

Normalization Overview

by Sophia



WHAT'S COVERED

This tutorial explores the importance of normalization in two parts:

1. Normal Forms and Normalization
2. Movie Rating Example

1. Normal Forms and Normalization

Normalization is something that was brought up a few times in prior tutorials, but it is time to dive into the details about this important process in database design. Normalization helps database designers evaluate and correct issues with table structures to minimize the amount of data redundancy and reduce storage needs. It also reduces the chance of potential data anomalies. The normalization process works through a series of stages, or steps, called **normal forms**.

The first three normal forms are first normal form (1NF), second normal form (2NF), and third normal form (3NF). From a table structure perspective, second normal form (2NF) is better than first normal form (1NF), and third normal form (3NF) is better than second normal form (2NF). Although there are higher normal forms, third normal form (3NF) is typically the highest level necessary for most applications.



KEY CONCEPT

Why not go higher? The higher the normal form, the more join operations are necessary, since with each normal form, more tables are added to reduce redundancy. As important as reducing data redundancy is, it is only important to a point. You must also ensure that the database offers fast performance. There may even be situations where a database needs to be denormalized (going down in normal form and adding data redundancy) to help improve performance.

Normalization is typically performed between conceptual and logical data modeling. Once you have the basic conceptual model created, you can then go through each stage of normalization and determine if the table structure can be improved. It's important to identify the business rules of the organization and the data constraints when performing normalization. You must also identify any functional dependencies, entities, relationships, and any multi-valued attributes. Then you can use normalization to validate and refine the data model for the database.



TERM TO KNOW

Normal Forms

The series of stages, or steps that the normalization process works through with each stage reducing redundancy. The first three normal forms are first normal form (1NF), second normal form (2NF), and third normal form (3NF).

2. Movie Rating Example

Let's go back to our movie rating database. If it had been set up in a spreadsheet, it would look something like this:

User	MovieTitle	NumericRating	TextualRating	Actor	Genre	ReleaseDate
Andy Joe	Toy Story	5	Such a classic!	Tom Hanks, Tim Allen	Animation, Adventure, Comedy	1995
Andy Joe	Titanic	4	Great movie!	Leonardo DiCaprio, Kate Winslet, Billy Zane, Kathy Bates	Drama	1997
Sophia McKenzie	Terminator 2: Judgement Day	5	Loved this movie, definitely a must watch!	Arnold Schwarzenegger, Linda Hamilton, Edward Furlong	Action, Science Fiction	1991
James Wang	Titanic	5	Best movie!	Leonardo DiCaprio, Kate Winslet, Billy Zane, Kathy Bates	Romance	1997
Barton Rafter	Titanic	3	It was OK, could be shorter.	Leonardo DiCaprio, Kate Winslet, Billy Zane, Kathy Bates	Drama, Romance	1997
Barton Rafter	Toy Story	4.5	Good movie! Watched a few times.	Tom Hanks, Tim Allen, Don Rickles, Jim Varney, Annie Potts, John Morris	Animation, Adventure, Comedy	1995

You can see some issues here. The actors listed for the movie *Toy Story* seem to be different, since they were entered in by different users. The genre has the same issue. Notice that the movie *Titanic* is listed as a drama by one user, as a romance by another user, and as a drama and a romance by a third user. Throughout the upcoming tutorials, we will walk through the normalization process to resolve these issues and define the final normalized database structure.



SUMMARY

In this tutorial, you learned that **normalization** is a process performed in stages by a database designer to reduce data anomalies and data redundancy in tables. These stages are called **normal forms**. You learned that utilizing higher normal forms can impact performance which is something to be considered. You also saw some poor examples of normalization using the **movie rating example**.

Source: Authored by Vincent Tran



TERMS TO KNOW

Normal Forms

The series of stages, or steps that the normalization process works through with each stage reducing redundancy. The first three normal forms are first normal form (1NF), second normal form (2NF), and third normal form (3NF).