

NORMALIZATION

NORMALIZATION

- **Normalization** is the process of splitting relations into well-structured relations that allow users to insert, delete, and update tuples without introducing database inconsistencies. Without normalization, many problems can occur when trying to load an integrated conceptual model into the DBMS. These problems arise from relations that are generated directly from user views are called anomalies. There are three types of anomalies: update, deletion, and insertion anomalies.

The process of producing a simpler and more reliable database structure is called **normalization**. It is used to create a suitable set of relations for storing data. This process works through different stages known as **normal forms**. These stages are 1NF, 2NF, 3NF and so on. Each normal form has certain requirements or condition. These conditions have to fulfilled to bring the database in that particular normal form. If a relation satisfies the conditions of a normal form, it is said to be in that normal form.

The task of database design starts with unnormalized set of relations. The process of normalization identifies and corrects the problems and complexities of database design. It produces a new set of relations. The new design is as free of processing problems as possible.

DATABASE ANOMALIES

- Anomalies are problems that can occur in poorly planned, unnormalized databases where all the data is stored in one table (a flat file database).
- Types Of Anomalies:
- The three type of anomalies that can arise in the database because of redundancy are
 1. Insertion
 2. Deletion
 3. Modification/ Updating anomalies.

PURPOSE OF NORMALIZATION

The purposes of normalization are as follows:

- It makes the database design efficient in performance.
- It reduces the amount of data if possible.
- It makes the database design free of update, insertion and deletion anomalies.
- It makes the design according to the rules of relational databases.
- It identifies relationship between entities.
- It makes a design that allows simple retrieval of data.
- It simplifies data maintenance and reduces the need to restructure data.

First Normal Form(1NF)

A relation is in first normal form (1NF) if it does not contain a **repeating group**. A repeating group is a set of one or more data items that may occur a variable number of times in a tuple. The value in each attribute value should be atomic and every tuple should be unique. Each cell in a relation should contain only one value. An example of un-normalized relation is as follows: -

Accountant Number	Skill Number	Skill Category	Proficiency Number	Accountant Name	Accountant Age	Group Number	Group City	Group Supervisor
21	113	Systems	3	Ali	55	52	ISD	Babar
35	113	Systems	5					
	179	Tax	1	Daud	32	44	LHR	Ghafoor
	204	Audit	6					
50	179	Tax	2	Chohan	40	44	LHR	Ghafoor
77	148	Consulting	6	Zahid	52	52	ISD	Babar
	179	Tax	6					

The above relation is un-normalized because it contains repeating groups of three attributes Skill Number, Skill Category and Proficiency Number. All three fields contain more than one value.

In order to convert this relation in first normal form, these repeating groups should be removed. The following relation is in first normal form:

Primary Key



Accountant Number	Skill Number	Skill Category	Proficiency	Accountant Name	Accountant Age	Group Number	Group City	Group Supervisor
21	113	Systems	3	Ali	55	52	ISD	Babar
35	113	Systems	5	Daud	32	44	LHR	Ghafoor
35	179	Tax	1	Daud	32	44	LHR	Ghafoor
35	204	Audit	6	Daud	32	44	LHR	Ghafoor
50	179	Tax	2	Chohan	40	44	LHR	Ghafoor
77	148	Consulting	6	Zahid	52	52	ISD	Babar
77	179	Tax	6	Zahid	52	52	ISD	Babar

Figure: First Normal Form (1NF)

PROBLEMS IN INF

1. Updating Problem

Suppose the user wants to change the name of Accountant Number 35 to "M. Daud". He has to change the name in all records in which Accountant number 35 appears. This process of updating can be very lengthy.

2. Inconsistent Data

The above table may contain inconsistent data. There are three records of Accountant Number 35. It is possible that there are two different names with Account Number 35 in two different records. The user can make this error during updating.

PROBLEMS IN INF

3. Addition Problem

Suppose the user wants to add another skill number in the table. It is not possible until an Accountant with that skill exists because both Skill Number and Accountant Number are used as primary key in the above table.

4. Deletion Problem

Suppose the user wants to delete the record of supervisor Ghafoor. If he deletes the whole record in which Ghafoor appears, the information about Accountants will also be lost.

Second Normal Form (2NF) functional Dependency

A relation is in Second Normal Form (2NF) if it is in 1NF and if all of its nonkey attributes are fully functionally dependent on the whole key. It means that none of non-key attributes are related to a part of key.

The above relation in 1NF has some attributes which are not depending on the whole primary key. For example, Accountant Name, Accountant Age and group information is determined by Accountant Number and is not dependent on Skill. The following relation can be created in which all attributes are fully dependent on primary key Account Number.

Primary Key



Accountant Number	Skill Number	Skill Category	Proficiency	Accountant Name	Accountant Age	Group Number	Group City	Group Supervisor
21	113	Systems	3	Ali	55	52	ISD	Babar
35	113	Systems	5	Daud	32	44	LHR	Ghafoor
35	179	Tax	1	Daud	32	44	LHR	Ghafoor
35	204	Audit	6	Daud	32	44	LHR	Ghafoor
50	179	Tax	2	Chohan	40	44	LHR	Ghafoor
77	148	Consulting	6	Zahid	52	52	ISD	Babar
77	179	Tax	6	Zahid	52	52	ISD	Babar

Figure: First Normal Form (1NF)

Primary Key



Accountant Number	Accountant Name	Accountant Age	Group Number	Group City	Group Supervisor
21	Ali	55	52	ISD	Babar
35	Daud	32	44	LHR	Ghafoor
50	Chohan	40	44	LHR	Ghafoor
77	Zahid	52	52	ISD	Babar

Figure: Accountant Table in 2NF

Similarly, another relation Skill can be created in which all fields are fully dependent on the primary key as follows:

Primary Key

↓

Skill Number	Skill Category
113	Systems
179	Tax
204	Audit
148	Consulting

Figure: Skill Table in 2NF

The attribute Proficiency in 1NF relation was fully dependent on the whole primary key. The Proficiency requires to know the accountant number and skill number. The third relation will be created as follows:

Primary Key

↓

Accountant Number	Skill Number	Proficiency
21	113	3
35	113	5
35	179	1
35	204	6
50	179	2
77	148	6
77	179	6

Figure: Proficiency Table in 2NF

Analysis of Second Normal Form

The following analysis indicates whether the problems are eliminated in 2NF or not.

- **Updating Problem:** If the user needs to change the name of Accountant Number 35 to "M. Daud" in 1NF, he must change the name in every record in which Accountant number 35 appears. But in 2NF, the record of one accountant appears only once. The updating problem is eliminated in 2NF.
- **Inconsistent Data:** The record of one accountant appears only once in the database, the possibility of inconsistent data is automatically eliminated.
- **Addition Problem:** In 1NF, it was not possible to enter a new Skill Number until an Accountant with that skill existed. In 2NF, any number of skills can be added in Skill relation without an accountant with that skill. It eliminates the addition problem.
- **Deletion Problem:** In 2NF, if the record of Ghafoor is deleted, it does not delete any other record.

The analysis shows that the second normal form has solved all problems of 1NF.

3 NF TRANSITIVE DEPENDENCY

A relation is in third normal form if it is in 2NF and if no non-key attribute is dependent on another non-key attribute. It means that all non-key attributes are functionally dependent only on primary key. There should be no transitive dependency in a relation.

In order to convert a relation to 3NF:

- Remove all attributes from the 2NF record that depend on another non-key field
- Place them into a new relation with the other attribute as the primary key.

The Accountant table in 2NF contains some attributes which are depending on non-key attributes. For example, Group City and Group Supervisor are depending on a non-key field Group Number. A new relation can be created as follows:

Primary Key



Accountant Number	Accountant Name	Accountant Age	Group Number
21	Ali	55	52
35	Daud	32	44
50	Chohan	40	44
77	Zahid	52	52

Figure: Accountant Table in 3NF

Primary Key



Group Number	Group City	Group Supervisor
52	ISD	Babar
44	LHR	Ghafoor

Figure: Group table in 3NF

Primary Key



Skill Number	Skill Category
113	Systems
179	Tax
204	Audit
148	Consulting

Figure: Skill Table in 3NF

Primary Key



Accountant Number	Skill Number	Proficiency
21	113	3
35	113	5
35	179	1
35	204	6
50	179	2
77	148	6
77	179	6

Figure: Proficiency Table in 3NF