

SQL BOOTCAMP

Mason Gallo

Instructor

COURSE

PRE-WORK

IF YOU HAVEN'T ALREADY...

- Read the README:

- <https://github.com/MasonGallo/sql-bootcamp>

- Pay particular attention to the cheatsheet and exercises

OPENING

GETTING STARTED

SQL BOOTCAMP

WHO AM I

- Data Scientist
- Open Source Contributor
- ML Researcher - Educational Technology
- Data Science Instructor @ GA



SQL BOOTCAMP

MY PHILOSOPHY

- If you're not sure, please ask!
- Be considerate of your fellow students
- Participate!!!
- NO ONE knows everything
- Anything worth knowing is hard
- Examples then theory
- Breaks



LEARNING OBJECTIVES

- Build a personalized roadmap for your career next steps with SQL
- Learn enough SQL to effectively query data from a single table
- Learn enough SQL to effectively query data from multiple tables
- Proficiency in importing/exporting data for your work

SQL BOOTCAMP

AGENDA

- Installation
- Databases 101
- Basic queries
- Aggregation
- Joins
- Next steps for your career / Personal plan

SQL BOOTCAMP

WHY ARE YOU HERE?

- New skills?
- New careers?

HELP ME ADAPT THE CURRICULUM TO YOU!

SQL BOOTCAMP

INSTALLATION

DOWNLOAD CLASS MATERIALS

▸ All materials are available*** [HERE](#)

Click “Clone or Download” -> “Download Zip”

Slides are available in the main directory with filename sql_bootcamp.pdf

***If you know git, you can always clone the repo

INSTALLATION

If for some reason you're not using the recommended installation, please raise your hand!

GET ACQUAINTED

- Take a few minutes to get acquainted
- Look for some common buttons you might need:
 - Run
 - Open
 - Import
 - Export

GETTING HELP

- How do you know where to get help?
- Is Googling cheating?
- How do I remember all this stuff?

SQL BOOTCAMP

DATABASES 101

YOUR THOUGHTS

How do you store your data? At work? Personal?

WHAT ARE THE OPTIONS

- Storing in a raw text file
- Storing in excel or google sheets

MY THOUGHTS

Excel just won't cut it here...

- Storing social performance for many brands over several years**
- Ad campaign with billions of impressions per month**

LIMITATIONS

- Slow and inefficient
- Hard to share with coworkers
- Difficult to use with other applications
- Doesn't scale (millions? billions?)

WHAT DO WE WANT

- A methodology that can be used by many people and machines
- Scalability
- Speed
- Safety

WHAT DO WE WANT

Databases are a **structured** data source optimized for efficient **retrieval** and **storage**

WHAT IS SQL?

SQL (Structured Query Language) is a query language designed to **Extract, Transform, Load** data in relational databases

WHO USES SQL?

Web Developers

Data Analysts

Data Scientists

Product Managers

Statisticians

System Administrators

Backend Developers

...And many more

THE TERMS NO ONE EXPLAINS

- Database: think of as a directory or folder of tables
- Table: a collection of related structured data, usually as rows / columns
- Record: SQL lingo for “row”
- Field: SQL lingo for “column”
- Localhost: this computer (ie the computer you’re using)
- Root: a user with FULL permission to do virtually anything
- UTF-8: capable of encoding all unicode characters

A database is
software that's
good at storing
lots of data

Database vs. Spreadsheet

- Many people can use it
 - Fast for lots of data
(1M, 1B, or 1T rows!)
 - It's always on
 - The source of truth
 - Good for complicated
analysis
-

SQL is the
language used to
extract data from
databases

- **Code uses SQL to get data**
 - **People use SQL to get data**
 - **Analytics**
 - **Operations**
 - **others?**
-

Databases live on computers

- **Cloud (e.g. Amazon)**
 - **Your data center**
 - **Local (your computer)**
-

MOTIVATING EXERCISE 1

Everyone together!

```
select *
```

```
from box_office
```

```
where movie_name = 'Toy Story 3';
```

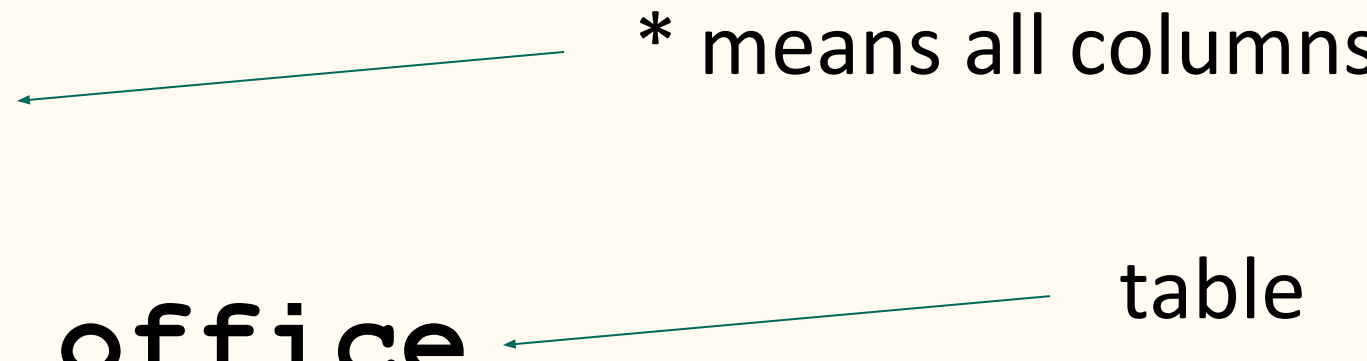
Databases are organized

- **Databases have tables**
 - **Tables have columns**
 - **Columns have types**
(e.g. date, number, text)
-

```
select *  
from box_office  
where movie_name = 'Toy Story 3';
```

* means all columns

table



```
select
```

```
    full_date,
```

columns

```
    sales
```

```
from box_office
```

table

```
where movie_name = 'Toy Story 3';
```


What just happened??

```
select
    full_date,
    sales
from box_office
where movie_name = 'Toy Story 3';
```

movie_name	full_date	sales	day_of_month	month_of_year	year	genre_id
The Twilight Saga: New Moon	11/21/2009	42,288,881	21	11	2009	2
Transformers: Age of Extinction	6/27/2014	41,868,135	27	6	2014	1
The Twilight Saga: Breaking Dawn - Part 2	11/17/2012	41,443,147	17	11	2012	2
Toy Story 3	6/18/2010	41,148,961	18	6	2010	2
The Hunger Games: Mockingjay - Part 1	11/22/2014	40,905,873	22	11	2014	2
Alice in Wonderland	3/5/2010	40,804,962	5	3	2010	2
Star Wars: Episode III - Revenge of the Sith	5/21/2005	40,693,760	21	5	2005	1

What just happened?

```
select
    full_date,
    sales
from box_office
where movie_name = 'Toy Story 3';
```

movie_name	full_date	sales	day_of_month	month_of_year	year	genre_id
The Twilight Saga: New Moon	11/21/2009	42,288,881	21	11	2009	2
Transformers: Age of Extinction	6/27/2014	41,868,135	27	6	2014	1
The Twilight Saga: Breaking Dawn - Part 2	11/17/2012	41,443,147	17	11	2012	2
Toy Story 3	6/18/2010	41,148,961	18	6	2010	2
The Hunger Games: Mockingjay - Part 1	11/22/2014	40,905,873	22	11	2014	2
Alice in Wonderland	3/5/2010	40,804,962	5	3	2010	2
Star Wars: Episode III - Revenge of the Sith	5/21/2005	40,693,760	21	5	2005	1

```
select full_date, sales  
from box_office  
where movie_name = 'Toy Story 3';
```

- Graph sales over time for *Dead Poets Society*
- Graph sales over time for your favorite movie

MOTIVATING EXERCISE 2

```
select
```

```
    movie_name,
```

```
    full_date,
```

```
    sales
```

```
from box_office
```

```
where movie_name like 'Toy Story%';
```

Operators

Text:

= equals

!= not equal

like `'%'` matches any sequence of zero or more characters
e.g. `'%wag%'` matches “wagon”, “swag”, “swagger”, etc.

ilike is the same as **like**, but is case insensitive

~ matches regular expressions

in included in a list, e.g. `movie_name in ('Titanic', 'Toy Story')`
matches Titanic and Toy Story

- Graph box office sales for all movies that include the word “good” (case insensitive)

hint: use the `ilike` operator

```
select *  
  
from box_office  
  
where sales > 40000000;
```


Operators

Numbers:

< less than

> greater than

<= less than or equal to

>= greater than or equal to

= equal

!= not equal

```
select
    full_date,
    sales
from box_office
where full_date > '01/01/2005';
```

Operators

Date:

< before

> after

<= before or same date

>= after or same date

= equal

!= not equal

```
select
    movie_name,
    full_date,
    sales
from box_office
where
    movie_name = 'Forrest Gump'
    and full_date < '01/01/2000';
```

```
select
    movie_name, full_date, sales
from box_office
where
    movie_name = 'Forrest Gump'
    and full_date < '01/01/2000';
```

1. “Titanic” was re-released in 2012 (why?).
Graph sales for the original release only.
2. Select the days that "Inception" had sales of
\$10 million or more

BASIC QUERIES

NOTE BEFORE WE GET STARTED

- Capitalization doesn't matter (SQL isn't case-sensitive)
- SQL ignores whitespace (but don't go crazy)
- Be OCD about your punctuation (it matters!)

BASIC QUERIES

SELECT *
FROM users

*Return all rows from all
columns*

BASIC QUERIES

```
SELECT title, first_name,  
last_name  
FROM users
```

*Return all rows from specific
columns*

BASIC QUERIES

```
SELECT DISTINCT title,  
first_name, last_name  
FROM users
```

*Return unique rows from the
specified columns*

BASIC QUERIES

SELECT *
FROM users
WHERE <condition>

*Return all columns under a
specified condition*

BASIC QUERIES

```
SELECT *  
FROM users  
WHERE state = "arizona"
```

BASIC QUERIES

```
SELECT *  
FROM users  
WHERE state LIKE "arizona"  
AND title LIKE "Miss"
```

BASIC QUERIES

```
SELECT *  
FROM users  
WHERE state IN  
("arizona", "florida")
```

BASIC QUERIES

SELECT *
FROM users
WHERE state IN
("arizona", "florida")

What happens if we put a
NOT in front of IN?

BASIC QUERIES

```
SELECT *  
FROM users  
WHERE zip = 10007
```

Notice the use of = here
instead of LIKE. Any ideas?

We can also use >, <, !=, >=,
<=

BASIC QUERIES

SELECT *

FROM users

WHERE street LIKE “%rd”

% matches anything

_ matches a single character

BASIC QUERIES

```
SELECT *  
FROM users  
WHERE street LIKE "%rd"  
ORDER BY first_name
```

What happens if you put
DESC after first_name?

BASIC QUERIES

```
SELECT *  
FROM users  
WHERE street LIKE "%rd"  
ORDER BY first_name  
LIMIT 5
```

What does LIMIT do?

YOUR TURN

Complete the warm up exercises [here](#)

MOTIVATING EXERCISE 3

ORDER BY

```
select
    full_date,
    sales
from box_office
where movie_name = 'Toy Story 3'
order by full_date asc;
```

- Comes after select, from, where, group by, having
- Can be ASC or DESC
- Default is ASC

LIMIT

`select * from table limit 10;` is the first query most people run!

- Comes after `select`, `from`, `where`, `group by`, `having`, `order by`
- Good way to quickly browse results

Find the top 10 highest grossing days of all time.

Hint: use `order by desc & limit!`

GROUP BY: aggregate many rows into 1

Example: `select * from box_office;`

<code>movie_name</code>	<code>full_date</code>	<code>sales</code>
<code>Titanic</code>	<code>1/1/1998</code>	<code>100</code>
<code>Titanic</code>	<code>1/2/1998</code>	<code>200</code>
<code>Titanic</code>	<code>1/3/1998</code>	<code>300</code>
<code>Good Will Hunting</code>	<code>1/1/1998</code>	<code>200</code>
<code>Good Will Hunting</code>	<code>1/2/1998</code>	<code>200</code>
<code>Good Will Hunting</code>	<code>1/3/1998</code>	<code>500</code>

Total Box Office sales per movie?

movie_name	full_date	sales
Titanic	1/1/1998	100
Titanic	1/2/1998	200
Titanic	1/3/1998	300
Good Will Hunting	1/1/1998	200
Good Will Hunting	1/2/1998	200
Good Will Hunting	1/3/1998	500

← 600

← 900

Total Box Office sales per movie?

```
select
    movie_name,
    sum(sales)
from box_office
group by movie_name;
```

movie_name	sum
Titanic	600
Good Will Hunting	900

```
select
    movie_name,
    sum(sales)
from box_office
group by movie_name;
```

What is the average sales per movie?

Hint: Use `avg(column_name)` instead
of `sum(column_name)`

Average Box Office sales per movie?

```
select
```

```
    movie_name,
```

```
    sum(sales) ,
```

```
    avg(sales)
```

```
from box_office
```

```
group by movie_name;
```

Average Box Office sales per movie?

```
select
```

```
    movie_name,
```

```
    sum(sales) as total_sales,
```

```
    avg(sales) as avg_sales
```

```
from box_office
```

```
group by movie_name;
```

Total Box Office sales per month?

movie_name	month_of_year	sales
------------	---------------	-------

Titanic	1	100
---------	---	-----

Titanic	2	200
---------	---	-----

Titanic	3	300
---------	---	-----

Good Will Hunting	1	200
-------------------	---	-----

Good Will Hunting	2	200
-------------------	---	-----

Good Will Hunting	3	500
-------------------	---	-----

Total Box Office sales per month?

movie_name	month_of_year	sales
------------	---------------	-------

Titanic	1	100
---------	---	-----

Titanic	2	200
---------	---	-----

Titanic	3	300
---------	---	-----

Good Will Hunting	1	200
-------------------	---	-----

Good Will Hunting	2	200
-------------------	---	-----

Good Will Hunting	3	500
-------------------	---	-----

Total Box Office sales per month?

movie_name	month_of_year	sales
------------	---------------	-------

Titanic	1	100
---------	---	-----

Titanic	2	200
---------	---	-----

Titanic	3	300
---------	---	-----

Good Will Hunting	1	200
-------------------	---	-----

Good Will Hunting	2	200
-------------------	---	-----

Good Will Hunting	3	500
-------------------	---	-----

Total Box Office sales per month?

```
select
```

```
    month_of_year,
```

```
    sum(sales) as total_sales
```

```
from box_office
```

```
group by month_of_year;
```

Total Box Office sales per month, per year?

month_of_year	year	total_sales
10	2003	100
11	2003	200
12	2003	300
1	2004	200
2	2004	200
3	2004	500

Total Box Office sales per month, per year?

```
select
```

```
    month_of_year,
```

```
    year,
```

```
    sum(sales) as total_sales
```

```
from box_office
```

```
group by month_of_year, year
```

```
order by year, month_of_year;
```

```
select
    movie_name, sum(sales)
from box_office
group by movie_name;
```

- Find the avg of sales per month_of_year
- Find & graph the total sales per day_of_month
- Find the sum of sales per day_of_month
per month_of_year

AGGREGATION

YOUR TURN

What if we want to compute overall statistics about the data?

Think: we want counts, averages, etc

AGGREGATION

SELECT *
FROM users

*Return all rows from
columns*

AGGREGATION

```
SELECT count(*)  
FROM users
```

*Return the count of all rows
from all columns*

AGGREGATION

```
SELECT distinct first_name  
FROM users
```

*Return the unique first
names*

AGGREGATION

```
SELECT count(distinct first_name)  
FROM users
```

*Return the count of unique
first names*

AGGREGATION

```
SELECT avg(zip)  
FROM users
```

*Return the average zip code
(yes, I know this doesn't
make sense)*

AGGREGATION

```
SELECT avg(zip) as “nonsense”  
FROM users
```

*Rename our avg zip to
“nonsense”*

AGGREGATION

SELECT gender, avg(zip) as
“nonsense”

FROM users

GROUP BY gender

*What if we want to know the
avg zip by gender?*

*We can use sum, avg, min,
max, count*

AGGREGATION

SELECT state, gender, avg(zip) as
“nonsense”

FROM users

GROUP BY state, gender

What happened here?

AGGREGATION

SELECT state, gender, avg(zip) as
“nonsense”

FROM users

GROUP BY state, gender

ORDER BY 1

What happened here?

AGGREGATION

SELECT state, gender, avg(zip) as
“nonsense”

FROM users

GROUP BY state, gender

ORDER BY 1

*What happened here?
Shorthand?*

What about where?

- Can still use **where**,
comes before **group by**
- To conditionally include
aggregates, use **having**

What about where?

```
select
```

```
    movie_name,
```

```
    sum(sales)
```

```
from box_office
```

```
group by movie_name
```

```
having sum(sales) > 4000000000;
```

Aggregate functions galore!

- `avg(column_name)`
- `count(column_name)`
- `count(distinct column_name)`
- `sum(column_name)`
- `max(column_name)`
- `min(column_name)`
- `corr(col_A, col_B)`
- many more!

```
select movie_name, sum(sales)
from box_office group by movie_name;
```

- Find the top 10 grossing movies of all time.
(Hint: use `order by`, `limit` and `group`!)
- Calculate the release date of each movie.
(hint: how can you express release date as an aggregate function of `full_date`?)
- Count how many days each movie was in theaters.

AGGREGATION

SELECT state, gender, avg(zip) as
nonsense
FROM users
GROUP BY state, gender
HAVING nonsense > 50000

*What happened here? What
happens if we use WHERE
instead?*

AGGREGATION

SELECT <columns>
FROM <table>
WHERE <condition>
GROUP BY <columns>
HAVING <condition on aggregates>
ORDER BY <columns>
LIMIT <number>

*General SQL structure
(putting it all together)*

YOUR TURN

Complete the aggregation exercises [here](#)

MOTIVATING EXERCISE 4

Why join?

- Not all data stored in all tables
- Relationships exist between tables

Example: Box office sales by genre

```
select
```

```
    genres.genre_name,
```

```
    box_office.movie_name,
```

```
    box_office.full_date,
```

```
    box_office.sales
```

```
from box_office
```

```
join genres on box_office.genre_id = genres.genre_id
```

```
where box_office.movie_name ilike '%new york%';
```

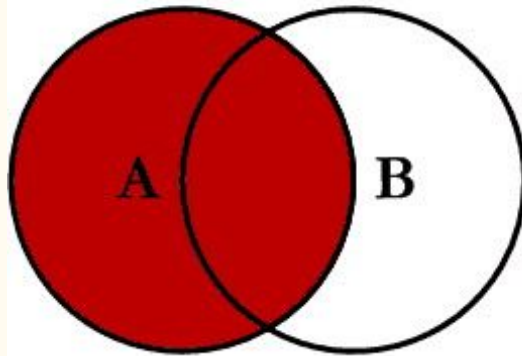
What if a movie doesn't have a genre_id?

```
select *
```

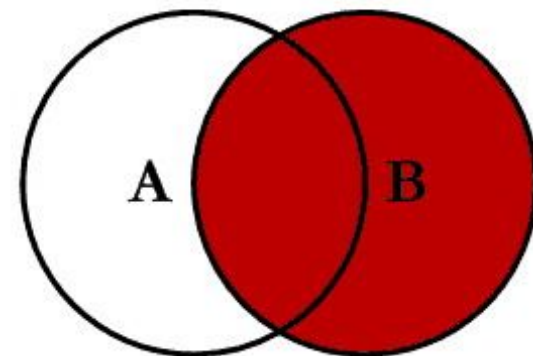
```
from box_office
```

```
where genre_id is null;
```

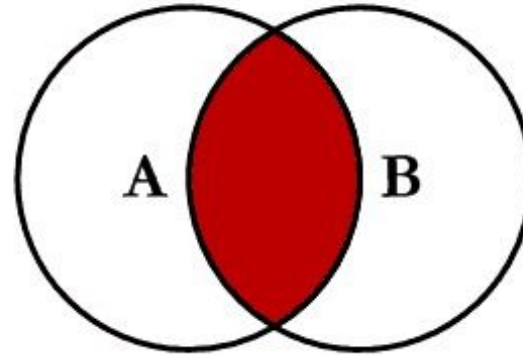
SQL JOINS



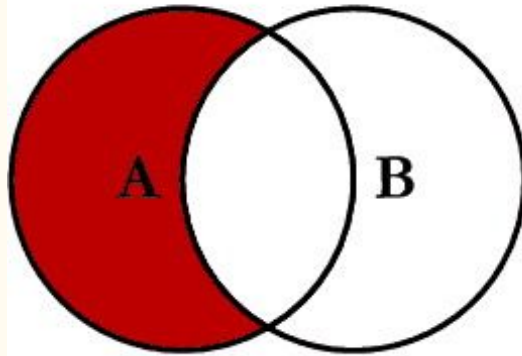
```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key
```



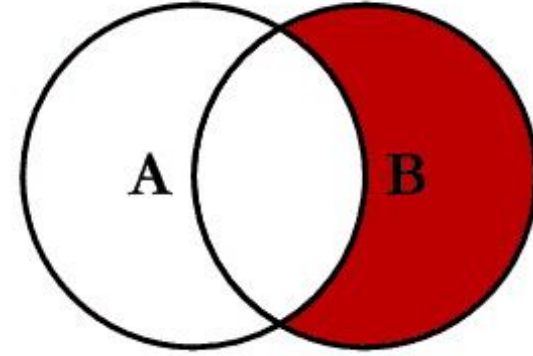
```
SELECT <select_list>  
FROM TableA A  
RIGHT JOIN TableB B  
ON A.Key = B.Key
```



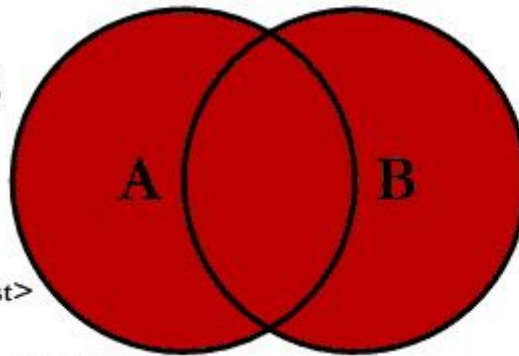
```
SELECT <select_list>  
FROM TableA A  
INNER JOIN TableB B  
ON A.Key = B.Key
```



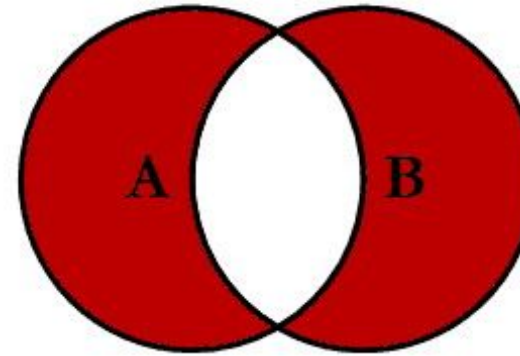
```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key  
WHERE B.Key IS NULL
```



```
SELECT <select_list>  
FROM TableA A  
RIGHT JOIN TableB B  
ON A.Key = B.Key  
WHERE A.Key IS NULL
```



```
SELECT <select_list>  
FROM TableA A  
FULL OUTER JOIN TableB B  
ON A.Key = B.Key
```



```
SELECT <select_list>  
FROM TableA A  
FULL OUTER JOIN TableB B  
ON A.Key = B.Key  
WHERE A.Key IS NULL  
OR B.Key IS NULL
```

```
select
    genres.genre_name,
    box_office.movie_name,
    box_office.full_date,
    box_office.sales
from box_office
join genres
    on box_office.genre_id = genres.genre_id;
```

- Use `group by` and `join` to find the sum of sales per `genre_name`

JOINS

JOINS

What if we want to deal with multiple tables?

Think: we want to capture information from 2 or more tables simultaneously

Import flights.sql

JOINS

A **relational database** is organized in the following manner:

- ▶ A database has **tables** which represent individual entities or objects
- ▶ Tables have a predefined **schema** – rules that tell it what columns exist and what they look like

JOINS

A **relational database** is organized in the following manner:

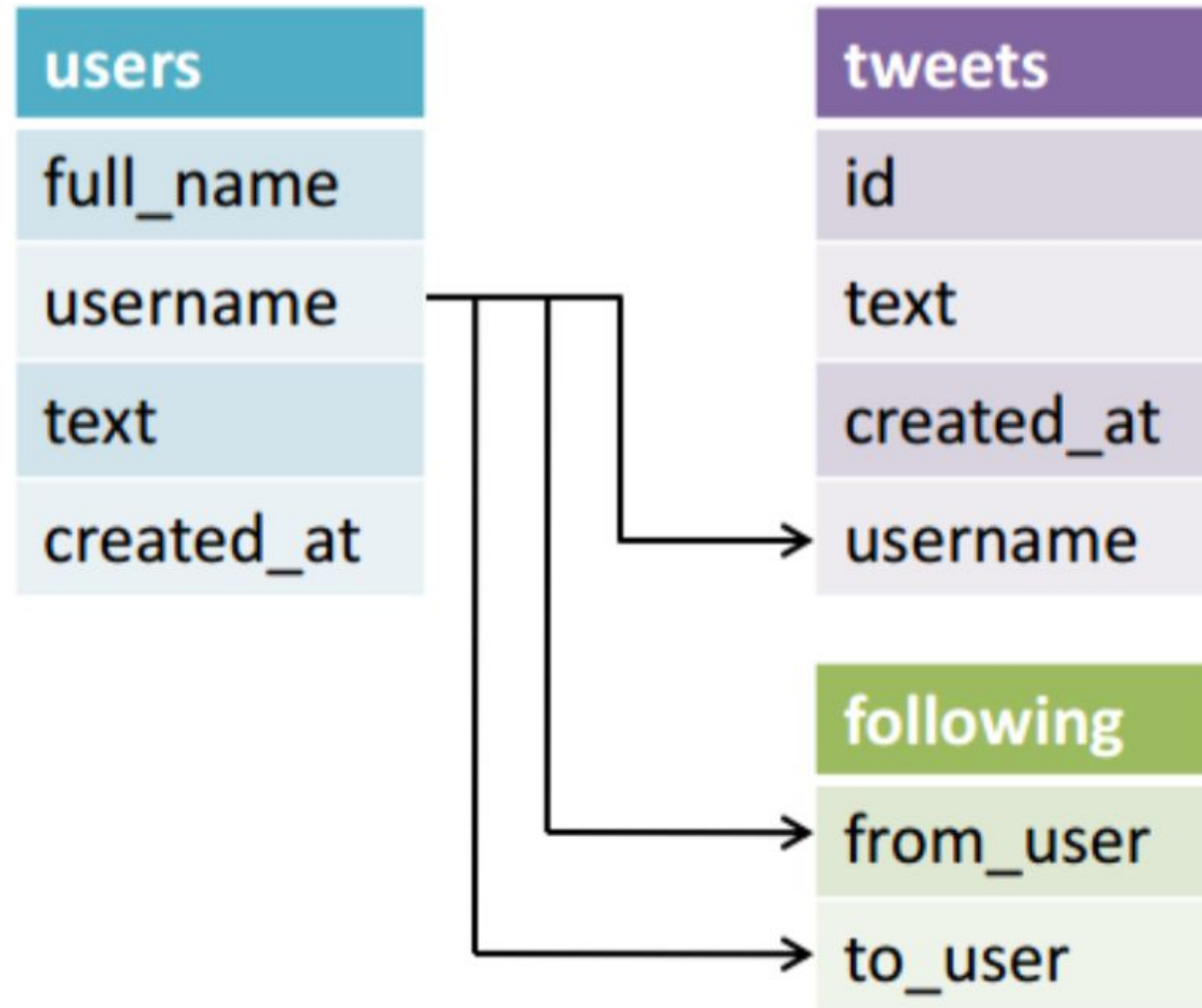
► **table**

id	first name	last name	date of birth
312	Joe	Smith	1980-12-24
1532	Michelle	Anderson	1973-03-12

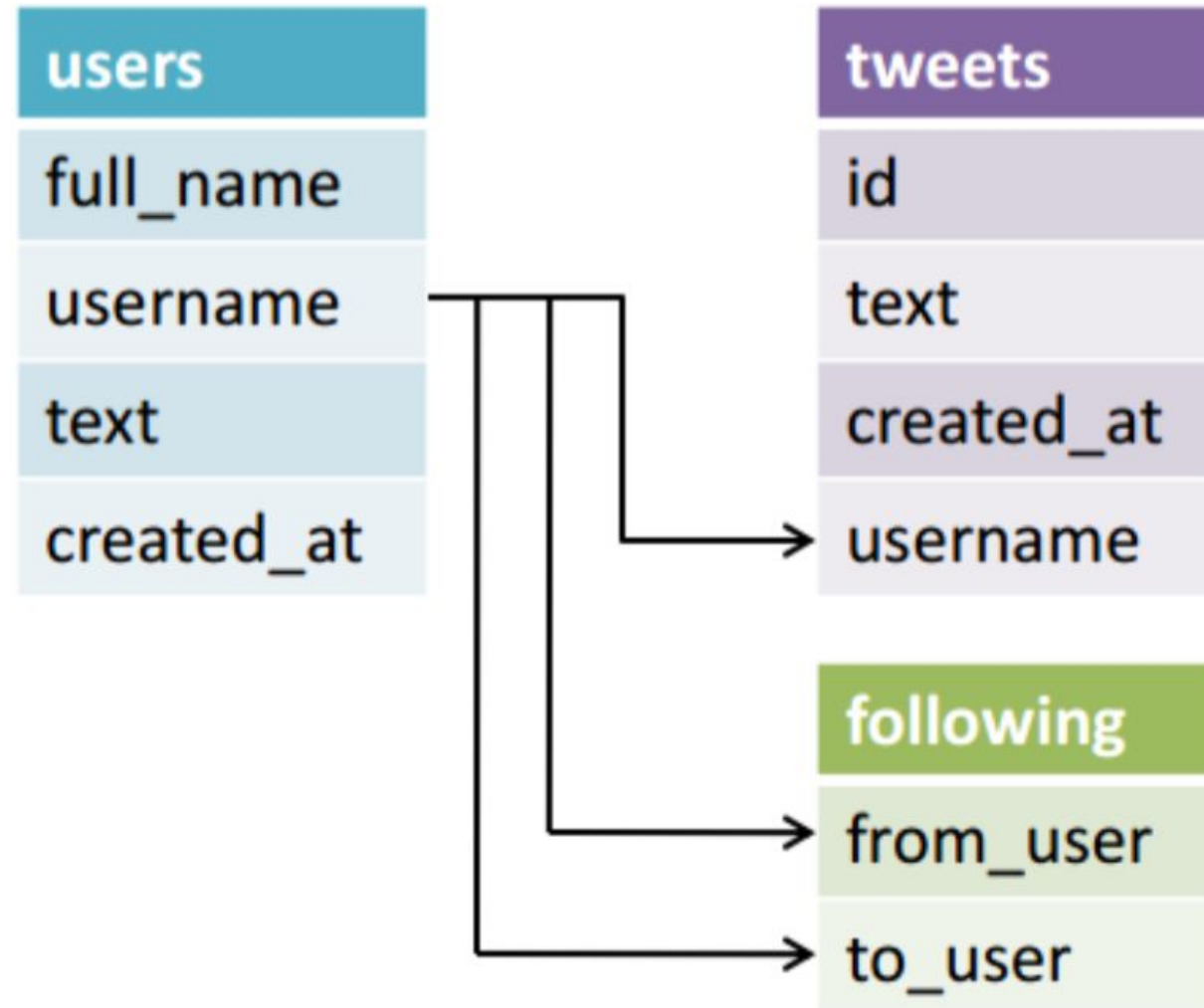
► **schema**

id	bigint
first_name	char(36)
last_name	char(36)
date_of_birth	timestamp

JOINS



JOINS



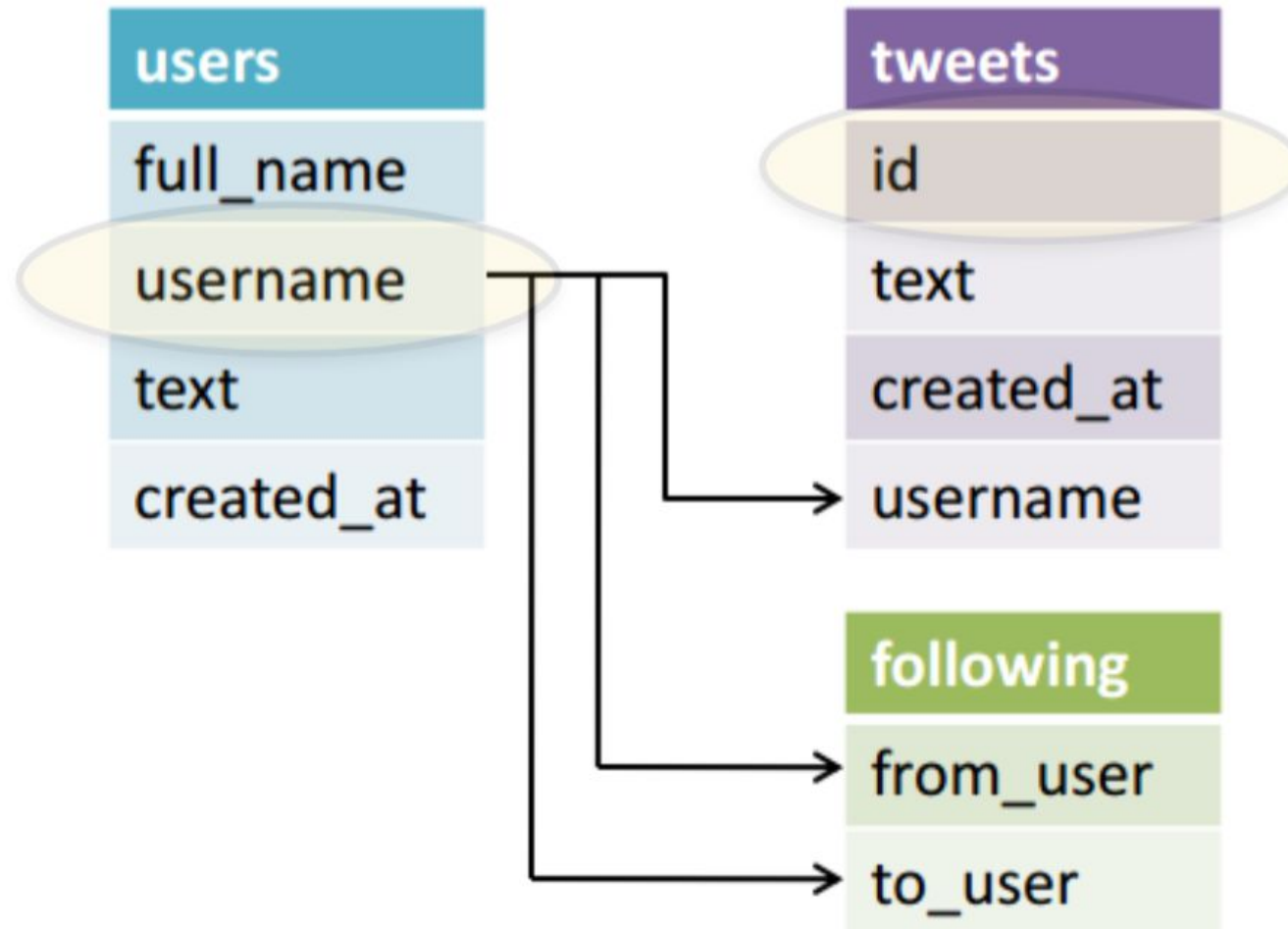
JOINS

Each table should have a PRIMARY KEY

Think: a unique identifier for each row

Ex: SSN

JOINS



JOINS

Each table should have a PRIMARY KEY

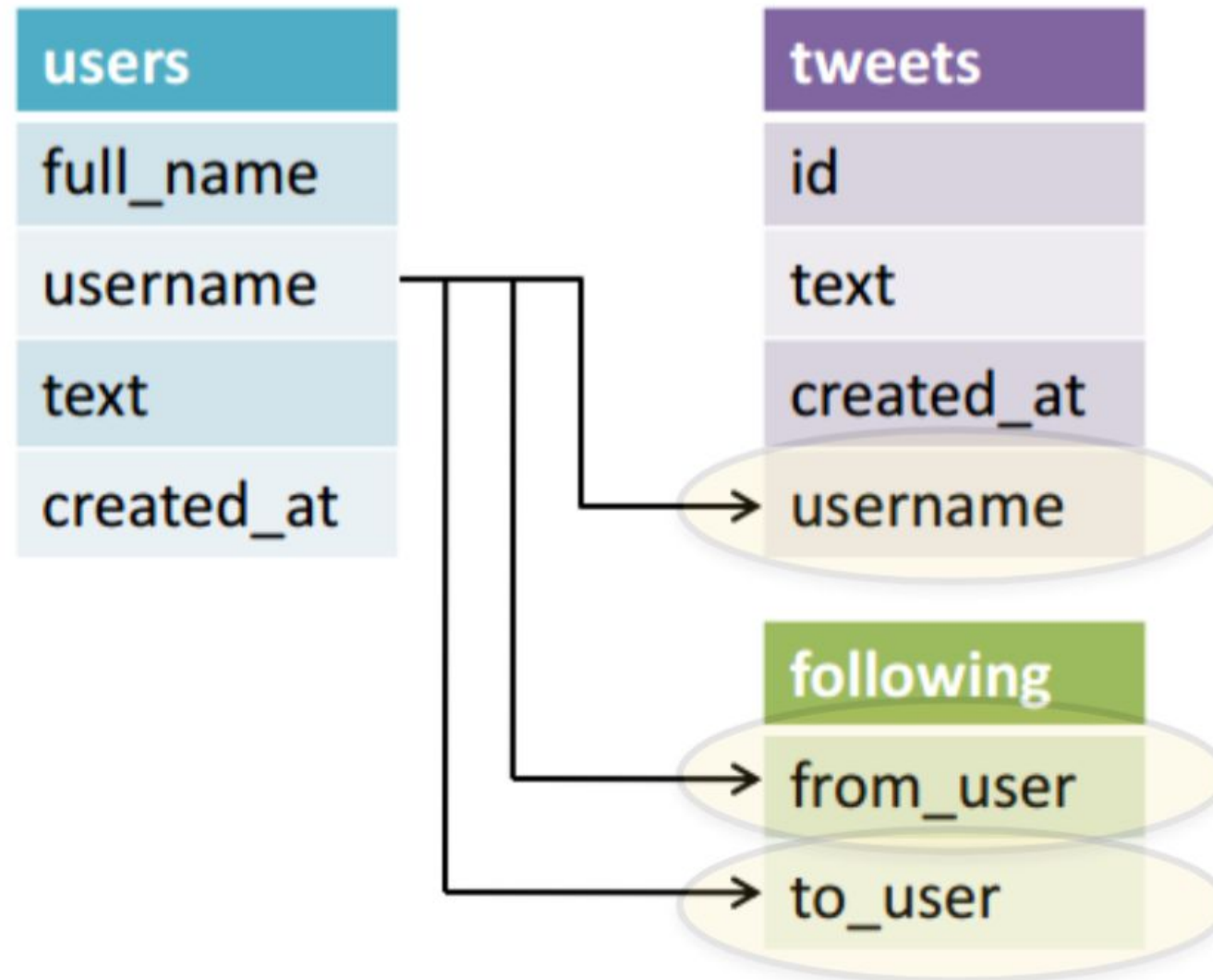
Think: a unique identifier for each row

Ex: SSN

Additionally, each table may have a FOREIGN KEY

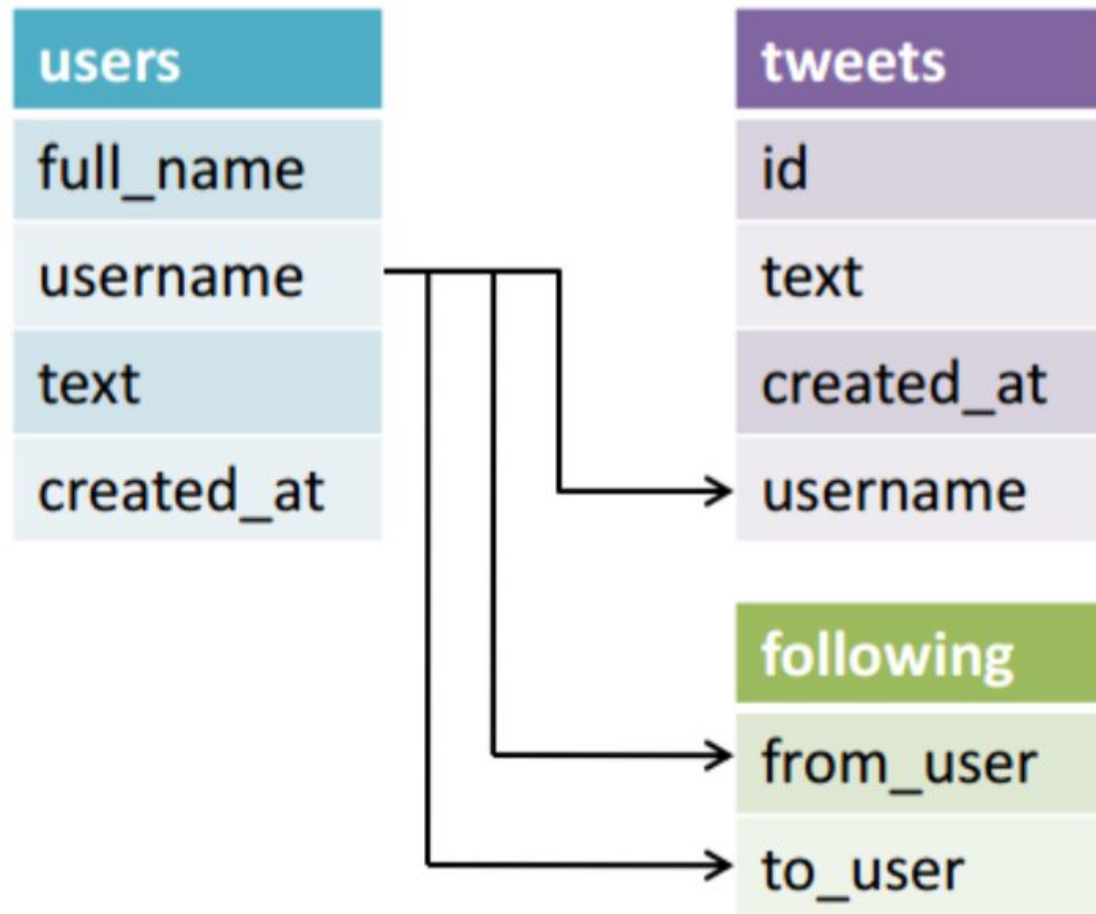
Think: an ID that links one table to another

JOINS



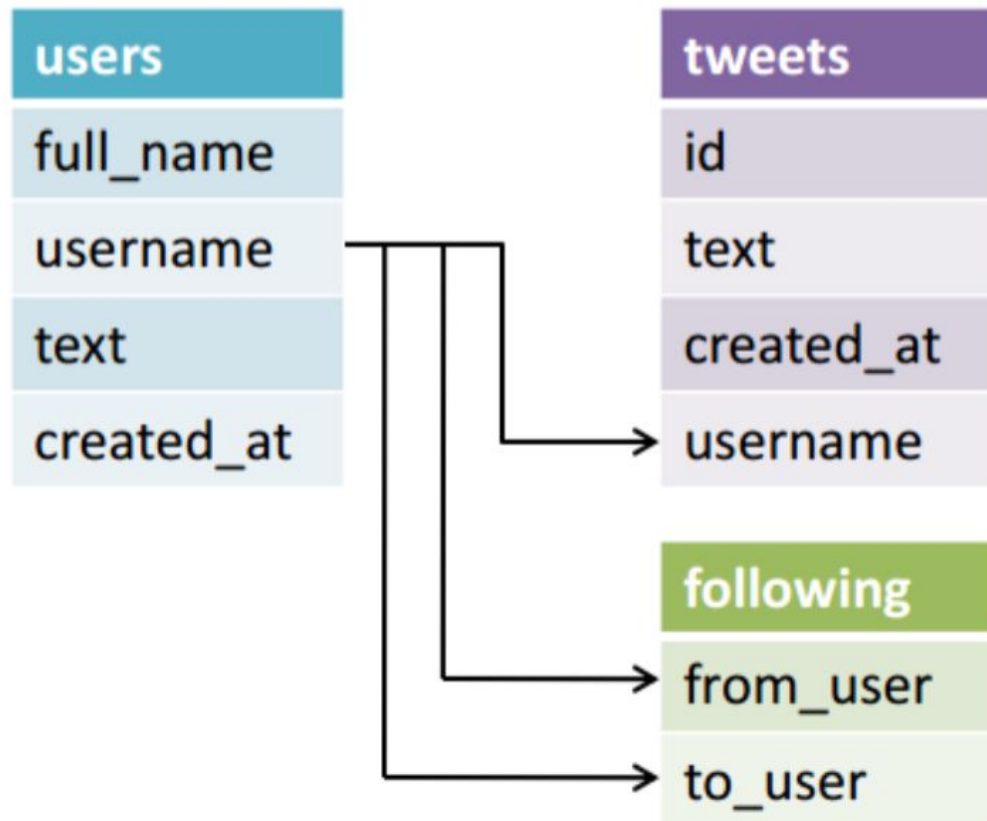
JOINS

Create a table with all the users' full names and their tweets



JOINS

Create a table with all the users' full names and their tweets



<u>full_name</u>	<u>tweet</u>
------------------	--------------

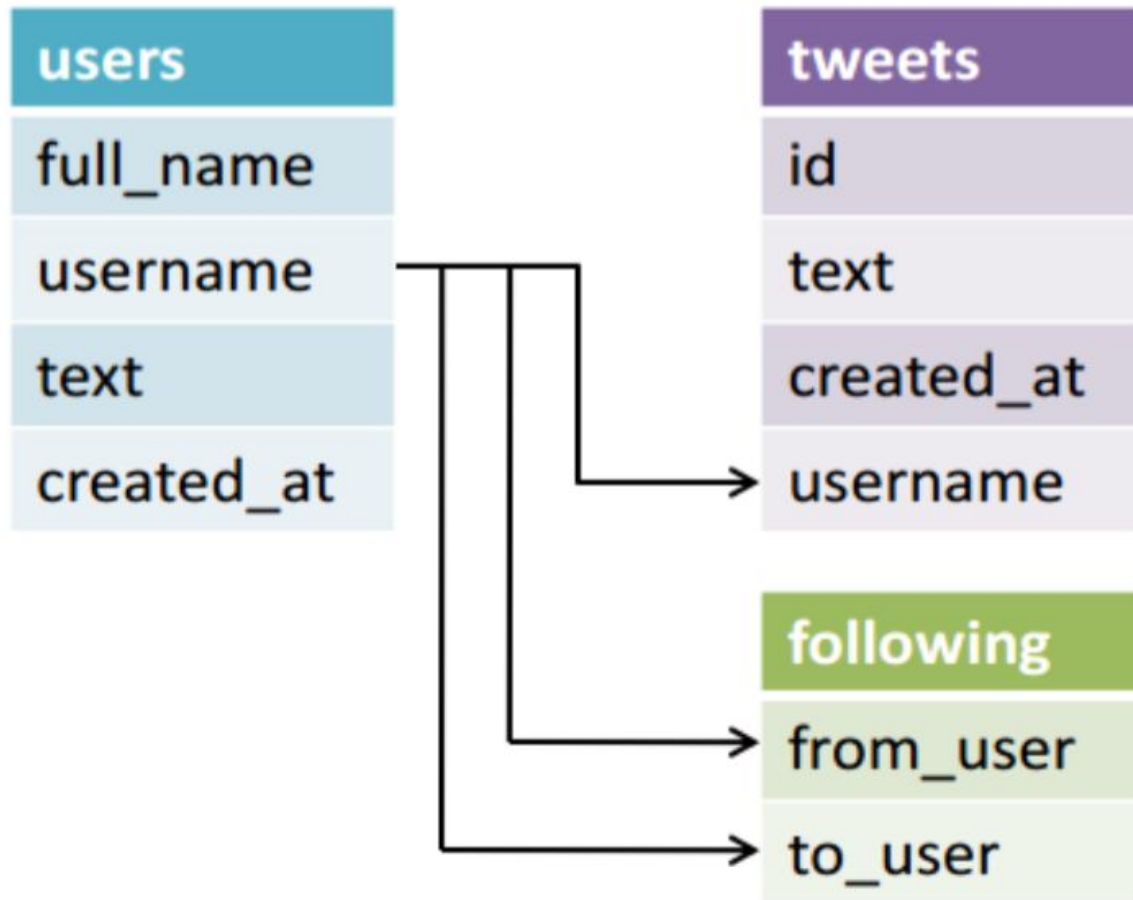
Joe Smith	Hello, world!
-----------	---------------

Joe Smith	Just tweetin'
-----------	---------------

Michelle Ng	I am eating pizza tonight
-------------	---------------------------

JOINS

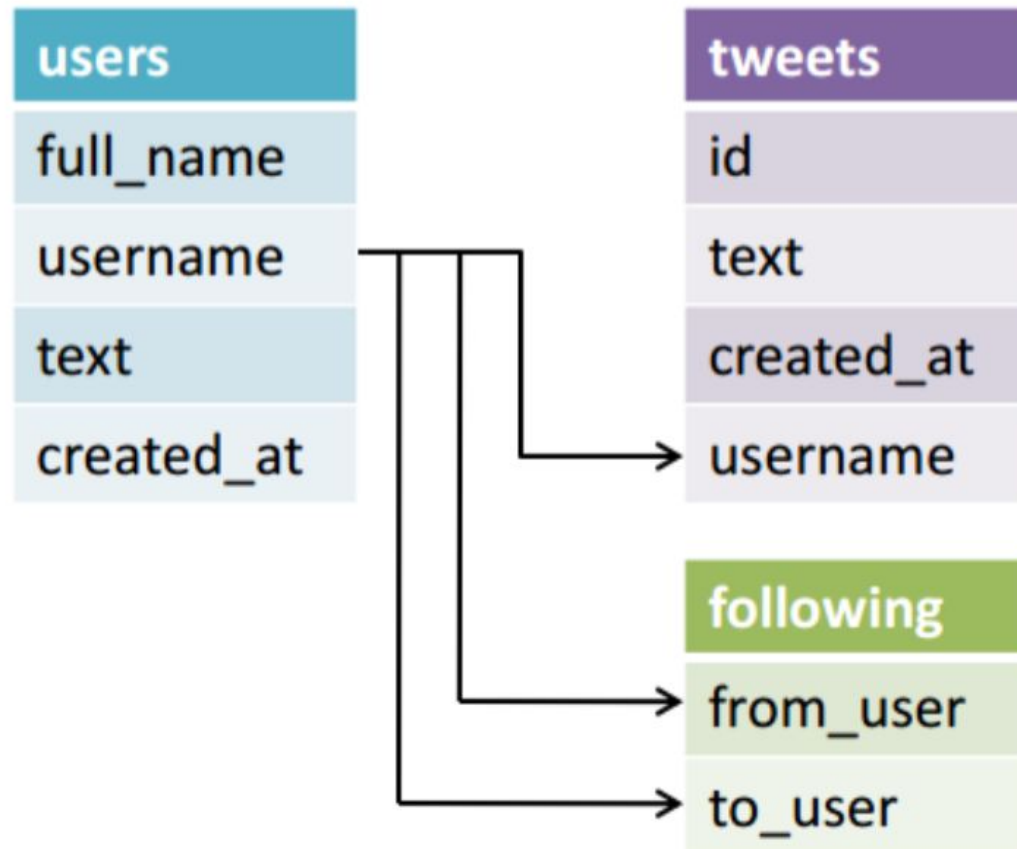
Create a table with all the users' full names and their tweets



```
SELECT
  users.full_name,
  tweets.text
```

JOINS

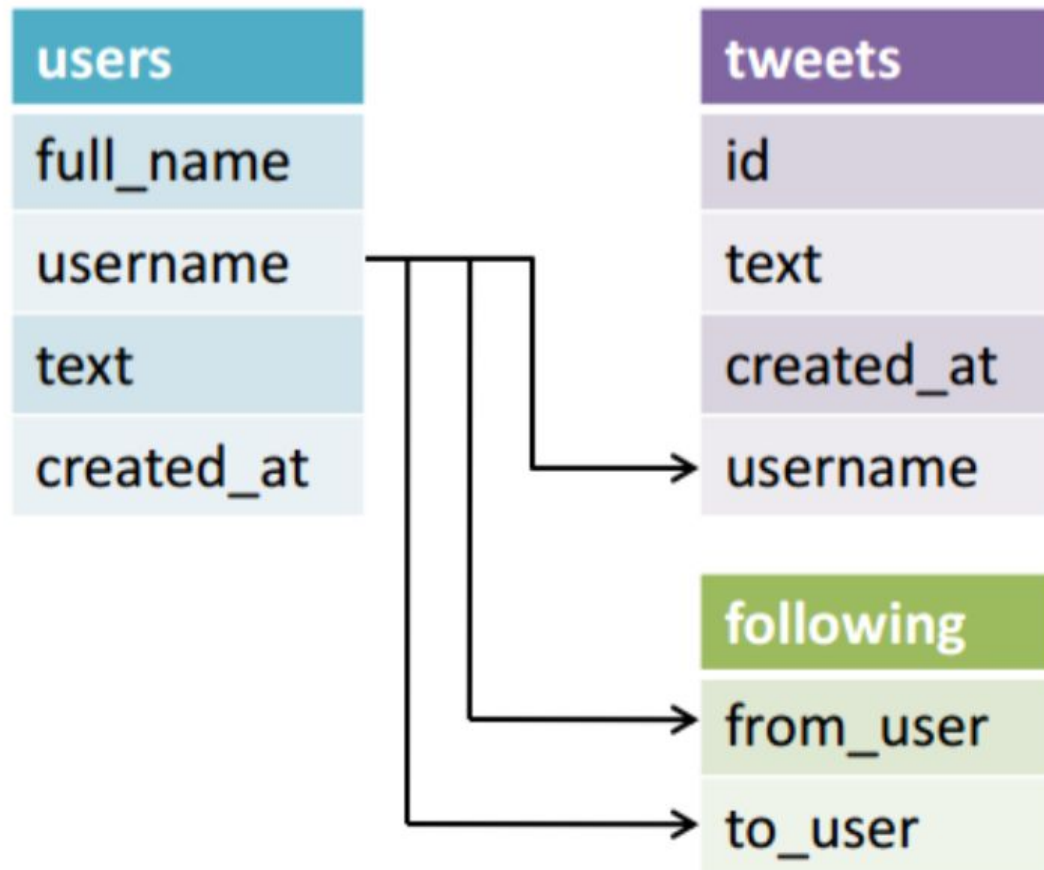
Create a table with all the users' full names and their tweets



```
SELECT
    users.full_name,
    tweets.text
FROM users
JOIN tweets
ON users.username = tweets.username
```

JOINS

Create a table with all the users' full names and their tweets



Will users who never tweeted appear in the list?

JOINS

JOIN *will only include entries that occur in both tables.*

```
SELECT
  full_name,
  text
FROM users
JOIN tweets
ON users.username = tweets.username
```

<u>full_name</u>	<u>tweet</u>
Joe Smith	Hello, world!
Joe Smith	Just tweetin'
Michelle Ng	I am eating pizza tonight

JOINS

LEFT JOIN *will always include all entries from the left table, even if there are no matches in the other table.*

```
SELECT
    full_name,
    text
FROM users
LEFT JOIN tweets
ON users.username = tweets.username
```

<u>full_name</u>	<u>tweet</u>
Joe Smith	Hello, world!
Joe Smith	Just tweetin'
Michelle Ng	I am eating pizza tonight
Jim Rogers	

JOINS

FULL OUTER JOIN *will always include all entries from both tables, even if there are no matches in the other table.*

```
SELECT
    full_name,
    text
FROM users
FULL OUTER JOIN tweets
ON users.username = tweets.username
```

<u>full_name</u>	<u>tweet</u>
Joe Smith	Hello, world!
Joe Smith	Just tweetin'
Michelle Ng	I am eating pizza tonight
Jim Rogers	OK, deleting my account

JOINS

The holes in the resulting table are called **NULLs**.

NULL indicates missing data.

Note that **NULL** is not the same as zero or an empty string "", it really means that there is no data.

<u>full_name</u>	<u>tweet</u>
Joe Smith	Hello, world!
Joe Smith	Just tweetin'
Michelle Ng	I am eating pizza tonight
Jim Rogers	NULL
NULL	OK, deleting my account

JOINS

For example, to print a list of users without tweets, we'd write

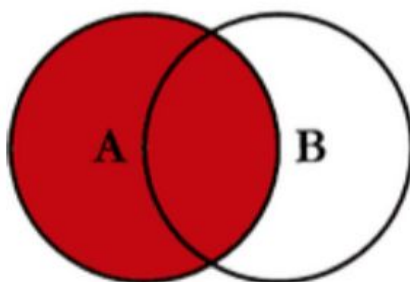
```
SELECT full_name  
FROM users  
FULL OUTER JOIN tweets  
ON users.username = tweets.username  
WHERE tweets.text IS NULL
```



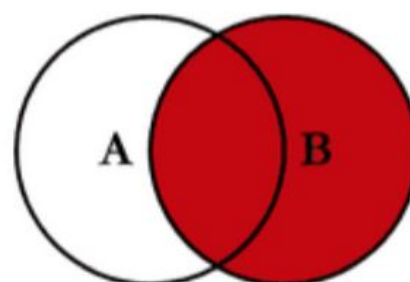
<u>full_name</u>
Jim Rogers

JOINS

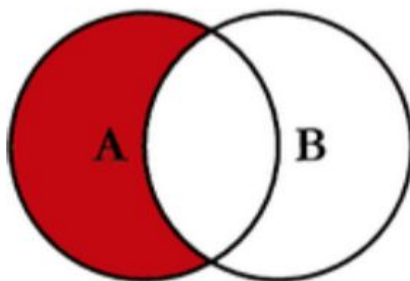
SQL JOINS



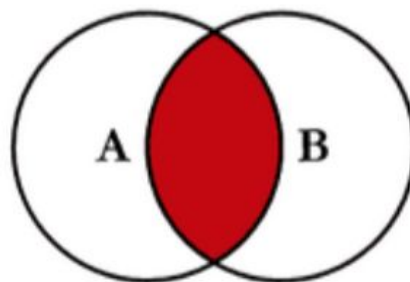
```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key
```



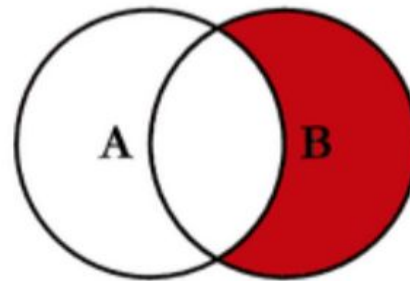
```
SELECT <select_list>  
FROM TableA A  
RIGHT JOIN TableB B  
ON A.Key = B.Key
```



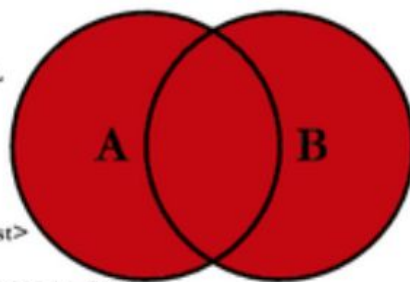
```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key  
WHERE B.Key IS NULL
```



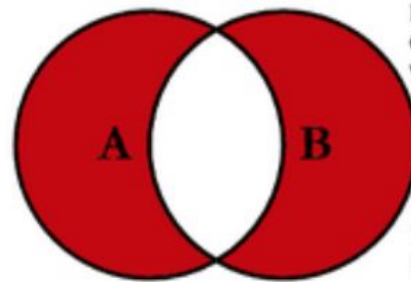
```
SELECT <select_list>  
FROM TableA A  
INNER JOIN TableB B  
ON A.Key = B.Key
```



```
SELECT <select_list>  
FROM TableA A  
RIGHT JOIN TableB B  
ON A.Key = B.Key  
WHERE A.Key IS NULL
```



```
SELECT <select_list>  
FROM TableA A  
FULL OUTER JOIN TableB B  
ON A.Key = B.Key
```



```
SELECT <select_list>  
FROM TableA A  
FULL OUTER JOIN TableB B  
ON A.Key = B.Key  
WHERE A.Key IS NULL  
OR B.Key IS NULL
```

JOINS

```
SELECT <columns>  
FROM <table>  
JOIN <otherTable>  
ON <table.key> = <otherTable.key>  
JOIN <yetAnotherTable>  
ON <otherTable.key> = <yetAnotherTable.key>
```

General SQL structure

NOTE

You can combine as many **JOINS** as you want!

JOINS

General SQL structure

SELECT <columns>
FROM <table>
[INNER|LEFT|RIGHT|FULL OUTER] JOIN <otherTable>
ON <table.key> = <otherTable.key>
WHERE <condition>
GROUP BY <columns>
HAVING <condition>
ORDER BY <columns> [DESC|ASC]
LIMIT <number>

JOINS

We could have had a table structure as follow:

Why is this different?

tweets
id
text
created_at
username
full_name
username
text
created_at

JOINS

We could have had a table structure as follow:

Why is this different?

We would repeat the user information on each row.

This is called
denormalization

tweets
id
text
created_at
username
full_name
username
text
created_at

JOINS

Normalized Data:

Many tables to reduce redundant or repeated data in a table

Denormalized Data:

Wide data, fields are often repeated

but removes the need to join together multiple tables

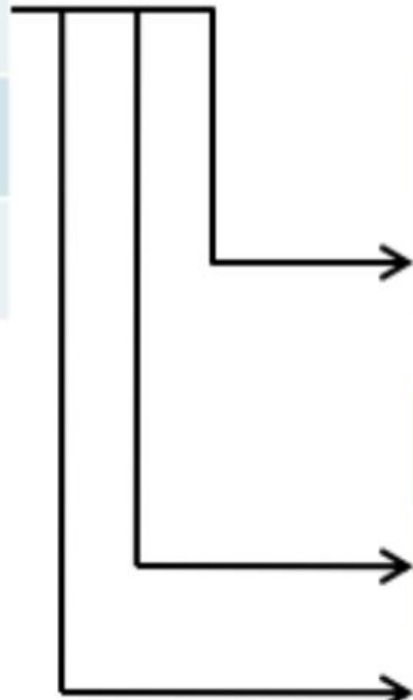
JOINS

users
full_name
username
text
created_at

tweets
id
text
created_at
username

following
from_user
to_user

tweets
id
text
created_at
username
full_name
username
text
created_at



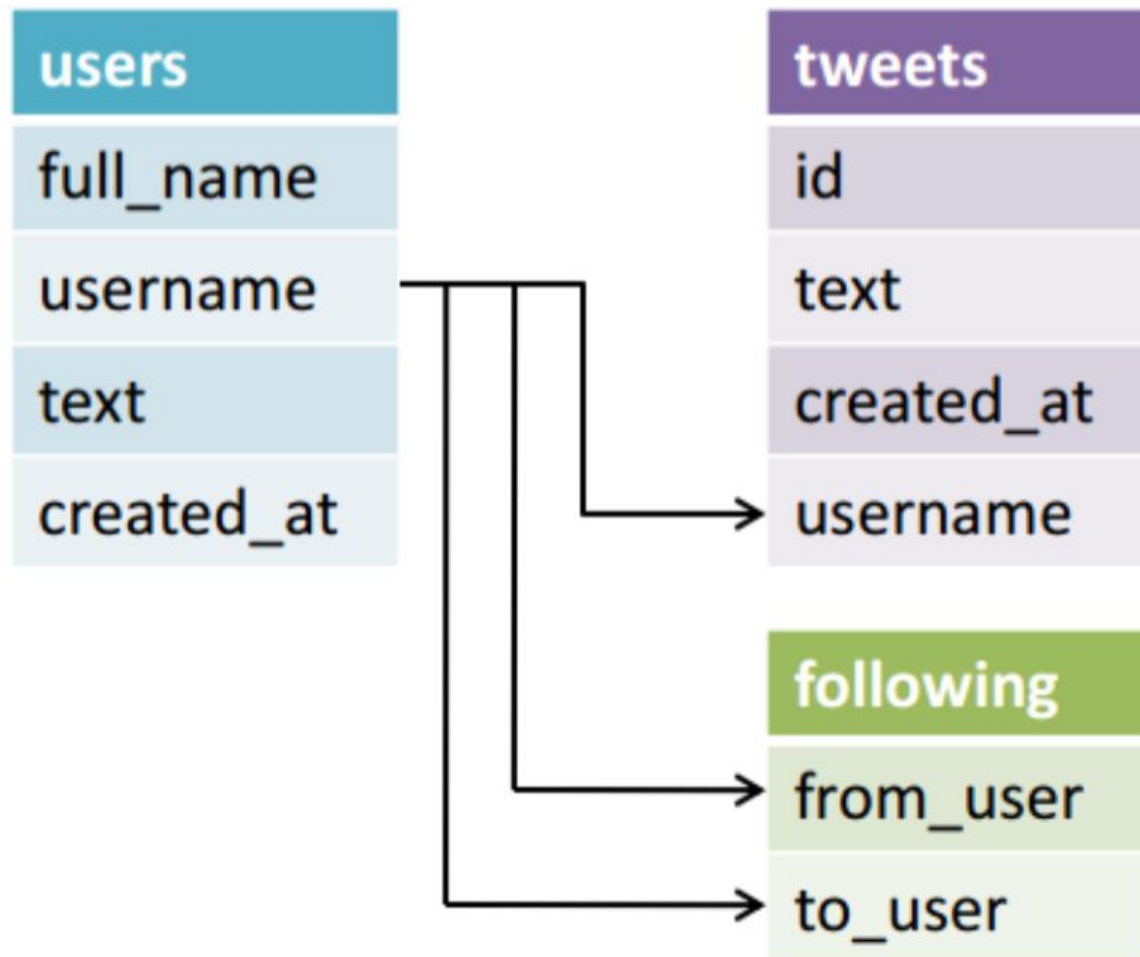
JOINS

Q: How do we commonly evaluate databases?

- ▶ *read-speed vs. write speed*
- ▶ *space considerations*
- ▶ *(...and many other criteria)*

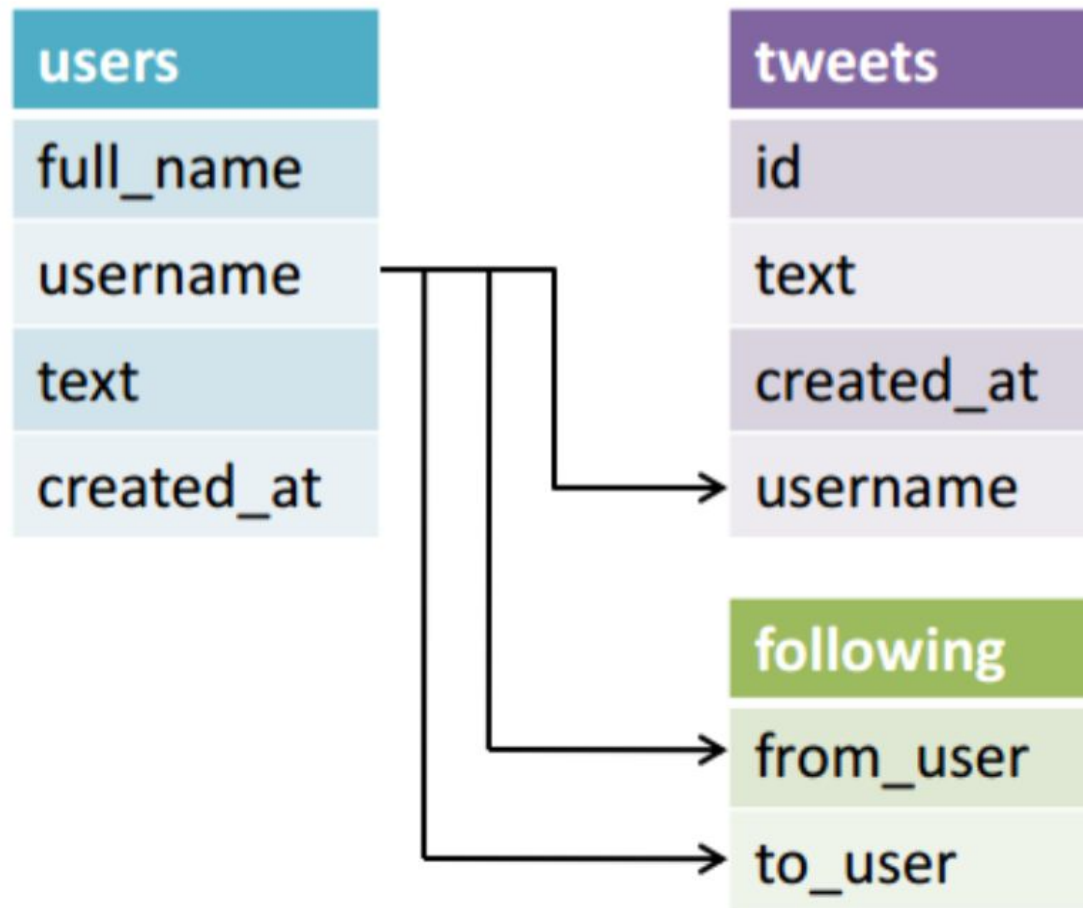
JOINS

Q: Why are normalized tables (possibly) slower to read?



JOINS

Q: Why are normalized tables (possibly) slower to read?



We'll have to get data from multiple tables to answer some questions

JOINS

Q: Why are denormalized tables (possibly) slower to write?

tweets
id
text
created_at
username
full_name
username
text
created_at

JOINS

Q: Why are denormalized tables (possibly) slower to write?

tweets
id
text
created_at
username
full_name
username
text
created_at

*We'll have to write more data
each time we store something*

NEXT STEPS

NEXT STEPS FOR WEB DEVELOPERS

Install the [LAMP Stack](#)

You now know the M in LAMP!

Try creating a site and using SQL to store your data

NEXT STEPS FOR DATA ANALYSTS / DATA SCIENTISTS

Get a toy dataset

Use in conjunction with a data visualization tool like Excel or Tableau

Excel has [Power Query](#) and Tableau has native SQL support!

NEXT STEPS FOR BACKEND DEVELOPERS

Get a toy dataset and manipulate using your favorite scripting language, like Python (I'm biased!)

OPINION

Be weary of people telling you that SQL is “too slow” or “too fast”

Make sure they’re not trying to sell you something!

OPINION

You should not be “cleaning” your data in SQL

You should not be “formatting” your data in SQL for visualization

SQL is only for storing and retrieving raw data!

OPINION

You now know probably 80% of the SQL you need

Many people will never need to use the remaining 20% EVER

Make sure your skills don't atrophy!

KEEP IN TOUCH

If you enjoyed my class, I am starting a blog:

Follow me on [Twitter](#)

If you're on Github:

Follow [Me](#)

If you're on LinkedIn:

Connect with [Me](#)