

Normalization: Friend or Foe

Beth Tucker Long

Who am I?

- Beth Tucker Long (@e3beth)
- Editor-in-Chief - php[architect] magazine
- Freelancer under Treeline Design, LLC
- Stay-at-home-mom



- User group organizer – Madison PHP

Disclaimer

The databases on these slides are not necessarily showing good database design. They are contrived examples meant to teach the definitions of the normal forms.

Database Normalization

A way to organize and structure **relational** databases.

Why?

- Reduce Redundancy
- Decrease Dependency
- Simplify Data Modification

A Brief History

Edgar F. Codd started what we know today as normalization.

1970 – Codd introduced the First Normal Form.

1971 – Codd introduced the Second and Third Normal Forms.

There are more forms, but generally, meeting the Third Normal Form is enough to be called "normalized".

First Normal Form

Briefly:

1. Each record has the same number of fields.
2. No duplicate records.
3. Each set of data must have a unique identifier.
4. Only one value per field and those values must be consistent.

Sample Database Table

lineName	transportationType	district	districtOffice
Red Line	Bus	1, 2	Main St., First Ave.
Yellow Line	Train	2, 3, 4, 5	First Ave., Quivey Rd., Pierce Ct., Second St.
Yellow Line	Bus	2, 3, 4, 5	First Ave., Quivey Rd., Pierce Ct., Second St.
Blue Line	Train	5, 6	Second St., Juniper Dr.

districtManager	Ext
Adams, Tully	325, 546
Tully, Sandstone, Emerson, Gladstone	546, 254, 643, 756
Tully, Sandstone, Emerson, Gladstone	546, 254, 643, 756
Gladstone, Smithwick	756, 934

Step 1

1. Each record has the same number of fields.
2. No duplicate records.
3. Each set of data must have a unique identifier.

Sample Database Table

lineName	transportationType	district	districtOffice
Red Line	Bus	1, 2	Main St., First Ave.
Yellow Line	Train	2, 3, 4, 5	First Ave., Quivey Rd., Pierce Ct., Second St.
Yellow Line	Bus	2, 3, 4, 5	First Ave., Quivey Rd., Pierce Ct., Second St.
Blue Line	Train	5, 6	Second St., Juniper Dr.

districtManager	Ext
Adams, Tully	325, 546
Tully, Sandstone, Emerson, Gladstone	546, 254, 643, 756
Tully, Sandstone, Emerson, Gladstone	546, 254, 643, 756
Gladstone, Smithwick	756, 934

Sample Database Table

lineName	transportationType	district	districtOffice
Red Line	Bus	1, 2	Main St., First Ave.
 Yellow Line	Train, Bus	2, 3, 4, 5	First Ave., Quivey Rd., Pierce Ct., Second St.
Blue Line	Train	5, 6	Second St., Juniper Dr.

districtManager	Ext
Adams, Tully	325, 546
Tully, Sandstone, Emerson, Gladstone	546, 254, 643, 756
Gladstone, Smithwick	756, 934

Sample Database Table



ID	lineName	transportationType	district
1	Red Line	Bus	1, 2
2	Yellow Line	Train	2, 3, 4, 5
3	Yellow Line	Bus	2, 3, 4, 5
4	Blue Line	Train	5, 6

districtOffice	districtManager	Ext
Main St., First Ave.	Adams, Tully	325, 546
First Ave., Quivey Rd., Pierce Ct., Second St.	Tully, Sandstone, Emerson, Gladstone	546, 254, 643, 756
First Ave., Quivey Rd., Pierce Ct., Second St.	Tully, Sandstone, Emerson, Gladstone	546, 254, 643, 756
Second St., Juniper Dr.	Gladstone, Smithwick	756, 934

Lastly

Only one value per field and those values must be consistent.

Sample Database Table

ID	lineName	transportationType	district
1	Red Line	Bus	1, 2
2	Yellow Line	Train	2, 3, 4, 5
3	Yellow Line	Bus	2, 3, 4, 5
4	Blue Line	Train	5, 6

districtOffice	districtManager	Ext
Main St., First Ave.	Adams, Tully	325, 546
First Ave., Quivey Rd., Pierce Ct., Second St.	Tully, Sandstone, Emerson, Gladstone	546, 254, 643, 756
First Ave., Quivey Rd., Pierce Ct., Second St.	Tully, Sandstone, Emerson, Gladstone	546, 254, 643, 756
Second St., Juniper Dr.	Gladstone, Smithwick	756, 934

Sample Database Table

ID	lineName	transportation Type	district	districtOffice	districtManager	Ext
1	Red Line	Bus	1	Main St.	Adams	325
2	Red Line	Bus	2	First Ave.	Tully	546
3	Yellow Line	Bus	2	First Ave.	Tully	546
4	Yellow Line	Bus	3	Quivey Rd.	Sandstone	254
5	Yellow Line	Bus	4	Pierce Ct.	Emerson	643
6	Yellow Line	Bus	5	Second St.	Gladstone	756
7	Yellow Line	Train	2	First Ave.	Tully	546
8	Yellow Line	Train	3	Quivey Rd.	Sandstone	254
9	Yellow Line	Train	4	Pierce Ct.	Emerson	643
10	Yellow Line	Train	5	Second St.	Gladstone	756
11	Blue Line	Train	5	Second St.	Gladstone	756
12	Blue Line	Train	6	Smithwick	Smithwick	934

Second Normal Form

Briefly:

1. Must meet First Normal Form requirements.
2. Everything must relate directly to the "key" or main value.

Sample Database Table

ID	lineName	transportation Type	district	districtOffice	districtManager	Ext
1	Red Line	Bus	1	Main St.	Adams	325
2	Red Line	Bus	2	First Ave.	Tully	546
3	Yellow Line	Bus	2	First Ave.	Tully	546
4	Yellow Line	Bus	3	Quivey Rd.	Sandstone	254
5	Yellow Line	Bus	4	Pierce Ct.	Emerson	643
6	Yellow Line	Bus	5	Second St.	Gladstone	756
7	Yellow Line	Train	2	First Ave.	Tully	546
8	Yellow Line	Train	3	Quivey Rd.	Sandstone	254
9	Yellow Line	Train	4	Pierce Ct.	Emerson	643
10	Yellow Line	Train	5	Second St.	Gladstone	756
11	Blue Line	Train	5	Second St.	Gladstone	756
12	Blue Line	Train	6	Smithwick	Smithwick	934

Sample Database Table

ID	lineName	transportationType	district
1	Red Line	Bus	1
2	Red Line	Bus	2
3	Yellow Line	Bus	2
4	Yellow Line	Bus	3
5	Yellow Line	Bus	4
6	Yellow Line	Bus	5
7	Yellow Line	Train	2
8	Yellow Line	Train	3
9	Yellow Line	Train	4
10	Yellow Line	Train	5
11	Blue Line	Train	5
12	Blue Line	Train	6

district	districtOffice	districtManager	Ext
1	Main St.	Adams	325
2	First Ave.	Tully	546
3	Quivey Rd.	Sandstone	254
4	Pierce Ct.	Emerson	643
5	Second St.	Gladstone	756
6	Smithwick	Smithwick	934

Third Normal Form

Briefly:

1. Must meet Second Normal Form requirements.
2. No transitive dependencies.

Sample Database Table

ID	lineName	transportationType	district
1	Red Line	Bus	1
2	Red Line	Bus	2
3	Yellow Line	Bus	2
4	Yellow Line	Bus	3
5	Yellow Line	Bus	4
6	Yellow Line	Bus	5
7	Yellow Line	Train	2
8	Yellow Line	Train	3
9	Yellow Line	Train	4
10	Yellow Line	Train	5
11	Blue Line	Train	5
12	Blue Line	Train	6

district	districtOffice	districtManager	Ext
1	Main St.	Adams	325
2	First Ave.	Tully	546
3	Quivey Rd.	Sandstone	254
4	Pierce Ct.	Emerson	643
5	Second St.	Gladstone	756
6	Smithwick	Smithwick	934

Sample Database Table

ID	lineName	transportationType	district
1	Red Line	Bus	1
2	Red Line	Bus	2
3	Yellow Line	Bus	2
4	Yellow Line	Bus	3
5	Yellow Line	Bus	4
6	Yellow Line	Bus	5
7	Yellow Line	Train	2
8	Yellow Line	Train	3
9	Yellow Line	Train	4
10	Yellow Line	Train	5
11	Blue Line	Train	5
12	Blue Line	Train	6

district	districtOffice
1	Main St.
2	First Ave.
3	Quivey Rd.
4	Pierce Ct.
5	Second St.
6	Smithwick

district	districtManager	Ext
1	Adams	325
2	Tully	546
3	Sandstone	254
4	Emerson	643
5	Gladstone	756
6	Smithwick	934

Another Example - Third Normal Form

1. Each record has the same number of fields.
2. No duplicate records.
3. Each set of data must have a unique identifier.
4. Only one value per field and those values must be consistent.
5. Everything must relate directly to the "key" or main value.
6. No transitive dependencies.

Sample Database Table

Name	Description	Serves	Serving Method
Fruit Salad	A side-dish made with various fruit pieces	4-5 people	Bowl
Cheese display	A variety of cheese	4-5 people	Plate
Havarti on Rye	A sandwich with cheese	1 person	Plate

Ingredients	Preparation Method	Preparation Description
Apples, Strawberries	Chopped, Wedges	Cut into small cubes, Cut into triangles
Brie, Havarti	Wedges, Sliced	Cut into triangles, Cut into slices
Rye Bread, Havarti	Sliced, Sliced	Cut into slices, Cut into slices

Normalizing to First Normal Form



ID	Name	Description	Serves	Serving Method
1	Fruit Salad	A side-dish made with various fruit pieces	4-5 people	Bowl
2	Cheese display	A variety of cheese	4-5 people	Plate
3	Havarti on Rye	A sandwich with cheese	1 person	Plate

Ingredients	Preparation Method	Preparation Description
Apples, Strawberries	Chopped, Wedges	Cut into small cubes, Cut into triangles
Brie, Havarti	Wedges, Sliced	Cut into triangles, Cut into slices
Rye Bread, Havarti	Sliced, Sliced	Cut into slices, Cut into slices

Normalizing to First Normal Form

ID	Name	Description	Serves	Serving Method
1	Fruit Salad	A side-dish made with various fruit pieces	4-5 people	Bowl
2	Cheese display	A variety of cheese	4-5 people	Plate
3	Havarti on Rye	A sandwich with cheese	1 person	Plate

Ingredient 1	Preparation Method 1	Preparation Description 1
Apples	Chopped	Cut into small cubes
Brie	Wedges	Cut into triangles
Rye Bread	Sliced	Cut into slices

Ingredient 2	Preparation Method 2	Preparation Description 2
Strawberries	Wedges	Cut into triangles
Havarti	Sliced	Cut into slices
Havarti	Sliced	Cut into slices



Normalized to Second Normal Form

ID	Name	Description	Serves	Serving Method
1	Fruit Salad	A side-dish made with various fruit pieces	4-5 people	Bowl
2	Cheese display	A variety of cheese	4-5 people	Plate
3	Havarti on Rye	A sandwich with cheese	1 person	Plate

ID	Name	Preparation Method	Preparation Description
1	Apples	Chopped	Cut into small cubes
2	Brie	Wedges	Cut into triangles
3	Rye Bread	Sliced	Cut into slices
4	Strawberries	Wedges	Cut into triangles
5	Havarti	Sliced	Cut into slices

ID	DishID	IngredientsID
1	1	1
2	1	4
3	2	2
4	2	5
5	3	3
6	3	5

Normalized to Third Normal Form

ID	Name	Description	Serves	ServingMethod
1	Fruit Salad	A side-dish made with various fruit pieces	4-5 people	Bowl
2	Cheese display	A variety of cheese	4-5 people	Plate
3	Havarti on Rye	A sandwich with cheese	1 person	Plate

ID	DishID	IngredientsID	ID	Name	PreparationID
1	1	1	1	Apples	1
2	1	4	2	Brie	2
3	2	2	3	Rye Bread	3
4	2	5	4	Strawberries	2
5	3	3	5	Havarti	3
6	3	5			

ID	PreparationMethod	PreparationDescription
1	Chopped	Cut into small cubes
2	Wedges	Cut into triangles
3	Sliced	Cut into slices

Which Normal Form?

ID	Name	JobTitle
1	Sue	CEO
2	Naheed	Project Manager
3	Julia	Developer
4	Bob	Developer

ID	Employee	Address	City	StateProv	PostalCode	Country
1	1	123 Main St.	Savannah	Georgia	84882	USA
2	2	345 Rue Mansa	Montreal	Quebec	L4J 4K3	Canada
3	3	456 Barbican	Toronto	Ontario	H8E K3H	Canada
4	4	678 Leng Kee Rd.	San Diego	California	92834	USA

Normalization Resources

- http://en.wikipedia.org/wiki/Database_normalization
- **Normalization Exercises**
<http://holowczak.com/database-normalization/13/>
- **A Simple Guide to Five Normal Forms in Relational Database Theory**
<http://www.bkent.net/Doc/simple5.htm>

More Normalization Resources

- <http://www.studytonight.com/dbms/database-normalization>
- <http://www.slideshare.net/jagaarij/database-design-normalization>
- <http://www.youtube.com/watch?v=fg7r3DgS3r>
[A](#)

When Could Normalization Hurt

When a commonly needed query now requires too many joins to produce.

Note: Yes, this may be fixed with views, etc. in some cases (though that can still be a performance hit), but it is something to consider.

Displaying a Our Data

Non-normalized:

```
select * from dish;
```

3NF:

```
select Dish.Name, Dish.Description, Dish.Serves,  
Dish.ServingMethod, Ingredients.Name,  
Preparation.PreparationMethod,  
Preparation.PreparationDescription  
from Dish  
left join DishIngredients on Dish.ID =  
DishIngredients.DishID  
left join Ingredients on DishIngredients.IngredientsID  
= Ingredients.ID  
left join Preparation on Ingredients.PreparationID =  
Preparation.ID
```

When Could Normalization Hurt

If you only do first normal form.

When Normalization is not Necessary

When you are using data that cannot be changed.

Denormalization Resources

- **Data Modeling 104: De-normalization**
<http://webandphp.com/DataModeling104%3ADenormalization>
- **Normalization Is for Sissies**
<http://blogs.msdn.com/b/pathelland/archive/2007/07/23/normalization-is-for-sissies.aspx>
- **Maybe Normalizing Isn't Normal**
<http://www.codinghorror.com/blog/2008/07/maybe-normalizing-isnt-normal.html>

Case Studies

- **A Detailed Five Step Twitter Scaling Plan**
<http://whydoeseverythingsuck.com/2008/05/detailed-five-step-twitter-scaling-plan.html>
- **When Not to Normalize your SQL Database (social networking)**
<http://www.25hoursaday.com/weblog/CommentView.aspx?guid=cc0e740c-a828-4b9d-b244-4ee96e2fad4b>

Find Me

- Twitter: e3betht
- Madison PHP
<http://www.madisonphp.com>
- Slides Available:
<http://www.TreeLineDesign.com/slides>

Want more? Take a PHP course! Visit:

www.phparch.com

and click on "TRAINING" for registration info.



php[architect]
AZ42-W1JJ-D57Z
25% off a new
subscription



Ask me about writing articles for the magazine!

<http://www.phparch.com>



Feedback or Questions

Twitter:

e3betht

E-mail:

Beth@Musketters.me