



NORMALIZATION

Normalization

A large database defined as a single relation may result in data duplication. This repetition of data may result in:

- Making relations very large.
- It isn't easy to maintain and update data as it would involve searching many records in relation.
- Wastage and poor utilization of disk space and resources.
- The likelihood of errors and inconsistencies increases.

Normalization

- So to handle these problems, we should analyze and decompose the relations with redundant data into smaller, simpler, and well-structured relations that are satisfy desirable properties.
- Normalization is a process of decomposing the relation into relations with fewer attributes.

What is Normalization??

- Normalization is the process of organizing the data in the database.
- Normalization is used to minimize the redundancy from a relation or set of relations.
- It is also used to eliminate undesirable characteristics like Insertion, Update, and Deletion Anomalies.
- Normalization divides the larger table into smaller and links them using relationships.
- The normal form is used to reduce redundancy from the database table.

Data Modification Anomalies

- **Insertion Anomaly:** Insertion Anomaly refers to when one cannot insert a new tuple into a relationship due to lack of data.
- **Deletion Anomaly:** The delete anomaly refers to the situation where the deletion of data results in the unintended loss of some other important data.
- **Updation Anomaly:** The update anomaly is when an update of a single data value requires multiple rows of data to be updated.

Insertion Anomaly

Student ID	Name	Age	Dept Name	Dept Code	Head of the Dept
1	Tom	19	Math	101	Kiani
2	Katie	19	Math	101	Kiani
3	Aleem	20	Math	101	Kiani
4	Nejati	18	Physics	102	Steve
5	John	21	Physics	102	Steve
6	Amer	19	Chemistry	103	Bill
			Geology	115	Gary

Not allowed

Deletion Anomaly

eno	ename	bdate	title	salary	supereno	dno	dname	mgreno
E1	J. Doe	01-05-75	EE	30000	E2	null	null	null
E2	M. Smith	06-04-66	SA	50000	E5	D3	Accounting	E5
E3	A. Lee	07-05-66	ME	40000	E7	D2	Consulting	E7
E4	J. Miller	09-01-50	PR	20000	E6	D3	Accounting	E5
E5	B. Casey	12-25-71	SA	50000	E8	D3	Accounting	E5
E6	L. Chu	11-30-65	EE	30000	E7	D2	Consulting	E7
E7	R. Davis	09-08-77	ME	40000	E8	D1	Management	E8
E8	J. Jones	10-11-72	SA	50000	null	D1	Management	E8

- Consider this deletion anomaly:
 - Delete employees E3 and E6 from the database.
 - Deleting those two employees removes them from the database, and we now have lost information about department D2!

Updation Anomaly

EMPLOYEES				
EID	Name	Salary	D-ID	Department
1	Poul	10,000	1-4	A/C
2	John	6,000	1	A/C
3	Shawn	7,000	3	Sales
4	Mike	6,000	2	Marketing
5	Hudson	10,000	2	Marketing
6	Smith	20,000	1	A/C
7	Paula	50,000	2	M/K
8	Matthew	15,000	1	A/C

Why Normalization is required?

- The main reason for normalizing the relations is removing these anomalies.
- Failure to eliminate anomalies leads to data redundancy and can cause data integrity and other problems as the database grows.
- Normalization consists of a series of guidelines that helps to guide you in creating a good database structure.

Types of Normal Forms

	1NF	2NF	3NF	4NF	5NF
Decomposition of Relation	R	R ₁₁ R ₁₂	R ₂₁ R ₂₂ R ₂₃	R ₃₁ R ₃₂ R ₃₃ R ₃₄	R ₄₁ R ₄₂ R ₄₃ R ₄₄ R ₄₅
Conditions	Eliminate Repeating Groups	Eliminate Partial Functional Dependency	Eliminate Transitive Dependency	Eliminate Multi-values Dependency	Eliminate Join Dependency

Types of Normal Forms

Normal Form	Description
1NF	A relation is in 1NF if it contains an atomic value.
2NF	A relation will be in 2NF if it is in 1NF and all non-key attributes are fully functional dependent on the primary key.
3NF	A relation will be in 3NF if it is in 2NF and no transition dependency exists.
BCNF	A stronger definition of 3NF is known as Boyce Codd's normal form.
4NF	A relation will be in 4NF if it is in Boyce Codd's normal form and has no multi-valued dependency.
5NF	A relation is in 5NF. If it is in 4NF and does not contain any join dependency, joining should be lossless.

Advantages of Normalization

- Normalization helps to minimize data redundancy.
- Data consistency within the database.
- Much more flexible database design.
- Enforces the concept of relational integrity.

Disadvantages of Normalization

- You cannot start building the database before knowing what the user needs.
- The performance degrades when normalizing the relations to higher normal forms, i.e., 4NF, 5NF.
- It is very time-consuming and difficult to normalize relations of a higher degree.
- Careless decomposition may lead to a bad database design, leading to serious problems.



Thank You