Normalization:

To restructure logical data model to :-

1. Eliminate Redundancy
2. Organize Data efficiency
3. Reduce potential for data anomalies.

Key: Gaurantees uniqueness, no two rows can be identical.

Data Anomalies:

Inconsistencies in data stored in a database as a result of an operation such as update, insertion and deletion.

Update anomaly:
Data inconsistency that results from data redundancy and a partial update.

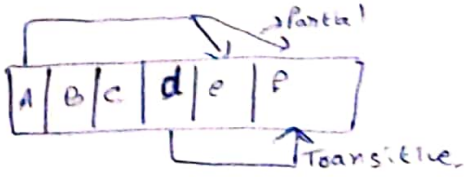
Eg: If fee increased for course, data must be updated for more than one row.

Insert Anomaly:
Occurs when certain attributes cannot be inserted into database without presence of other attributes.

Delet Anomaly:
If we delete a record that may contain attributes that shouldn't be deleted.

1 NF:

1. Each table has a primary key.
2. Values in each column of a table are atomic (No multi-valued).
3. There are no repeating groups.



2 NF

1. It should be in 1NF
2. There must be no partial dependency

Partial Dependency:
All non-prime attributes are not fully dependent on prime key attribute.

Functional Dependency:

Relationship that exists when one attribute uniquely determines another attribute.

$X \rightarrow Y$ (Y is dependent on X)

(X determines Y)

X ↙ Determinant

Amstrong axioms:

1. Reflexive Rule:
   $X \rightarrow X$
2. Augmentation:
   $X \rightarrow Y$ then $XZ \rightarrow Y$
3. Transitive:
   $X \rightarrow Y$, $Y \rightarrow Z$ then $X \rightarrow Z$
4. Union:
   $X \rightarrow Y$, $X \rightarrow Z$, then $X \rightarrow YZ$
5. Decomposition:
   $X \rightarrow Y$ then $X \rightarrow Z$ (if Z is subset of Y)
6. Pseudotransitivity:
   $X \rightarrow Y$, $YZ \rightarrow W$ then $XZ \rightarrow W$.

3NF:

1. It should be in 2NF
2. No non key attribute is transitively dependent on primary key.

Transitive Dependency:

No non-prime attribute should determine another non-prime attribute.

Trivial FD:

$A \rightarrow B$ is trivial if B is subset of A.

FD closure:

Denoted by F^+ ; Certain other FD's that are logically implied by F.

=) Using axioms, if we add more FD's. they are called as FD closure

Attribute Closure:

Set of all attributes which can be functionally determined from an attribute set.

Uses:

1. Testing for Superkey
2. Testing FD's.
3. Computing closure of F.

BCNF (Boyce and Codd Normal Form)

1. Every determinant is a super key.
2. It should be in 3NF.

→ 3NF table which doesn't have multiple overlapping candidate keys is said to be in BCNF.

Canonical Cover:

It is a set of FD's such that all following are satisfied:

1. F logically implies all dependencies in F_c.
2. F_c logically implies all dependencies in F.
3. No FD in F_c contains an extraneous attribute
4. Each left side of FD in F_c is unique.

*Split each right side into single one's, then combine.

Lossless Decomposition:

1. Union of decomposed table's attributes should be original table's attributes
2. Intersection of tables attributes should'nt be null.
3. Intersection of tables must be equal to set of super keys.

→ This property ensures no spurious tuples are generated when a Natural join operation is applied to relations in decomposition.

Spurious Tuples:

Extra tuples which might not be required.

It can be solved using ~~Rocchi~~ Chase method.

→ If it is not in lossless decomposition, then it is lossy.

Dependency Preservation:

If union of projections of F on each R_i in D is equalent to F:

→ If it is not dependency-preserving, ∴ dependency is lost in decomposition

* Each FD specified in F either appeared directly in one of relation schemas in decomposed D (or) could be inferred from dependencies.

→ Non-trivial FD's is allowed in 3NF but not in BCNF.

→ If no FD in F cause violation, then no violation in F+ (In BCNF).

→ 3NF : No need of losslessness (or) dependency preservation.

Disadvantage: we may have to use null values to represent some of possible meaningful relationships among data items, and there is prob. of repetition of info.