

Higher Normalization

by Sophia

WHAT'S COVERED

This tutorial explores the higher levels of normalization: Boyce-Codd normal form, fourth normal form (4NF), and fifth normal form (5NF). It does so in four parts:

- 1. Introduction
- 2. Boyce-Codd Normal Form
- 3. Fourth Normal Form
- 4. Fifth Normal Form

1. Introduction

With transactional databases, the third normal form (3NF) is generally the furthest you would take normalization. It does not make sense to go beyond it for most situations. There are certain cases, though, where it could potentially be useful.

2. Boyce-Codd Normal Form

Boyce-Codd normal form, also known as BCNF, is viewed as 3.5 NF. BCNF requires that the database design first fulfills the requirements of the third normal form (3NF), but also has every determinant in a table as a candidate key. A candidate key has the same characteristics as a primary key, but for some reason it may not have been chosen as the primary key. If a table only contains one candidate key, then 3NF and BCNF are equivalent to one another. It is only when a table contains more than one candidate key that you would need to address it.

E TERM TO KNOW

Boyce-Codd Normal Form

The Boyce-Codd normal form, also known as BCNF, is viewed as 3.5 NF. BCNF requires that the database design first fulfills the requirements of the third normal form (3NF), but also has every determinant in a table as a candidate key.

3. Fourth Normal Form

With the fourth normal form (4NF), the database design will satisfy all of the properties of the 3NF and BCNF,

and additionally should not have any multivalued dependencies. A multivalued dependency is when a single primary key can determine multiple values of two other attributes and those attributes are independent of one another. To eliminate multivalued dependencies, you have to create new tables for the components of the multivalued dependencies. By doing so, no row should contain two or more multivalued facts about an entity.

E TERM TO KNOW

Fourth Normal Form (4NF)

In the fourth normal form (4NF), the database design will satisfy all of the properties of the third normal form (3NF) and Boyce-Codd normal form (BCNF), and additionally should not have any multivalued dependencies.

4. Fifth Normal Form

Fifth normal form (5NF) is mostly conceptual. A relation between tables is in 5NF if it is in the fourth normal form (4NF) and does not contain any join dependencies. The joining should be completely lossless. 5NF breaks down tables into as many tables as possible to avoid redundancy. It is also known as the project-join normal form. In our database in PostgreSQL, it would be similar to taking the track table and splitting up each of the combined items:

| track | |
|---------------|---------------|
| milliseconds | INTEGER |
| composer | VARCHAR (220) |
| genre_id | INTEGER |
| media_type_id | INTEGER |
| album_id | INTEGER |
| name | VARCHAR (200) |
| track_id | INTEGER |
| unit_price | NUMERIC |
| bytes | INTEGER |

You would have a separate table for track_id and genre_id, track_id and album_id, track_id and media_type_id, genre_id, and album_id, and lastly media_type_id and album_id. This way, every combination would be set up as its own table to avoid redundancy. However, you can easily see why this would be impractical in a real-world setting.

TERM TO KNOW

Fifth Normal Form (5NF)

The fifth normal form (5NF) is mostly conceptual. A relation between tables is in 5NF if it is in the fourth normal form (4NF) and does not contain any join dependencies.

🖯 SUMMARY

In this tutorial, in the **introduction** you learned that the third normal form (3NF) is generally the furthest you would take normalization, however, there are some exceptions. The **Boyce-Codd normal form**, also known as BCNF, is viewed as 3.5 NF. This database design first fulfills the requirements of the third normal form (3NF) but also has every determinant in a table as a candidate key. You learned that the **fourth normal form** (4NF) adds the concept that there should not be any multivalued

dependencies. And the **fifth normal form** (5NF) is mostly conceptual and does not contain any join dependencies. Remember, higher levels of normalization like Boyce-Codd normal form, fourth normal form (4NF), and fifth normal form (5NF) are generally not used in real-world databases.

Source: Authored by Vincent Tran

TERMS TO KNOW

Boyce-Codd Normal Form

The Boyce-Codd normal form, also known as BCNF, is viewed as 3.5 NF. BCNF requires that the database design first fulfills the requirements of the third normal form (3NF), but also has every determinant in a table as a candidate key.

Fifth Normal Form (5NF)

The fifth normal form (5NF) is mostly conceptual. A relation between tables is in 5NF if it is in the fourth normal form (4NF) and does not contain any join dependencies.

Fourth Normal Form (4NF)

In the fourth normal form (4NF), the database design will satisfy all of the properties of the third normal form (3NF) and Boyce-Codd normal form (BCNF), and additionally should not have any multivalued dependencies.