CHAPTER 6: NORMALIZATION

Introduction and meaning of Database normalization

Database normalization is the process of organizing the fields and tables of a relational database to minimize redundancy. Normalization usually involves dividing large tables into smaller (and less redundant) tables and defining relationships between them.

Normalization of data can be defined as a process during which the existing tables of a database are tested to find certain data dependency between the column and the rows or normalizing of data can be referred to a formal technique of making preliminary data structures into an easy to maintain and make efficient data structure

With data normalization any table dependency detected, the table is restructured into multiple tables (two tables) which eliminate any column dependency. Incase data dependency is still exhibited the process is repeated till such dependency are eliminated. The process of eliminating data redundancy is based upon a theory called functional dependency

Importance of normalization

- It highlights constraints and dependency in the data and hence aid the understanding the nature of the data
- Normalization controls data redundancy to reduce storage requirement and standard maintenance
- Normalization provide unique identification for records in a database
- Each stage of normalization process eliminate a particular type of undesirable dependency
- Normalization permits simple data retrieval in response to reports and queries
- The third normalization form produces well designed database which provides a higher degree of independency
- Normalization helps define efficient data structures
- Normalized data structures are used for file and database design
- Normalization eliminate unnecessary dependency relationship within a database file

Forms of normalization/Normalization rules

First normal form (1NF)

Refers to the first step where preliminary data structures are transforming into the first normal form by eliminating any repeating sets of data elements. A relation table is said to be on the first normal form, if and only if it contains no repeating groups that is it has no repeated value for a particular attribute with a single record. Any repeated group of attribute is isolated to form a new relation. In other words first normal form (1nf) means that a table has no multiple value attribute or composite attribute, In the 1nf, each column holds one attribute and each row holds a single occurrence of the entity

Second normal form (2NF)

2nf concentrated on records with concatenated keys, they check the non key attribute for dependency on the entire key, and any data element that dependent only on part of the key is moved to a new entity

Third normal form (3NF)

All data element in the third normal form must be a function of the key. To reach the 3nf, you need to review the structure's non-key data elements and identify any data element dependent on an attribute other than the key, if there is all these data elements should be moved to a new entity

Fourth normal form (4NF)

In data normalization, the fourth normal form deals with data element with issues of multi-value dependency (when one attributes determine another attribute sets). A relation is said to be in the 4nf formal form if and if only all existing multi-value dependency is converted into functional dependency

Fifth normal form (5NF)

Here is where the join dependency is removed, the 5nf is also known as the projection join normal form(PJNF), and refers to the separation of one relation into any sub-relations or having sub-relations into one relation and can produce join dependencies

Performing Normalization by example

While designing a database out of an entity–relationship model, the main problem existing in that "raw" database is redundancy. Redundancy is storing the same data item in more one place. A redundancy creates several problems like the following:

1. Extra storage space: storing the same data in many places takes large amount of disk space.
2. Entering same data more than once during data insertion.
3. Deleting data from more than one place during deletion.
4. Modifying data in more than one place.
5. Anomalies may occur in the database if insertion, deletion, modification etc are no done properly. It creates inconsistency and unreliability in the database.

To solve this problem, the "raw" database needs to be normalized. This is a step by step process of removing different kinds of redundancy and anomaly at each step. At each step a specific rule is followed to remove specific kind of impurity in order to give the database a slim and clean look.

Un-Normalized Form (UNF)

If a table contains non-atomic values at each row, it is said to be in UNF. An atomic value is something that can not be further decomposed. A non-atomic value, as the name suggests, can be further decomposed and simplified. Consider the following table:

Emp-Id	Emp-Name	Month	Sales	Bank-Id	Bank-Name
E01	AA	Jan	1000	B01	SBI
		Feb	1200		
		Mar	850		
E02	BB	Jan	2200	B02	UTI
		Feb	2500		
E03	CC	Jan	1700	B01	SBI
		Feb	1800		
		Mar	1850		
		Apr	1725		