# everything you always wanted to know about SE\* but were afraid to ask

(\*software engineering of web apps)

Daniel Jackson

## "Everything you always wanted to know about sex-\*

BUT WERE AFRAID TO ASK 99



#### your goals for today's class

understand some basic notions

client/server systems, APIs, authentication

fill in gaps in your understanding of web apps

HTTP protocol, asynchronous JS

hold your own at cocktail parties

be able to talk REST, Jekyll, local first

## what's client-server?

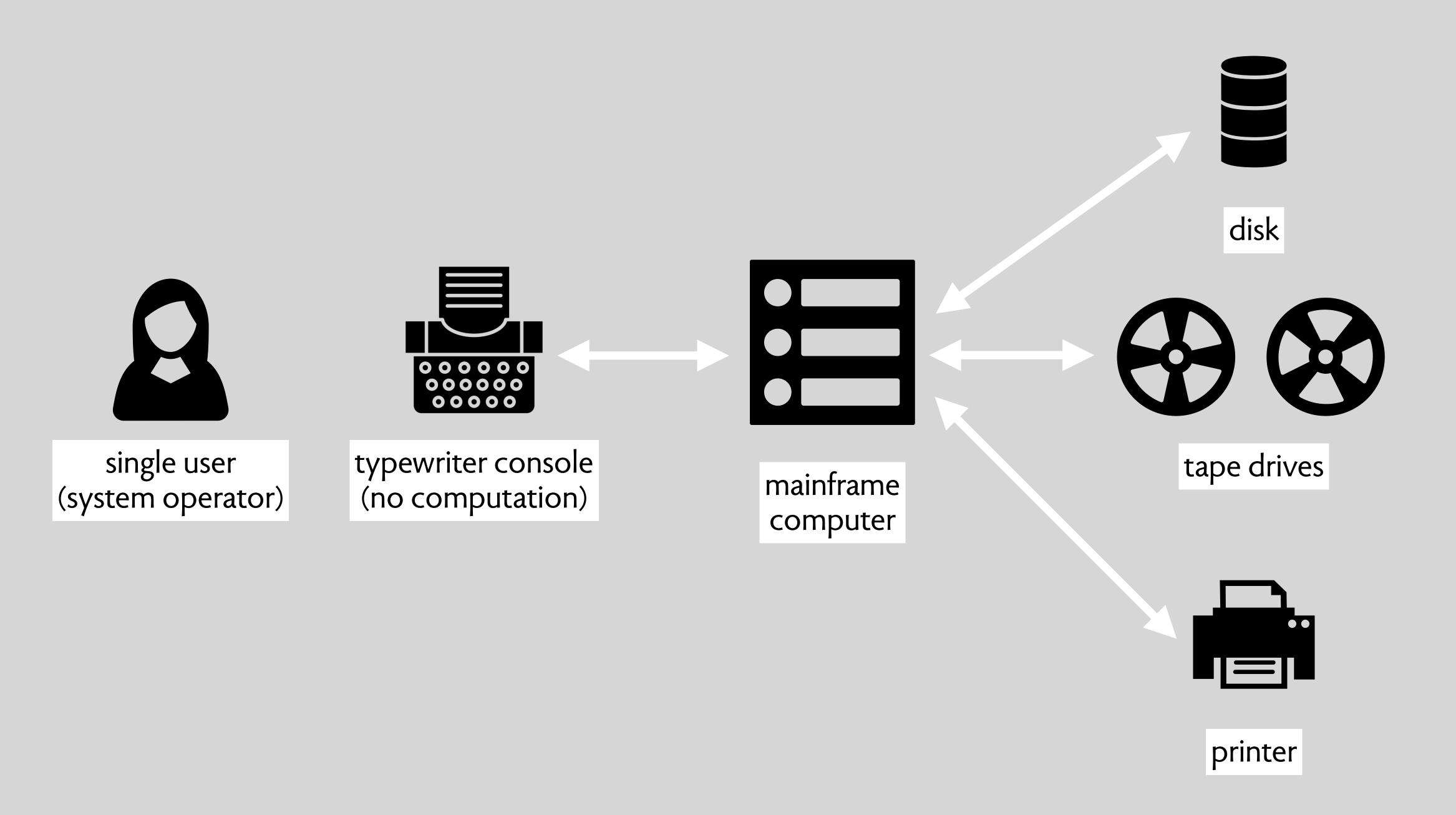
#### mainframe computer



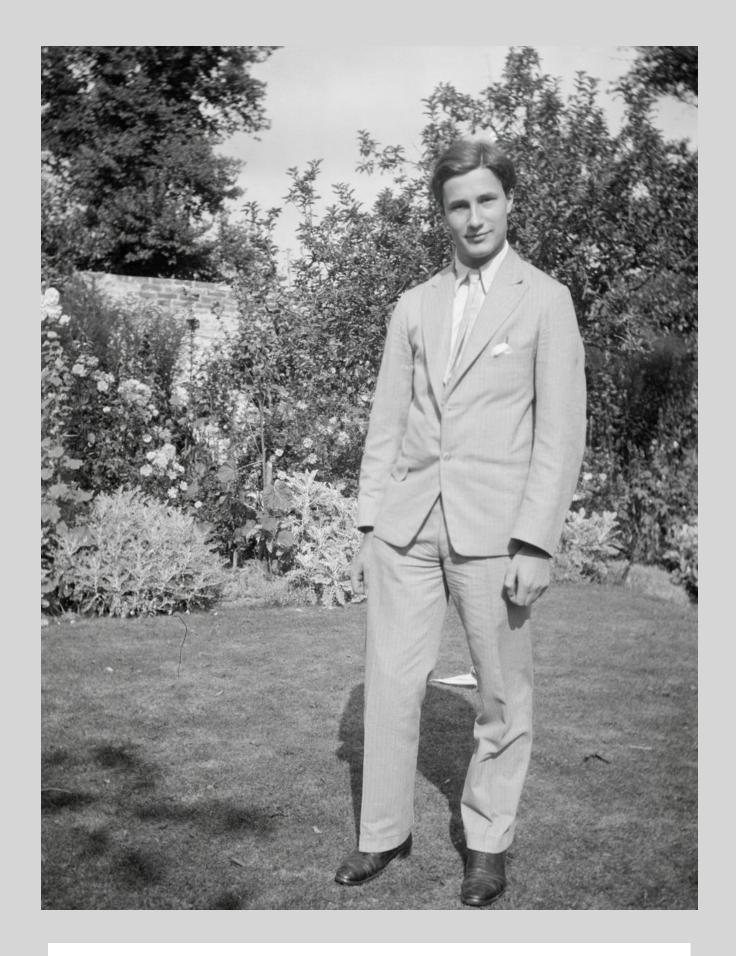
no real client console is wired in part of machine

Univac 1 at Census Bureau: first commercial computer for civilian government agency (1951)

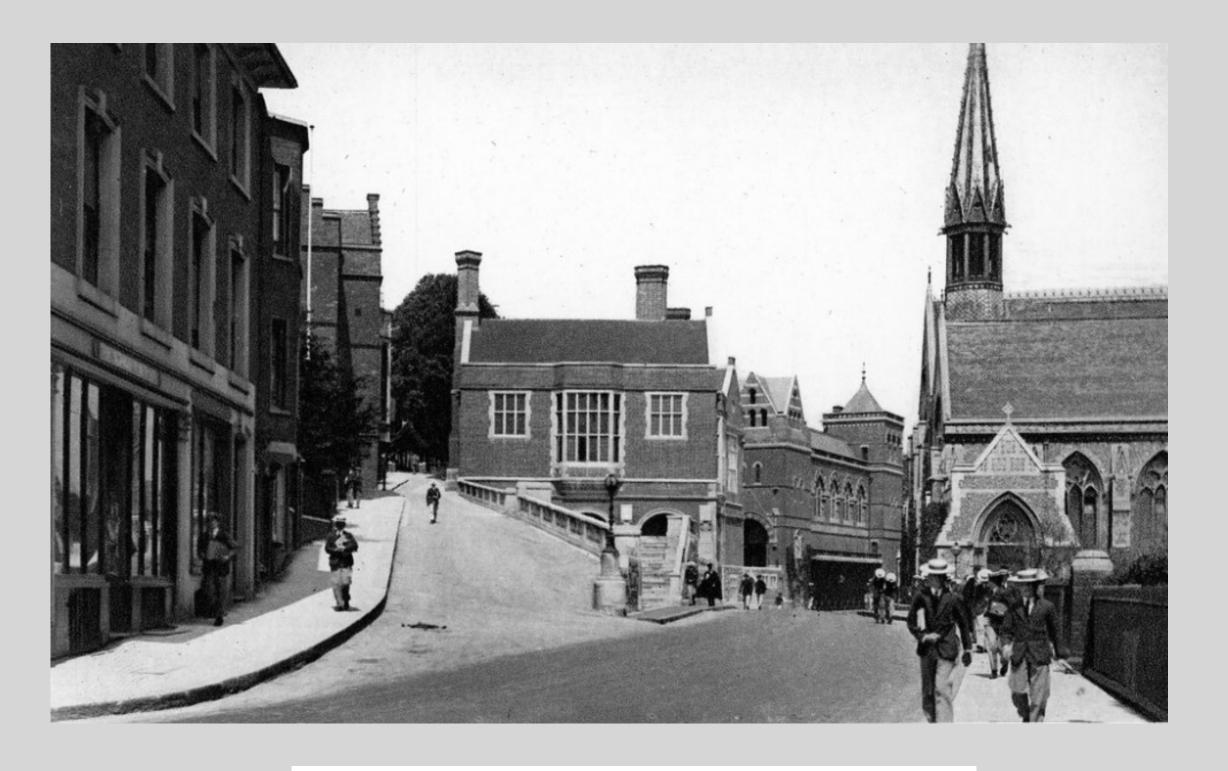
#### early mainframe architecture (1962)



#### the invention of time sharing



Christopher Strachey in 1936 patented time sharing (1959) denotational semantics (1966)



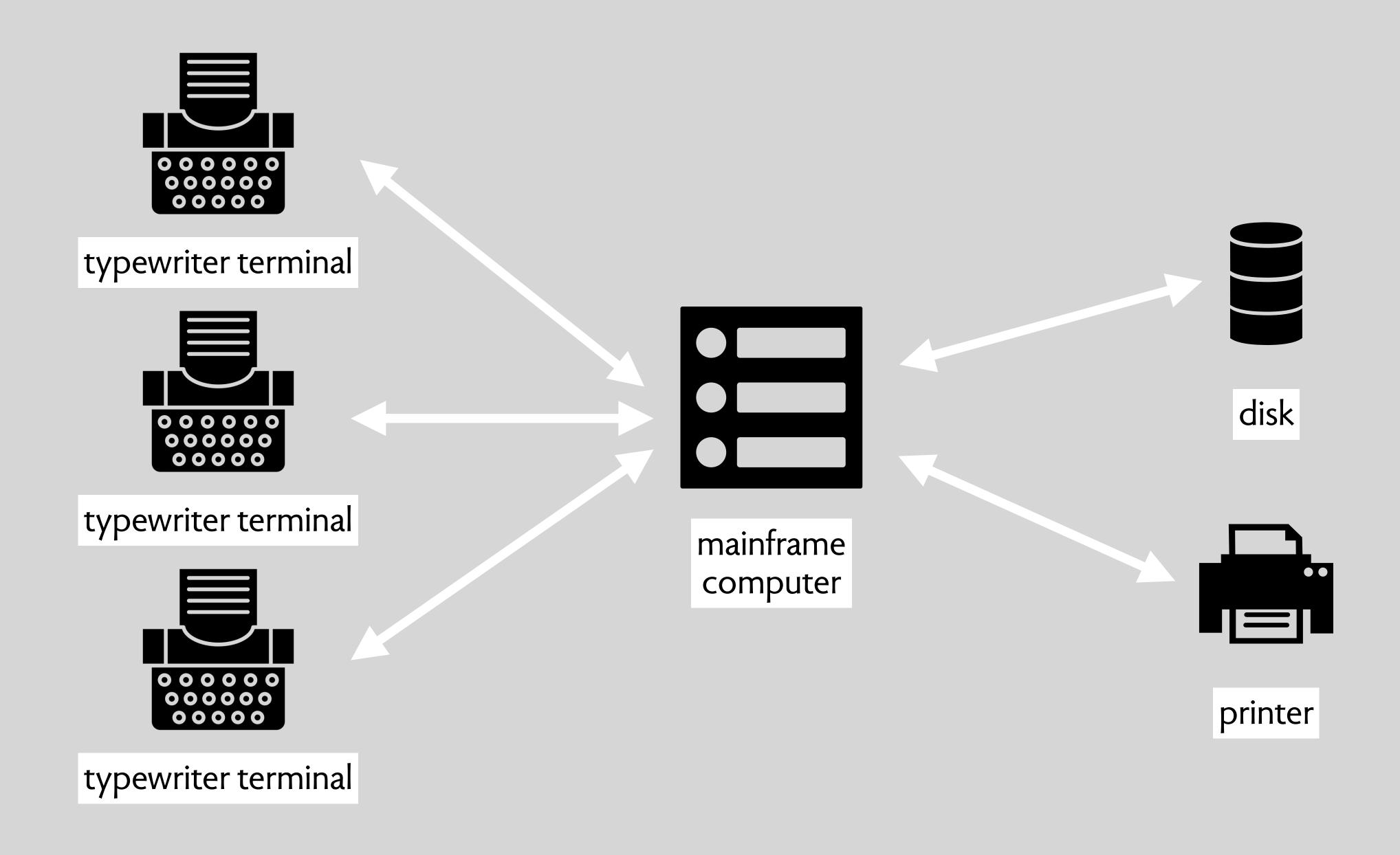
my father's high school teacher at Harrow School in London wrote checkers program (1951) with help from Alan Turing

#### timesharing at MIT: Project Mac and CTSS



Fernando J. Corbató with IBM 7090 (1962)

#### time sharing architecture (early 1960s)



#### first personal computer



Altair 8800 (1974)
hobbyist kit, required assembly



Paul Allen and Bill Gates (1975) wrote BASIC interpreter for Altair

#### first mass-market personal computers



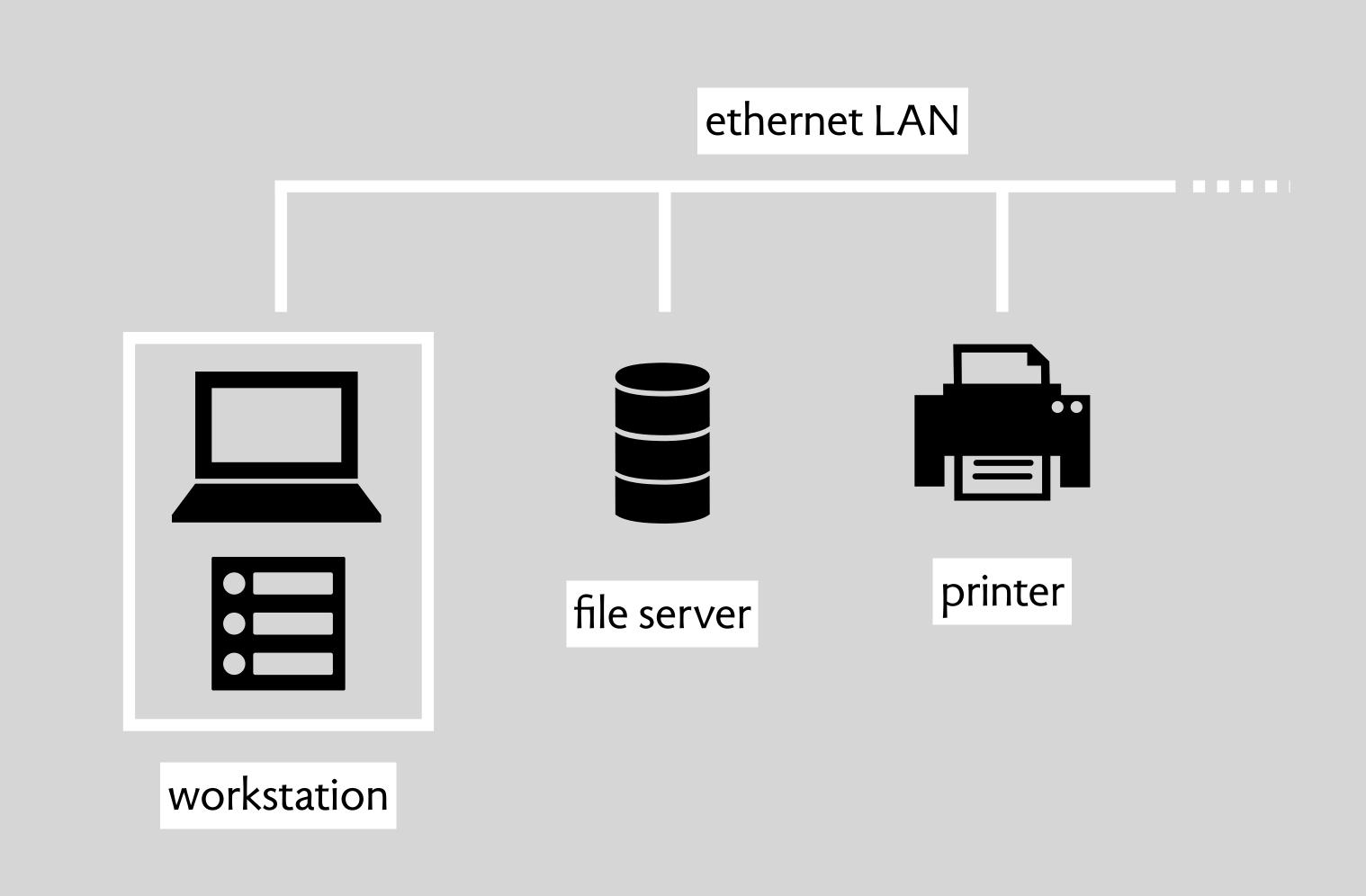
Commodore PET 2001 (1977)

all in one computer
for schools, home users, businesses

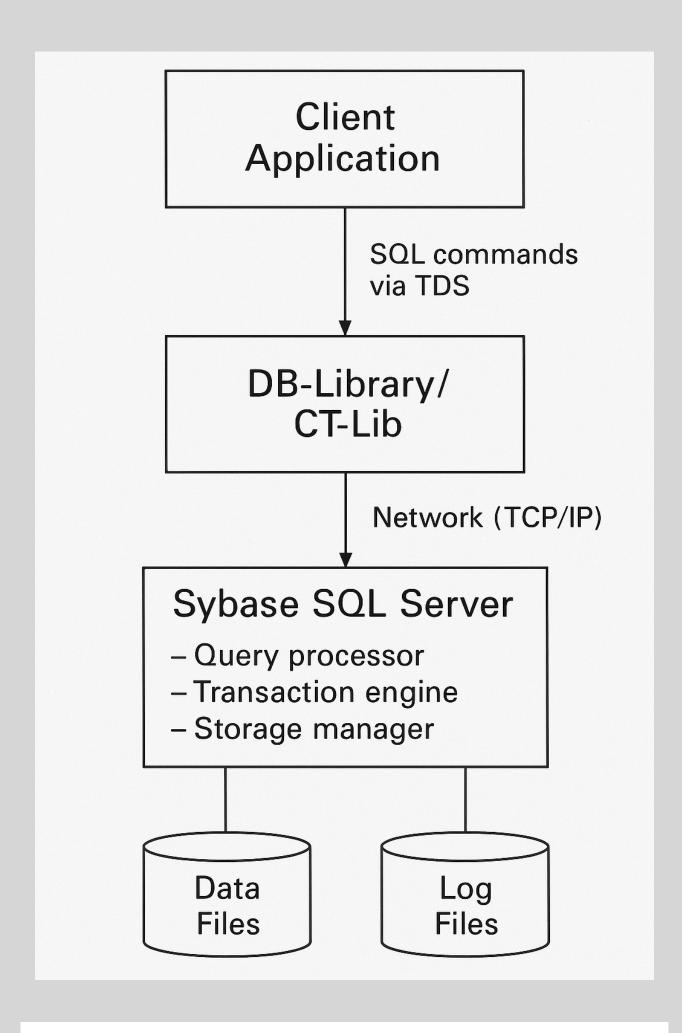
#### first client-server systems (1970s)



Xerox Alto (1973) first "client-server" system client is not just a terminal



#### first client-server database systems (1980s)

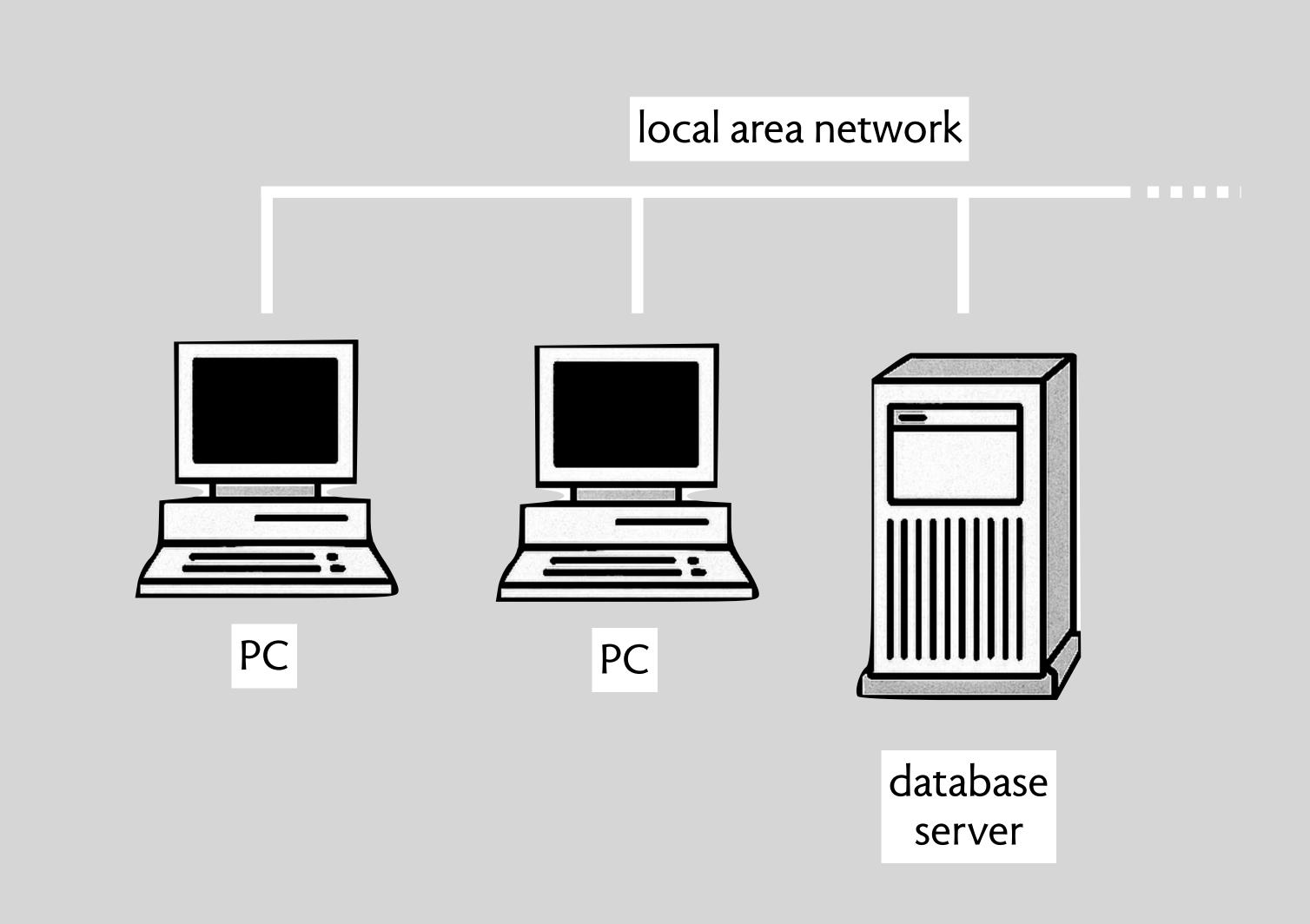


Sybase SQL Server (1986)

apps running on PCs

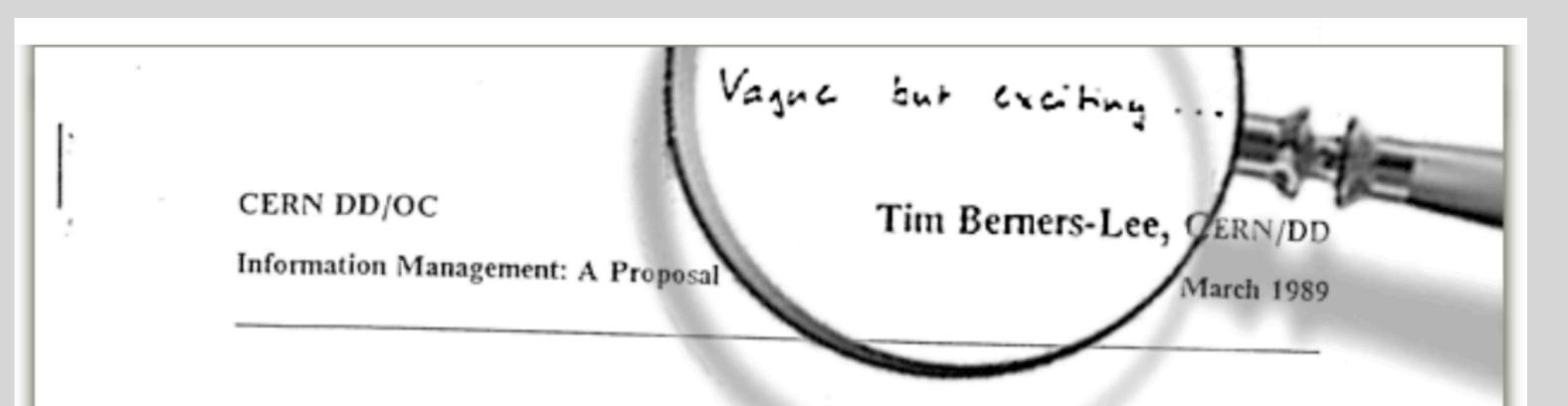
SQL queries to database server

no internet yet



### what's HTTP?

#### TBL's proposal (1989)

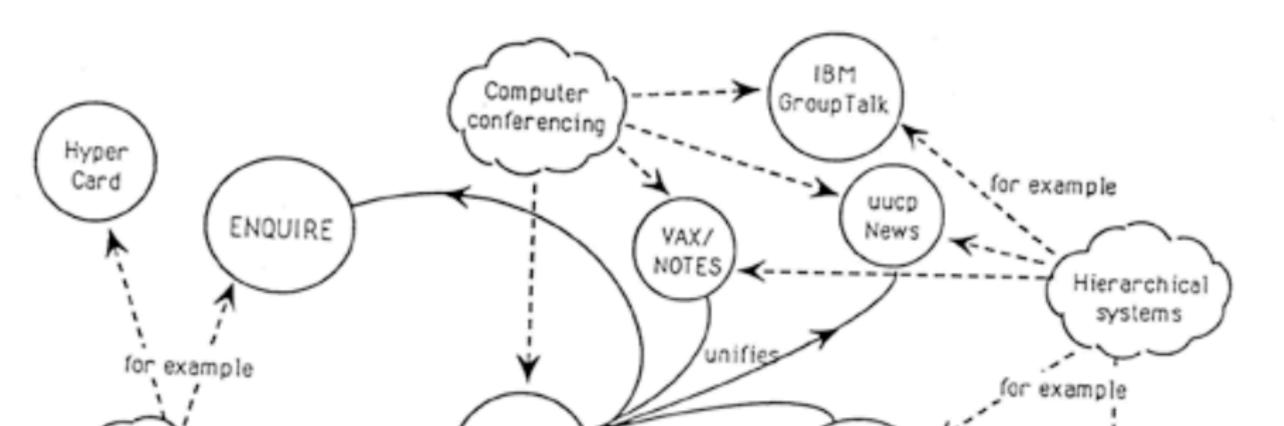


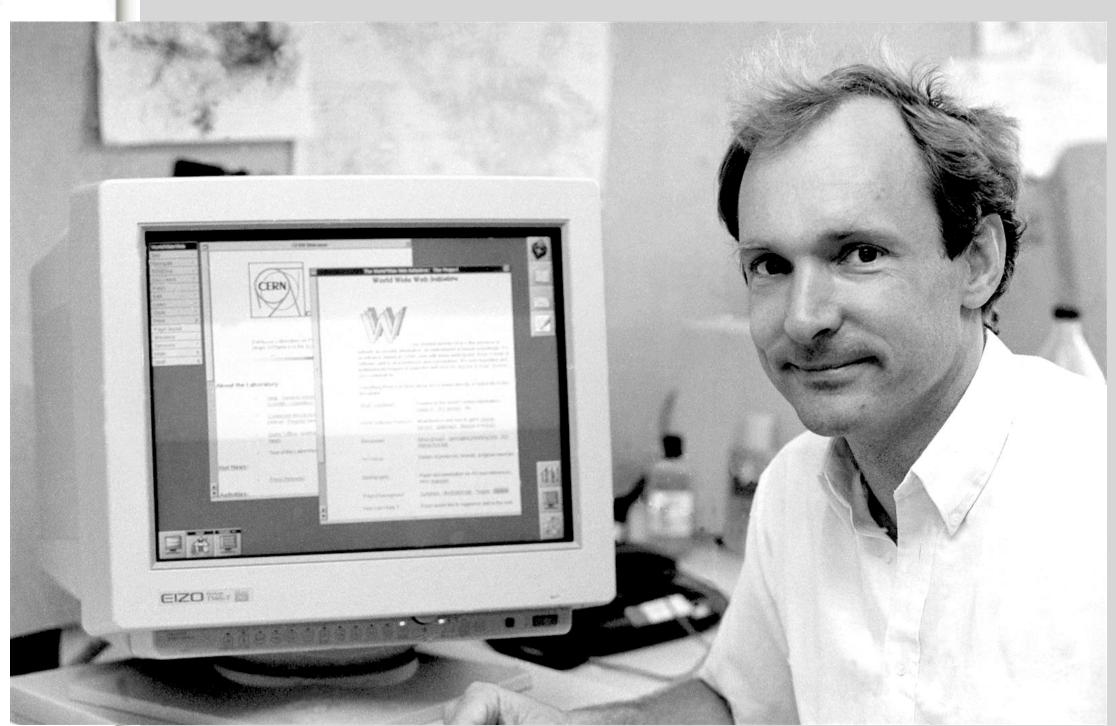
#### Information Management: A Proposal

#### Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext sytstem.

Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control





#### the first web page (1991)

#### World Wide Web

The WorldWideWeb (W3) is a wide-area <u>hypermedia</u> information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including an <a href="mailto:executive summary">executive summary</a> of the project, <a href="mailto:Mailing lists">Mailing lists</a> , <a href="Molecutive summary">Policy</a> , <a href="Molecutive summary">November's W3 news</a> , <a href="Molecutive summary">Frequently Asked Questions</a> .

#### What's out there?

Pointers to the world's online information, <u>subjects</u>, <u>W3 servers</u>, etc.

#### <u>Help</u>

on the browser you are using

#### **Software Products**

A list of W3 project components and their current state. (e.g. <u>Line Mode</u>, X11 <u>Viola</u>, <u>NeXTStep</u>, <u>Servers</u>, <u>Tools</u>, <u>Mail robot</u>, <u>Library</u>)

#### **Technical**

Details of protocols, formats, program internals etc

#### **Bibliography**

Paper documentation on W3 and references.

#### **People**

A list of some people involved in the project.

#### <u>History</u>

A summary of the history of the project.

#### How can I help?

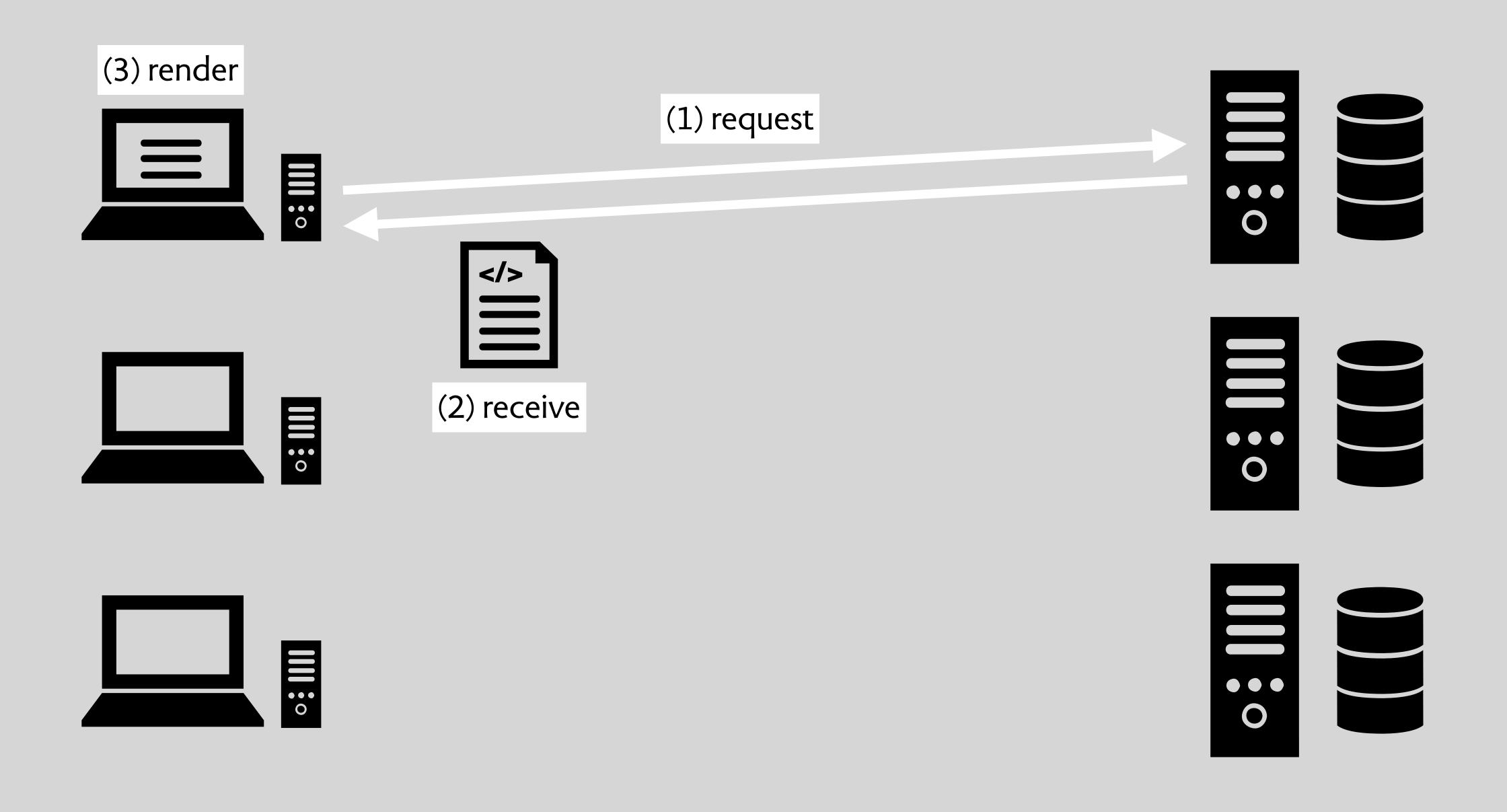
If you would like to support the web..

#### Getting code

Getting the code by <u>anonymous FTP</u>, etc.

**essential ingredients**URLs + HTML + HTTP

#### TBL's web (1991)



#### the web as a document editing system: HTTP verbs

Create → POST

Read → GET

Update → PUT

Delete → DELETE

#### **HTTP verbs**

many originally proposed (1992) eventually settles on these

#### distributed document editing

TBL's original concept of the web never happens in this form

#### structure of a URL

https://api.example.com/users/123/profile?view=full&include=posts&lang=en

Component	Example	Description	5
Scheme / Protocol	https://	Specifies secure HTTP (HTTPS)	
Domain name	api.example.com	The server's hostname	
Path	/users/123/profile	The route to a specific resource	
Query parameters	?view=full&include=posts⟨=en	Key–value pairs providing extra info for the request	
Full meaning	"Request the full profile of user 123 from the API, including their posts, in English."		

#### elements of an HTTP request

Element	Example	Description
Method / Verb	GET, POST, PUT, PATCH, DELETE	Specifies the action the client wants the server to perform on the resource.
Path	/api/users/42? verbose=true	The resource path and optional query parameters — relative to the host. (In most requests, this replaces the full URL.)
HTTP Version	HTTP/1.1	Indicates the HTTP protocol version being used.
Headers	Content-Type: application/json Authorization: Bearer abc123	Key-value pairs carrying metadata about the request: content type, authentication, caching, cookies, etc.
Host	Host: api.example.com	Specifies the target host (required in HTTP/1.1 and later).
Body / Payload	<pre>{ "email": "new@example.com" }</pre>	Optional data sent with the request — used mainly in POST, PUT, and PATCH methods.

#### REST: a standard pattern for routes

HTTP Request Type	URL Path	Description		
GET	'/articles'	display a list of all articles		
GET	'/articles/new'	return an HTML form for creating a new article		
POST	'/articles'	create a new article		
GET	'/articles/:id'	display a specific article		
GET	'/articles/:id/edit'	return an HTML form for editing a article		
PUT	'/articles/:id'	update a specific article		
DELETE	'/articles/:id'	delete a specific article	essence	
			nouns	

Representational State Transfer (Roy Fielding, 2000) example here for Ruby on Rails style

verbs (HTTP method)

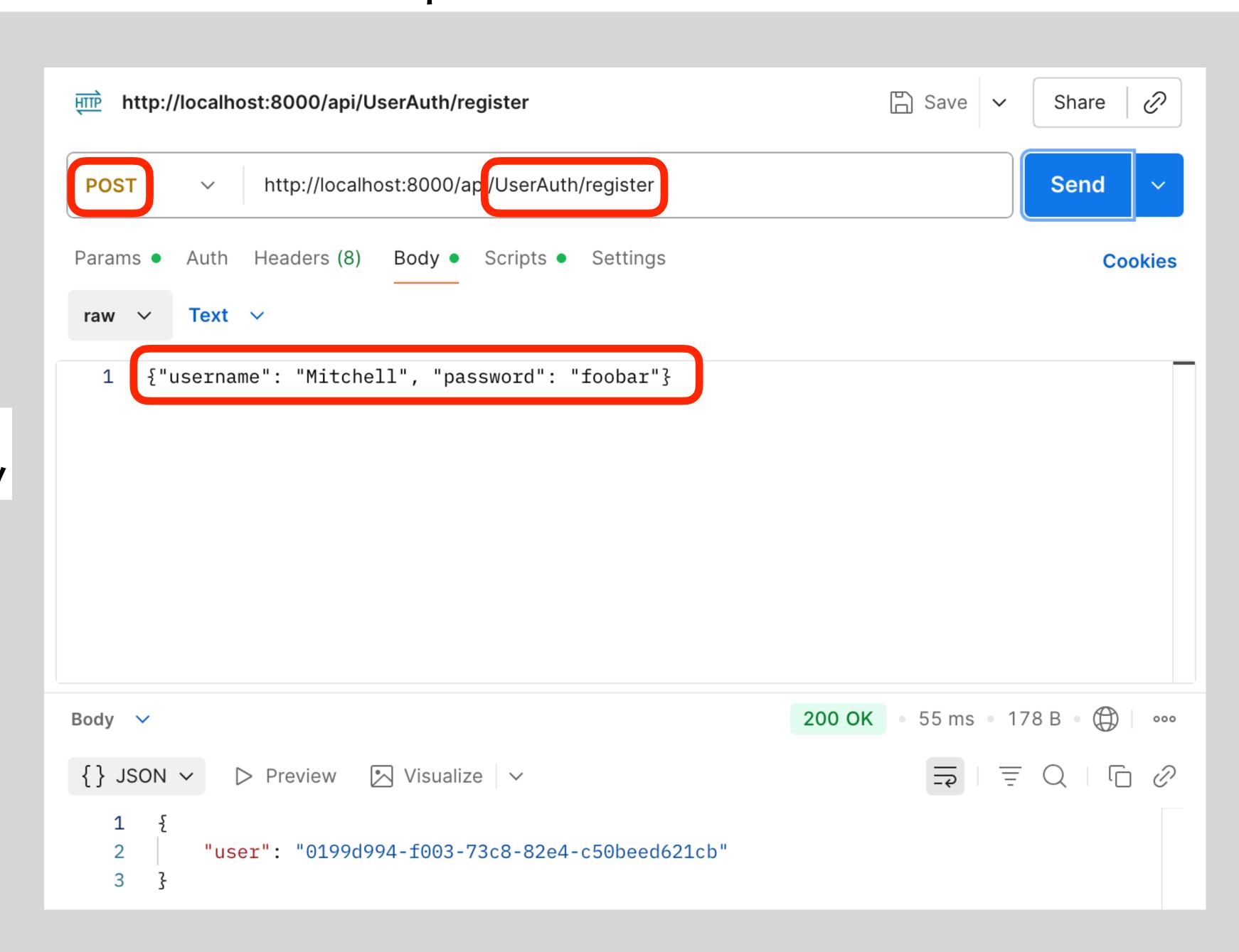
#### our simple scheme

verb is always **POST** 

path is concept/action

args passed as raw JSON in body

good for this class in industrial setting use GETs for reads (caching, reloads)



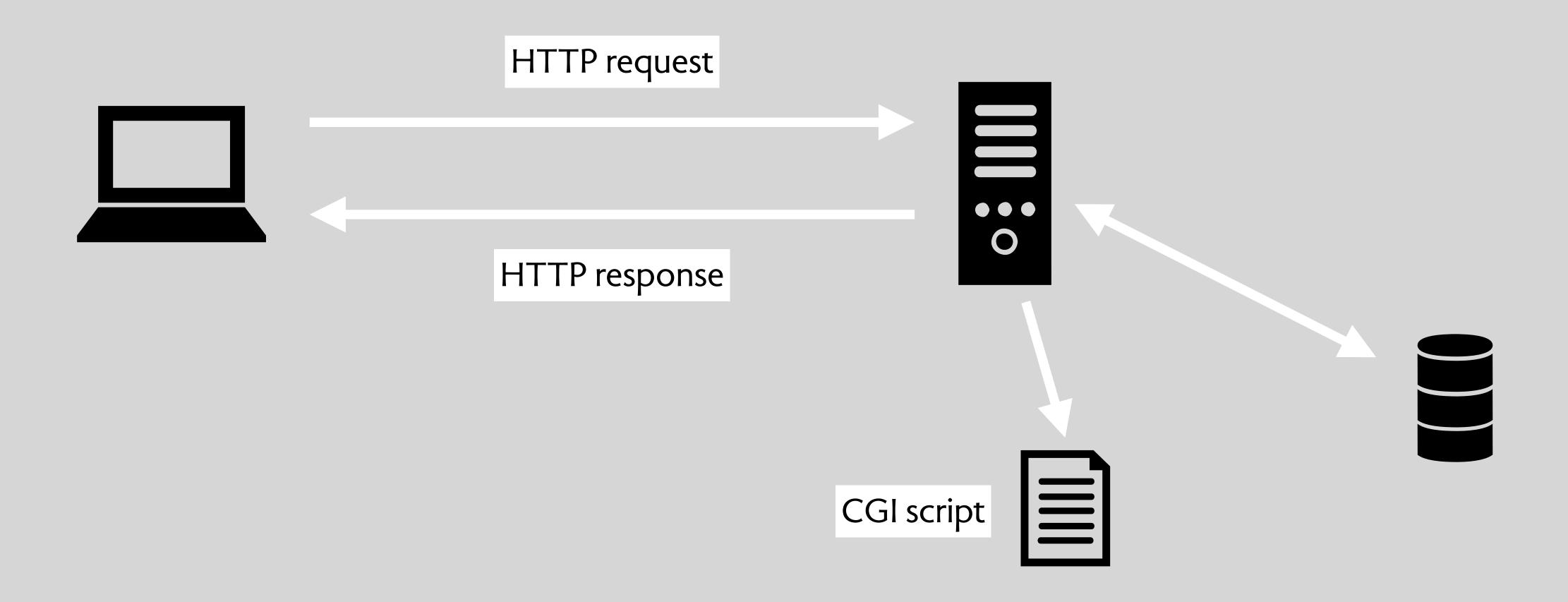
### comedic interlude

#### Mitchell and Webb: downloading the internet



## how did web sites become web apps?

#### common gateway interface (1993)



#### hello world CGI script

```
#!/usr/bin/perl
print "Content-type: text/html\n\n";
print "Hello, World.";
```

#### a famous CGI script

```
#!/usr/bin/perl
# Read form input
read(STDIN, $buffer, $ENV{'CONTENT_LENGTH'});
@pairs = split(/&/, $buffer);
foreach $pair (@pairs) {
    (\text{name, } \text{value}) = \text{split}(/=/, \text{spair});
    value =  tr/+//;
    value =  s/\%(...)/pack("C", hex($1))/eg;
    $FORM{$name} = $value;
# Get recipient from the form field
$recipient = $FORM{'recipient'};
# Send mail
open(MAIL, "|/usr/lib/sendmail -t");
print MAIL "To: $recipient\n";
print MAIL "From: $FORM{'email'}\n";
print MAIL "Subject: Form submission\n\n";
print MAIL "Here are the results:\n";
foreach $key (keys %FORM) {
    print MAIL "$key = $FORM{$key}\n";
close(MAIL);
```

#### FormMail (1995)

contact form email written by Matt Wright a high school student



recipient=admin@example.com%0ABcc:spamlist@example.org

script didn't sanitize input client can add extra fields so use as spam relay by early 2000s, blacklisted

#### a CGI-powered web app (1995)



#### Welcome to Amazon.com Books!

One million titles, consistently low prices.

(If you explore just one thing, make it our <u>personal notification service</u>. We think it's very cool! Also, check out <u>what our customers are saying about us</u>.)

#### SPOTLIGHT! -- JANUARY 16TH

These are books we love, offered at Amazon.com low prices. The spotlight moves **EVERY** day so please come often.



#### ONE MILLION TITLES

Search Amazon.com's <u>million title catalog</u> by author, subject, title, keyword, and more... Or take a look at the <u>books we</u> <u>recommend</u> in over 20 categories... Check out our <u>customer reviews</u> and the <u>award winners</u> from the Hugo and Nebula to the Pulitzer and Nobel... and <u>bestsellers</u> are 30% off the publisher's list...

#### EYES & EDITORS, A PERSONAL NOTIFICATION SERVICE

Like to know when that book you want comes out in paperback or when your favorite author releases a new title? Eyes, our tireless, automated search agent, will send you mail. Meanwhile, our human editors are busy previewing galleys and reading advance reviews. They can let you know when especially wonderful works are published in particular genres or subject areas. Come in, meet Eyes, and have it all explained.

#### YOUR ACCOUNT

Check the status of your orders or change the email address and password you have on file with us. Please note that you

#### running code in the browser

```
<div id="box" style="position:absolute; left:0px; top:100px;">\(\color{1}{\infty} < \cdot \)
<pre><script>
  let x = 0;
  function move() {
    x += 5;
    document.getElementById("box").style.left = x + "px";
    if (x < 300) setTimeout(move, 30);
}
  move();
</script>
```

JavaScript (1995) and standard DOM (1998)



#### AJAX: asynchronous JavaScript and XML

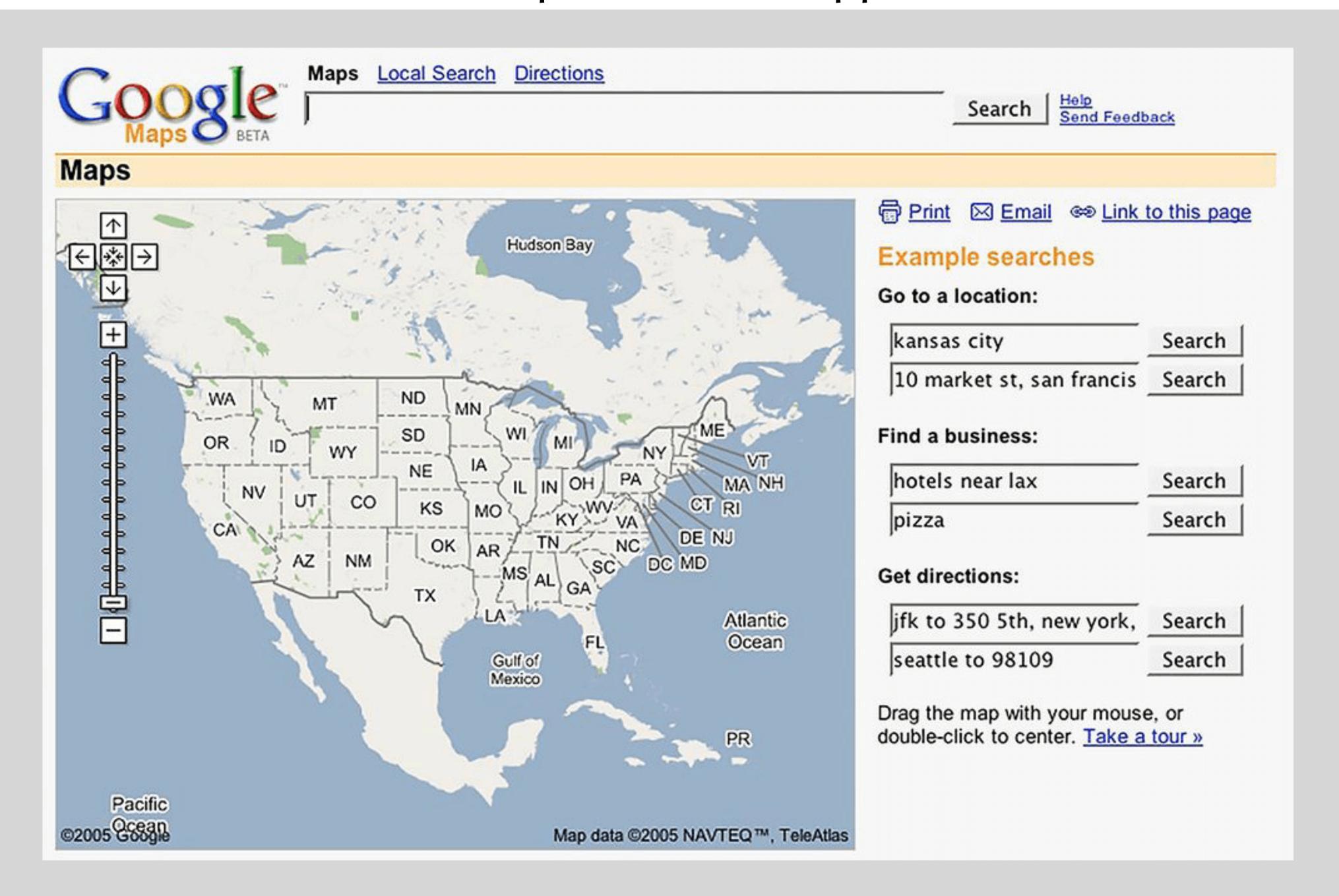
```
var xhr = new XMLHttpRequest();
xhr.open("GET", "example.txt", true);
xhr.onreadystatechange = function() {
  if (xhr.readyState === 4 && xhr.status === 200) {
    console.log(xhr.responseText);
  }
};
xhr.send();
```

```
var xhr = new XMLHttpRequest();
xhr.open("POST", "/submit", true);
xhr.setRequestHeader("Content-Type", "application/x-www-form-urlencoded");
xhr.send("name=Daniel&message=Hello");
```

#### XMLHttpRequest (1998)

calling server inside a script in the browser introduced by Microsoft later standardized for all browsers

#### an AJAX-powered web app (2005)



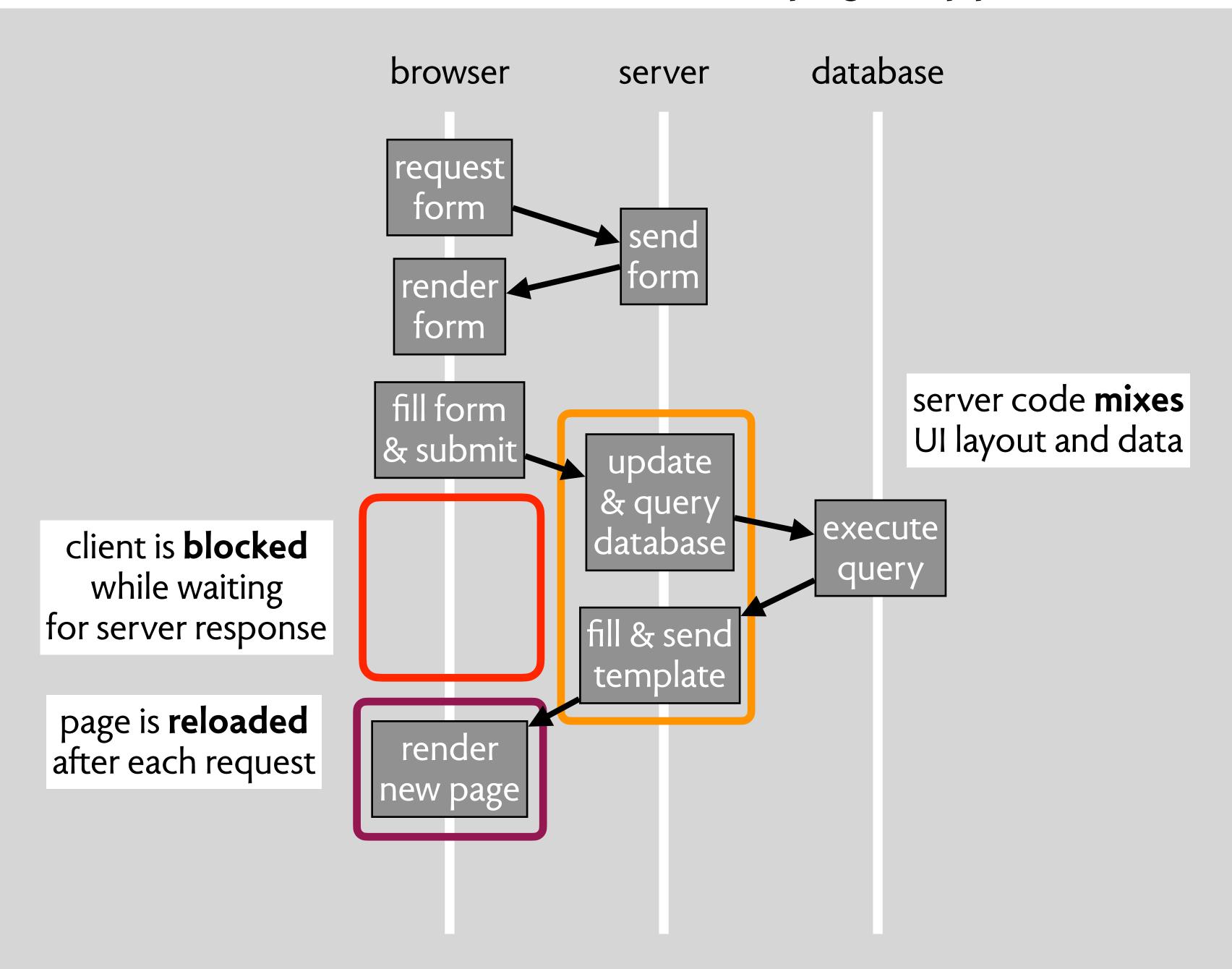
#### what server calls look like now

```
const api = axios.create({
  baseURL: '/api',
  headers: {
     'Content-Type': 'application/json',
    'Accept': 'application/json'
  timeout: 10000
async register(username, password) {
  const response = await api.post('/UserAuth/register', {
    username,
    password
  return response.data
},
const result = await register(form.value.username, form.value.password)
if (result.success) {
 // Redirect to home page after successful registration
  router.push('/')
```

now like a local call but asynchronous

### what's an SPA?

#### the flow for a "multi-page" app



#### using a template to render HTML in the server

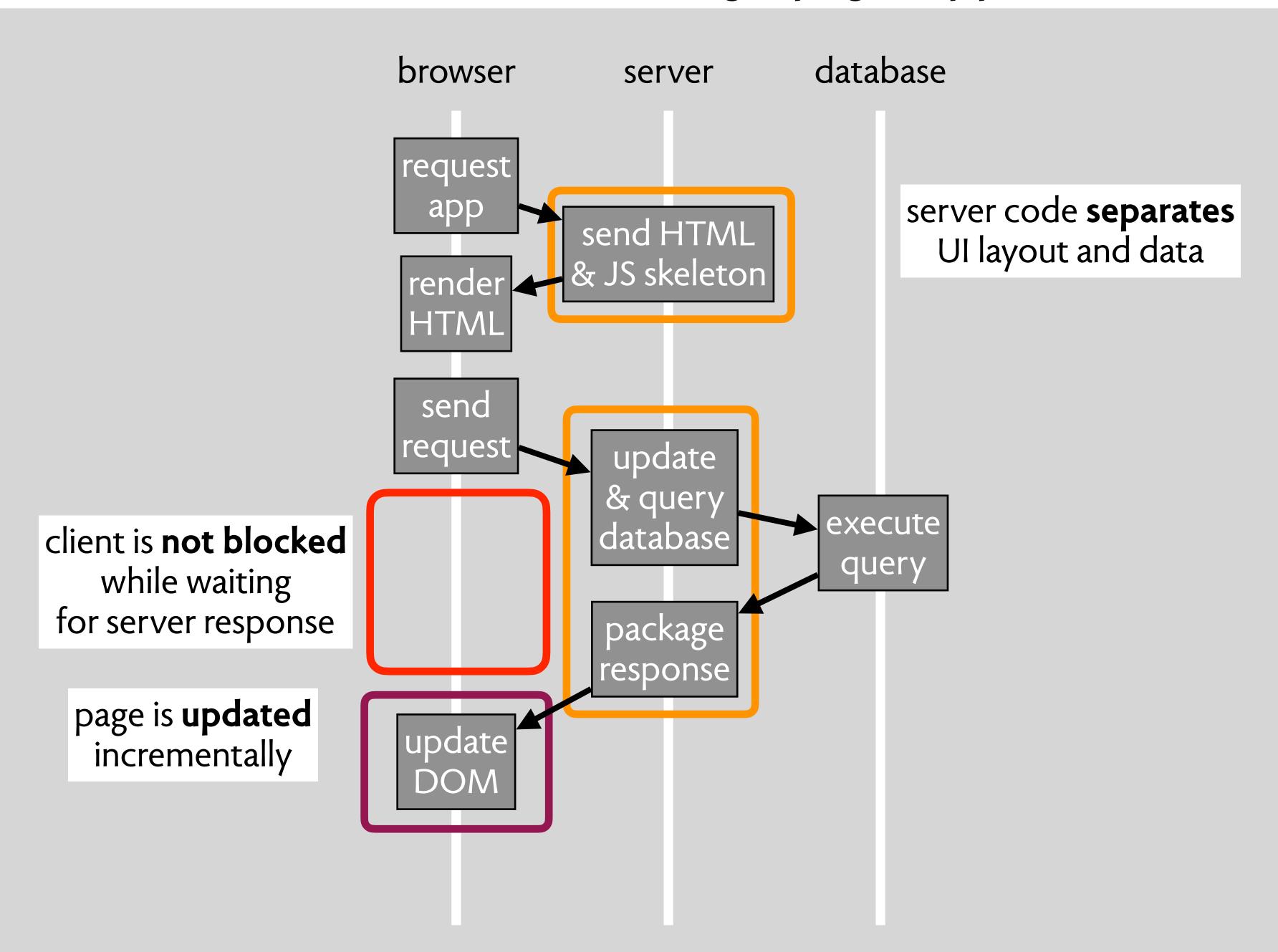
```
@app.route("/user/<name>")
def user_profile(name):
    # Data fetched from a database, for example
    user_info = {
        "name": name,
        "age": 32,
        "hobbies": ["music", "sailing", "AI"]
    return render_template(
        "profile.html",
        user=user_info
```

app.py

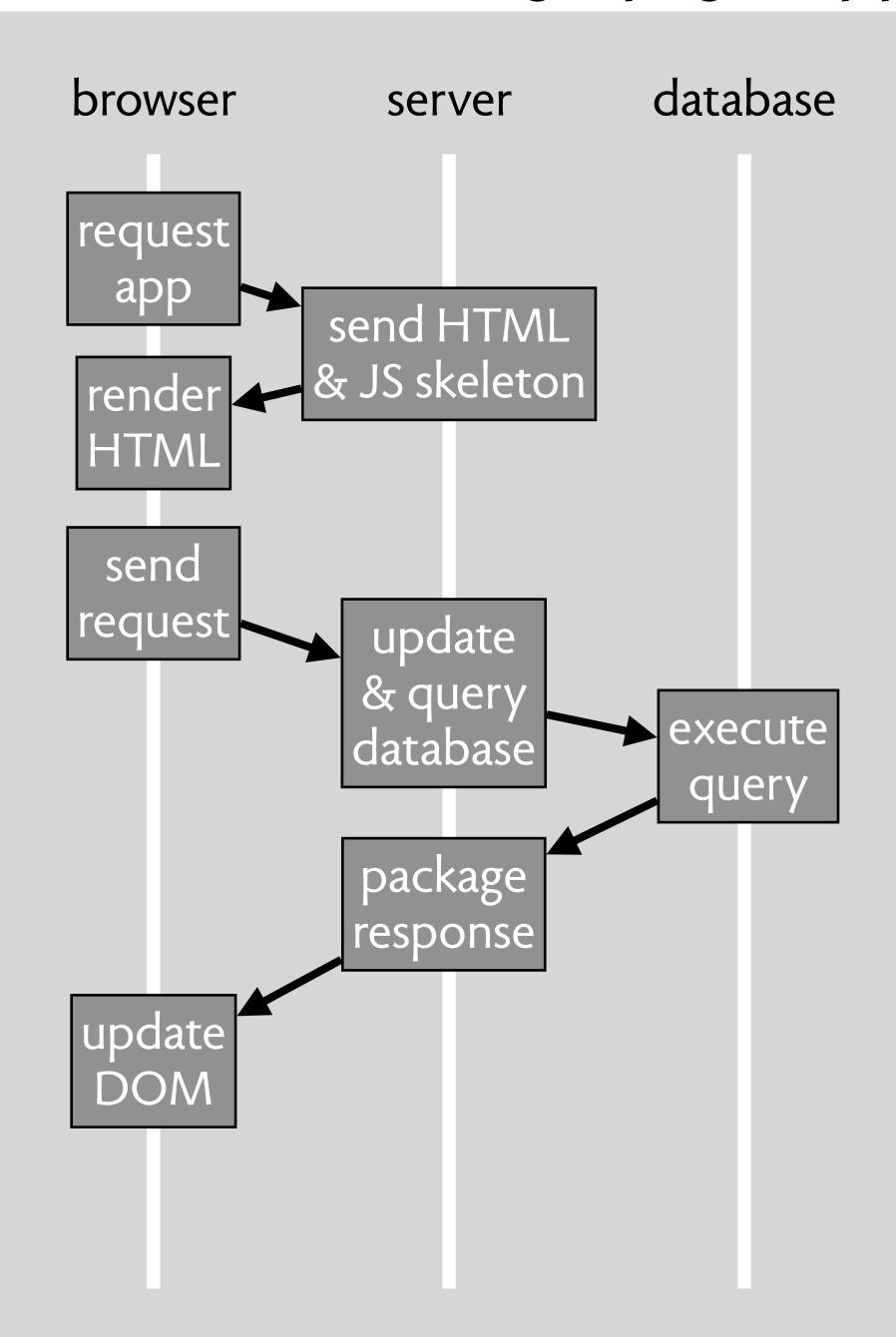
```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8" />
  <title>{{ user.name }}'s Profile</title>
</head>
<body>
  <h1>Welcome, {{ user.name }}!</h1>
  Age: {{ user.age }}
  <h2>Hobbies</h2>
  <l
    {% for hobby in user.hobbies %}
     {\li>{\li hobby \}}
    {% endfor %}
  </body>
</html>
```

profile.html

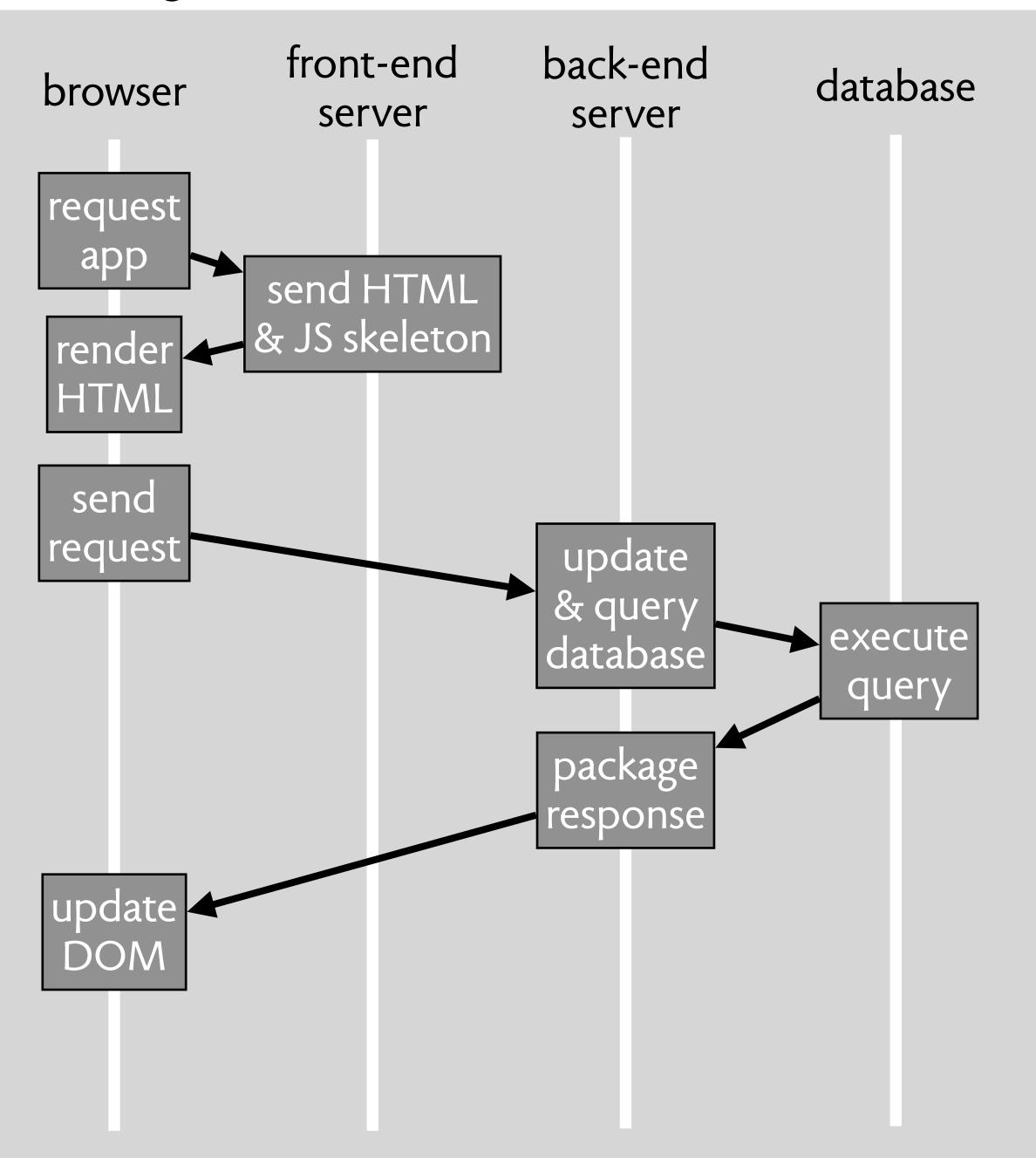
# the flow for a "single page" app



# the flow for a "single page" app



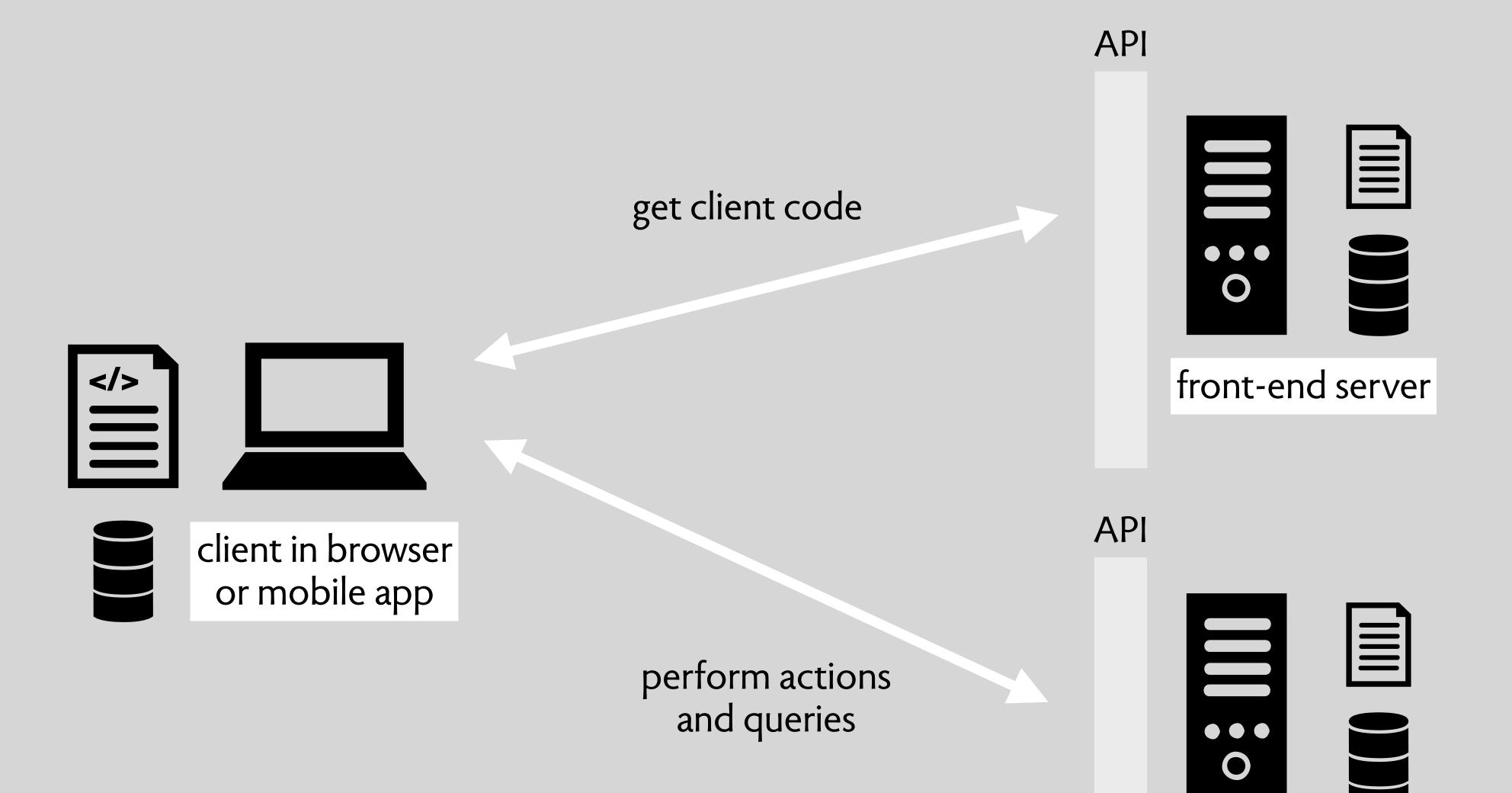
# separating the roles of an SPA into two servers



a client-server web app!

# what's an API?

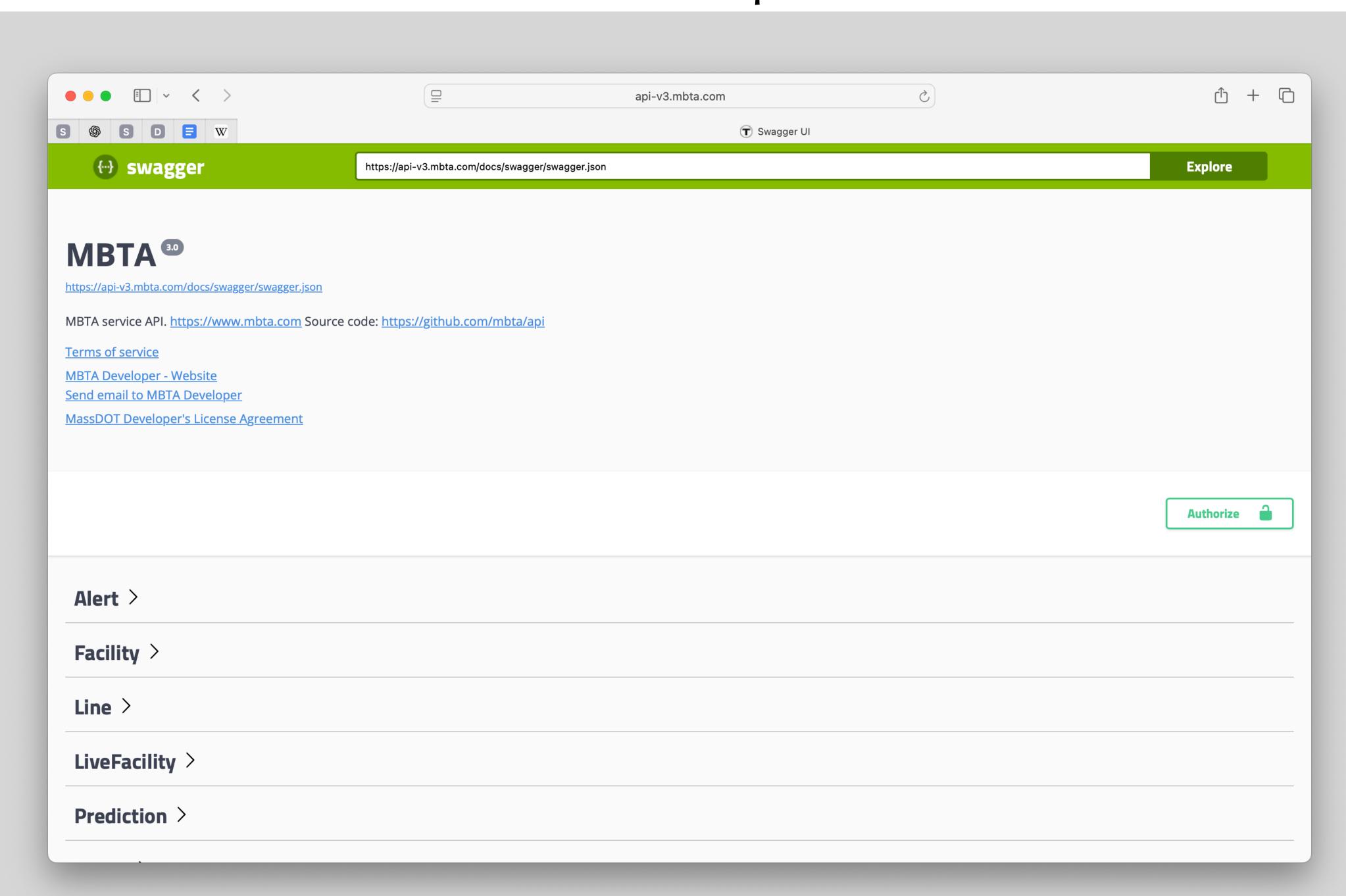
# a typical setup



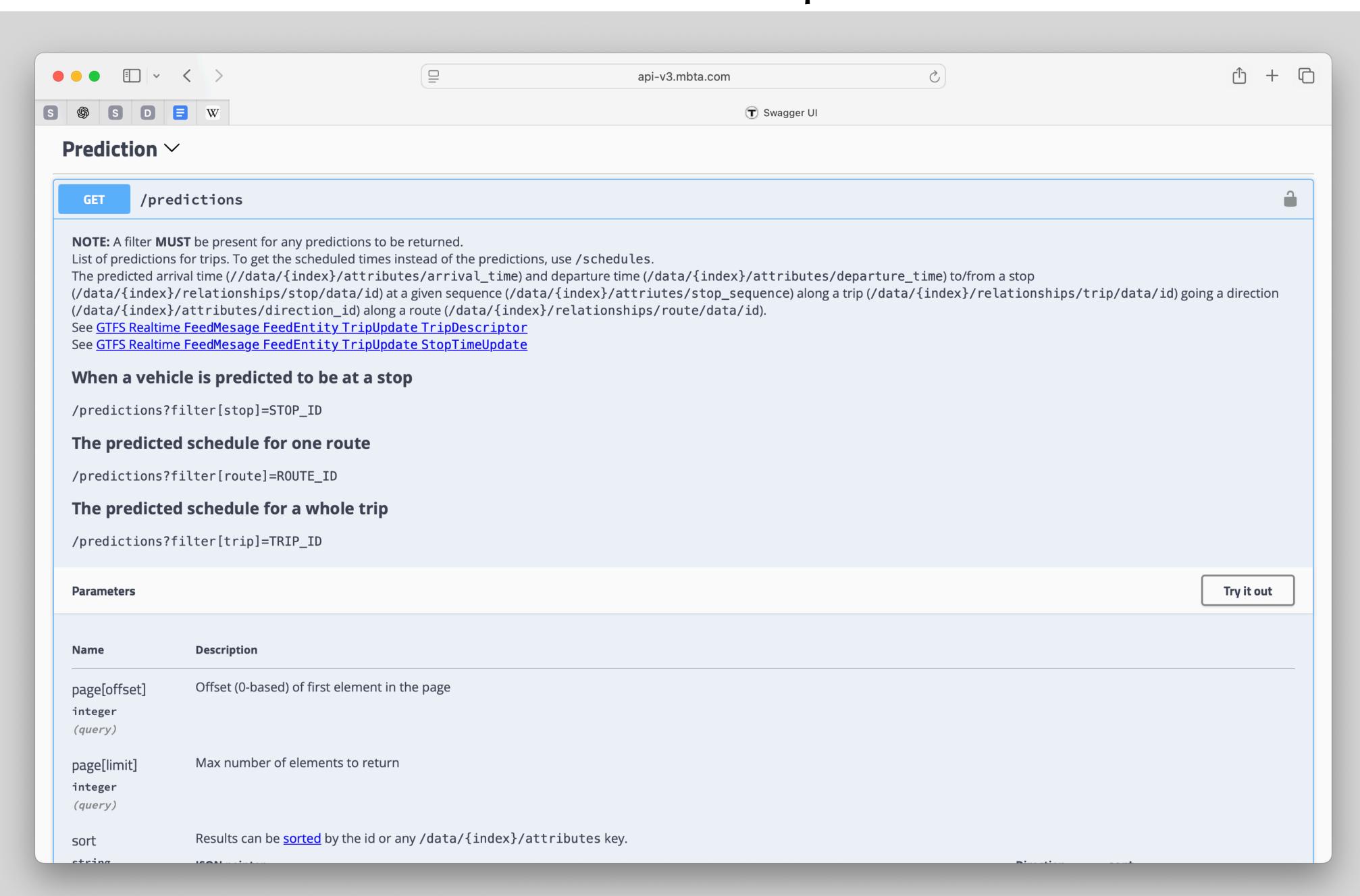
APIs give decoupling minimize knowledge hide changes to server like all specs...

back-end server

# MBTA's API for live transport information



## structure of a request



# sample request & response

#### Example

https://api-v3.mbta.com/predictions?filter%5Bstop%5D=place-sstat&filter%5Bdirection\_id%5D=0&include=stop returns predictions from South Station with direction\_id=0, below is a truncated response with only relevant fields displayed:

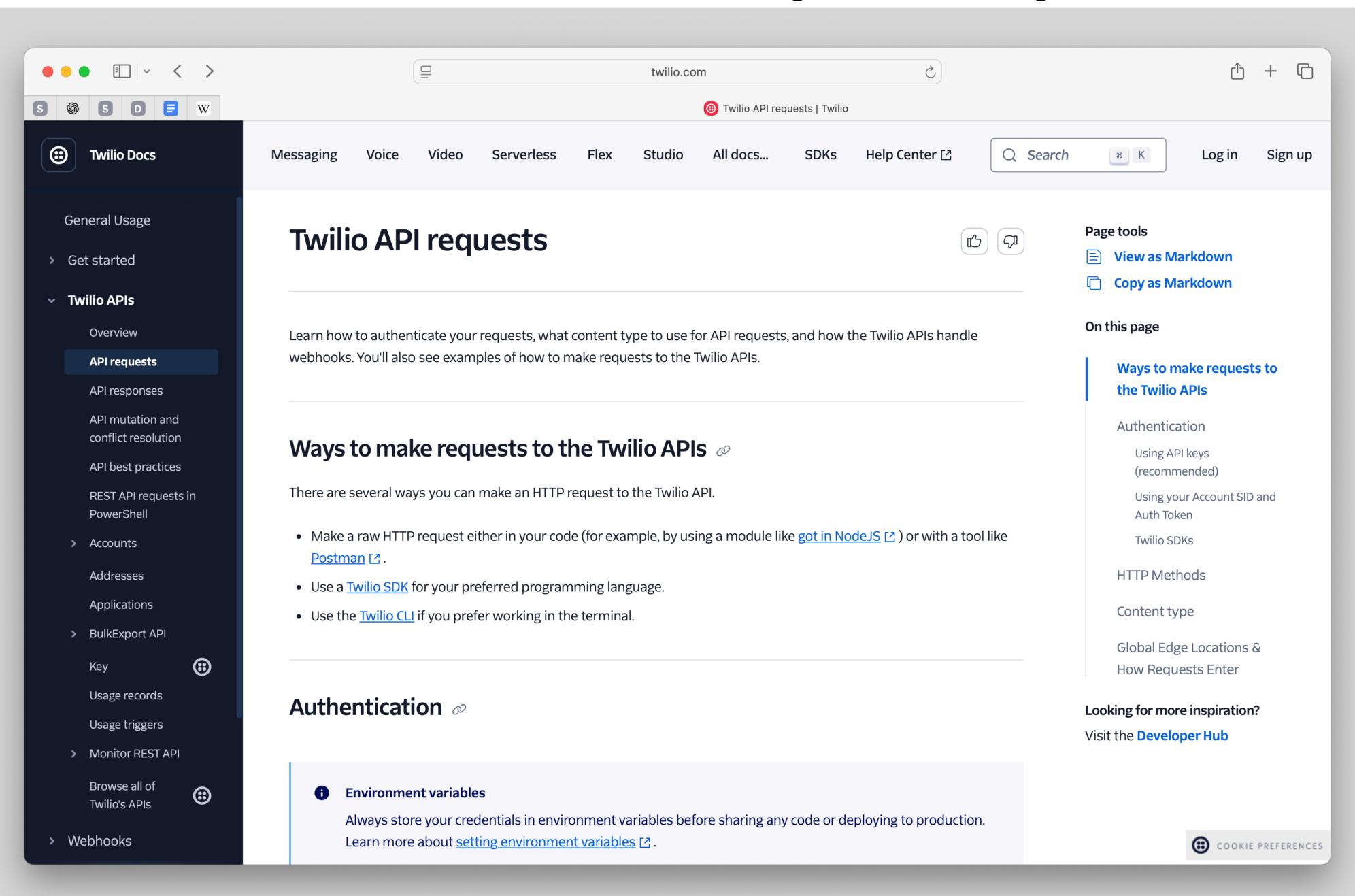
```
"data": [
    "id": "prediction-CR-Weekday-Fall-18-743-South Station-02-1",
    "relationships": {
      "stop": {
        "data": {
          "id": "South Station-02",
          "type": "stop"
    "type": "prediction"
"included": [
    "attributes": {
      "platform_code": "2",
    "id": "South Station-02",
    "type": "stop"
1,
```

Note the stop relationship; use it to cross-reference stop-id with the included stops to retrieve the platform\_code for the given prediction.

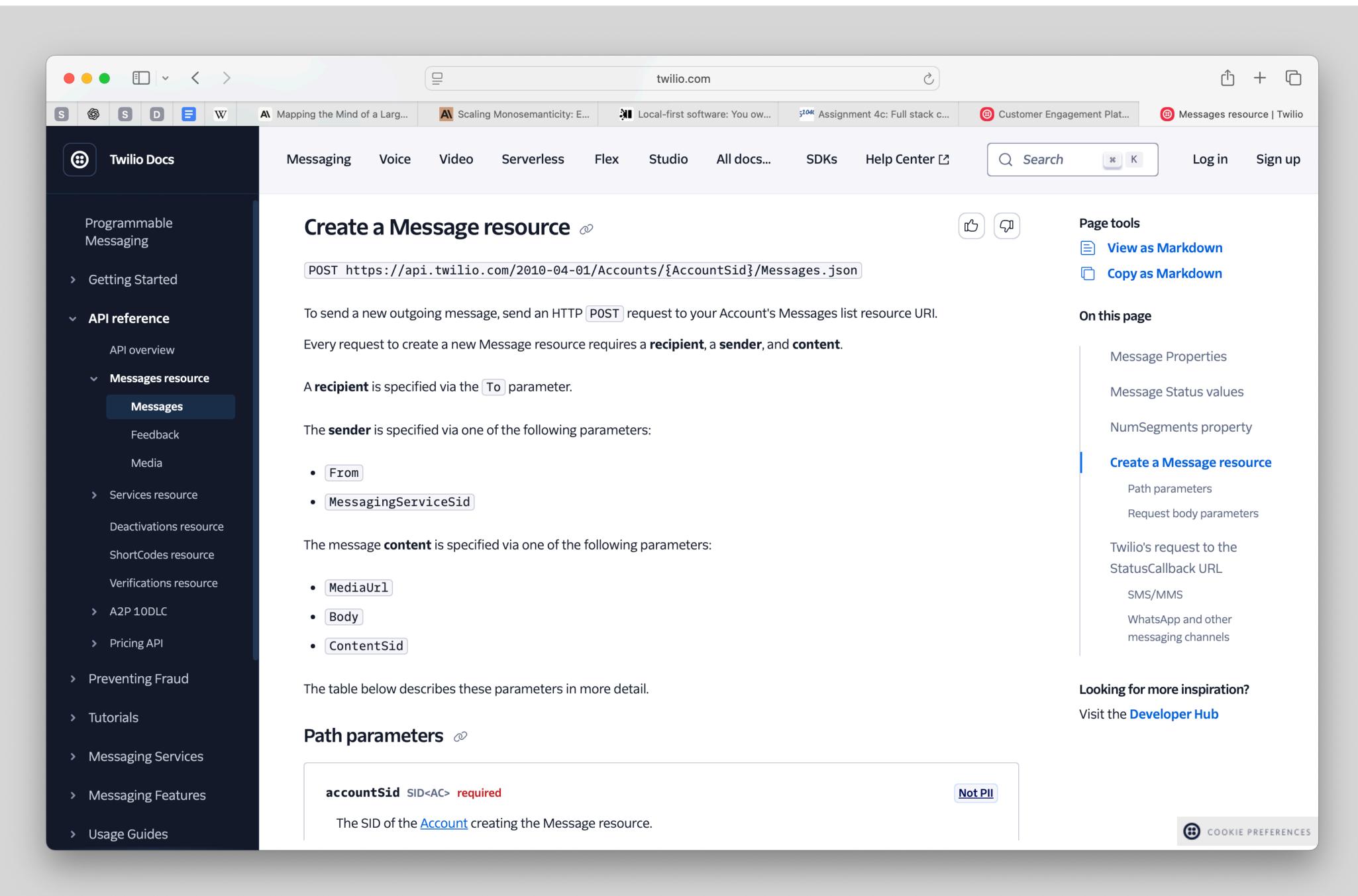
# JSON pointers

sort	Results can be <u>sorted</u> by the id or any /data/{index}/attributes key.		
string (query)	JSON pointer	Direction	sort
	/data/{index}/attributes/arrival_time	ascending	arrival_time
	/data/{index}/attributes/arrival_time	descending	-arrival_time
	/data/{index}/attributes/arrival_uncertainty	ascending	arrival_uncertainty
	/data/{index}/attributes/arrival_uncertainty	descending	-arrival_uncertainty
	/data/{index}/attributes/departure_time	ascending	departure_time
	/data/{index}/attributes/departure_time	descending	-departure_time
	/data/{index}/attributes/departure_uncertainty	ascending	departure_uncertainty
	/data/{index}/attributes/departure_uncertainty	descending	-departure_uncertainty
	/data/{index}/attributes/direction_id	ascending	direction_id
	/data/{index}/attributes/direction_id	descending	-direction_id

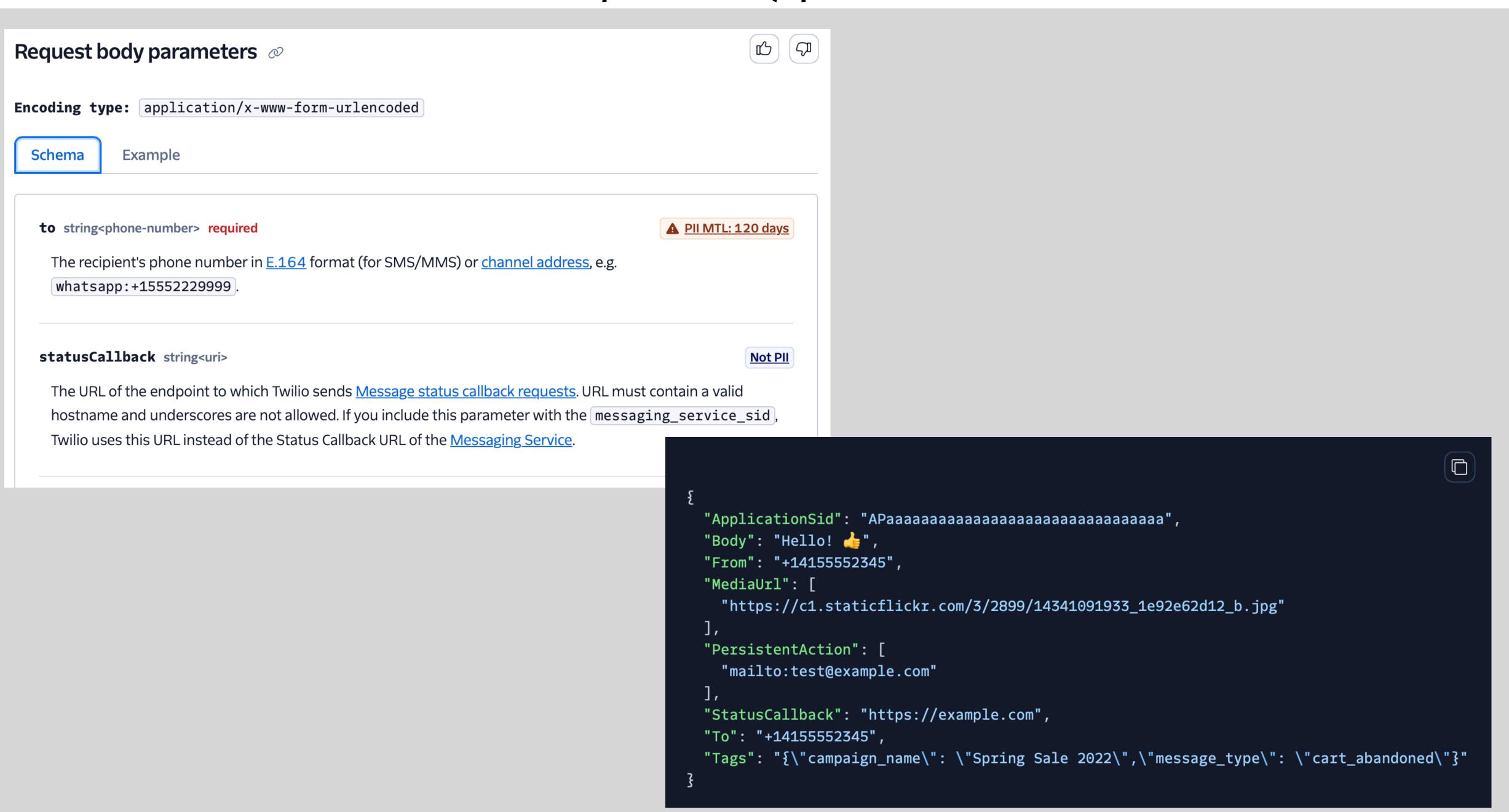
# Twilio: a service for sending text messages



# creating a message



# request body parameters



# using the Twilio client-side library

```
Send an SMS message 🚿
                                                                  Node.js
1 // Download the helper library from https://www.twilio.com/docs/node/install
2 const twilio = require("twilio"); // Or, for ESM: import twilio from "twilio";
4 // Find your Account SID and Auth Token at twilio.com/console
5 // and set the environment variables. See http://twil.io/secure
6 const accountSid = process.env.TWILIO_ACCOUNT_SID;
7 const authToken = process.env.TWILIO_AUTH_TOKEN;
8 const client = twilio(accountSid, authToken);
10 async function createMessage() {
const message = await client.messages.create({
       body: "Hi there",
12
13
       from: "+15557122661",
       to: "+15558675310",
14
     });
15
16
     console.log(message.body);
18 }
19
20 createMessage();
```

```
"api_version": "2010-04-01",
    "body": "Hi there",
   "date_created": "Thu, 24 Aug 2023 05:01:45 +0000",
    "date_sent": "Thu, 24 Aug 2023 05:01:45 +0000",
    "date_updated": "Thu, 24 Aug 2023 05:01:45 +0000",
    "direction": "outbound-api",
    "error_code": null,
    "error_message": null,
    "from": "+15557122661",
    "num_media": "0",
    "num_segments": "1",
   "price": null,
    "price_unit": null,
    "messaging_service_sid": "MGaaaaaaaaaaaaaaaaaaaaaaaaaaaaa",
   "status": "queued",
   "subresource_uris": {
     "media": "/2010-04-01/Accounts/ACaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa/Messages/SMaaaaaaaaaaaaaaaaaaaaaaaaaa
21
22 "to": "+15558675310",
```

# what's a static site?

#### static web sites

#### what the user sees

resource returned depends on HTTP request and not on previous HTTP requests

#### what the developer sees

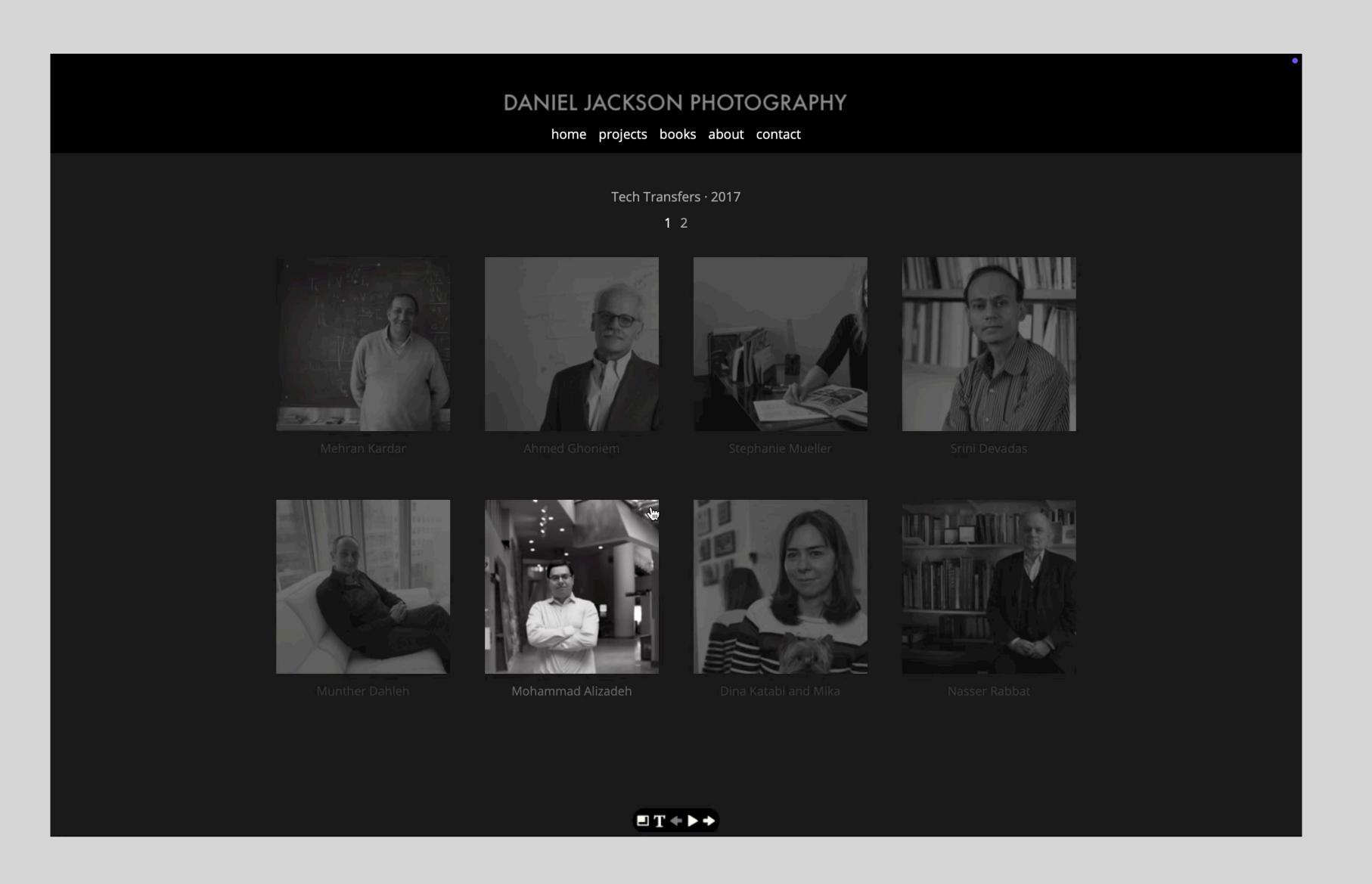
site is just a collection of files, not programs or databases served directly by a web server can include JavaScript that runs in the browser

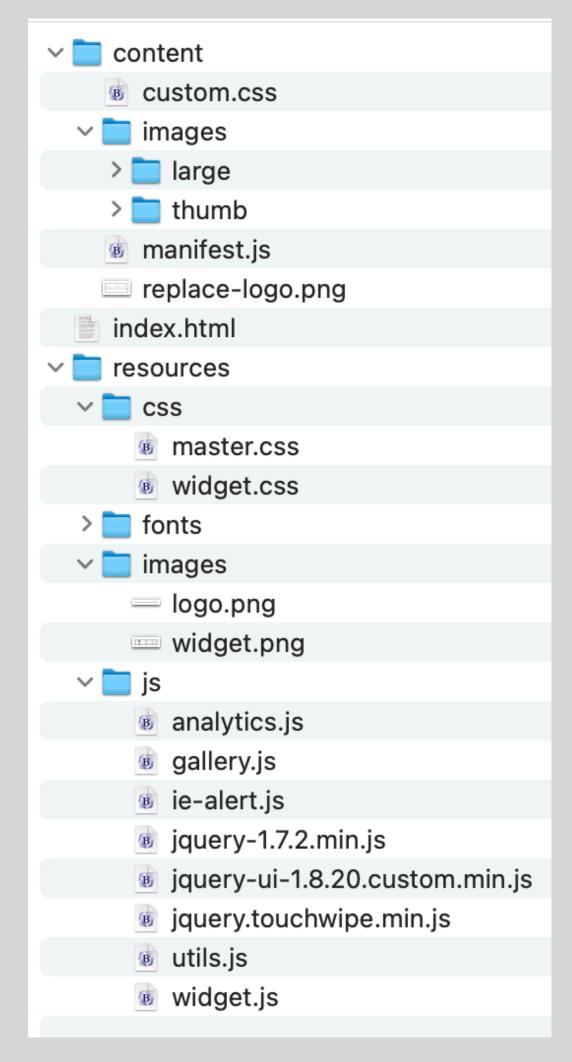
#### how to create a static site

set up a server or find a hosting service (eg, GitHub Pages) edit files locally and push to server (eg, with Git or Unison) or, use a static site builder or CMS

front-end server is a static website!

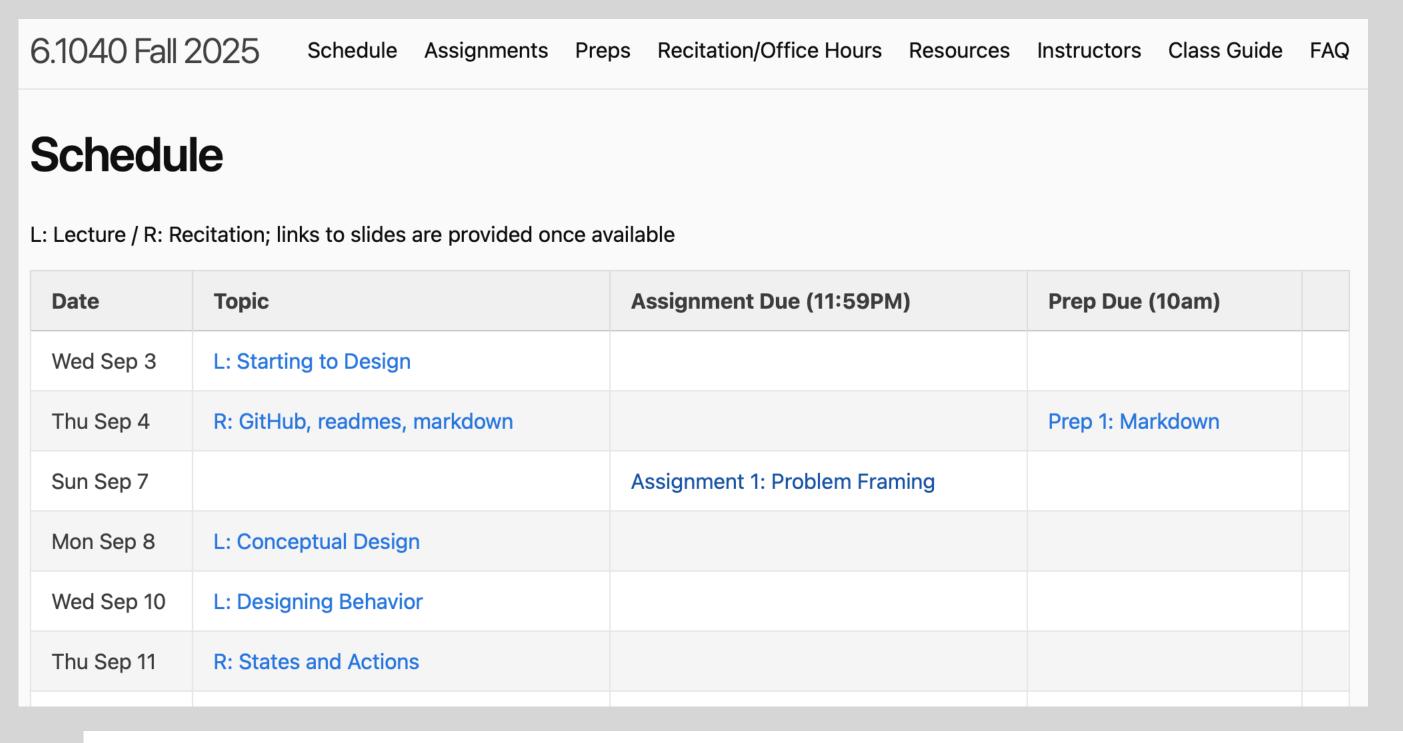
# example: my photo web pages





files that produce this page

#### static website builders





#### what they offer

editing in markdown, not HTML consistent styles, packaged in themes index pages (eg, of blog posts by date) easy management of nav bars live view while editing locally

#### how they work

you edit in markdown run build command push to repo

#### popular examples

Jekyll (Ruby) Hugo (Go)

#### disadvantages

sometimes obscure, can be limiting themes aren't truly interchangeable

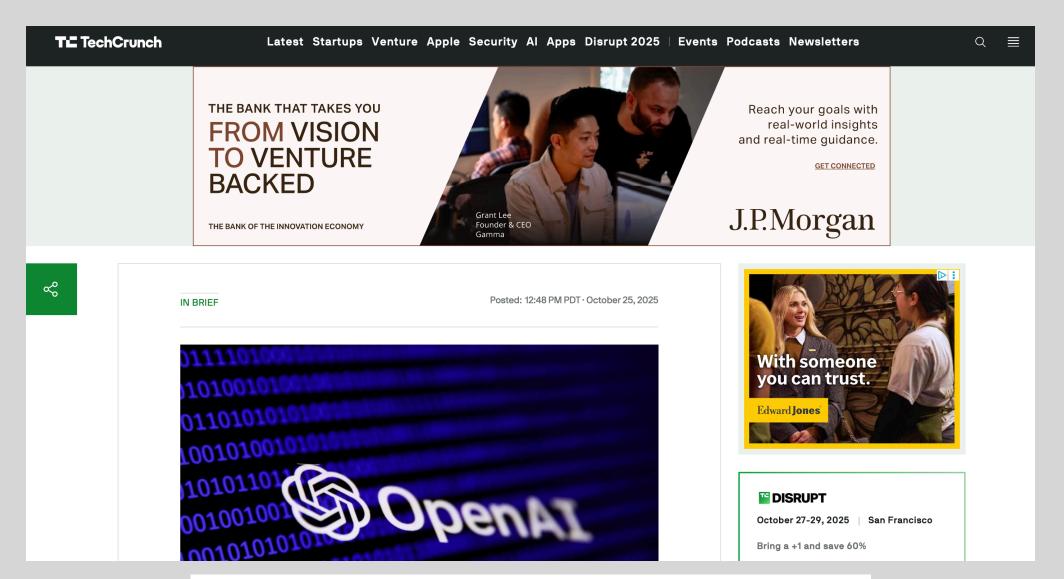
#### content management systems

#### what they are

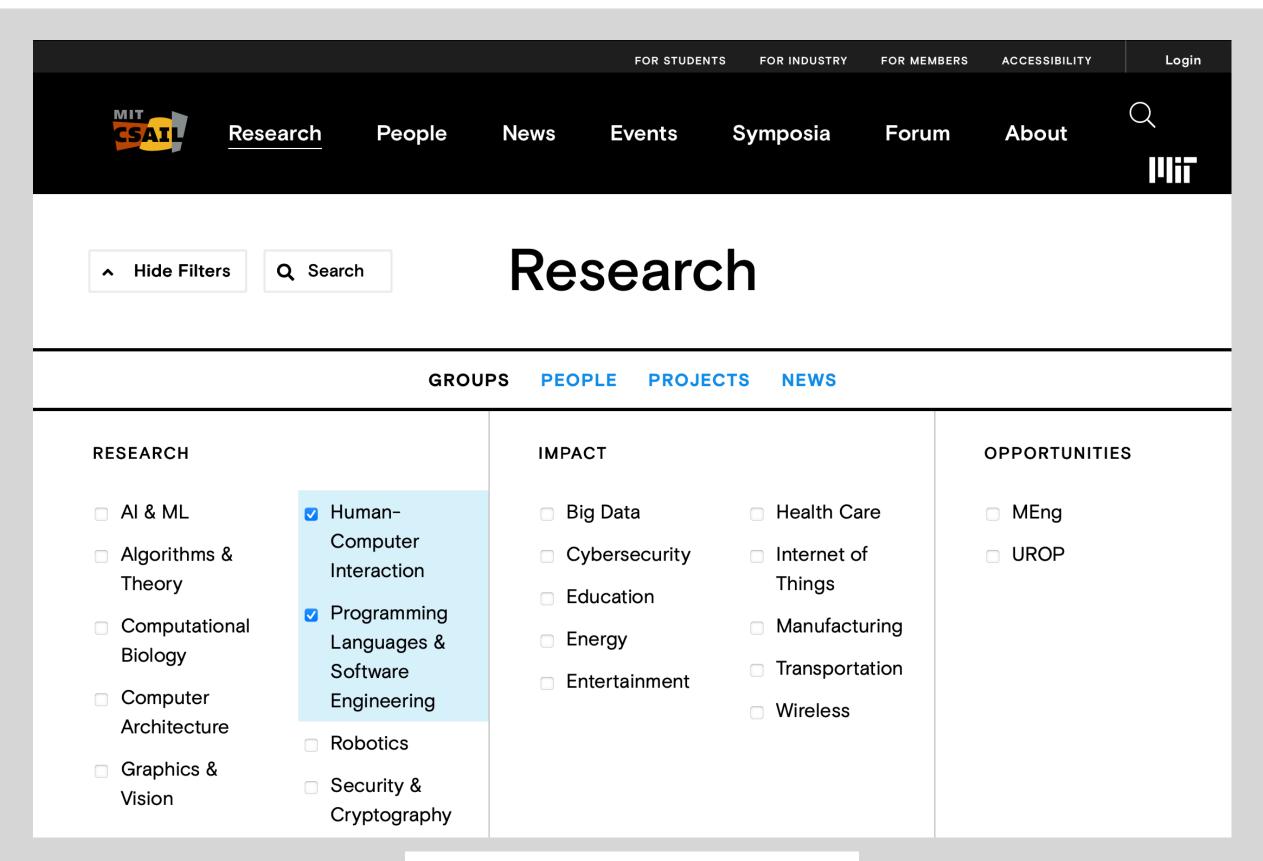
static to users, dynamic to admins updates via HTTP requests (ie, edit site in browser)

#### examples

WordPress: popular CMS for small sites
Wix, Google Sites: combined site builder and CMS
SquareSpace: site builder/CMS for small businesses
Drupal: CMS for complex sites, often built by pros



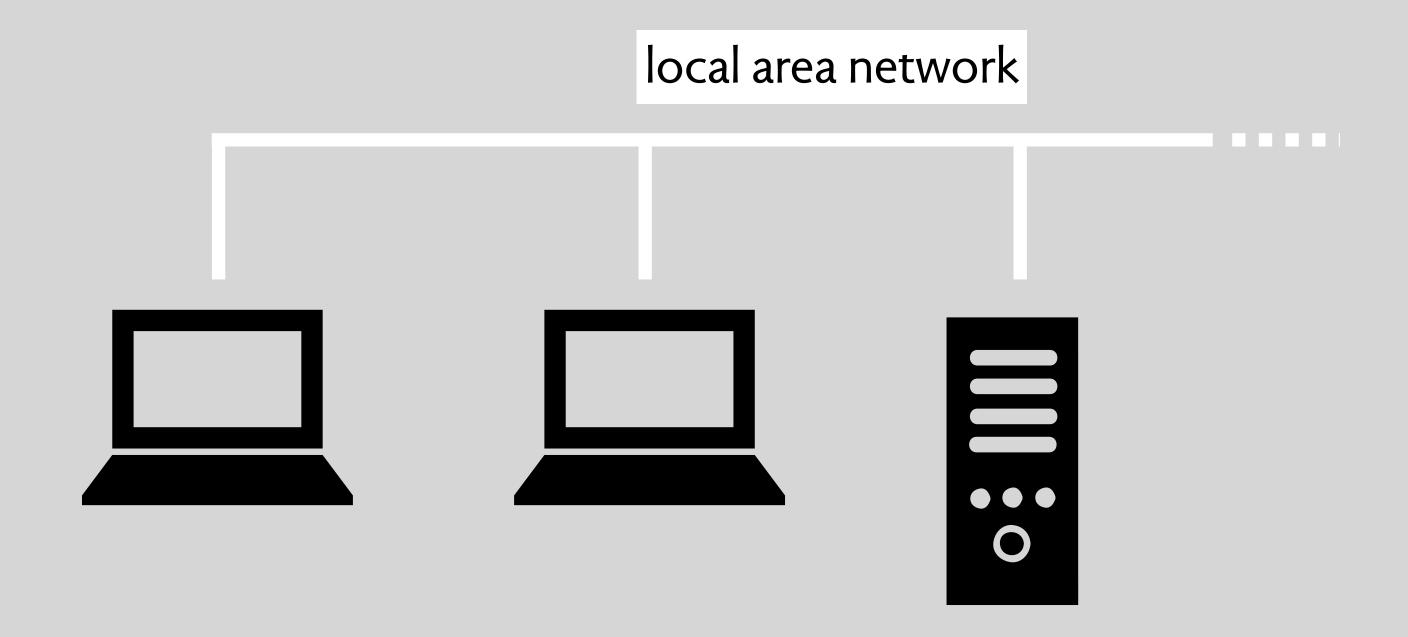
techcrunch.mit.edu (WordPress)



<u>csail.mit.edu</u> (Drupal)

# how are clients authenticated?

#### how to authenticate clients?

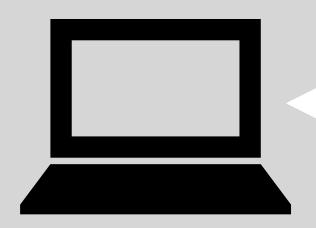


"trust by wire" in early LANs
if you could access the physical network
you were considered authenticated

password auth in early LANs server opens connection, user logs in & access files passwords sent in clear, easy to hijack connection too

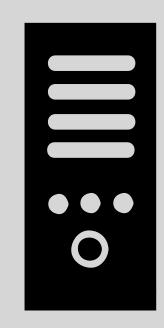
dial-up modems were a bit safer client gets next available modem line modem connection was a private circuit

#### a better scheme: session tokens



client and server establish shared key to prevent sniffing

client sends login credentials checked by server



server sends back session token which client stores in server state
session abc123 is
active and belongs
to user Alice

in client state session token is abc123

client sends token with all subsequent requests

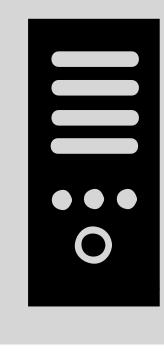
# cookies: a way to store session tokens

login ("dj", "password")



Set-Cookie: id=abc123; Domain=nytimes.com

get (nytimes.com/news)
Cookie: id=abc123

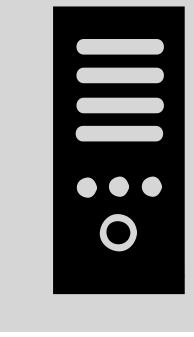


nytimes.com

login ("dj", "password")

Set-Cookie: id=def456; Domain=instagram.com

get (instagram.com/feed) Cookie: id=def456



instagram.com

other uses of cookies shopping carts eg

nice for programmer cookies sent by default

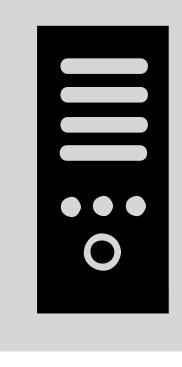
security model domains stay separate

# third-party cookies

get (nytimes.com/news)

get (analytics.com/pixel.gif)

Set-Cookie: id=xyz123; Domain=<u>analytics.com</u>

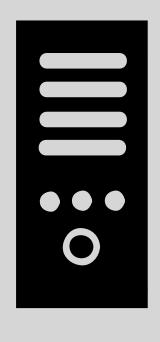


nytimes.com

get (instagram.com/feed)

get (analytics.com/pixel.gif)

Cookie: id=xyz123



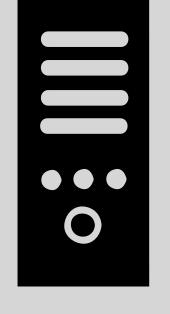
instagram.com

cross site tracking
GDPR requires consent

### JSON web tokens (JWTs)



client and server establish shared key to prevent sniffing



client sends login credentials checked by server

server sends back web token which client stores

in server state nothing!

in client state token says "this is Alice, an admin"

client sends web token with all subsequent requests

# refreshing web tokens

#### a problem: long or short?

with conventional sessions, server can always close a session but with JWTs, can access until token expires insecure if token is long lived, but inconvenient if short lived!

#### solution: two kinds of token

long-lived refresh tokens: rarely used, safe in browser (eg 14 days) short-lived access tokens: used for all accesses (eg 10 mins) when access token expires, client uses refresh token to get a new one when refresh token expires, have to log in again

# what goes on the client or server?

### security considerations

code and data in the browser are visible to and modifiable by the user with developer tools

user can issue any HTTP requests
by modifying JS in the browser document
by commands in the browser JS console
by using curl or Postman



#### so which of these are good strategies?

to prevent access to another user's data ensure client code passes username with request

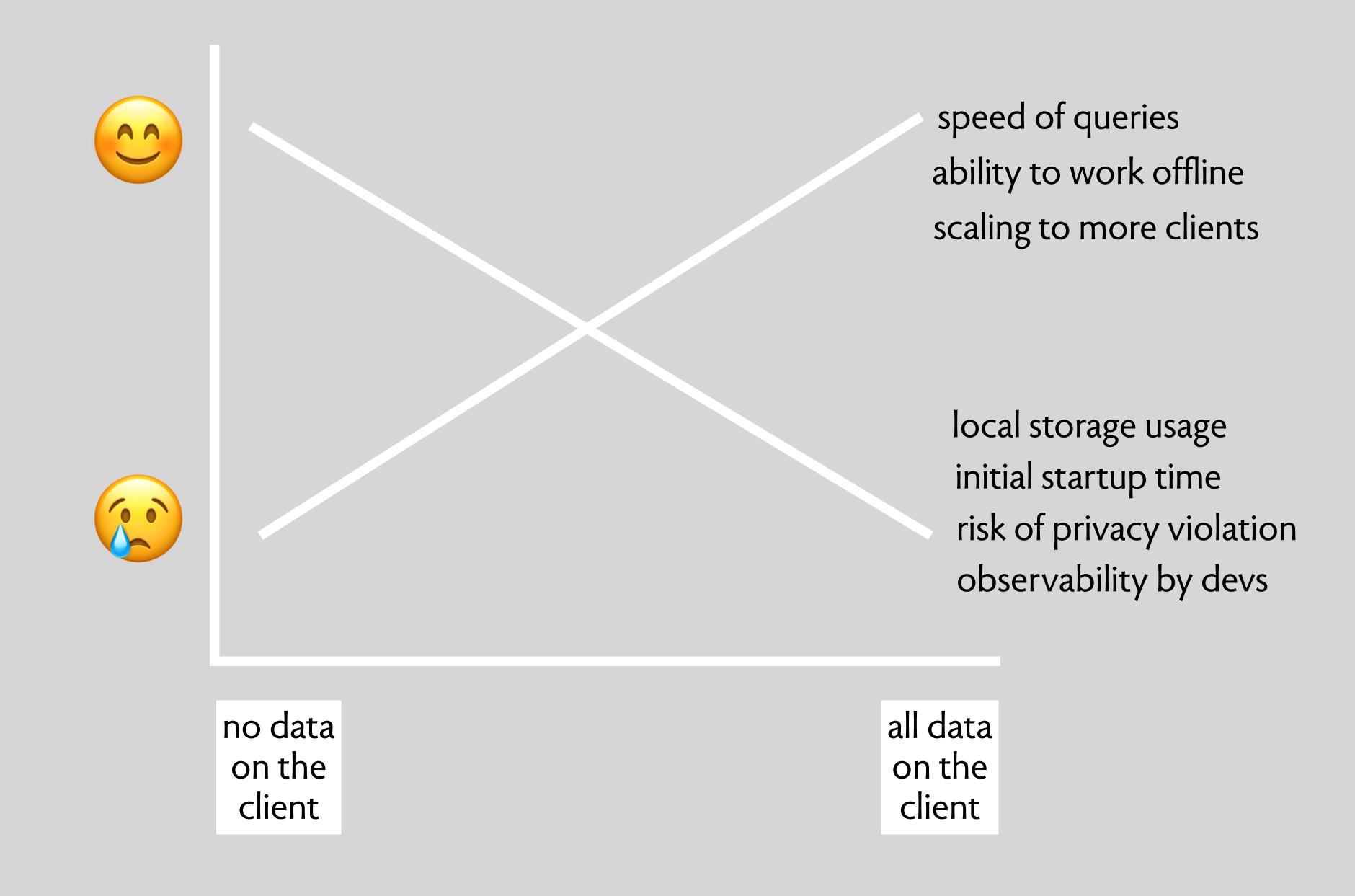
to prevent access to sensitive pages navigate first through login page

to prevent access to another user's data use counter for session ids and store id in cookie

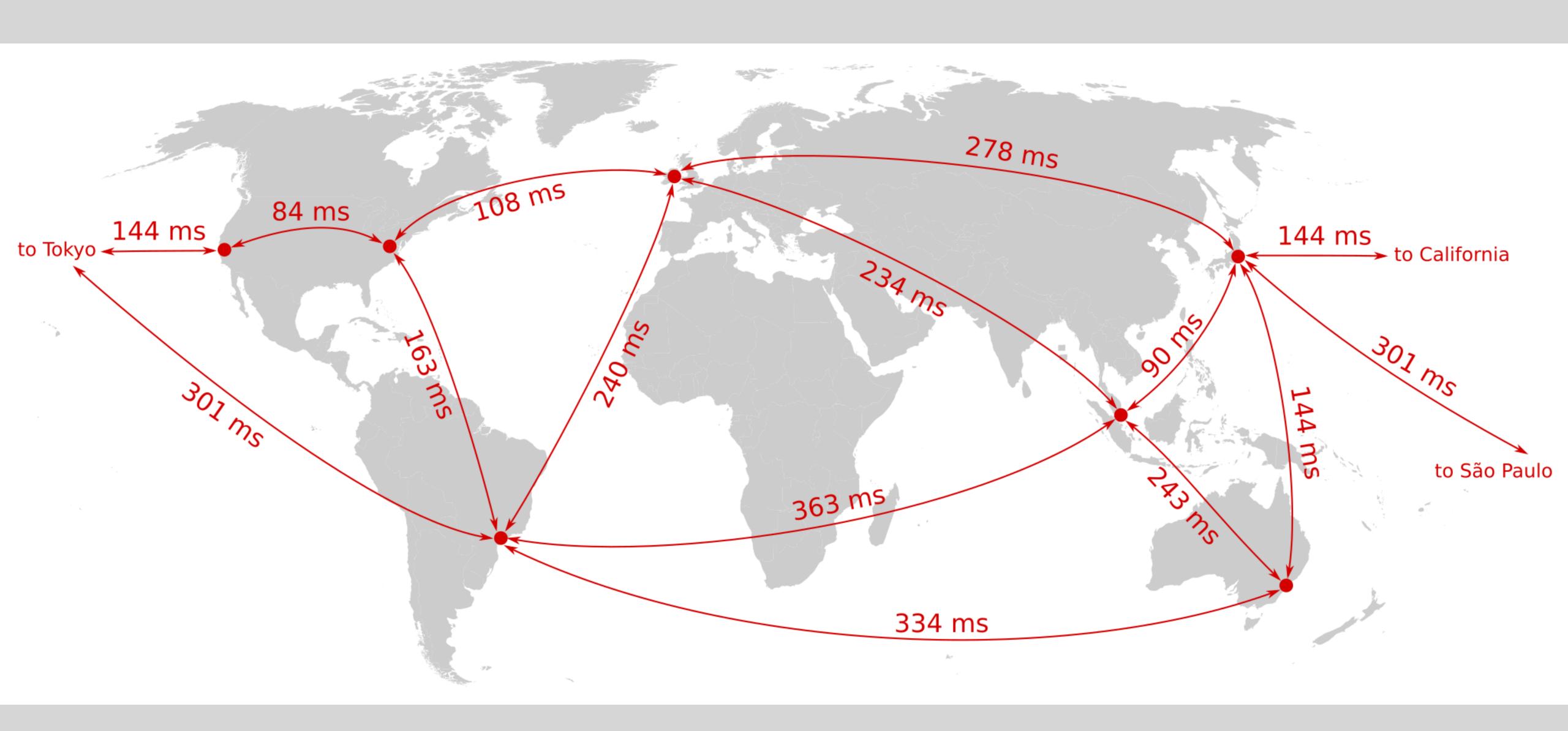
to prevent access to another user's data generate random session if and store in cookie



# performance considerations

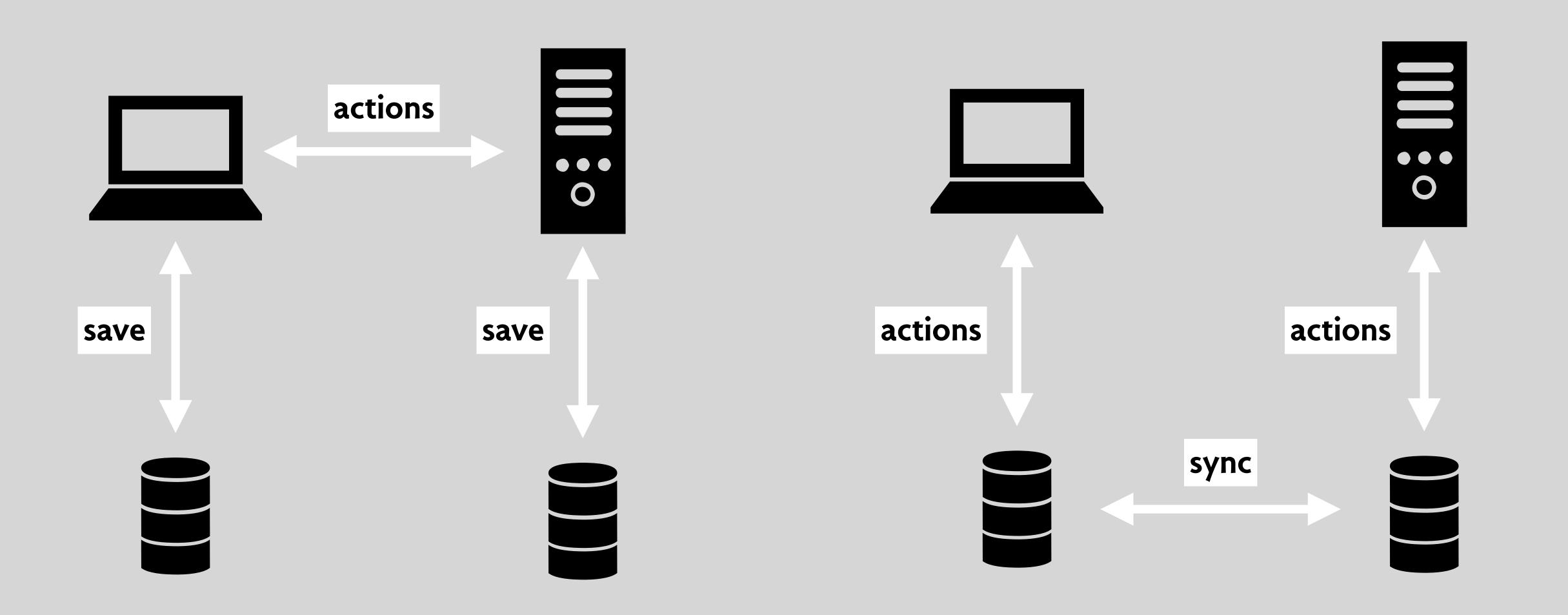


# online apps are slow!



server-to-server round trip times between AWS data centers (Ink & Switch)

# local first: a proposal for a new kind of app



# takeaways

# key ideas from this lecture

client server architecture web apps just one example

APIs specs, our old friend

tricky challenges
many inherent to all apps

web-specific notions
HTTP protocol, DOM

web challenges
how good a sandbox?