

Introduction to Databases

Spatial Data Capture
Data Science for Spatial Systems



Ollie Ballinger

First, an exercise

- You are the head of the UK's National Health Service (NHS).
 - There are 1000 COVID testing centers, each keeping track of positive and negative results.
 - You need to collect data from these centers, and share them with other government departments
-
1. What program do you use?
 2. What file format do you use?
 3. How do you collect data from test centers, and how do you share them to other parts of the government?

Covid: how Excel may have caused loss of 16,000 test results in England

A million-row limit on Microsoft's Excel spreadsheet software may have led to Public Health England misplacing nearly 16,000 Covid test results, it is understood.

The data error, which led to **15,841 positive tests being left off the official daily figures**, means than 50,000 potentially infectious people may have been missed by contact tracers and not told to self-isolate.

In this case, the Guardian understands, one lab had sent its daily test report to PHE in the form of a CSV file - the simplest possible database format, just a list of values separated by commas. That report was then loaded into Microsoft Excel, and the new tests at the bottom were added to the main database.

Three key advantages of Databases

1. Access

- Many people can connect to the same database
 - You can set privileges so that some people are read only, and others can edit.
- Since everyone's working on the same version, your data is always up to date

2. Structure

- You can store data in a relational way, avoiding unnecessary repetition
- You can store lots of data

3. Efficient Querying

- Syntax is very simple and human-readable
- SQL is a 4th generation programming language
 - Rather than telling it what to *do*, you tell it what you *want*. It then figures out the best way of acquiring that information.

Outline

1. What is a Database?
2. Access
3. Structure
4. Querying

1. What is a Database?

Imagine a library

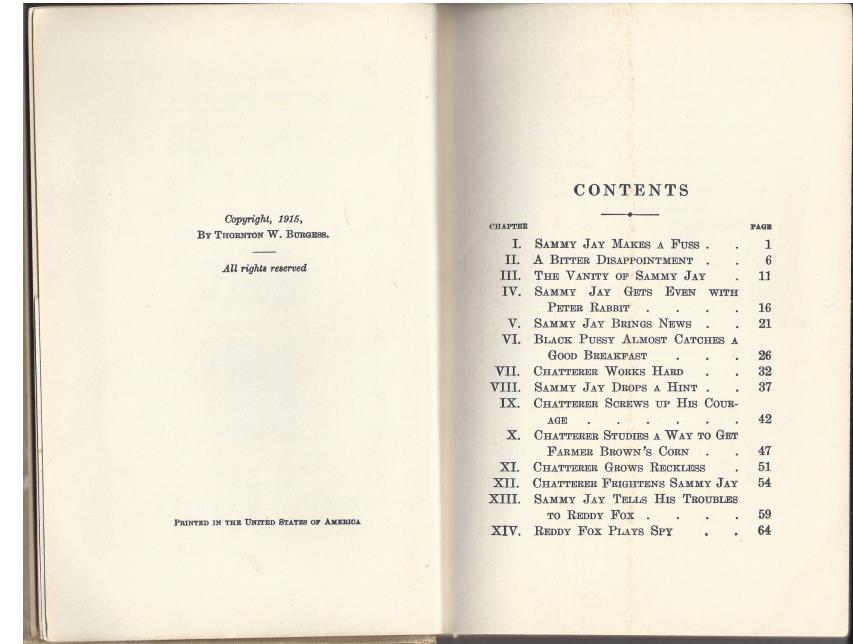
Librarian



Shelves



Books



Imagine a library

Librarian



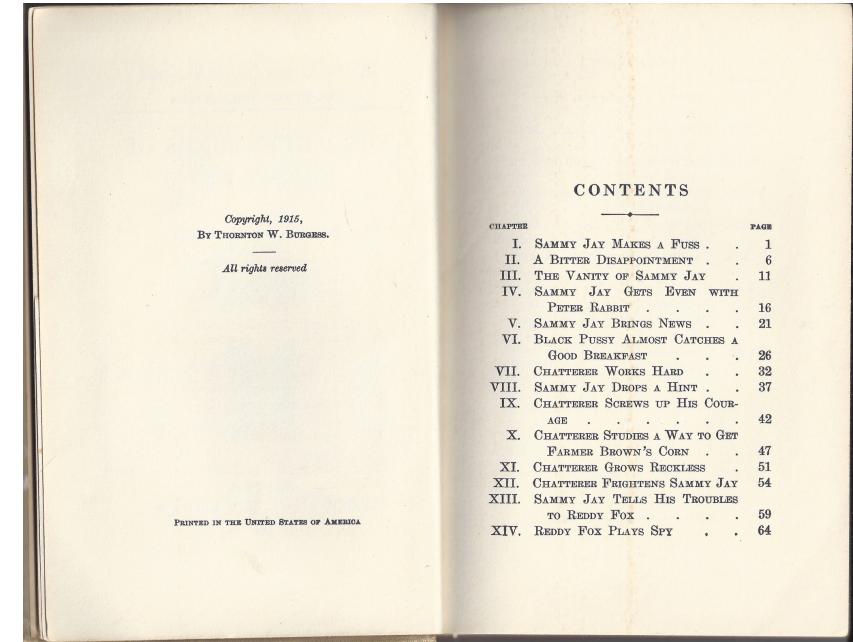
DBMS

Shelves



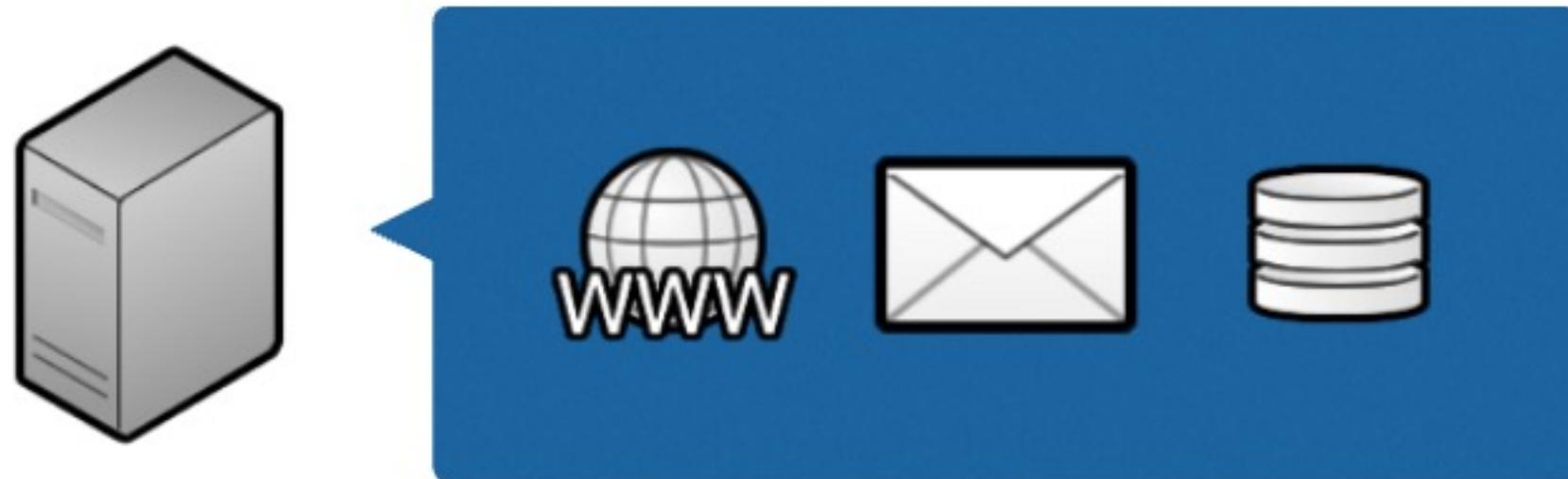
Databases

Books



Tables

What is a Database?



Servers run applications

What's in a database?

- Databases contain multiple tables

Customers		
Model: Person		
Customers have many Orders		
customerID	integer	PK
customer_name	varchar	
contact_name	varchar	
address	varchar	
city	varchar	
county	varchar	
postcode	varchar	
country	varchar	

Orders		
Model: Orders		
An order has 1 customer		
OrderID	int	PK
CustomerID	int	SK
OrderDate	datetime	
OrderTotalCost	integer	
OrderTax	integer	

Tweets		
Model: Tweets		
A user has many tweets		
twitterID	int	PK
twitterpost	data_type	
name	data_type	

Database Management System (DBMS)

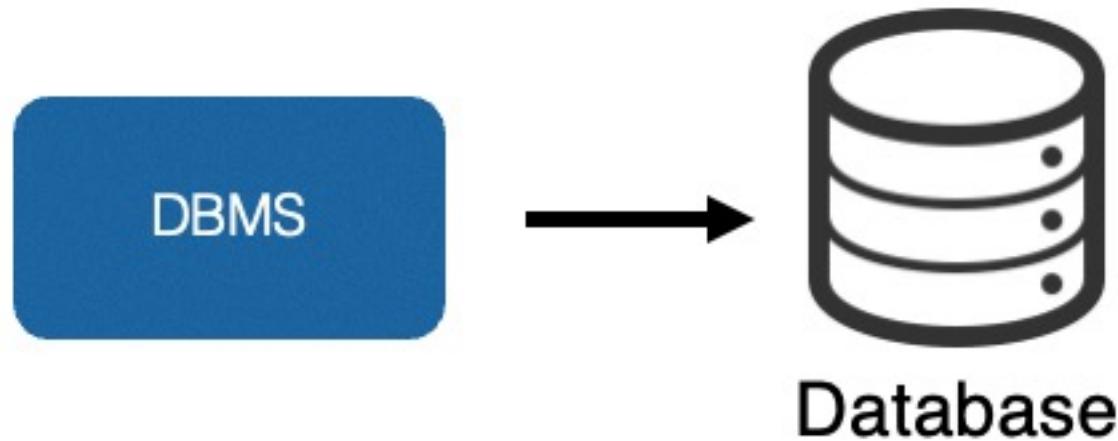
- Our Librarian is our window to the data stored in the database



Database

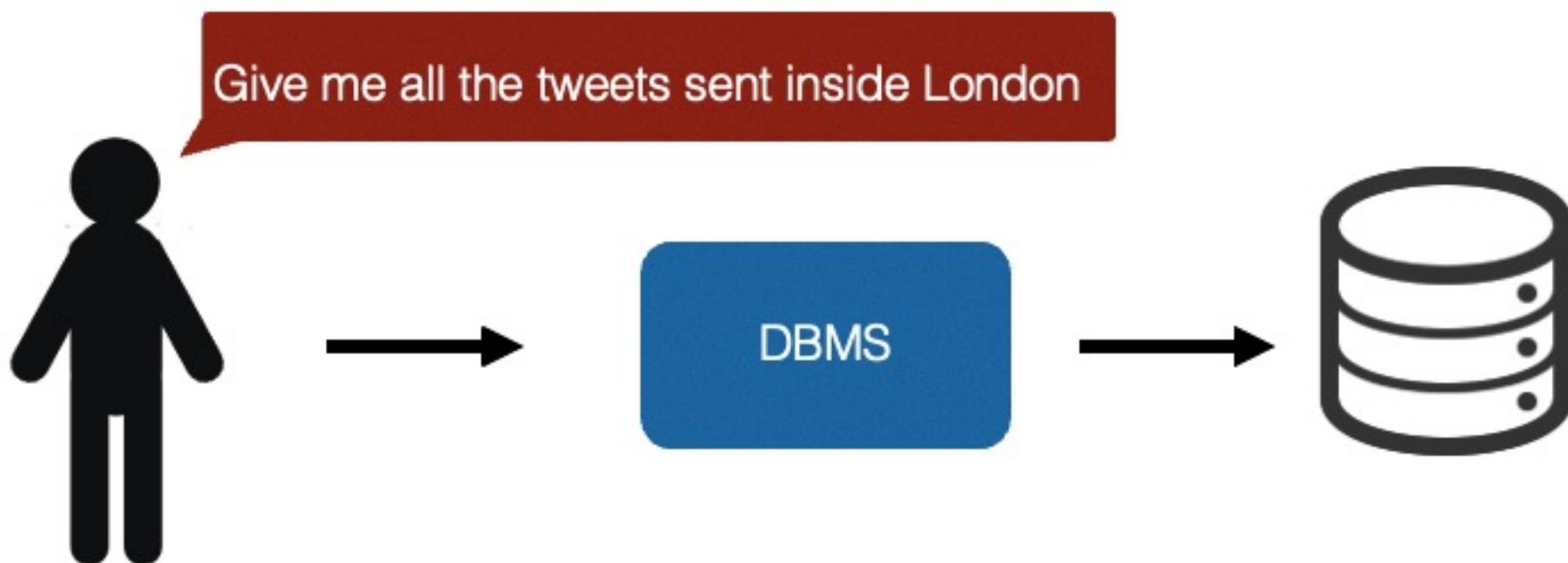
DBMS: what happens when you ask for data

- The DBMS sits in front of the database



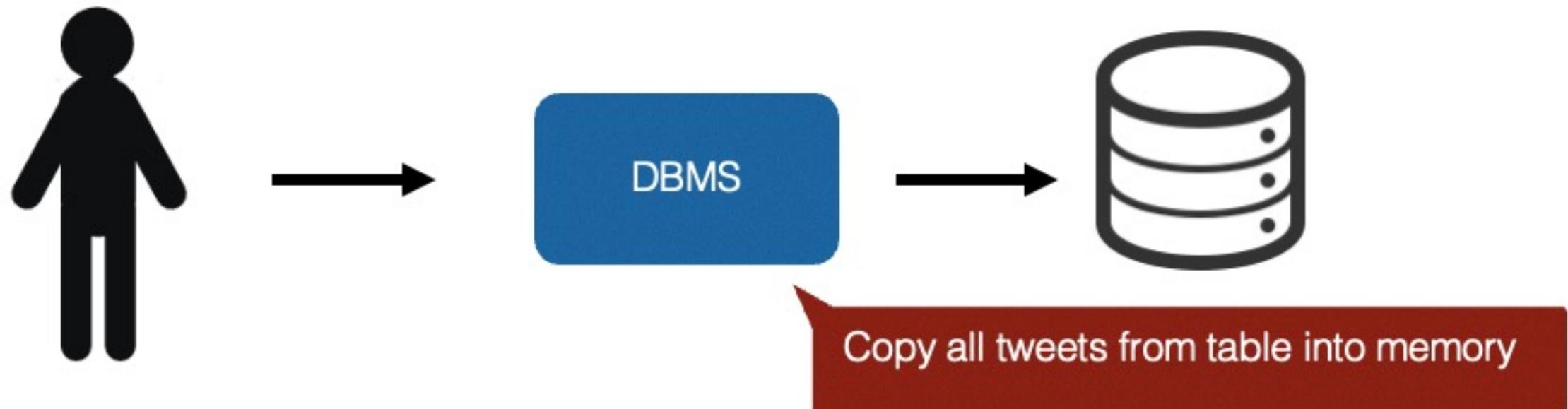
DBMS: what happens when you ask for data

- You ask it something



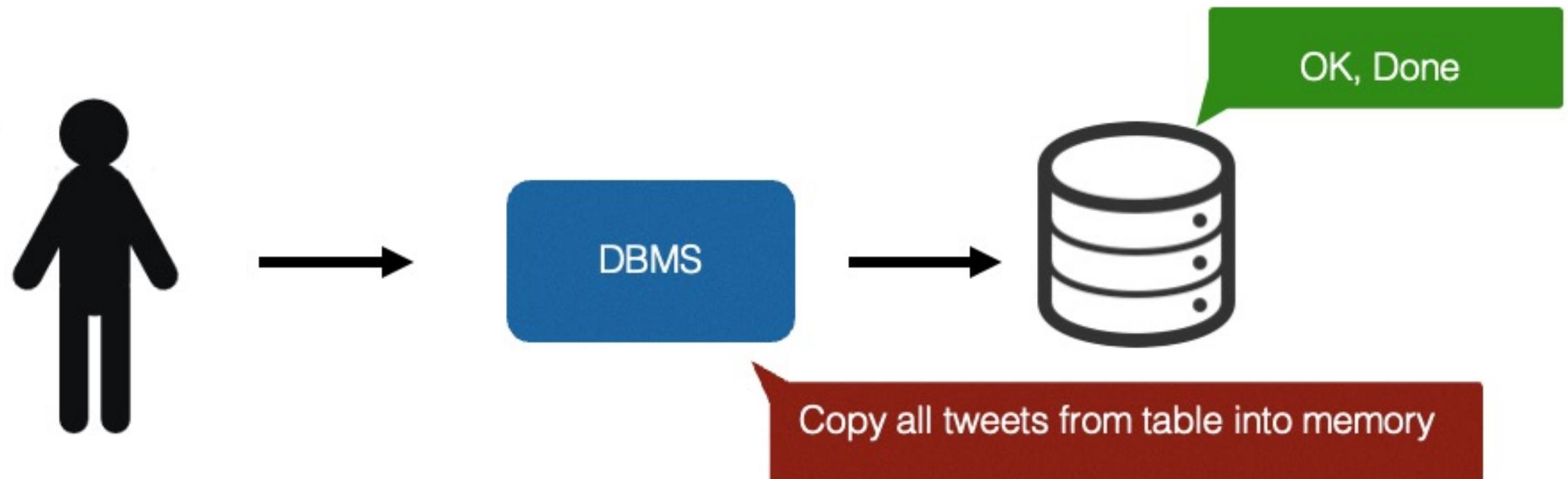
DBMS: what happens when you ask for data

- The DBMS will translate your query



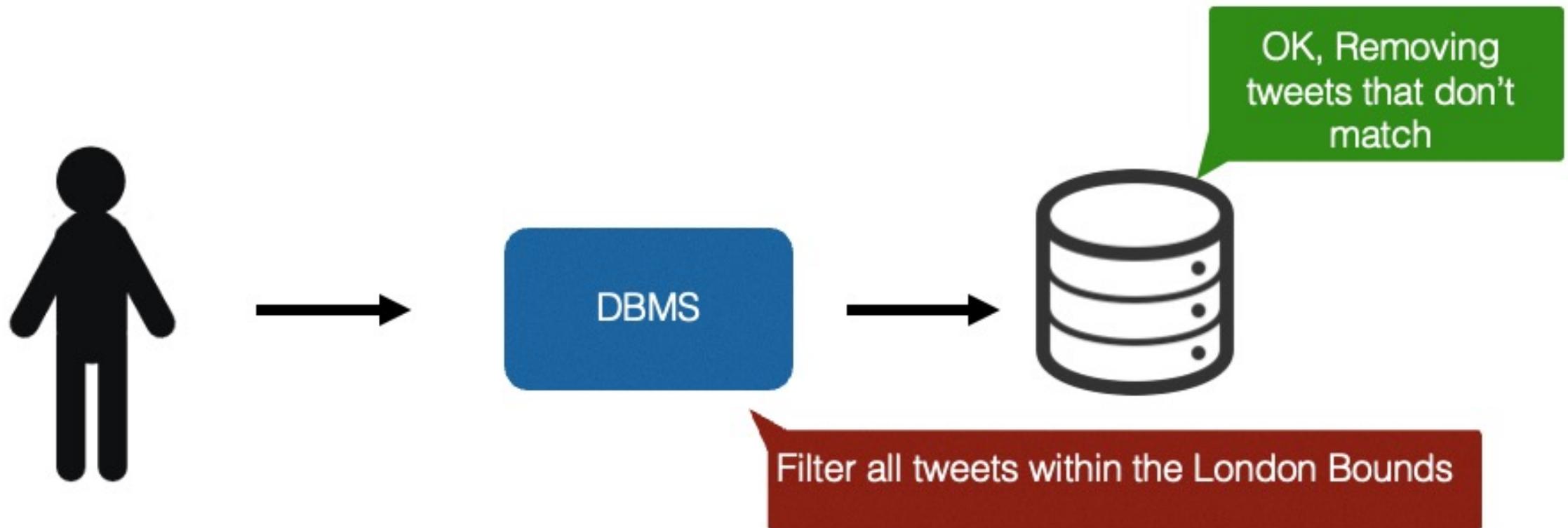
DBMS: what happens when you ask for data

- The DBMS will translate your query



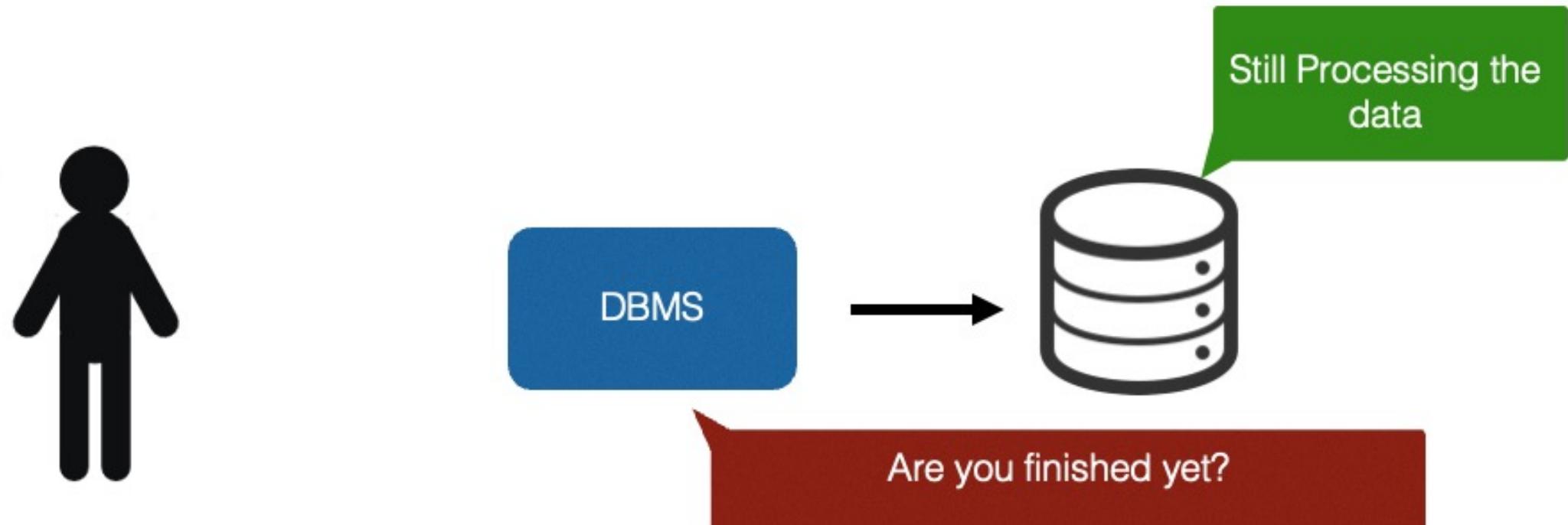
DBMS: what happens when you ask for data

- The DBMS will translate your query



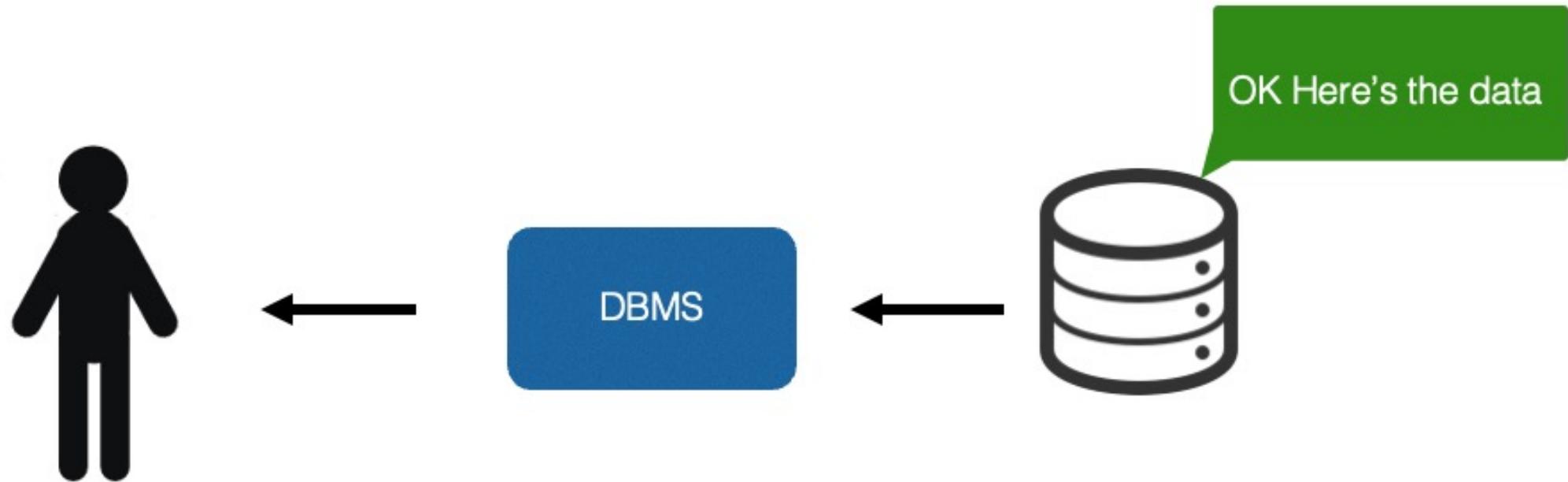
DBMS: what happens when you ask for data

- The DB looks for the data on the disk and hands it back to the DBMS



DBMS: what happens when you ask for data

- The user gets the data back



Different Database Management Systems

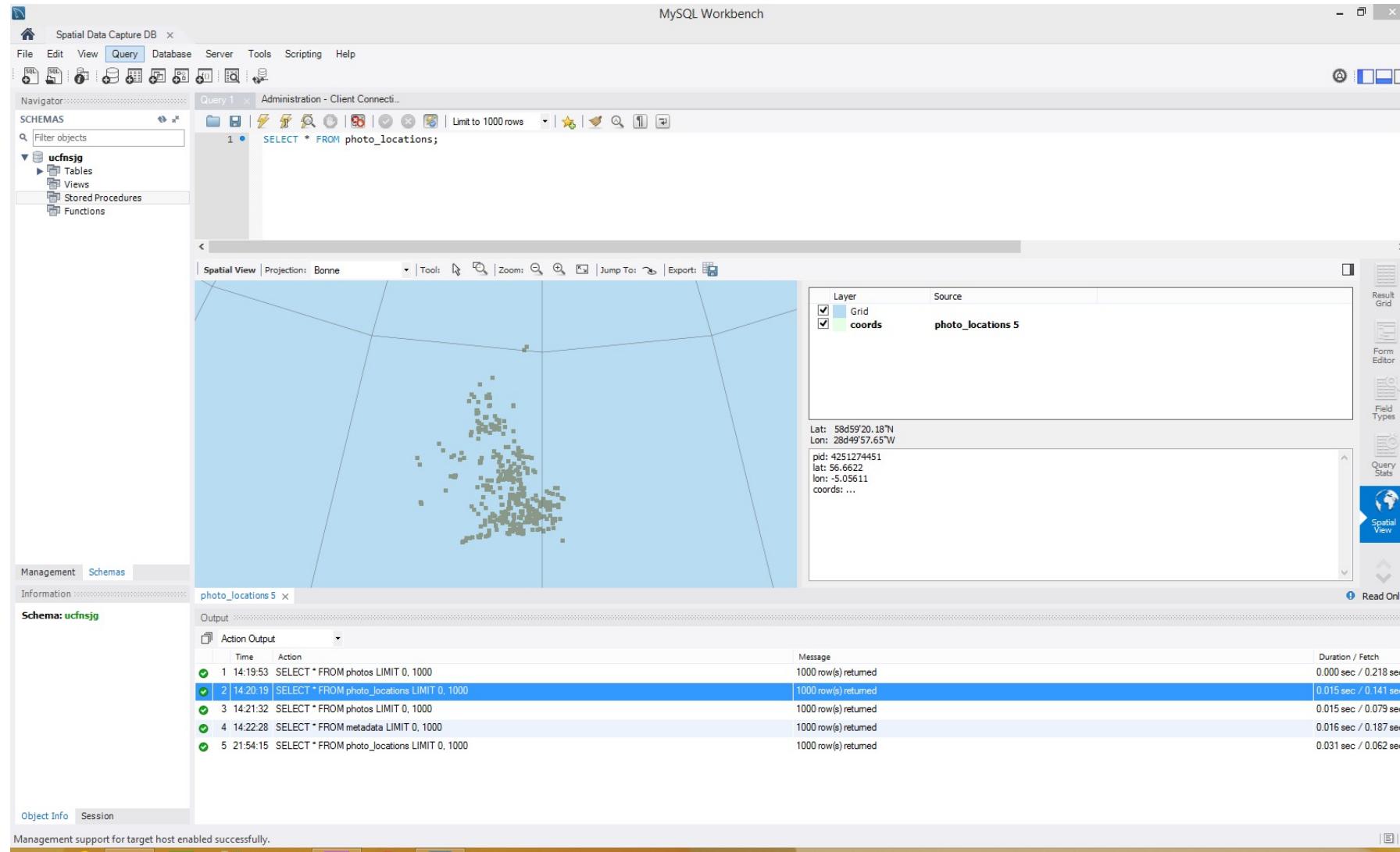
- MySQL
 - The most common database system on the web
- MSSQL / Oracle
 - Most common enterprise DB system
- Postgres
 - With PostGIS Extensions for spatial data
- NoSQL Servers: Big Query, MongoDB
 - In Memory Databases that respond quicker than traditional databases
- We'll be working with MySQL in this module

2. Database Access

Finding the Library

- Addressing System
 - IP or Domain Name -> Address to the Library
- Port Numbers
 - Large Apartment Block => Got the find the right door
- A database is software running on a computer or server
- Our server lives at 128.40.150.34:3306
 - bart150-34.bartlett.ucl.ac.uk:3306
 - dev.spatialdatacapture.org:3306

Database Clients

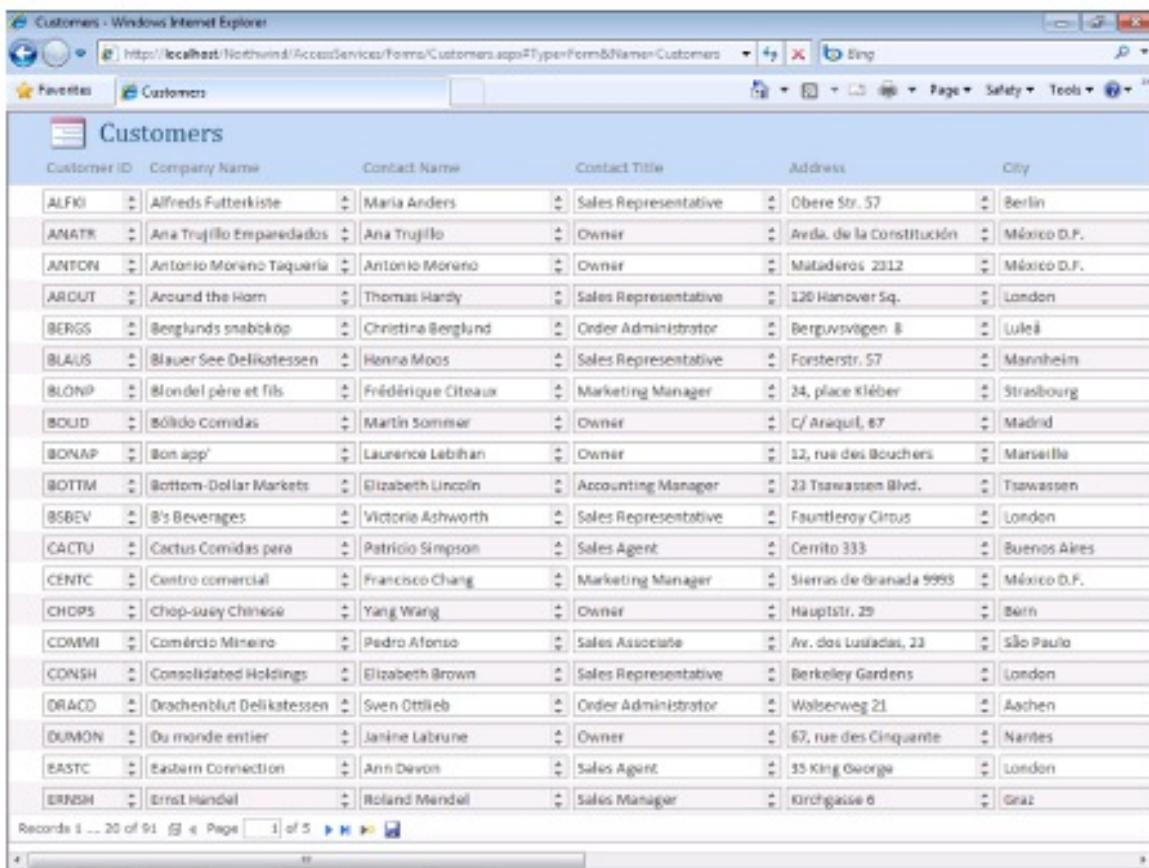


3. Database Structure

Tables

- How is our data stored?

Customers		
Model: Person		
Customers have many Orders		
customerID	integer	PK
customer_name	varchar	
contact_name	varchar	
address	varchar	
city	varchar	
county	varchar	
postcode	varchar	
country	varchar	



The screenshot shows a Microsoft Access application window titled "Customers". The form has a header section with columns for Customer ID, Company Name, Contact Name, Contact Title, Address, and City. Below this is a data grid containing 91 rows of customer data. Each row includes a small icon, a customer ID (e.g., ALFKI, ANATR, ANTON), a company name (e.g., Alfreds Futterkiste, Ana Trujillo Emparedados), a contact name (e.g., Maria Anders, Ana Trujillo), a contact title (e.g., Sales Representative, Owner), an address (e.g., Obere Str. 57, Avenida de la Constitución), and a city (e.g., Berlin, México D.F.). The data grid has scroll bars on the right and bottom, and a status bar at the bottom.

Customer ID	Company Name	Contact Name	Contact Title	Address	City
ALFKI	Alfreds Futterkiste	Maria Anders	Sales Representative	Obere Str. 57	Berlin
ANATR	Ana Trujillo Emparedados	Ana Trujillo	Owner	Avenida de la Constitución	México D.F.
ANTON	Antonio Moreno Taquería	Antonio Moreno	Owner	Mutadero 2312	México D.F.
AROUT	Around the Horn	Thomas Hardy	Sales Representative	120 Hanover Sq.	London
BERGS	Berglunds snabbköp	Christina Berglund	Order Administrator	Berguvsvägen 8	Luleå
BLAUS	Blauer See Delikatessen	Hanna Moos	Sales Representative	Forsterstr. 57	Mannheim
BLONP	Blondel père et fils	Fridérique Côteaux	Marketing Manager	24, place Kléber	Strasbourg
BOLID	Bólido Comidas	Martin Sommer	Owner	C/ Araquil, 67	Madrid
BONAP	Bon app'	Laurence Lebihan	Owner	12, rue des Bouchers	Marseille
BITTM	Bottom-Dollar Markets	Elizabeth Lincoln	Accounting Manager	23 Tsawassen Blvd.	Tsawassen
BSBEV	B's Beverages	Victoria Ashworth	Sales Representative	Fauntleroy Circus	London
CACTU	Cactus Comidas para	Patricia Simpson	Sales Agent	Cerrito 333	Buenos Aires
CENTC	Centro comercial	Francisco Chang	Marketing Manager	Sierras de Granada 9999	México D.F.
CHOPS	Chop-suey Chinese	Yang Wang	Owner	Hauptstr. 29	Bern
COMMI	Comércio Mineiro	Pedro Alonso	Sales Associate	Av. dos Luisias, 23	São Paulo
CONSH	Consolidated Holdings	Elizabeth Brown	Sales Representative	Berkley Gardens	London
DRAID	Drschenblat Delikatessen	Sven Ottieb	Order Administrator	Wolfsweg 21	Aachen
DUMON	Du monde entier	Janine Labrune	Owner	67, rue des Cinquante	Nantes
EASTC	Eastern Connection	Arin Devon	Sales Agent	35 King George	London
ERINSH	Ernst Handel	Roland Mendel	Sales Manager	Kirchgasse 6	Graz

Tables

- We can have indexes to fetch data quickly and link different tables together

Customers		
Model: Person		
Customers have many Orders		
customerID	integer	PK
customer_name	varchar	
contact_name	varchar	
address	varchar	
city	varchar	
county	varchar	
postcode	varchar	
country	varchar	

Loyalty Card		
Model: Card		
One customer has one card		
customer_id	int	SK
cardnumber	int	
expires	datetime	

Orders		
Model: Orders		
An order has 1 customer		
OrderID	int	PK
CustomerID	int	SK
OrderDate	datetime	
OrderTotalCost	integer	
OrderTax	integer	

Tables

- We can have a one-to-many relationship between tables

Customers		
Model: Person		
Customers have many Orders		
customerID	integer	PK
customer_name	varchar	
contact_name	varchar	
address	varchar	
city	varchar	
county	varchar	
postcode	varchar	
country	varchar	



Orders		
Model: Orders		
An order has 1 customer		
OrderID	int	PK
CustomerID	int	SK
OrderDate	datetime	
OrderTotalCost	integer	
OrderTax	integer	

Tables

- We can have a one-to-one relationship between tables



4. Querying

Talking to the Librarian

STRUCTURED QUERY LANGUAGE

Types of tasks in SQL

- Admin Tasks
 - Creating Databases
 - Creating Users and Passwords
- Data Tasks
 - Getting Data
 - Inserting Data
 - Sorting Data

AND OR ALTER TABLE AS BETWEEN
CREATE DATABASE CREATE TABLE CREATE
INDEX CREATE VIEW DELETE DROP
DATABASE DROP INDEX DROP TABLE
EXISTS GROUP BY HAVING IN INSERT INTO
INNER JOIN LEFT JOIN RIGHT JOIN FULL
JOIN LIKE ORDER BY SELECT SELECT *
SELECT DISTINCT SELECT INTO SELECT
TOP TRUNCATE TABLE UNION UNION ALL
UPDATE WHERE

AND OR ALTER TABLE AS BETWEEN
CREATE DATABASE CREATE TABLE CREATE
INDEX CREATE VIEW DELETE DROP
DATABASE DROP INDEX DROP TABLE
EXISTS GROUP BY HAVING IN INSERT INTO
INNER JOIN LEFT JOIN RIGHT JOIN FULL
JOIN LIKE ORDER BY **SELECT** SELECT *
SELECT DISTINCT SELECT INTO SELECT
TOP TRUNCATE TABLE UNION UNION ALL
UPDATE WHERE

AND OR ALTER TABLE AS BETWEEN
CREATE DATABASE CREATE TABLE CREATE
INDEX CREATE VIEW DELETE DROP
DATABASE DROP INDEX DROP TABLE
EXISTS GROUP BY HAVING IN INSERT INTO
INNER JOIN LEFT JOIN RIGHT JOIN FULL
JOIN LIKE ORDER BY SELECT SELECT *
SELECT DISTINCT SELECT INTO SELECT
TOP TRUNCATE TABLE UNION UNION ALL
UPDATE WHERE

Connected to the Database Server

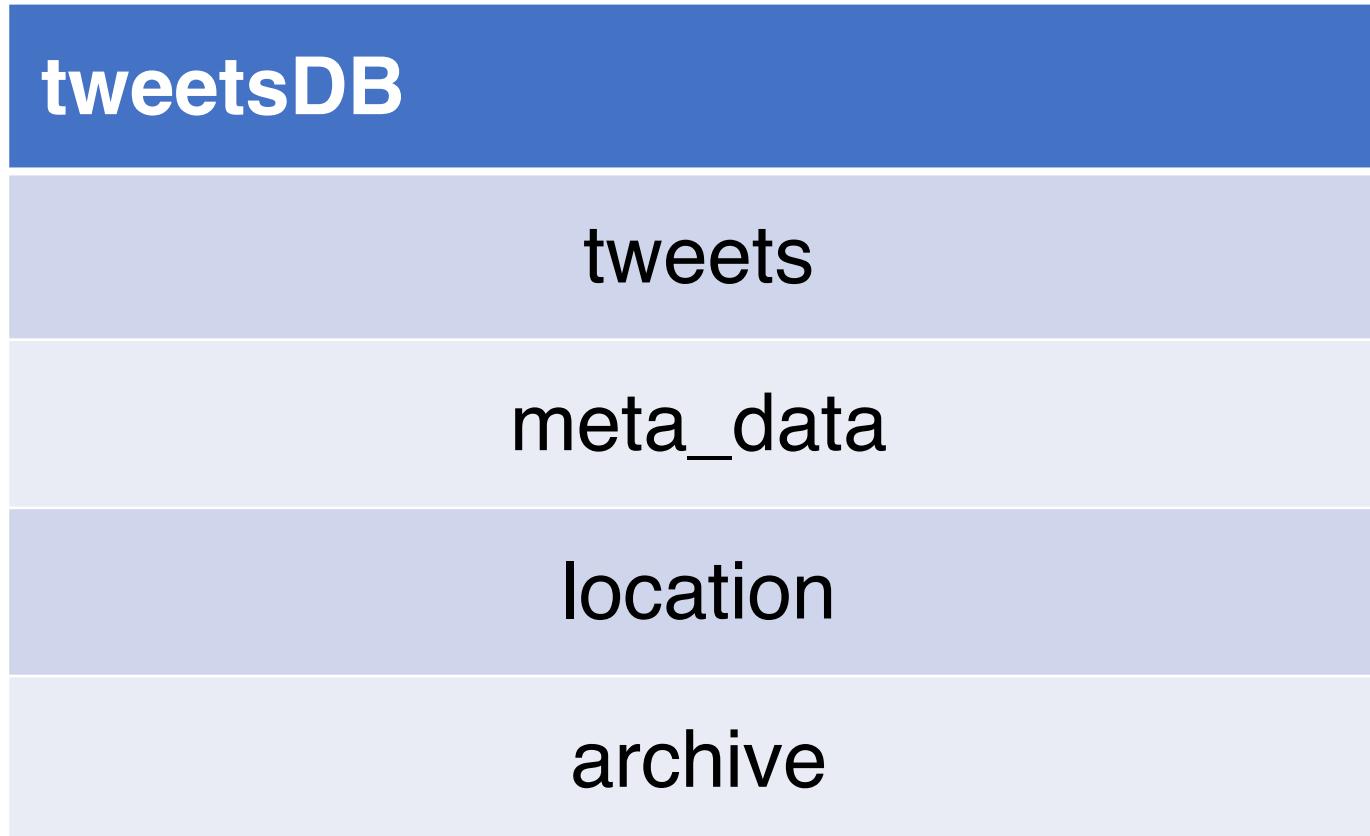
dev.spatialdatacapture.org:3306

CASA_Students

tweetsDB

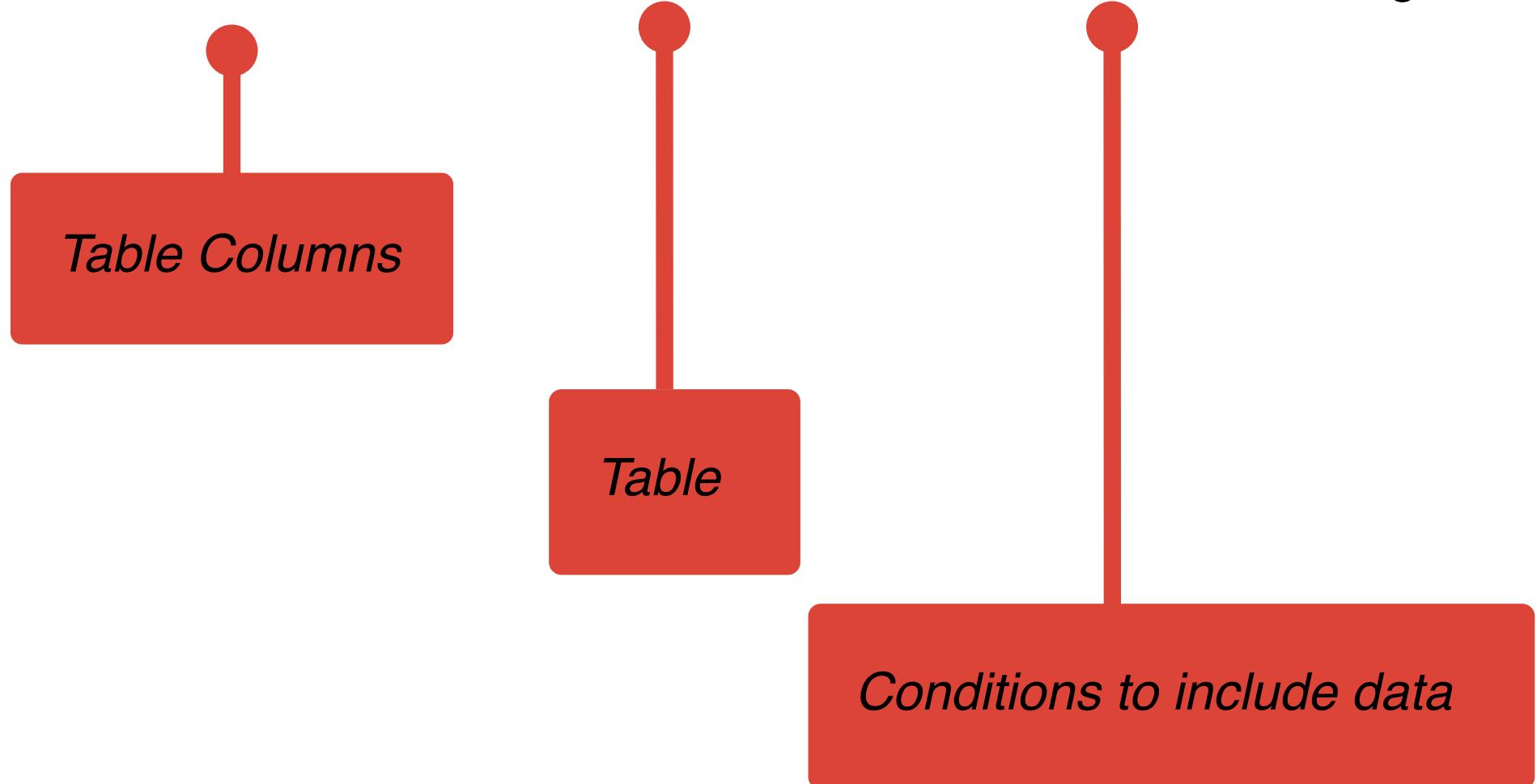
flickr_photos

Inside the Database



Let's build a query

SELECT twitter_id, twitterPost, name **FROM** tweets **WHERE** name = 'frogo'



Output

SELECT twitter_id, twitterPost, name **FROM** tweets **WHERE** name = 'frogo'

twitterID	twitterPost	name
519886361264812032	@HansLovesPixie purple hair would look good ☺	frogo
519886361256796160	Interesting article http://t.co/VCzbdA0wDI	frogo
519886361223258113	@rob_the_bob86 don't think so :-(frogo
519886361076056064	@IamAnooshay lol lol lol	frogo

> No errors: 4 rows affected taking 0.5 ms

Different flavours of SQL

- What happens if I run this command on MySQL?

```
SELECT * FROM tweets LIMIT 10
```

```
> No errors: 10 rows affected taking 0.15 ms
```

Different flavours of SQL

- What happens if I run this command on MSSQL?

```
SELECT * FROM tweets LIMIT 10
```

> No errors: 10 rows affected taking 0.15 ms

Different flavours of SQL

- What happens if I run this command on MSSQL?

SELECT * FROM tweets LIMIT 10

> No errors: 10 rows affected taking 0.15 ms

SELECT TOP 10 * FROM tweets

SQL Create User

```
CREATE USER "ollie@localhost" IDENTIFIED BY 'password';
FLUSH PRIVILEGES;
```

SQL Data Types

Numbers	Text	Dates	Spatial
int	char	date	geometry
float	text	datetime	point
double	tinytext	time	linestring
tinyint	mediumtext	timestamp	polygon
bigint	longtext	year	multipoint
decimal	varchar		multipolygon
...

Workshop: Week 1

- Connect to a Database
 - Understand the GUI
 - Explore the client, learn how to use it
- Explore simple SQL Queries
 - Create Database
 - Import data into your new database
 - Explore imported tables with the GUI
- **Make sure you connect to the UCL VPN!**
- **Make sure you've completed the Database Signup Form (on Moodle)**

Next week: Introduction to SQL

- Any Questions?
- o.ballinger@ucl.ac.uk