

How To Be Normal - A Guide for Developers

Presented By:
Mike Hillyer

Who Am I?

- Mike Hillyer, BSc
- Former MySQL Documentation Team Member
- Senior Engineer at Message Systems

- More at <http://www.mikehillyer.com/about/>

Who Are You?

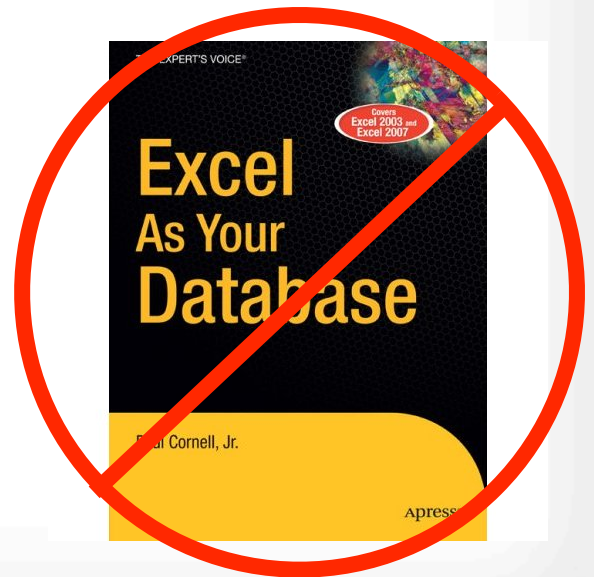
- How Many of You:
 - Are DBAs?
 - Are Developers?
 - Thought this was a session on fashion and social etiquette?
 - Know whether BCNF and DKNF are fashion brands or normal forms?
 - Know what a normal form is?

What Are You Watching?

- What is normalization and what are its benefits?
- What are the normal forms?
 - First Normal Form
 - Second Normal Form
 - Third Normal Form
 - And so on...
- Relationships and Joins
- By Example
- Over-Normalization & De-Normalization

You May Be Infected...

- Excellus Databasicus (Spreadsheet Syndrome)
- Symptoms
 - Asking “How many columns can a table have?”
 - Using few tables, often only one.
 - Ballooning storage needs.
 - Redundant data.
- Treatment



What is Normalization?

- Introduced by E.F. Codd.
- The modification of a schema so that it conforms to defined normal forms.
- Ensuring that every non-key column relates to “The Key, The Whole Key, and Nothing But the Key”
 - So Help Me Codd
- Makes data atomic.
- Reduces redundancy.

What Are The Benefits?

- Decreased storage consumption.
 - Removed redundancy means less data.
- Better/Faster(/Stronger) searches.
 - Less data to scan.
 - Easier searches on (previously) mixed data.
- Improved data integrity.
 - When data is only in one place you only have to get it right/fix it once.

What Are The Normal Forms?

- First Normal Form (1NF)
- Second Normal Form (2NF)
- Third Normal Form (3NF)

What is First Normal Form? (1NF)

- Requires a Primary Key.

(The Key)

- Requires that all data is atomic.
- Also work on removing horizontal and vertical redundancies from your table.

Name
Mike Hillyer

—

ID	First_Name	Last_Name
1	Mike	Hillyer

What is Second Normal Form? (2NF)

- Requires that when a composite key is present, all fields relate to the entire composite key.

(The Whole Key)

Reviewer_ID	First_Name	Last_Name	ISBN	Score
1	Mike	Hillyer	2233	4.5

Reviewer_ID	ISBN	Score
1	2233	4.5

What is Third Normal Form? (3NF)

- Requires that all fields depend directly on the primary key, and not on other non-key fields.

(And Nothing But The Key)

Address_ID	Address	City	State	Zip
1	123 Main St	Santa Clara	California	90221

Zip	City	State
90221	Santa Clara	California

Three's Not Enough, Are There More?

- Boyce-Codd Normal Form
- Fourth Normal Form
- Fifth Normal Form
- Domain/Key Normal Form
- Sixth Normal Form
- Non-First Normal Form

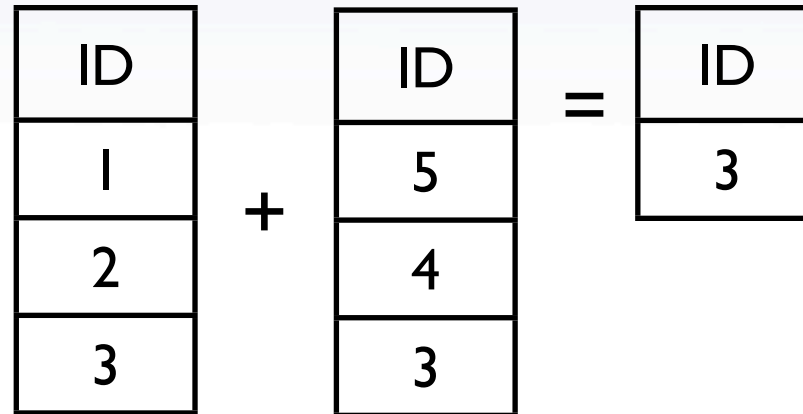
- http://en.wikipedia.org/wiki/Database_normalization

How Are My Entities Related?

- Three Forms
- One to One
 - Same Table?
- One to Many
 - Place PK of the One in the Many
- Many to Many
 - Create a joining table

How Do I Put Things Back Together?

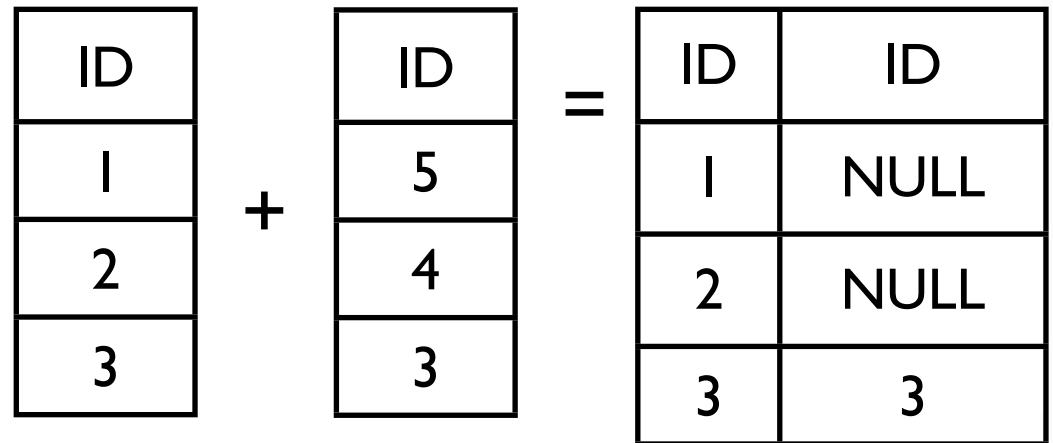
- INNER JOIN



- OUTER JOIN

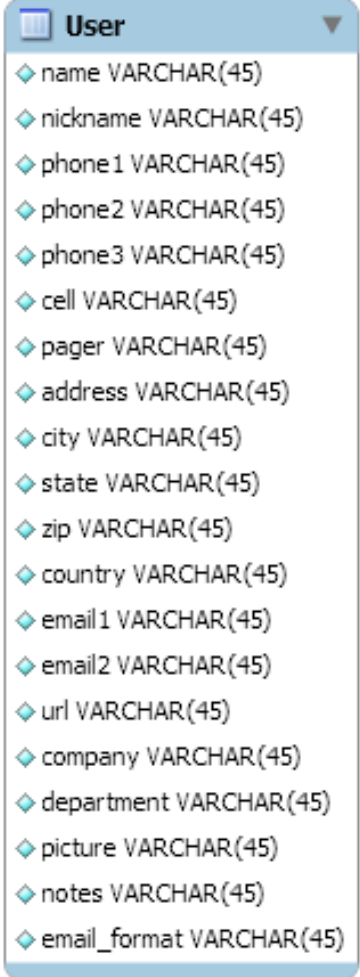
- LEFT JOIN

- RIGHT JOIN



By Example: User Tracking

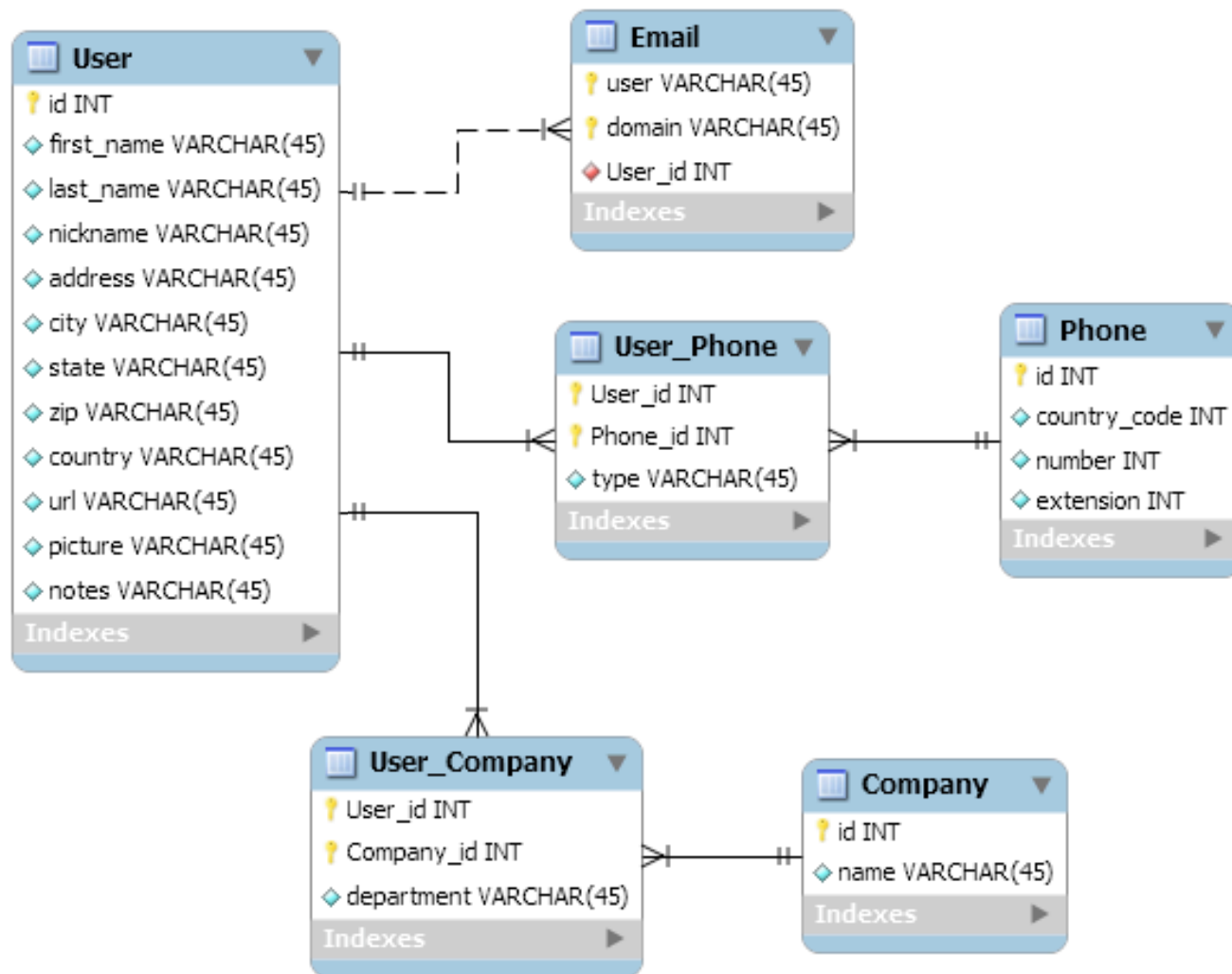
- Table has no Primary Key.
- Name is not atomic.
- Table starts with all possible user information in a single table.
- Phone numbers and email addresses are horizontally redundant.
- Company, department, city, state, zip are vertically redundant.



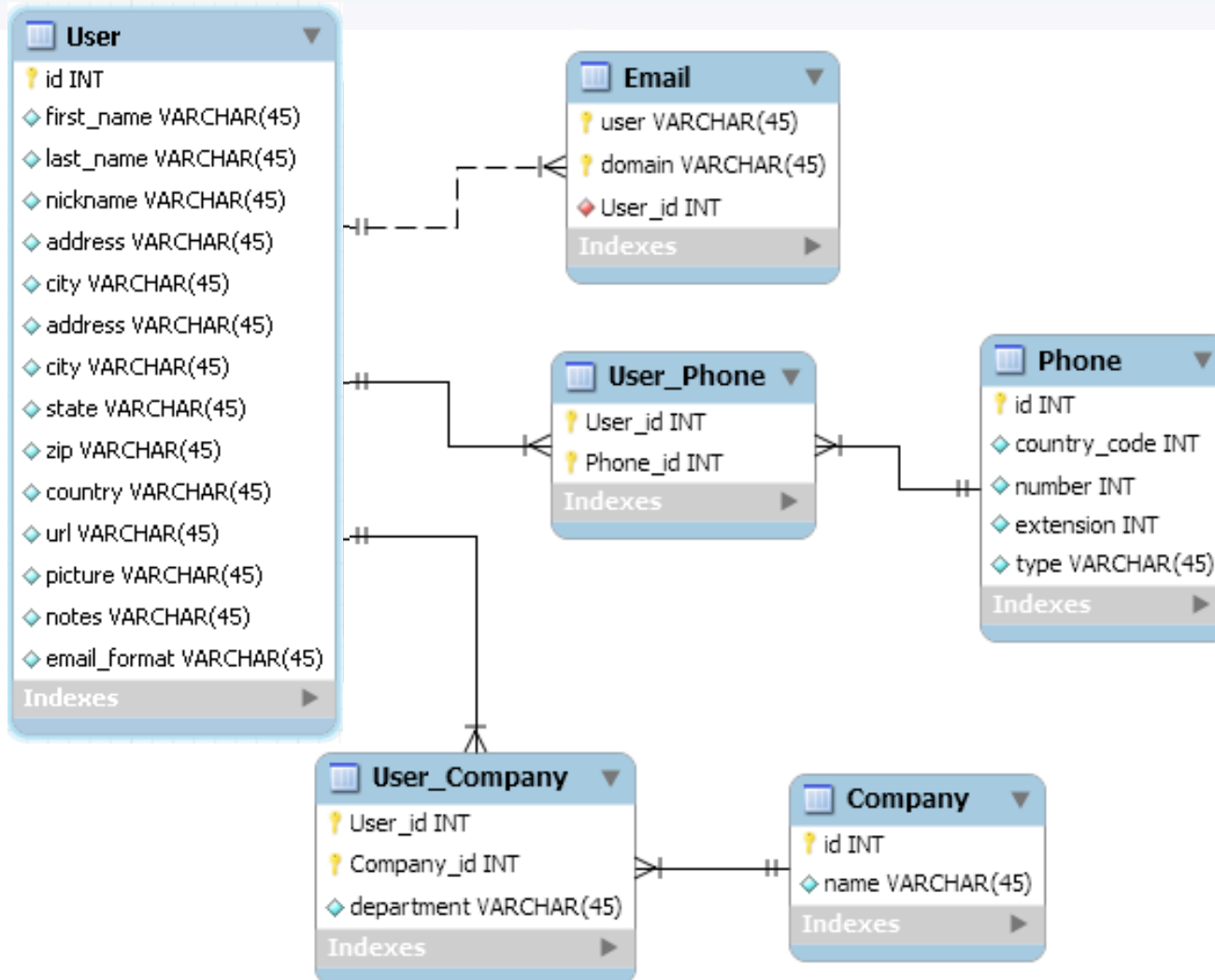
A screenshot of a database table named "User". The table contains the following columns, each with a VARCHAR(45) data type:

Column Name	Data Type
name	VARCHAR(45)
nickname	VARCHAR(45)
phone1	VARCHAR(45)
phone2	VARCHAR(45)
phone3	VARCHAR(45)
cell	VARCHAR(45)
pager	VARCHAR(45)
address	VARCHAR(45)
city	VARCHAR(45)
state	VARCHAR(45)
zip	VARCHAR(45)
country	VARCHAR(45)
email1	VARCHAR(45)
email2	VARCHAR(45)
url	VARCHAR(45)
company	VARCHAR(45)
department	VARCHAR(45)
picture	VARCHAR(45)
notes	VARCHAR(45)
email_format	VARCHAR(45)

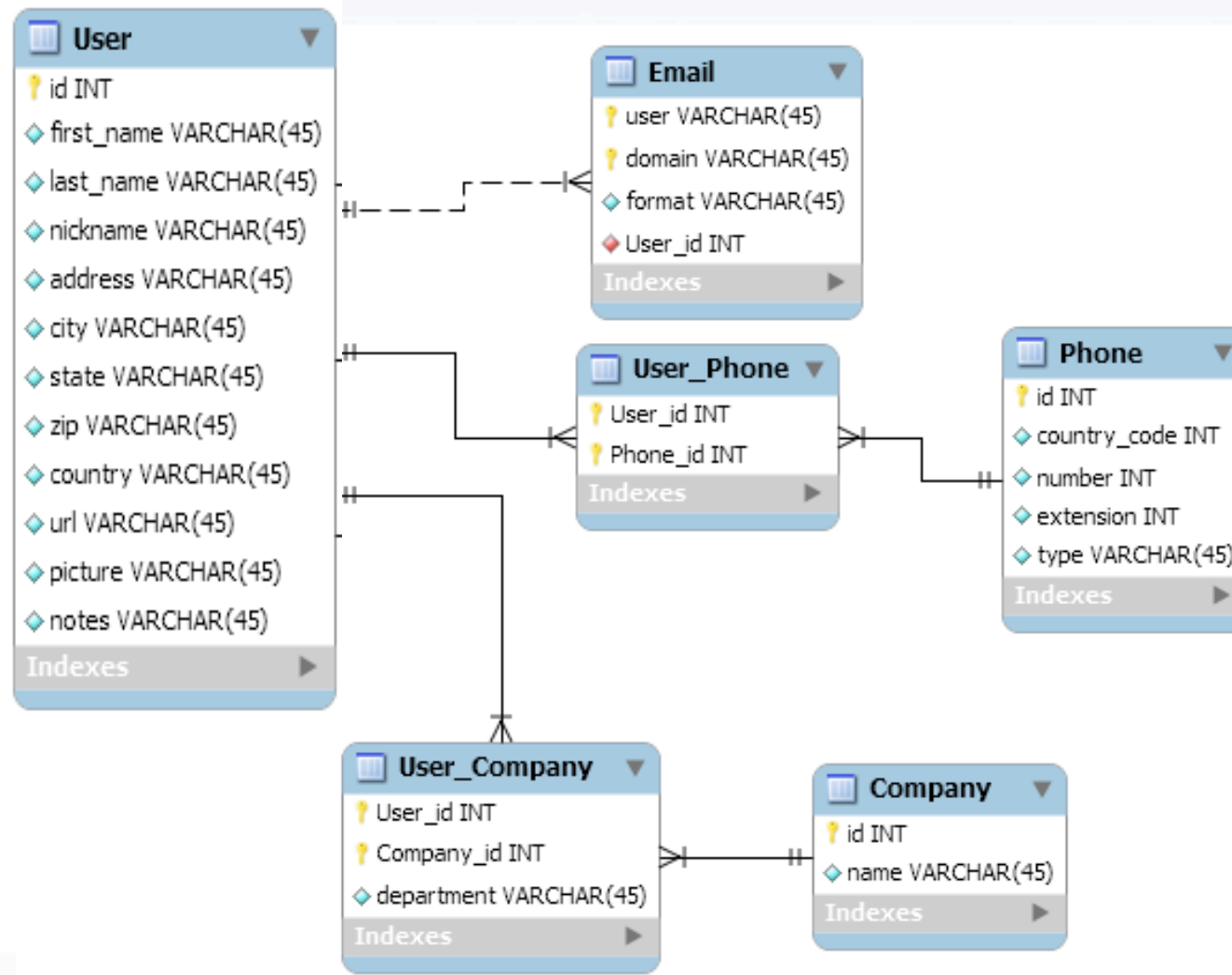
1NF Satisfied



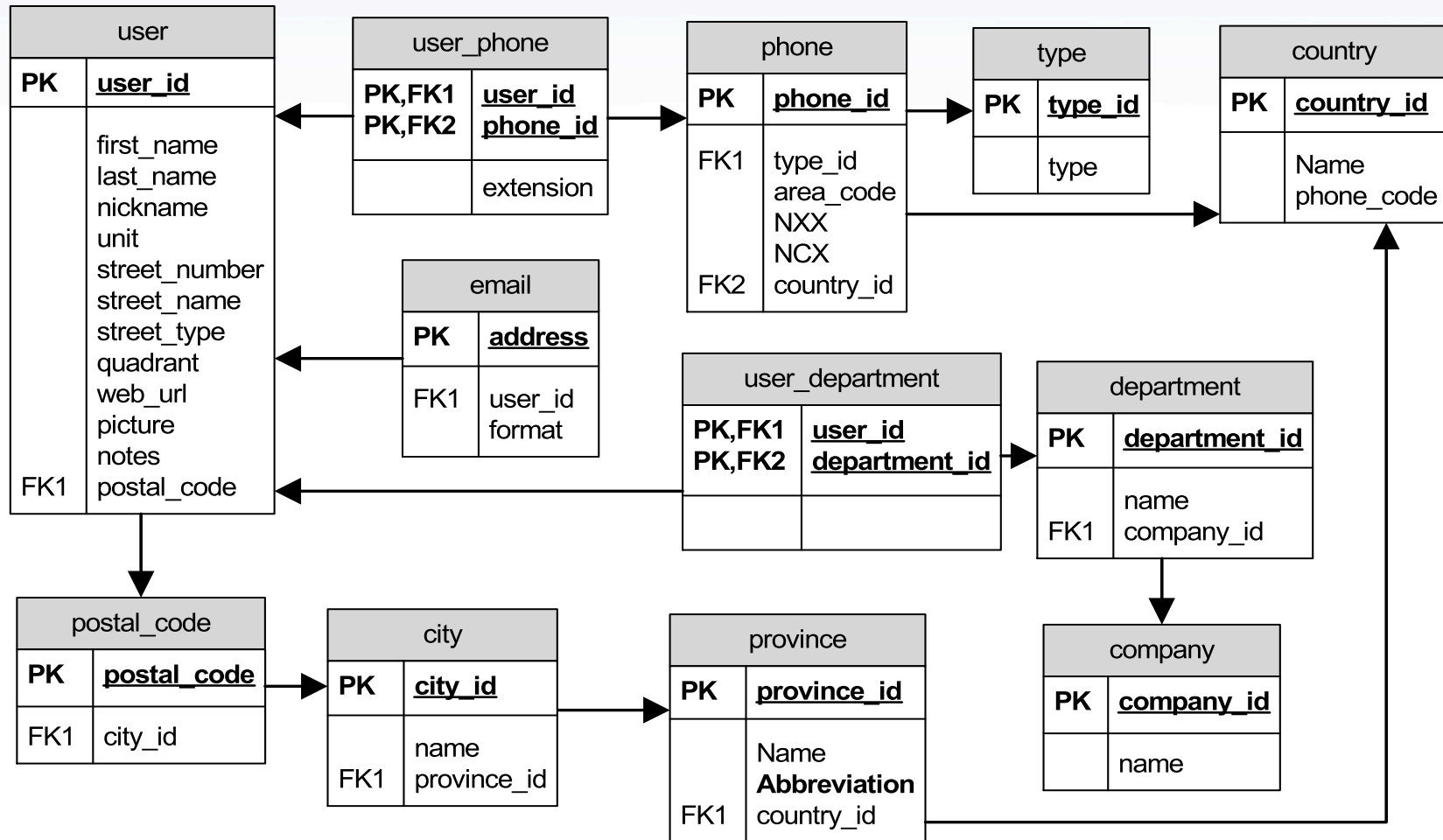
2NF Satisfied



3NF Satisfied



Have I Over-Normalized?



When Should I De-Normalize?

- Start by normalizing, then watch your slow query logs and run EXPLAIN.
- De-Normalization may be needed on certain queries (joining and sorting).
- Maintain data integrity with triggers.

Are We Done Yet?

- Slides and audio available soon.
- <http://www.mikehillyer.com/presentations/>
- <http://dev.mysql.com/tech-resources/articles/intro-to-normalization.html>
- mike@mikehillyer.com

Are You Hiring?

- Sales Engineer
- Software Engineer
- Web Engineer (Perl)
- Web Engineer (PHP)
- Site Reliability Engineer (Ops)

- Contact me for more info.