

Normalization

Objectives

- Present the idea of functional dependence.
- Define first normal form (1NF), second normal form (2NF), and third normal form (3NF).
- Describe the problems associated with tables (relations) that are not in 1NF, 2NF, or 3NF, along with the mechanism for converting to all three.

The Normalization Process

- A set of steps that enables you to identify the existence of potential problems, called anomalies, in the design of a relational database.
- The goal of normal forms is to allow you to take a table or collection of tables and produce a new collection of tables that represents the same information but is free of problems.
- At the end of the normalization process, you have a set of data groups which may become entities, each with attributes and UIDs, and with foreign keys also defined.
- Can use these to create an ERD
- Overall aim is a comprehensive & accurate understanding of the data requirements that support the desired functionality
- It is critical to have a thorough understanding of the data

Ultimately:

- All attributes should be dependent on the key, the whole key, and nothing but the key.

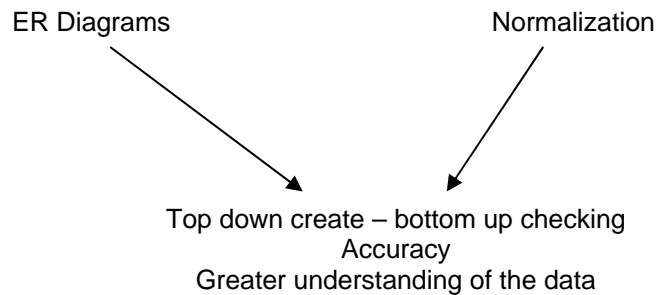
ERDs vs Normalization

ER Diagramming

- ♦ Top down approach
- ♦ Fast
- ♦ Examine requirements
- ♦ Business knowledge

Normalization

- ♦ Bottom up approach
- ♦ Very slow
- ♦ Examine existing data
- ♦ Mathematically based



Functional Dependencies

A column (attribute), B, is functionally dependent on another column, A (or possibly a collection of columns), if a value for A determines a single value for B at any one time.

Given A, a single value for B can be determined

- ♦ Functional Dependency:
 - Is a relationship between or among attributes
 - Given the value of one, we can obtain the other
- ♦ Equations can represent functional dependencies

$\text{TotalPrice} := \text{ItemPrice} * \text{Quantity};$

TotalPrice is functionally dependent on ItemPrice & Quantity

- Computers in a computer lab. Each computer has a serial number and a processor.

ComputerSerialNumber \rightarrow processor

- ♦ Assume students have a unique SID and each student has one, and only one major.

SID \rightarrow major

- Read the above as: “SID functionally determines major”, “SID determines major”, or “Major is dependent on SID.”
- SID is called *the determinant*.
- A particular value of SID is paired with only 1 major, but a major can be paired with one or more different values of SID.

- Can also involve groups of attributes;

(SID, ClassName) \rightarrow Grade

- ♦ Returning to the Definition: A column (attribute), B, is functionally dependent on another column, A (or possibly a collection of columns), if a value for A determines a single value for B at any one time.

A \rightarrow B

But does A give you a single value for B?

- Check business rules.

Keys

- ♦ Column (attribute) C (or a collection of columns) is the primary key for a table (relation), T, if:
 - All columns in T are functionally dependent on C
 - No subcollection of the columns in C (assuming C is a collection of columns and not just a single column) also has Property 1.

A key:

- Must have a unique value within the thing being normalized
- Should use as few data items as possible
- Should be non-textual if possible
- Should be the best candidate based on context
- Should not be invented if a practical alternative exists

Candidate key

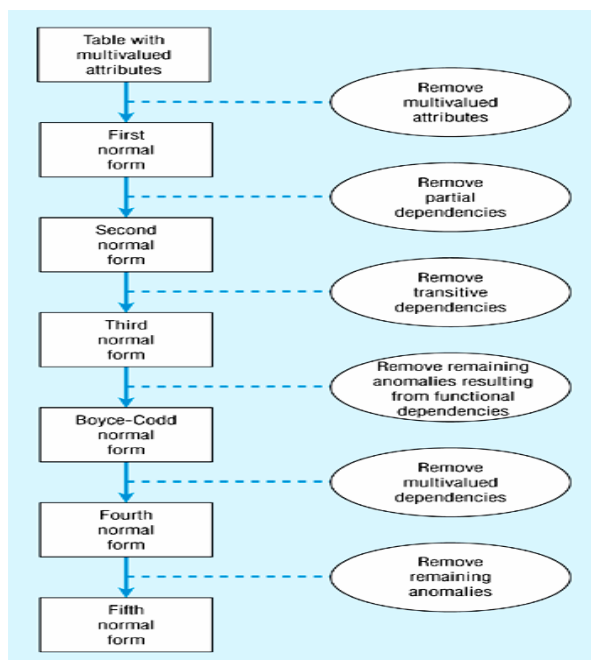
- Column(s) on which all other columns in table are functionally dependent
- There might be more than one possibility for the primary key (e.g. Employee number and SIN)
- The possibilities are referred to as candidate keys
- Candidate keys not chosen to be the primary key are often referred to as alternate keys

Other “key” terminology:

- Composite key
 - If have more than one column making up the primary key.
- Surrogate key
 - If use a non data item (e.g. a generated sequence number) as the primary key

NORMAL FORMS

Steps in Normalization



0 Normal Form

- ♦ First collect and list the raw data.
 - Then select a key

First Normal Form

- ♦ No multi-valued attributes.
- ♦ Every attribute value is atomic.
- ♦ A relation is said to be in 1NF if
 - cells are single valued
 - all entries in any column are of the same kind
 - each column has a unique name, but the order of the columns is not important
 - no two rows are identical, but the order of the rows is not important
- In general, converting a non-1NF table to 1NF, the primary key will typically include the original primary key concatenated with the primary key of the repeating group.

Second Normal Form

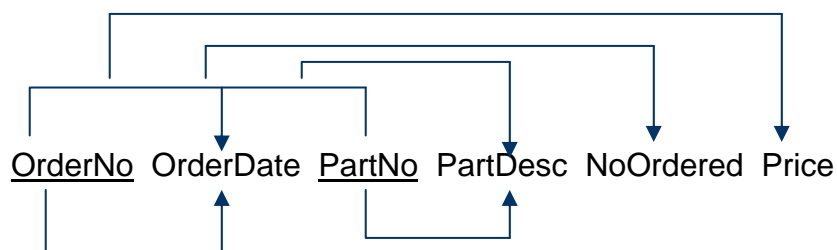
- ♦ 1NF and every non-key attribute is fully functionally dependent on the primary key.
- ♦ Every non-key attribute must be defined by the entire key, not by only part of the key.
- ♦ No partial functional dependencies.
- ♦ If the primary key of a table contains only a single column, the table is automatically in second normal form.

Modification Anomalies

- Deletion anomaly
 - Deleting one fact about an entity deletes a fact about another entity
- Insertion anomaly
 - Cannot insert one fact about an entity unless a fact about another entity is also added
- Update anomaly
 - Changing one fact about an entity requires multiple changes to a table

Partial Functional Dependencies

- Partial functional dependencies = where one or more nonkey attributes are functionally dependent on part (but not all) of the primary key.



Second normal form problems

- 2NF Tables may still contain problems
 - Redundancy and wasted space
 - Update Anomalies
 - Update, inconsistent data, additions, deletions
 - Occur because a column is dependent on a portion of a multi-column primary key

Third Normal Form

- 2NF and no transitive dependencies (functional dependency between non-key attributes.)
- Similar modification anomalies: Updates, inconsistent data, additions, deletions.

Transitive Dependencies

| SID | BLDG | FEE |
|-----|-----------|------|
| 100 | Randolph | 1200 |
| 150 | Ingersoll | 1100 |
| 200 | Randolph | 1200 |
| 250 | Pilkin | 1100 |
| 300 | Randolph | 1200 |

Normalization Example

| | | |
|---|-----------------------|----------------------------|
| Customer No: 327 | Order No: C662 | Order Date: 28/1/02 |
| Customer Name: I. Bea Normal | | |
| Customer Address: 555 Some St.. Victoria | | |
| Product Number | | Quantitv |
| XI298 | | 5 |
| C337 | | 5 |
| C338 | | 1 |
| A614 | | 2 |