

NORMALIZATION

By **MATTIE MACGREGOR**

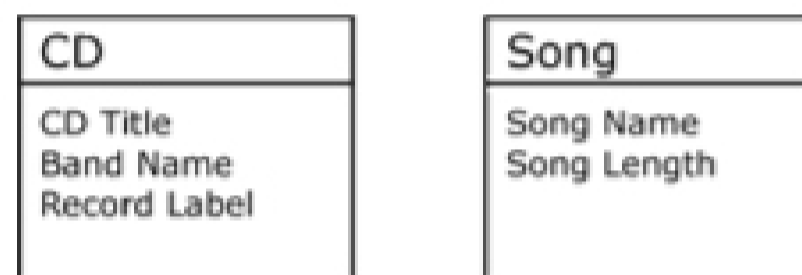
The aim of **normalization** is to remove certain kinds of data redundancy, and therefore avoid update anomalies.

Update anomalies are difficulties with the insert, update and delete operations on a database due to the data structure.

FIRST NORMAL FORM (1NF)

An entity is said to be in **first normal form (1NF)** when all attributes are single-valued.

To apply the first normal form to an entity, verify that each attribute in the entity has a single value for each instance of the entity. If any attribute has repeating values, it is not in 1NF.



The data model with two entities

UNIQUE IDENTIFIERS

Each entity must have a **unique identifier**. An identifier is an attribute of an entity that meets the following rules:

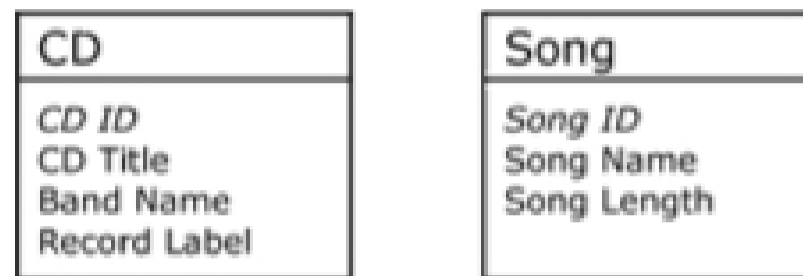
- It is unique across all instances of the entity.
- It has a non-NULL value for each instance of the entity for the lifetime of the instance.
- It has a value that never changes for the lifetime of the instance.

The unique identifier is important because it designates which instance of an entity is being dealt with. Identifier selection is critical because it is also used to model relationships. If an identifier for an entity doesn't meet one of the above rules, it could affect the whole data model.

Designers often invent an identifying attribute that has no other meaning except to serve as an identifying attribute. Because this attribute is invented and is completely unrelated

to the entity, the designer has full control over it and a guarantee that it meets the rules of unique identifiers.

A unique identifier is diagrammed as an underlined or italicized attribute.



The entities with unique identifiers

RELATIONSHIPS

Entity identifiers enable their relationships to be modeled.

A **relationship** describes a binary association between two entities. A relationship may also exist between an entity and itself. Such a relationship is a **recursive relationship**. Each entity within a relationship describes and is described by the other.

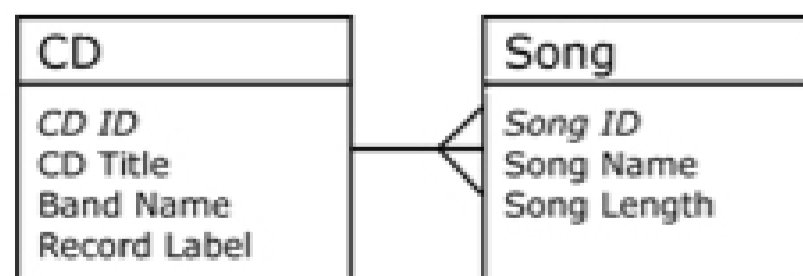
Each side of the relationship has two components: a name and a degree. Each side of the relationship has a name that describes the relationship. Degree, also referred to as cardinality, states how many instances of the describing entity must describe one instance of the described entity.

Degree is expressed using two different values: "one and only one" (1) and "one or many" (M).

Sometimes it is helpful to express a relationship verbally. One way of doing so is to plug the various components of a direction of the relationship into the formula:

Entity_A has [one and only one | one or many] Entity_B

In data models, a relationship can be shown by drawing a line between the two entities. Degree is expressed with a straight line for "one and only one" relationships or crows feet for "one or many" relationships.



A relationship between two entities

SECOND NORMAL FORM (2NF)

An entity is said to be in the **second normal form (2NF)** if it is already in 1NF and all non-identifying attributes are dependent on the entity's entire unique identifier.

If any attribute is not dependent entirely upon the entity's unique identifier, that attribute has been misplaced and must be removed.

Normalize these attributes either by finding the entity where it belongs or by creating an additional entity where the attribute should be placed.



The data model with a new entity

TYPES OF RELATIONSHIPS

There are three main kinds of relationships, and it is important to determine both directions of the relationship when modeling the relationships between entities:

- **1-to-1:** Both sides of the relationship have a degree of one and only one. (1-to-1 relationships are rare.)
- **1-to-M:** One of the sides has a degree of one or many and the other side has a degree of one and only one.
- **M-to-M:** Both sides of the relationship are one or many relationships.

REFINING RELATIONSHIPS

A 1-to-1 relationship may imply that two entities are the same entity. If they are, fold them into a single entity.

There is often data to be captured about M-to-M relationships.



Entity relationships in the data model

The data model below is another representation of this M-to-M relationship.