

Normalization – Why it is important?

BY

Vijay Anand Madhuranayagam



Content

- Normalization
- Why it is important?
- Stages/Levels: 1NF, 2NF, 3NF & BCNF
Normal Forms

Normalization

- Normalization is the process of efficiently organizing data in a database with two goals in mind
- First goal: Eliminate redundant data
 - for example, storing the same data in more than one table
- Second Goal: Ensure data dependencies make sense
 - for example, only storing related data in a table

Why it is important?

- Less storage space
- Quicker updates
- Less data inconsistency
- Clearer data relationships
- Easier to add data
- Flexible Structure
- Bad database design results in:
 - redundancy: inefficient storage.
 - anomalies: data inconsistency, difficulties in maintenance

Stages/Levels (Normal Forms)

- 1NF – First Normal Form
- 2NF – Second Normal Form
- 3NF – Third Normal Form
- BCNF – Boyce-Codd Normal Form

These are some of the important Normal Forms

First Normal Form (1NF)

- Eliminate duplicate values (columns) from the same table
- Create separate tables for each group of related data and identify each row with a unique column or set of columns (Primary key)

First Normal Form - 1NF

(Example)

Un-normalized Students table:

<u>Student#</u>	AdvID	AdvName	AdvRoom	Class1	Class2
123	123A	James	555	102-8	104-9
124	123B	Smith	467	209-0	102-8

Normalized Students table:

<u>Student#</u>	AdvID	AdvName	AdvRoom	Class#
123	123A	James	555	102-8
123	123A	James	555	104-9
124	123B	Smith	467	209-0
124	123B	Smith	467	102-8

Second Normal Form (2NF)

- A relation R is in 2NF if
 - (a) R is 1NF , and
 - (b) all non-prime attributes are fully dependent on the candidate keys. Which is creating relationships between these new tables and their predecessors through the use of foreign keys.
- A prime attribute appears in a candidate key.
- There is no partial dependency in 2NF.

Second Normal Form (2NF) Example

Students table:

<u>Student#</u>	AdvID	AdvName	AdvRoom
123	123A	James	555
124	123B	Smith	467

Registration table:

<u>Student#</u>	Class#
123	102-8
123	104-9
124	209-0
124	102-8

Third Normal Form (3NF)

- Meet all the requirements of the second normal form.
- Remove columns that are not dependent upon the primary key.

Third Normal Form - 3NF

(Example)

Students table:

<u>Student#</u>	AdvID	AdvName	AdvRoom
123	123A	James	555
124	123B	Smith	467

Student table:

<u>Student#</u>	<u>AdvID</u>
123	123A
124	123B

Advisor table:

<u>AdvID</u>	AdvName	AdvRoom
123A	James	555
123B	Smith	467

Example with first three forms

Suppose we have this Invoice Table

Invoice Table				Violate's Normalization Form 1								
Invoice#	Customer Information			Quant1	Part1	Amt1	Quant2	Part2	Amt2	Quant3	Part3	Amt3
	Cust#	Name	Addr									
1001	43	Jones	121 1st	200	Screw	2.00	300	Nut	2.25	100	Washr	0.75
1002	55	Smith	222 2nd	1	Motor	52.00	5	Brace	44.44			
1003	43	Jones	121 1st	10	Saw	121.00						

First Normal Form: No repeating groups.

- The above table violates 1NF because it has columns for the first, second, and third line item.

- Solution: you make a separate line item table, with it's own key, in this case the combination of invoice number and line number

Table now in 1NF

Complies with Normalization Form 1, Violate's Normalization Form 2

Line item table

Invoice table

Invoice#	Customer Information							
	Invoice#	Line#	Cust#	Name	Address	Quant1	Part1	Amt1
1001	1001	1	43	Jones	121 1st	200	Screw	2.00
1002	1001	2	43	Jones	121 1st	300	Nut	2.25
1003	1001	3	43	Jones	121 1st	100	Washr	0.75
	1002	1	55	Smith	222 2nd	1	Motor	52.00
	1002	2	55	Smith	222 2nd	10	Saw	121.00
	1003	1	43	Jones	121 1st	5	Brace	44.44

Second Normal Form:
Each column must depend on the *entire* primary key.

Complies with Normalization Form 2, Violate's Normalization Form 3

Invoice table

Invoice#	Customer Information		
	Cust#	Name	Address
1001	43	Jones	121 1st
1002	55	Smith	222 2nd
1003	43	Jones	121 1st

Line item table

Invoice#	Line#	Quant1	Part1	Amt1
1001	1	200	Screw	2.00
1001	2	300	Nut	2.25
1001	3	100	Washr	0.75
1002	1	1	Motor	52.00
1002	2	10	Saw	121.00
1003	1	5	Brace	44.44

Third Normal Form:

Each column must depend on ***directly*** on the primary key.

Complies with Normalization Form 3

Invoice table

Invoice#	Cust#
1001	43
1002	55
1003	43

Customer table

Cust#	Name	Address
43	Jones	121 1st
55	Smith	222 2nd

Line item table

Invoice#	Line#	Quant1	Part1	Amt1
1001	1	200	Screw	2.00
1001	2	300	Nut	2.25
1001	3	100	Washr	0.75
1002	1	1	Motor	52.00
1002	2	10	Saw	121.00
1003	1	5	Brace	44.44