer` : C_id is the primary key
  - In `Order` : O_no is the primary key
  - In the Relationship table, `Places` :
    - Foreign keys:

        - C_id: referencing to C_id of `Customer`.
        - O_no: referencing to O_no of `Order`.
      - Relationship (1-Many): A customer can place multiple orders.
        - C_id: Value can repeat
        - O_no: Value is unique
      - **Primary Key**: Primary key of the RHS table.
        > `O_no`, since `C_id` can repeat.
    - We **can merge** (reduce) `Places` & `Order` into a single `Order` table, ie we're merging the Relationship table with the RHS table.

Many to One

- Representation: M-1 | Many-One
- Example:
  - 3 tables: `Order`, `Placed-by`, `Customer`
  - In `Order` : O_no is the primary key
  - In `Customer` : C_id is the primary key
  - In the Relationship table, `Placed-by` :
    - Foreign keys:
      - O_no: referencing to O_no of `Order`.
      - C_id: referencing to C_id of `Customer`.
    - Relationship (Many-1): Many orders can be placed by a single customer.
      - O_no: Value is unique
      - C_id: Value can repeat
    - **Primary Key**: Primary key of the LHS table.
      > `O_no`, since `C_id` can repeat.
  - We **can merge** (reduce) `Placed-by` & `Order` into a single `Order` table, ie we're merging with the LHS table.

Many to Many

- Representation: M-N | Many-Many
- Example:
  - 3 tables: `Student`, `Study`, `Course`

- In `Study` : RollNo is the primary key.
- In `Course` : Cid is the primary key.
- In the Relationship table, `Study` :
  - Foreign keys:
    - RollNo: referencing to RollNo of `Student` .
    - C_id: referencing to C_id of `Course` .
  - Relationship (May-Many): A student can take up multiple courses, and a course can be taken up by multiple students.
    - `RollNo` : Value can repeat
    - `C_id` : Value can repeat
  - **Primary Key**: A composite key, including both the primary keys from the LHS & RHS tables.

    > `RollNo + C_id` , since both `RollNo` & `C_id` can repeat.

- We **cannot merge** (reduce) the tables.

Data Models: ER Model, Relational Model, Object-Oriented Model, Network Model, Hierarchical Model

Basics of keys

Basics of Keys (especially foreign keys)

Normalization

Transaction Control & Concurrency:

ACID property, R-W W-R W-W locking, Conflict Serializability, Recoverability, 2-Phase locking, timestamp ordering for concurrency

---

SQL & Relational algebra: DDL, DML & DCL commands, constraints of various keys, Aggregate functions, Joins (natural, outer join, inner join), nested query

---

Indexing: single-level (primary, cluster, secondary), multi-level (b-tree, b+ tree)
