The

DBMS&SQLNOTES

Database: A database is a collection of related data which represents some aspect of the real world. A database system is designed to be built and populated with data for a certain task.

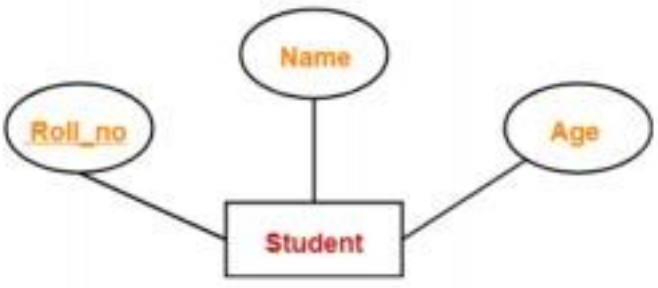
Database Management System (DBMS) is a software for storing and retrieving users' data while considering appropriate security measures. It consists of a group of programs which manipulate the database. The DBMS accepts the request for data from an application and instructs the operating system to provide the specific data. In large systems, a DBMS helps users and other third-party software to store and retrieve data.

Database management systems were developed to handle the following difficulties of typical File-processing systems supported by conventional operating systems.

1. Data redundancy and inconsistency
2. Difficulty in accessing data
3. Data isolation – multiple files and formats
4. Integrity problems
5. Atomicity of updates
6. Concurrent access by multiple users
7. Security problems

ER diagram:

- ER diagram or **Entity Relationship diagram** is a conceptual model that gives the graphical representation of the logical structure of the database.
- It shows all the constraints and relationships that exist among the different components.
- An ER diagram is mainly composed of following three components- Entity Sets, Attributes and Relationship Set.



- Roll_no is a primary key that can identify each entity uniquely.
- Thus, by using a student's roll number, a student can be identified uniquely.

Entity Set:

An entity set is a set of the same type of entities.

- **Strong Entity Set:**
  - A strong entity set is an entity set that contains sufficient attributes to uniquely identify all its entities.
  - In other words, a primary key exists for a strong entity set.
  - Primary key of a strong entity set is represented by underlining it.
- **Weak Entity Set:**
  - A weak entity set is an entity set that does not contain sufficient attributes to uniquely identify its entities.
  - In other words, a primary key does not exist for a weak entity set.
  - However, it contains a partial key called a discriminator. ○ Discriminator can identify a group of entities from the entity set.
  - Discriminator is represented by underlining with a dashed line.

Relationship:

A relationship is defined as an association among several entities.

- **Unary Relationship Set -** Unary relationship set is a relationship set where only one entity set participates in a relationship set.
- **Binary Relationship Set -** Binary relationship set is a relationship set where two entity sets participate in a relationship set.
- **Ternary Relationship Set -** Ternary relationship set is a relationship set where three entity sets participate in a relationship set.
- **N-ary Relationship Set -** N-ary relationship set is a relationship set where 'n' entity sets participate in a relationship set.

Cardinality Constraint:

Cardinality constraint defines the maximum number of relationship instances in which an entity can participate.

The

- **One-to-One Cardinality -** An entity in set A can be associated with at most one entity in set B. An entity in set B can be associated with at most one entity in set A.
- **One-to-Many Cardinality -** An entity in set A can be associated with any number (zero or more) of entities in set B. An entity in set B can be associated with at most one entity in set A.
- **Many-to-One Cardinality -** An entity in set A can be associated with at most one entity in set B. An entity in set B can be associated with any number of entities in set A. ● **Many-to-Many Cardinality -** An entity in set A can be associated with any number (zero or more) of entities in set B. An entity in set B can be associated with any number (zero or more) of entities in set A.

Attributes:

Attributes are the descriptive properties which are owned by each entity of an Entity Set.

Types of Attributes:

- **Simple Attributes -** Simple attributes are those attributes which cannot be divided further. Ex. Age
- **Composite Attributes -** Composite attributes are those attributes which are composed of many other simple attributes. Ex. Name, Address
- **Multi Valued Attributes -** Multi valued attributes are those attributes which can take more than one value for a given entity from an entity set. Ex. Mobile No, Email ID ● **Derived Attributes -** Derived attributes are those attributes which can be derived from other attribute(s). Ex. Age can be derived from DOB.
- **Key Attributes -** Key attributes are those attributes which can identify an entity uniquely in an entity set. Ex. Roll No.

Constraints:

Relational constraints are the restrictions imposed on the database contents and operations. They ensure the correctness of data in the database.

- **Domain Constraint -** Domain constraint defines the domain or set of values for an attribute. It specifies that the value taken by the attribute must be the atomic value from its domain.
- **Tuple Uniqueness Constraint -** Tuple Uniqueness constraint specifies that all the tuples must be necessarily unique in any relation.

- **Key Constraint -** All the values of the primary key must be unique. The value of the primary key must not be null.
- **Entity Integrity Constraint -** Entity integrity constraint specifies that no attribute of primary key must contain a null value in any relation.
- **Referential Integrity Constraint -** It specifies that all the values taken by the foreign key must either be available in the relation of the primary key or be null.

Closure of an Attribute Set:

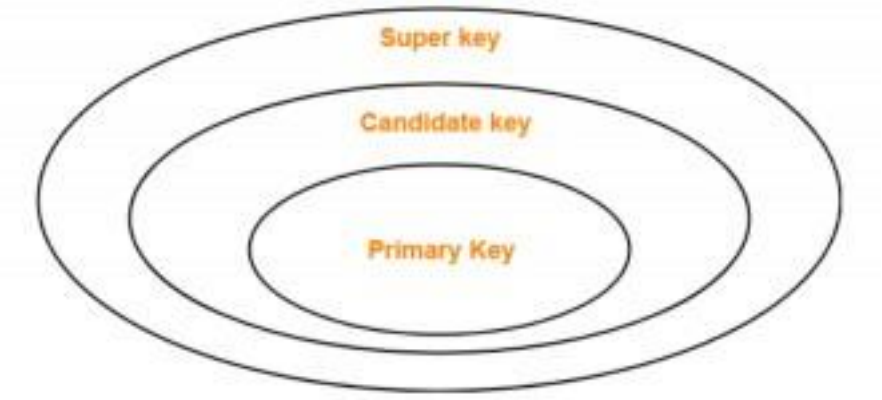
The set of all those attributes which can be functionally determined from an attribute set is called a closure of that attribute set.

Keys:

A key is a set of attributes that can identify each tuple uniquely in the given relation**.**

Types of Keys:

- **Super Key -** A superkey is a set of attributes that can identify each tuple uniquely in the given relation. A super key may consist of any number of attributes.
- **Candidate Key -** A set of minimal attribute(s) that can identify each tuple uniquely in the given relation is called a candidate key.
- **Primary Key -** A primary key is a candidate key that the database designer selects while designing the database. Primary Keys are unique and NOT NULL.



- **Alternate Key -** Candidate keys that are left unimplemented or unused after implementing the primary key are called as alternate keys.
- **Foreign Key -** An attribute ‘X’ is called as a foreign key to some other attribute ‘Y’ when its values are dependent on the values of attribute ‘Y’. The relation in which attribute ‘Y’ is present is called as the referenced relation. The relation in which attribute ‘X’ is present is called as the referencing relation.

The

- **Composite Key -** A primary key composed of multiple attributes and not just a single attribute is called a composite key.
- **Unique Key -** It is unique for all the records of the table. Once assigned, its value cannot be changed i.e. it is non-updatable. It may have a NULL value. **Functional Dependency:**

In any relation, a functional dependency $\alpha \rightarrow \beta$ holds if- Two tuples having same value of attribute α also have same value for attribute β.

Types of Functional Dependency:

- **Trivial Functional Dependencies –**
  Thus, if RHS of a functional dependency is a subset of LHS, then it is called$\subseteq$
  a ○ A functional dependency X ⟶ Y is said to be trivial if and only if Y
  X. ○
  trivial functional dependency.
- **Non-Trivial Functional Dependencies –**
  ○ A functional dependency X ⟶ Y is said to be non-trivial if and only if Y X. ○
  that is not a part of LHS, then it is called a non-trivial functional dependency.
  Thus, if there exists at least one attribute in the RHS of a functional dependency$\not\subset$

Decomposition of a Relation:

The process of breaking up or dividing a single relation into two or more sub relations is called the decomposition of a relation.

Properties of Decomposition:

- **Lossless Decomposition -** Lossless decomposition ensures ○ No information is lost from the original relation during decomposition. ○ When
  the sub relations are joined back, the same relation is obtained that was decomposed.
- **Dependency Preservation -** Dependency preservation ensures ○ None of the functional dependencies that hold on the original relation are lost. ○ The sub relations still hold or satisfy the functional dependencies of the original relation.

Types of Decomposition:

- **Lossless Join Decomposition:**

- Consider there is a relation R which is decomposed into sub relations R1, R2, …., Rn.
- This decomposition is called lossless join decomposition when the join of the sub relations results in the same relation R that was decomposed.
- For lossless join decomposition, we always have- R1 R2 R3 ……. Rn = R

  where ⋈ is a natural join operator ⋈ ⋈ ⋈

- **Lossy Join Decomposition:**
  - Consider there is a relation R which is decomposed into sub relations R1, R2, …., Rn.
  - This decomposition is called lossy join decomposition when the join of the sub relations does not result in the same relation R that was decomposed.
  - For lossy join decomposition, we always have- R1 R2 R3 ……. Rn R

    where ⋈ is a natural join operator ⋈ ⋈ ⋈ ⊃

<u>Normalization:</u>

In DBMS, database normalization is a process of making the database consistent by-

- Reducing the redundancies
- Ensuring the integrity of data through lossless decomposition

Normal Forms:

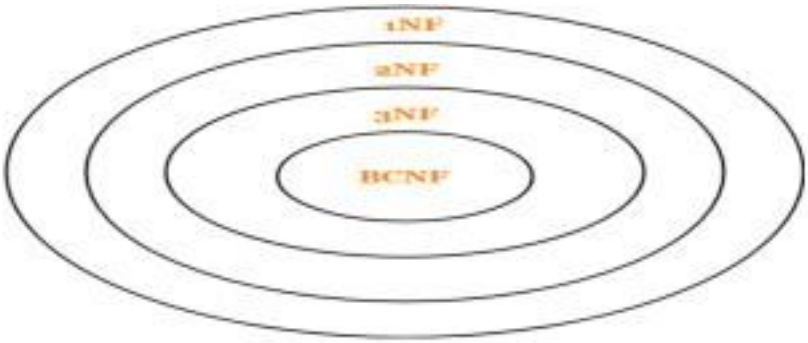
- **First Normal Form (1NF) -** A given relation is called in First Normal Form (1NF) if each cell of the table contains only an atomic value i.e. if the attribute of every tuple is either single valued or a null value.
- **Second Normal Form (2NF) -** A given relation is called in Second Normal Form (2NF) if and only if ○ Relation already exists in 1NF.
  - No partial dependency exists in the relation.

    A ⟶ B is called a **partial dependency** if and only if- A is a subset of some candidate key and B is a non-prime attribute.
- **Third Normal Form (3NF) -** A given relation is called in Third Normal Form (3NF) if and only if ○ Relation already exists in 2NF.

The

- No transitive dependency exists for non-prime attributes.
  A → B is called a **transitive dependency** if and only if- A is not a super key and B is a non-prime attribute.

- **Boyce-Codd Normal Form -** A given relation is called in BCNF if and only if ○ Relation already exists in 3NF.
  - For each non-trivial functional dependency 'A → B', A is a super key of the relation.



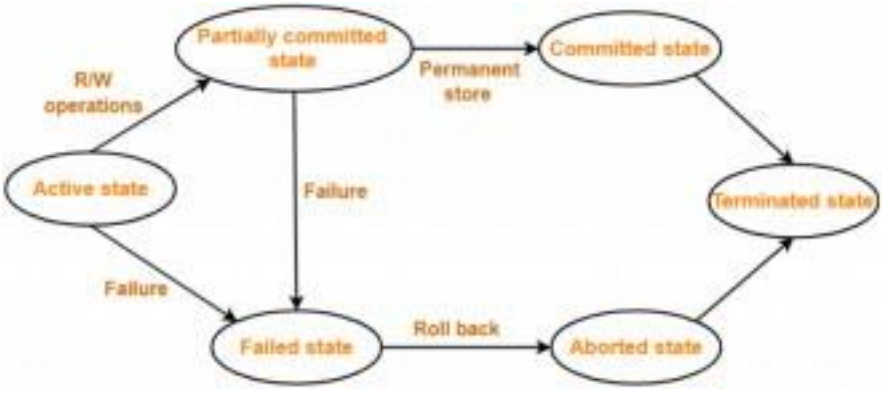
Transaction:

Transaction is a single logical unit of work formed by a set of operations.

Operations in Transaction:

- **Read Operation - Read(A)** instruction will read the value of 'A' from the database and will store it in the buffer in main memory.
- **Write Operation – Write(A)** will write the updated value of 'A' from the buffer to the database.

Transaction States:



- **Active State –** ○ This is the first state in the life cycle of a transaction.
  - A transaction is called in an active state as long as its instructions are getting executed. o All the changes made by the transaction now are stored in the buffer in main memory.

- **Partially Committed State –** ○ After the last instruction of the transaction has been executed, it enters into a partially committed state.
    - ○ After entering this state, the transaction is considered to be partially committed.
    - ○ It is not considered fully committed because all the changes made by the transaction are still stored in the buffer in main memory.
- **Committed State –** ○ After all the changes made by the transaction have been successfully stored into the database, it enters into a committed state.
    - ○ Now, the transaction is considered to be fully committed.
- **Failed State –** ○ When a transaction is getting executed in the active state or partially committed state and some failure occurs due to which it becomes impossible to continue the execution, it enters into a failed state.
- **Aborted State –** ○ After the transaction has failed and entered into a failed state, all the changes made by it have to be undone.
    - ○ To undo the changes made by the transaction, it becomes necessary to roll back the transaction.
    - ○ After the transaction has rolled back completely, it enters into an aborted state.
- **Terminated State –** ○ This is the last state in the life cycle of a transaction.
    - ○ After entering the committed state or aborted state, the transaction finally enters into a terminated state where its life cycle finally comes to an end.

ACID Properties:

To ensure the consistency of the database, certain properties are followed by all the transactions occurring in the system. These properties are called as **ACID Properties** of a transaction.

- **Atomicity –** ○ This property ensures that either the transaction occurs completely or it does not occur at all.
    - ○ In other words, it ensures that no transaction occurs partially.
- **Consistency –** ○ This property ensures that integrity constraints are maintained.
    - ○ In other words, it ensures that the database remains consistent before and after the transaction.
- **Isolation –** ○ This property ensures that multiple transactions can occur simultaneously without causing any inconsistency.

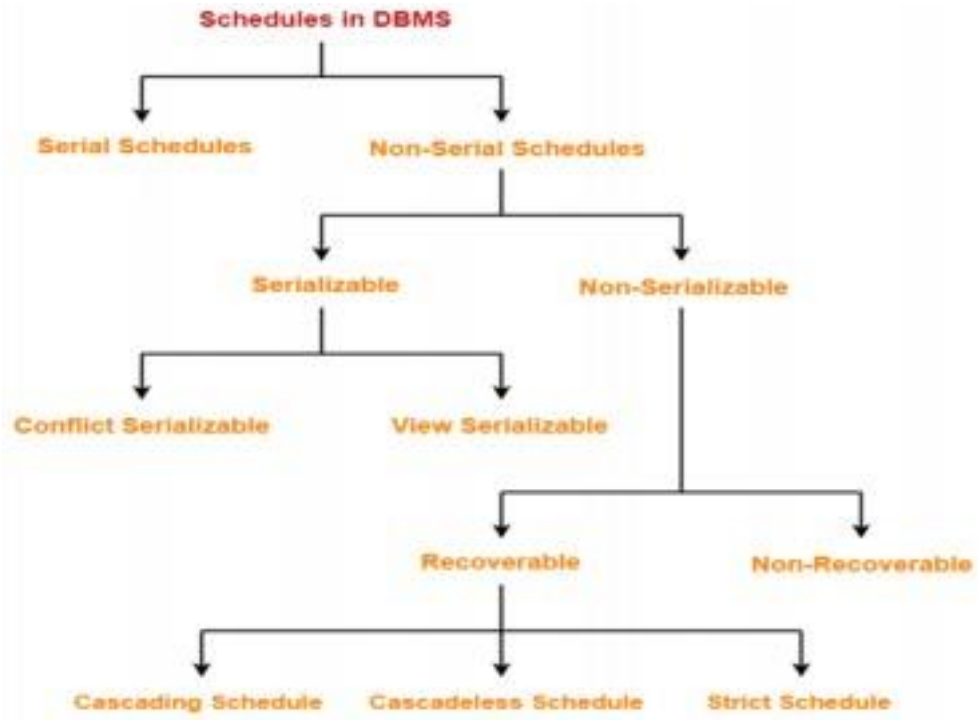
The

- The resultant state of the system after executing all the transactions is the same as the state that would be achieved if the transactions were executed serially one after the other.

- **Durability –** ○ This property ensures that all the changes made by a transaction after its successful execution are written successfully to the disk.
  - It also ensures that these changes exist permanently and are never lost even if there occurs a failure of any kind.

Schedules:

The order in which the operations of multiple transactions appear for execution is called as a schedule.



- **Serial Schedules –** ○ All the transactions execute serially one after the other.
  - When one transaction executes, no other transaction is allowed to execute.
  - Serial schedules are always- Consistent, Recoverable, Cascadeless and Strict. ●

Non-Serial Schedules – ○ Multiple transactions execute concurrently. o Operations of all the transactions are inter leaved or mixed with each other.

- Non-serial schedules are **not** always- Consistent, Recoverable, Cascadeless and Strict.

Serializability –

- Some non-serial schedules may lead to inconsistency of the database. ● Serializability is a concept that helps to identify which non-serial schedules are correct and will maintain the consistency of the database.
- **Serializable Schedules –** ○ If a given non-serial schedule of 'n' transactions is equivalent to some serial schedule of 'n' transactions, then it is called as a serializable schedule.
    - Serializable schedules are always- Consistent, Recoverable, Cascadeless and Strict.

Types of Serializability –

- **Conflict Serializability** - If a given non-serial schedule can be converted into a serial schedule by swapping its non-conflicting operations, then it is called a conflict serializable schedule.
- **View Serializability -** If a given schedule is found to be viewed as equivalent to some serial schedule, then it is called a view serializable schedule.

Non-Serializable Schedules –

- A non-serial schedule which is not serializable is called a non-serializable schedule. ● A non-serializable schedule is not guaranteed to produce the same effect as produced by some serial schedule on any consistent database.
- Non-serializable schedules- may or may not be consistent, may or may not be recoverable.
- **Irrecoverable Schedules –**

  If in a schedule, ○ A transaction performs a dirty read operation from an uncommitted transaction

  ○ And commits before the transaction from which it has read the value then such a schedule is known as an Irrecoverable Schedule.
- **Recoverable Schedules –**

  If in a schedule, ○ A transaction performs a dirty read operation from an uncommitted transaction ○ And its commit operation is delayed till the uncommitted transaction either commits or roll backs

  then such a schedule is known as a Recoverable Schedule.

The

Types of Recoverable Schedules –

- **Cascading Schedule -** If in a schedule, failure of one transaction causes several other dependent transactions to rollback or abort, then such a schedule is called as a Cascading Schedule or Cascading Rollback or Cascading Abort.
- **Cascadeless Schedule -** If in a schedule, a transaction is not allowed to read a data item until the last transaction that has written it is committed or aborted, then such a schedule is called as a Cascadeless Schedule.
- **Strict Schedule -** If in a schedule, a transaction is neither allowed to read nor write a data item until the last transaction that has written it is committed or aborted, then such a schedule is called as a Strict Schedule.

Relational Algebra:

Relational Algebra is a procedural query language which takes a relation as an input and generates a relation as an output.

Basic Operator	**Semantic**
σ(Selection)	Select rows based on given condition
∏(Projection)	Project some columns
X (Cross Product)	Cross product of relations, returns **m*n** rows where m and n are number of rows in R1 and R2 respectively.
U (Union)	Return those tuples which are either in R1 or in R2. Max no. of rows returned **= m+n** and Min no. of rows returned = **max(m,n)**
−(Minus)	R1-R2 returns those tuples which are in R1 but not in R2. Max no. of rows returned = **m** and Min no. of rows returned = **m-n**
ρ(Rename)	Renaming a relation to another relation.
Extended Operator	**Semantic**

∩ (Intersection)	Returns those tuples which are in both R1 and R2. Max no. of rows returned = min(m,n) and Min no. of rows returned = 0
⋈θ(Conditional Join)	Selection from two or more tables based on some condition (Cross product followed by selection)
⋈(Equi Join)	It is a special case of conditional join when only equality conditions are applied between attributes.
⋈(Natural Join)	In natural join, equality conditions on common attributes hold and duplicate attributes are removed by default. **Note:** Natural Join is equivalent to cross product if two relations have no attribute in common and natural join of a relation R with itself will return R only.
⟕(Left Outer Join)	When applying join on two relations R and S, some tuples of R or S do not appear in the result set which does not satisfy the join conditions. But Left Outer Joins gives all tuples of R in the result set. The tuples of R which do not satisfy the join condition will have values as NULL for attributes of S.
⟖(Right Outer Join)	When applying join on two relations R and S, some tuples of R or S do not appear in the result set which does not satisfy the join conditions. But Right Outer Joins gives all tuples of S in the result set. The tuples of S which do not satisfy the join condition will have values as NULL for attributes of R.
⟗(Full Outer Join) /(Division Operator)	When applying join on two relations R and S, some tuples of R or S do not appear in the result set which does not satisfy the join conditions. But Full Outer Joins gives all tuples of S and all tuples of R in the result set. The tuples of S which do not satisfy the join condition will have values as NULL for attributes of R and vice versa. Division operator A/B will return those tuples in A which are associated with every tuple of B. **Note:** Attributes of B should be a proper subset of attributes of A. The attributes in A/B will be Attributes of A- Attribute of B.

File Structures:

The

- **Primary Index:** A primary index is an ordered file, records of fixed length with two fields. First field is the same as the primary key as a data file and the second field is a pointer to the data block, where the key is available. The average number of block accesses using index = $\mathbf{\log_2 Bi + 1}$, where Bi = number of index blocks.
- **Clustering Index:** Clustering index is created on data file whose records are physically ordered on a non-key field (called Clustering field).
- **Secondary Index:** Secondary index provides secondary means of accessing a file for which primary access already exists.

B Trees

At every level , we have Key and Data Pointer and data pointer points to either block or record.

Properties of B-Trees:

Root of B-tree can have children between **2** and **P**, where P is Order of tree.

Order of tree – Maximum number of children a node can have.

Internal node can have keys between **P/2**⌈ ⌉ **– 1**⌉and **P-1**
Internal node can have children between **P/2** and **P**

B+ Trees ⌈

In B+ trees, the structure of leaf and non-leaf are different, so their order is. Order of non-leaf will be higher as compared to leaf nodes.

Searching time will be less in B+ trees, since it doesn't have record pointers in non-leaf because of which depth will decrease.