**Database Normalization**

* It is a database design principle for organizing data in an organized and consistent way.
* It helps you avoid redundancy and maintain the integrity of the database.
* It also helps you eliminate undesirable characteristics associated with insertion, deletion, and updating.

**Purpose of Normalization**

* The main purpose of database normalization is to avoid complexities, eliminate duplicates, and organize data in a consistent way.
* In normalization, the data is divided into several tables linked together with relationships.
* Database administrators are able to achieve these relationships by using primary keys, foreign keys, and composite keys.
* To get it done, a primary key in one table, for example, employee\_wages is related to the value from another table, for instance, employee\_data.

**N.B.**: **A primary key** is a column that uniquely identifies the rows of data in that table. It’s a unique identifier such as an employee ID, student ID, voter’s identification number (VIN), and so on.

**A foreign key** is a field that relates to the primary key in another table.

**A composite key** is just like a primary key, but instead of having a column, it has multiple columns.

**TYPES**

* 1NF, 2NF, and 3NF are the first three types of database normalization.
* They stand for **first normal form**, **second normal form**, and **third normal form**, respectively.
* There are also 4NF (fourth normal form) and 5NF (fifth normal form). There’s even 6NF (sixth normal form), but the commonest normal form you’ll see out there is 3NF (third normal form).
* All the types of database normalization are cumulative – meaning each one builds on top of those beneath it. So all the concepts in 1NF also carry over to 2NF, and so on.

**The First Normal Form – 1NF**

For a table to be in the first normal form, it must meet the following criteria:

* a single cell must not hold more than one value (atomicity)
* there must be a primary key for identification
* no duplicated rows or columns
* each column must have only one value for each row in the table

**The Second Normal Form – 2NF**

The 1NF only eliminates repeating groups, not redundancy. That’s why there is 2NF.

A table is said to be in 2NF if it meets the following criteria:

* it’s already in 1NF
* has no partial dependency. That is, all non-key attributes are fully dependent on a primary key.

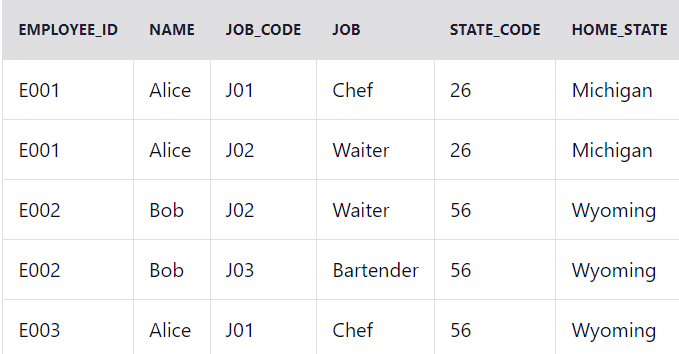
**The Third Normal Form – 3NF**

When a table is in 2NF, it eliminates repeating groups and redundancy, but it does not eliminate transitive partial dependency. This means a non-prime attribute (an attribute that is not part of the candidate’s key) is dependent on another non-prime attribute. This is what the third normal form (3NF) eliminates.So, for a table to be in 3NF, it must:

* be in 2NF
* have no transitive partial dependency.

## Examples

Database normalization is quite technical, but we will illustrate each of the normal forms with examples.

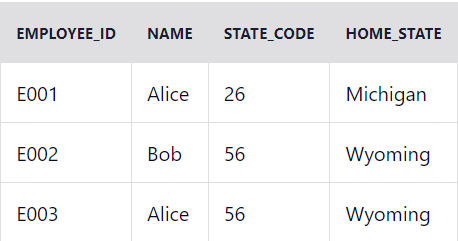
Imagine we're building a restaurant management application. That application needs to store data about the company's employees and it starts out by creating the following table of employees:

All the entries are atomic and there is a composite primary key (employee\_id, job\_code) so the table is in the **first normal form (1NF)**.

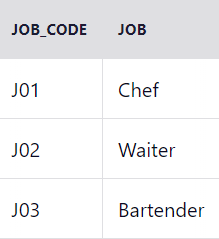
But even if you only know someone's employee\_id, then you can determine their name, home\_state, and state\_code (because they should be the same person). This means name, home\_state, and state\_code are dependent on employee\_id (a part of primary composite key). So, the table is not in **2NF**. We should separate them to a different table to make it 2NF.

### Example of Second Normal Form (2NF)

#### employee\_roles Table employees Table

#### **jobs** table

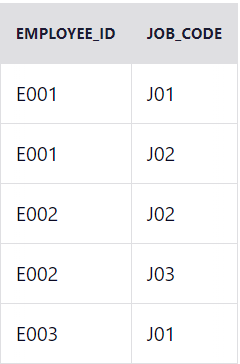
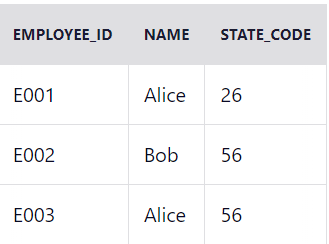


home\_state is now dependent on state\_code. So, if you know the state\_code, then you can find the home\_state value.

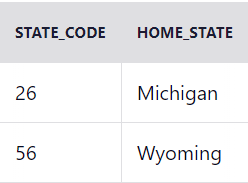
To take this a step further, we should separate them again to a different table to make it 3NF.

### Example of Third Normal Form (3NF)

#### **employee\_roles** Table **employees** Table

Jos table States table

Now our database is in 3NF.