The Normalisation Process
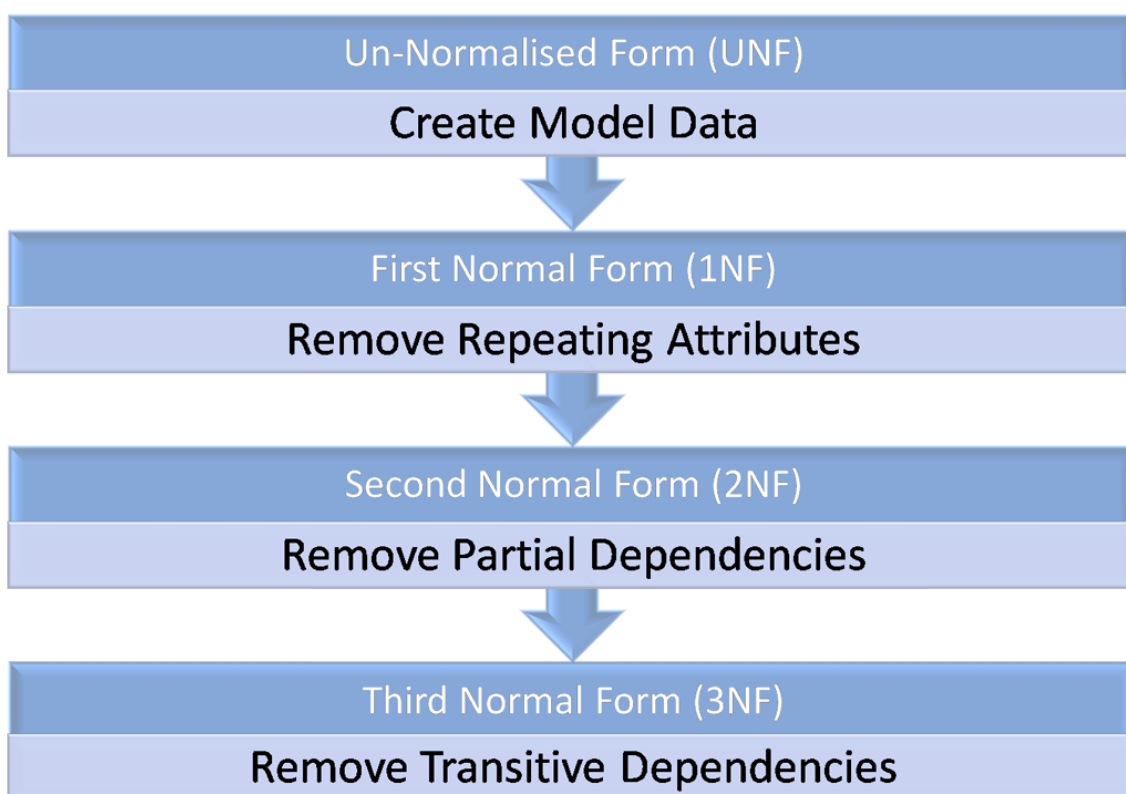
Normalisation was developed by Dr. E.F.Codd in 1972 as part of the Relational Database Theory as a means of breaking data into its related groups and defining the relationships between those groups. It is said the name Normalisation was initially a political gag taken from President Nixon and his initiative for 'Normalising' relations with China. Codd figured if you can Normalise relations with a country you should be able to normalise the relations with data as well.

Normalisation is a specific relational database analysis and design technique used to model groups of related data within an organisation. Its purpose is to ensure data stored within the database adheres to best practices by following a set of rules with the purpose of eliminating redundancies and optimising the process of information retrieval. Normalisation leaves us with a structure that groups like data into relational models referenced by keys and linked to other relational models to form a relational database schema.

Normalisation is represented by a logical set of steps that follow simple rules that are applied to each stage of the modelling process. At the highest level the stages are separated into something called Normal Forms, identified by a particular named process.

Un-Normalised Form (UNF)
Create Model Data

First Normal Form (1NF)
Remove Repeating Attributes

Second Normal Form (2NF)
Remove Partial Dependencies

Third Normal Form (3NF)
Remove Transitive Dependencies

Initially there were only three normal forms, First Normal Form (1NF), Second Normal Form (2NF) and Third Normal Form (3NF), but over time three more were added. In

general terms the first three are more commonly used in database modelling. The additional three are identification of potential redundancies that could be considered but however when applied practically can lead to inefficiencies in performance and tend to be used under special circumstances or for consideration with complex data structures.

In addition we have something called Un-Normalised Form (UNF), though not generally considered as part of the normalisation rules, is representative of the very first stages of the normalisation process.

We can identify each of the normal forms as follows and will define each in detail thereafter:

Un-Normalised Form (UNF) – Data Modelling
First Normal Form (1NF) – Repeating Groups
Second Normal Form (2NF) – Partial Dependencies
Third Normal Form (3NF) – Transitive Dependencies