

# Introduction to Web Technologies and e-Services

# Contents

1. Internet, Web
2. HTTP
3. URL
4. Web Browser
5. Web Application
6. Web Application Architecture
7. Web Developer Roadmap

# Reasonable Questions

- What is the World Wide Web?
- Is it the same thing as the Internet?
- Who invented it?
- How old is it?
- How does it work?
- What kinds of things can it do?
- What does it have to do with programming?

# Web $\neq$ Internet

- Internet: a physical network connecting millions of computers using the same protocols for sharing/transmitting information (TCP/IP)
  - Internet is a network of smaller networks
- World Wide Web: a collection of interlinked multimedia documents that are stored on the Internet and accessed using a common protocol (HTTP)
- Key distinction: Internet is hardware; Web is software along with data, documents, and other media
- Many other Internet-based applications exist e.g., email, telnet, ftp, usenet, instant messaging services, file-sharing services, ...

# (A Very Brief) History of the Internet

- The idea of a long-distance computer network traces back to early 60's
  - Joseph Licklider at M.I.T. (a “time-sharing network of computers”)
  - Paul Baran at Rand (tasked with designing a “survivable” communications system that could maintain communication between end points even after damage from a nuclear attack)
  - Donald Davies at National Physics Laboratory in U.K.
- In particular, the US Department of Defense was interested in the development of distributed, decentralized networks
  - survivability (i.e., network still functions despite a local attack)
  - fault-tolerance (i.e., network still functions despite local failure)
  - contrast with phone system, electrical system which are highly centralized services

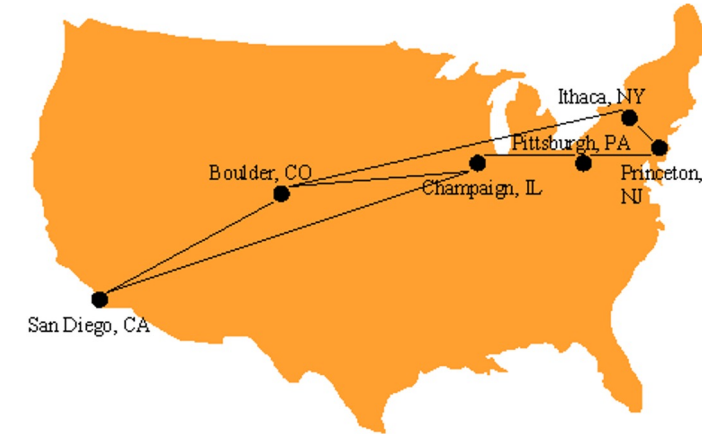
# The Internet

- In 1969, Advanced Research Project Agency funded the ARPANET
  - Connected computers at UC Los Angeles, UC Santa Barbara, Stanford Research Institute, and University of Utah
  - Allowed researchers to share data, communicate
- Technical origin
  - One of earliest attempts to network heterogeneous, geographically dispersed computers
  - Email first available on ARPANET in 1971 (and quickly very popular!)

# The Internet

- Open-access networks
  - Regional university networks (e.g., SURAnet)
  - CSNET for CS departments not on ARPANET
- NSFNET (1985-1995)
  - Primary purpose: connect supercomputer centers
  - Secondary purpose: provide backbone to connect regional networks

The 6 supercomputer centers connected by the early NSFNET backbone



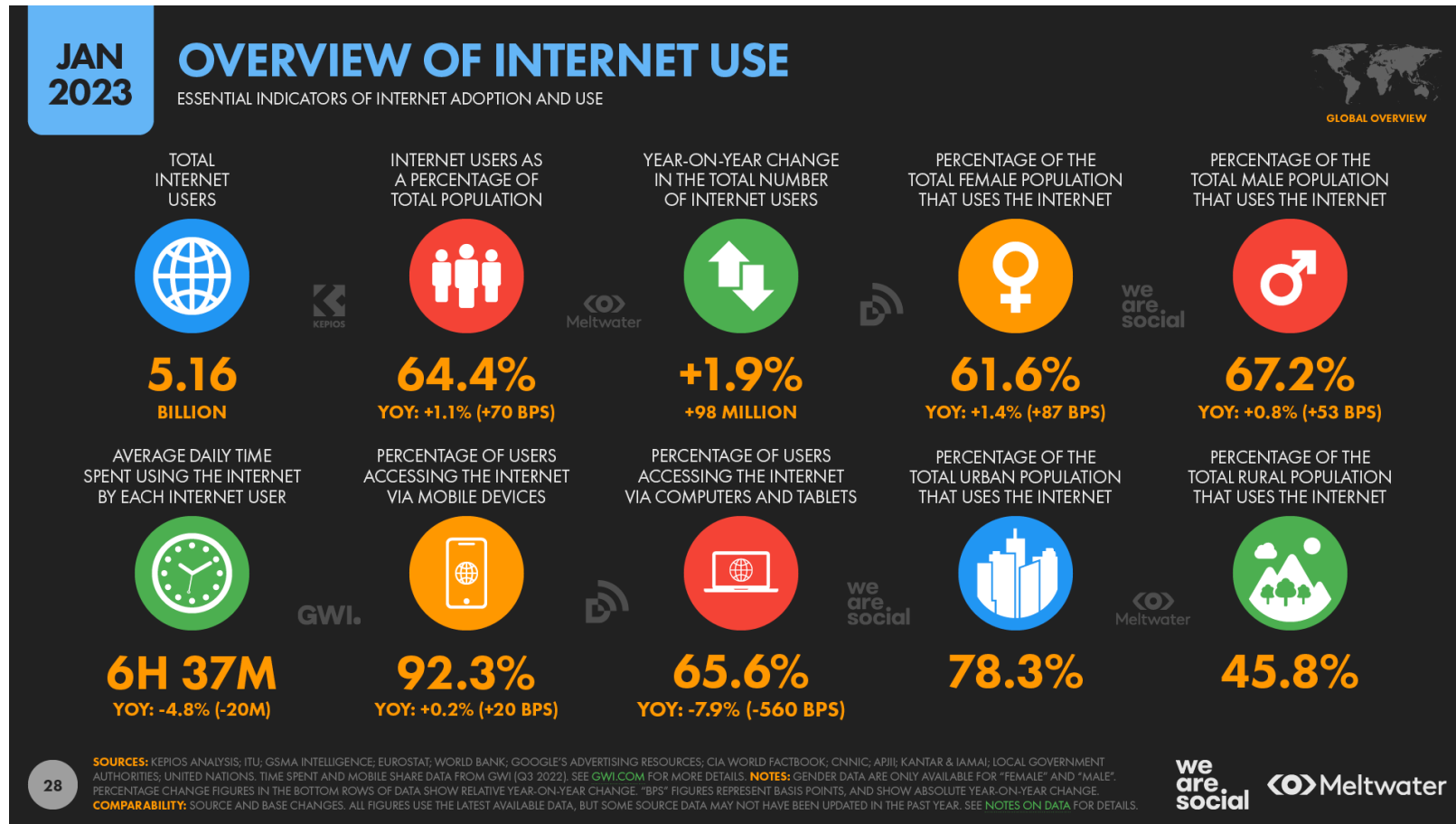
# Internet Growth

- Throughout the 70's, the size of the ARPANET doubled every year
  - First ARPANET e-mail sent in 1971
  - Decentralization makes adding new computers easy
  - TCP/IP developed in the mid 1970s for more efficient packet routing
  - Migration of ARPANET to TCP/IP completed 1 January, 1983
  - ~1000 military & academic host computers connected by 1984
- In 80's, U.S. government took a larger role in Internet development
  - Created NSFNET for academic research in 1986
  - ARPANET was retained for military & government computers
- By 90's, Internet connected virtually all colleges & universities
  - Businesses and individuals also connecting as computing costs fell
  - ~1,000,000 computers by 1992
- In 1992, control of the Internet was transferred to a non-profit organizations
  - Internet society: Internet engineering task force, Internet architecture board, Internet assigned number authority, World-wide-web consortium (W3C)



# Internet Growth (cont.)

- Internet has exhibited exponential growth, doubling in size every 1-2 years (stats from Internet Software Consortium)



# (A Very Brief) History of the Web

- The idea of hypertext (cross-linked and inter-linked documents) traces back to Vannevar Bush in the 1940's
  - Online hypertext systems began to be developed in 1960's
  - In 1987, Apple introduced hypercard (a hypermedia system that predated the WWW)
- In 1989, Tim Berners-lee at the European particle physics laboratory (CERN) designed a hypertext system for linking documents over the internet
  - Designed a (non-wysiwyg) language for specifying document content => Evolved into hypertext markup language (HTML)
  - Designed a protocol for downloading documents and interpreting the content => Evolved into hypertext transfer protocol (HTTP)
  - Implemented the first browser -- text-based, no embedded media

The web was born!

# History of the Web (cont.)

- The Web was an obscure, European research tool until 1993
- In 1993, Marc Andreessen and Eric Bina (at the National Center for Supercomputing Applications, a unit of the University of Illinois) developed Mosaic, one of the early graphical Web browsers that popularized the WWW for the general public
- Andreessen left NCSA to found Netscape in 1994
  - Cheap/free browser further popularized the Web (75% market share in 1996)
- In 1995, Microsoft came out with Internet Explorer
- Opera web browser released in 1996
  - Firefox web browser, version 1.0, released in 2004
  - Google Chrome released in 2008
- Today, the Web is the most visible aspect of the Internet

# World Wide Web

- **The Web** is the collection of machines (Web servers) on the Internet that provide information, particularly HTML documents, via HTTP.
- Machines that access information on the Web are known as **Web clients**. A **Web browser** is software used by an end user to access the Web.

# Hypertext Transport Protocol (HTTP)

- HTTP is based on the request-response communication model:
  - Client sends a request
  - Server sends a response
- HTTP is a stateless protocol:
  - The protocol does not require the server to remember anything about the client between requests.

# HTTP

- Normally implemented over a TCP connection (80 is standard port number for HTTP)
- Typical browser-server interaction:
  - User enters Web address in browser
  - Browser uses DNS to locate IP address
  - Browser opens TCP connection to server
  - Browser sends HTTP request over connection
  - Server sends HTTP response to browser over connection
  - Browser displays body of response in the client area of the browser window

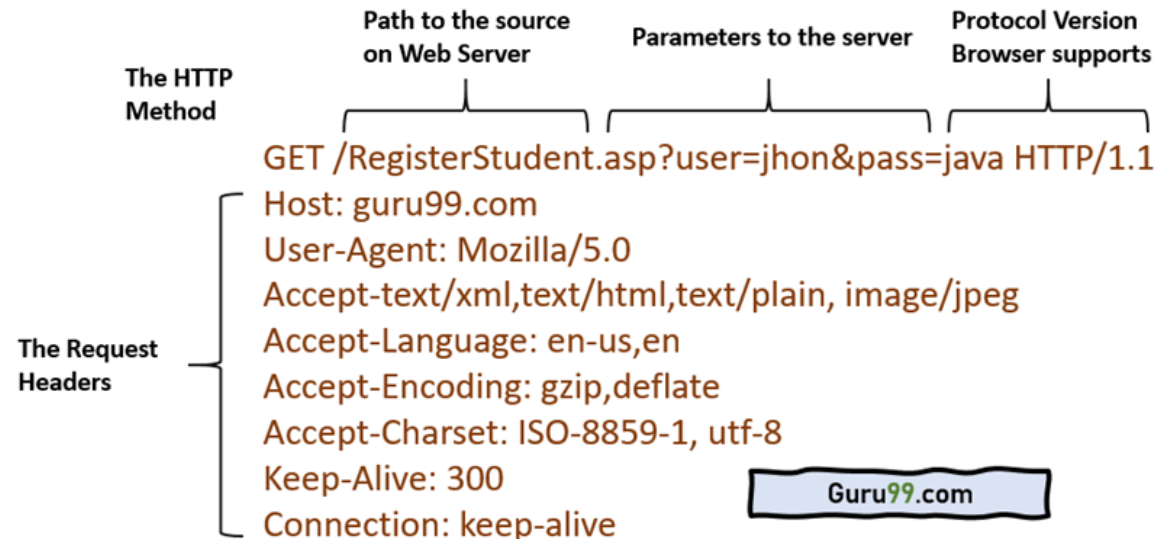
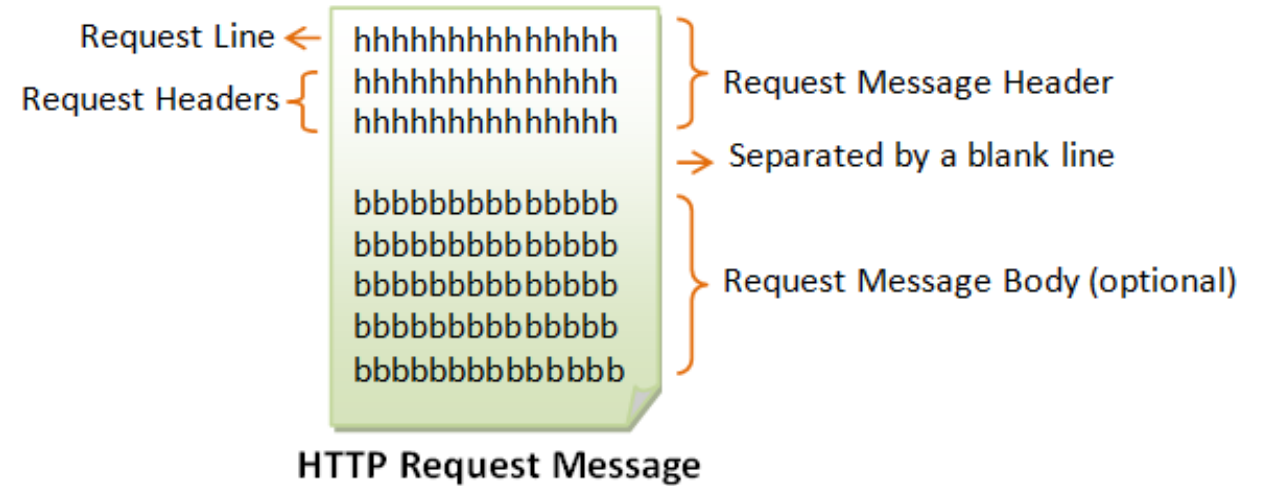
# HTTP Request

- Structure of the request:
  - request line
  - header field(s)
  - blank line
  - optional body

# HTTP Request

- Structure of the request:

- request line
- request headers
- blank line
- optional request body





# HTTP Request

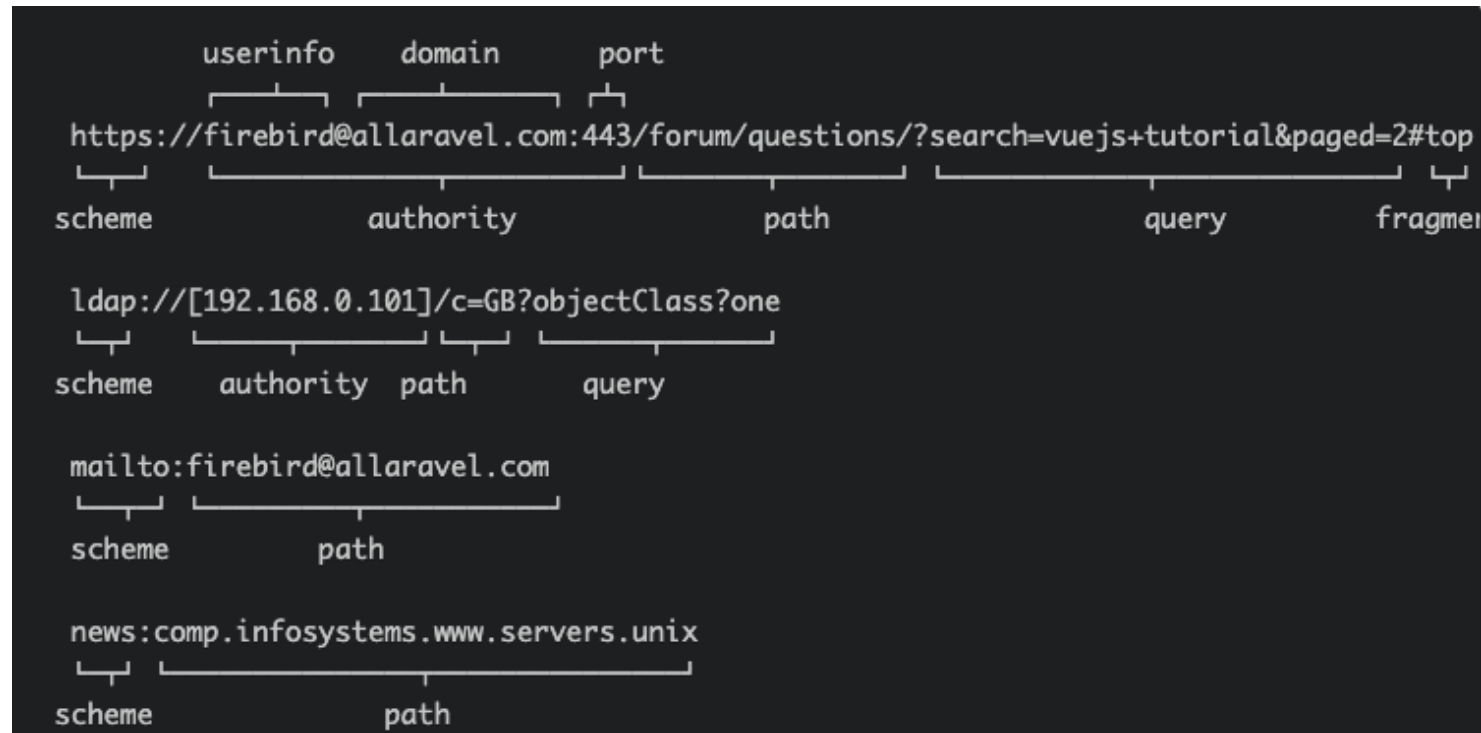
- Request line
  - GET /test.html HTTP/1.1
  - POST /index.html HTTP/1.1
- Three space-separated parts:
  - HTTP request method
  - Request-URI ([Uniform Resource Identifier](#))
  - HTTP version

# HTTP Request

- Request line
  - GET /test.html HTTP/1.1
  - POST /index.html HTTP/1.1
- Three space-separated parts:
  - HTTP request method
  - Request-URI
  - **HTTP version**
    - 1.1: 1997
    - 2: 2015
    - 3: 2022

# URI

- Uniform Resource Identifier
- URI = scheme:[//authority]path[?query][#fragment]

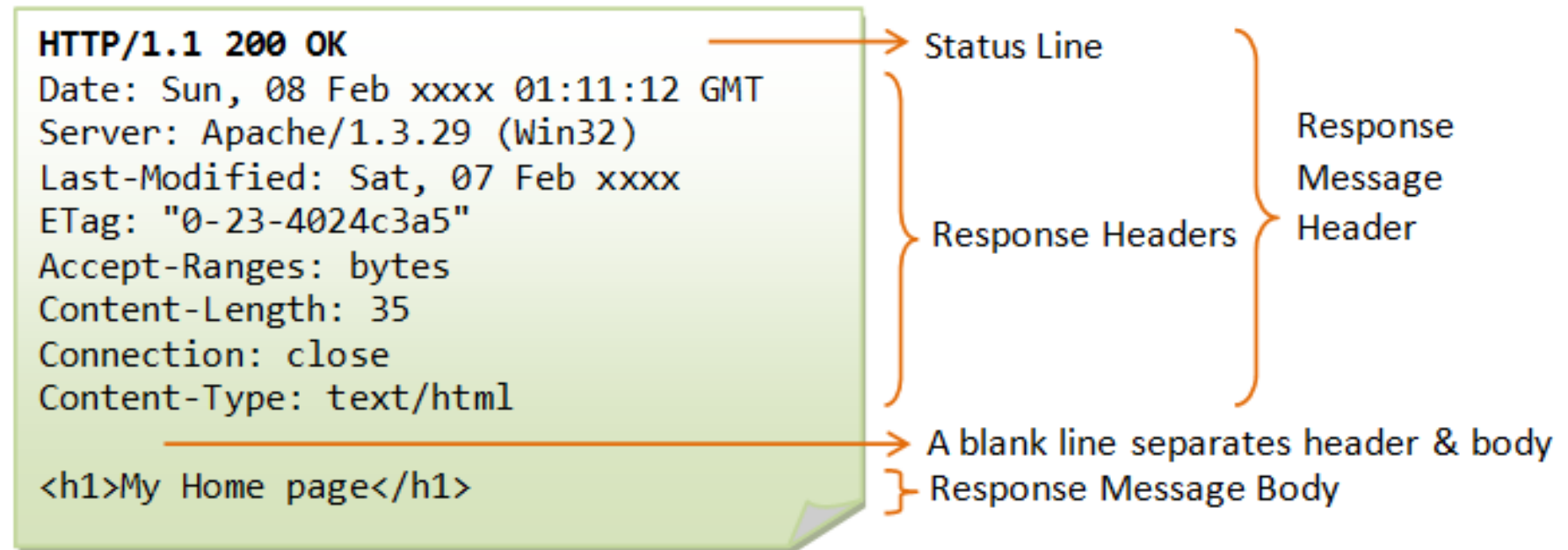


# URI

- URI's are of two types:
- Uniform Resource Name (URN)
  - Can be used to identify resources with unique names, such as books (which have unique ISBN's)
  - Scheme is urn
- Uniform Resource Locator (URL)
  - Specifies location at which a resource can be found
  - In addition to http, some other URL schemes are https, ftp, mailto, and file

# HTTP Response

- Structure of the response:
  - status line
  - header field(s)
  - blank line
  - optional body



# HTTP Response

- Structure of the response:
  - **status line**
  - header field(s)
  - blank line
  - optional body

# HTTP Response

- Status line
  - Example: HTTP/1.1 200 OK
- Three space-separated parts:
  - HTTP version
  - status code
  - reason phrase (intended for human use)

# HTTP Response

- Status code
  - Three-digit number
  - First digit is class of the status code:
    - 1=Informational
    - 2=Success
    - 3=Redirection (alternate URL is supplied)
    - 4=Client Error
    - 5=Server Error
  - Other two digits provide additional information
  - See <http://www.w3.org/Protocols/rfc2616/rfc2616-sec10.html>



# HTTP Response

- Structure of the response:
  - status line
  - **header field(s)**
  - blank line
  - optional body

# HTTP Response

- Common header fields:
  - **Connection**, **Content-Type**, **Content-Length**
  - **Date**: date and time at which response was generated (required)
  - **Location**: alternate URI if status is redirection
  - **Last-Modified**: date and time the requested resource was last modified on the server
  - **Expires**: date and time after which the client's copy of the resource will be out-of-date
  - **ETag**: a unique identifier for this version of the requested resource (changes if resource changes)

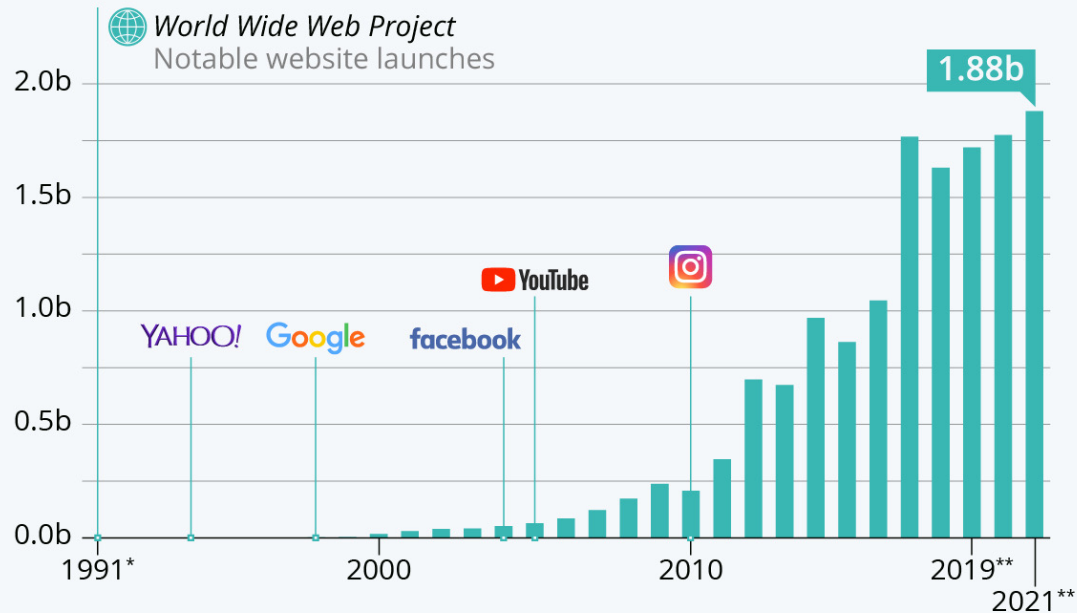
# HTTP Request/Response Examples

Connect	{	<b>\$ telnet www.example.org 80</b> Trying 192.0.34.166... Connected to www.example.com (192.0.34.166). Escape character is '^['.
Send Request	{	<b>GET / HTTP/1.1</b> <b>Host: www.example.org</b>
Receive Response	{	HTTP/1.1 200 OK Date: Tue, 11 Oct 2022 20:30:49 GMT ...

# Web Growth

## How Many Websites Are There?

Number of websites online from 1991 to 2021



\* As of August 1, 1991.

\*\* Latest available data for 2019: October 28, for 2020: June 2, for 2021: August 6.

Source: Internet Live Stats



statista

Google Chrome

Firefox

Safari

IE, Opera

Netscape

Mosaic

Year	Computers on the Internet	Web Servers on the Internet
2011	~605,000,000	~250,000,000
2008		172,338,726
2006	439,286,364	85,541,228
2004	285,139,107	56,923,737
2002	162,128,493	33,082,657
2000	93,047,785	18,169,498
1998	36,739,000	4,279,000
1996	12,881,000	300,000
1994	3,212,000	3,000
1992	992,000	50

# Web growth (cont.)

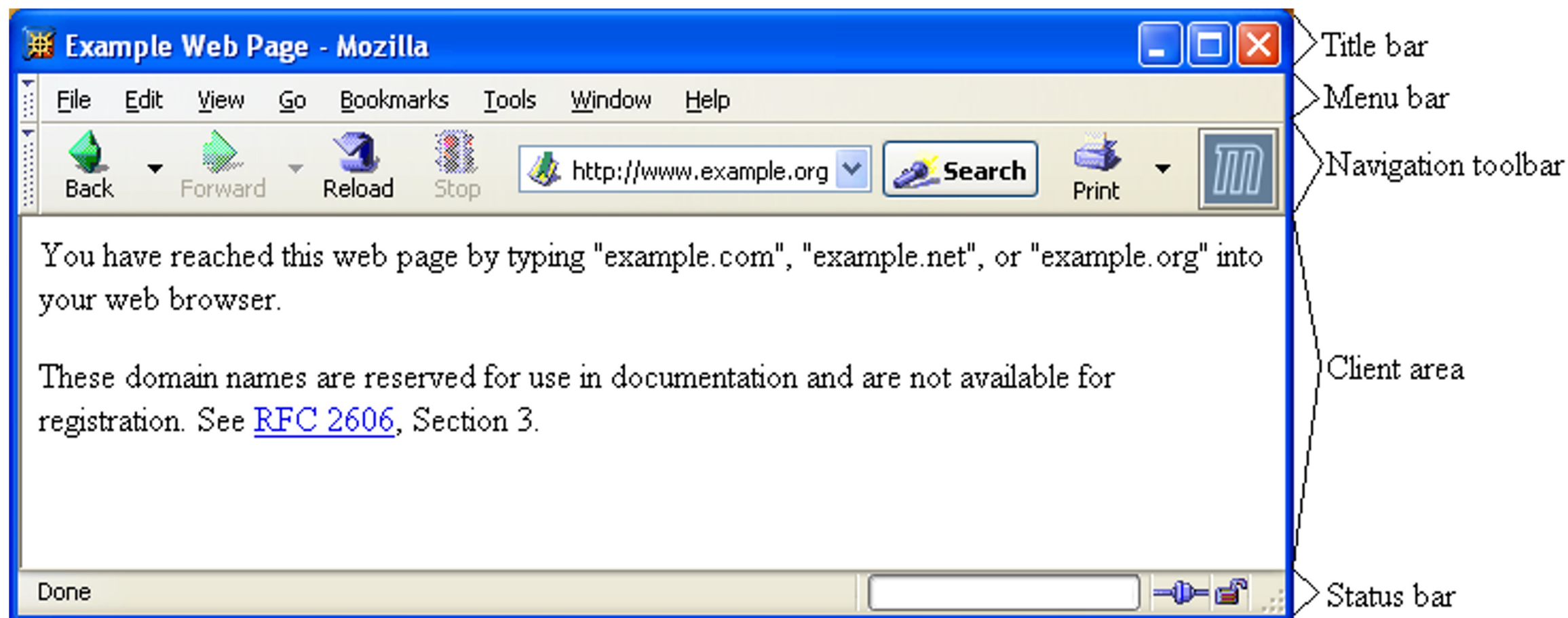
- Internet addresses are used to identify computers on the internet.
- Internet Protocol version 4 (IPv4) was first defined in 1981 and is still in use today, but this uses a 32-bit number to specify addresses.
- IPv4 provides around 4.29 billion addresses that are in use (or reserved).
- IPv6 had been deployed since the mid-2000s and uses 128 bit addresses, but also redesigned to allow more efficient routing, network aggregation, and ease of network reconfiguration.

# Web Browsers

- First graphical browser running on general-purpose platforms:



# Web Browsers



# Web Browsers

- Primary tasks:
  - Convert web addresses (URL's) to HTTP requests
  - Communicate with web servers via HTTP
  - Render (appropriately display) documents returned by a server



# Web Browsers - History

- 1990. WordWideWeb, Tim Berners-Lee
- 1993. Mosaic 1.0
- 1994. Netscape Navigator 1.0
- 1995. Microsoft Internet Explorer 1.0
- 1996. Opera 2.0
- 2002. Mozilla Phoenix 0.1
- 2003. Apple Safari Public Beta
- 2004. Mozilla Firefox 1.0
- 2008. Google Chrome Beta

# Web Browsers - History



IE crew in front of Netscape building  
IE 4.0, 1997



The next day

# Static vs. Dynamic pages

- Most Web pages are *static*
  - contents (text/links/images) are the same each time it is accessed  
*e.g., online documents, most homepages*
  - HyperText Markup Language (HTML) is used to specify text/image format
- Online services and e-commerce continues to grow, Web pages must also provide dynamic content
  - pages can be fluid, changeable (e.g., rotating banners)
  - must be able to react to the user's actions, request and process info, tailor services, e.g., amazon.com

# Web Clients

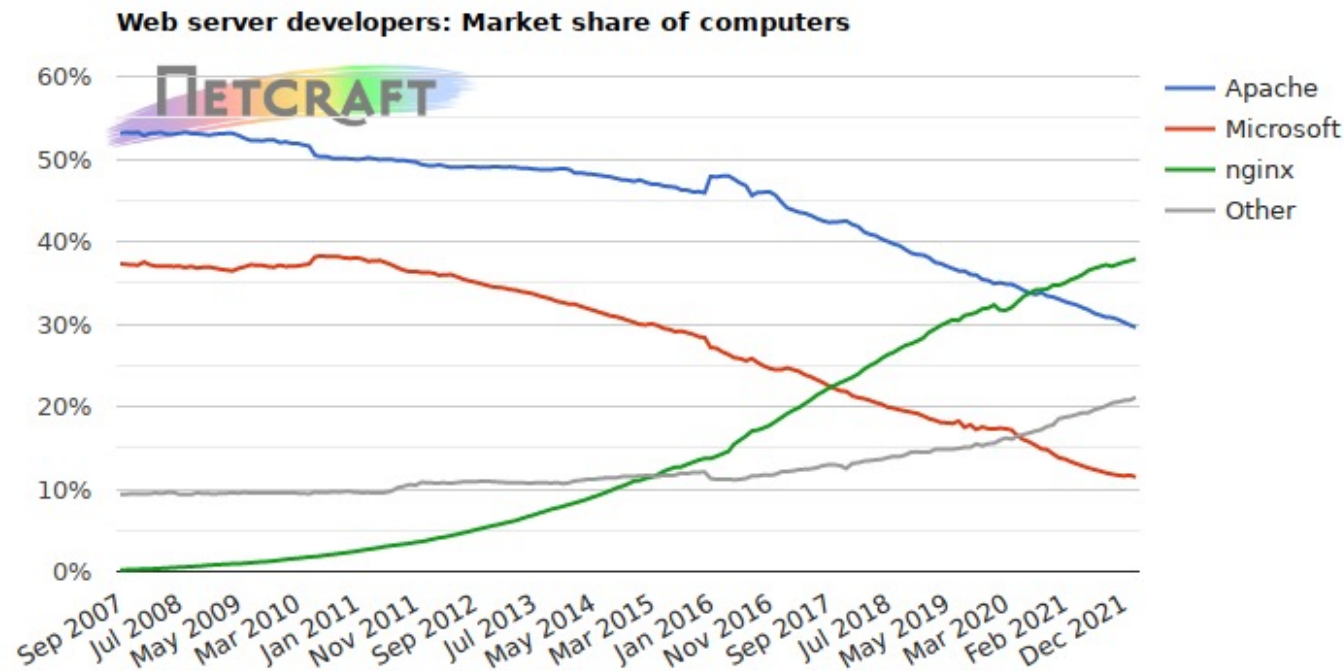
- Many possible web clients:
  - Text-only “browser” (lynx)
  - Mobile phones
  - Robots (software-only clients, e.g., search engine “crawlers”)
  - etc.

# Web Servers

- Receive HTTP request via TCP
- Map host header (domain name) to specific **virtual host** (one of many host names sharing an IP address)
- Map Request-URI to specific resource associated with the virtual host
  - File: Return file in HTTP response
  - Program: Run program and return output in HTTP response
- Map type of resource to appropriate MIME type and use to set Content-Type header in HTTP response
- Log information about the request and response

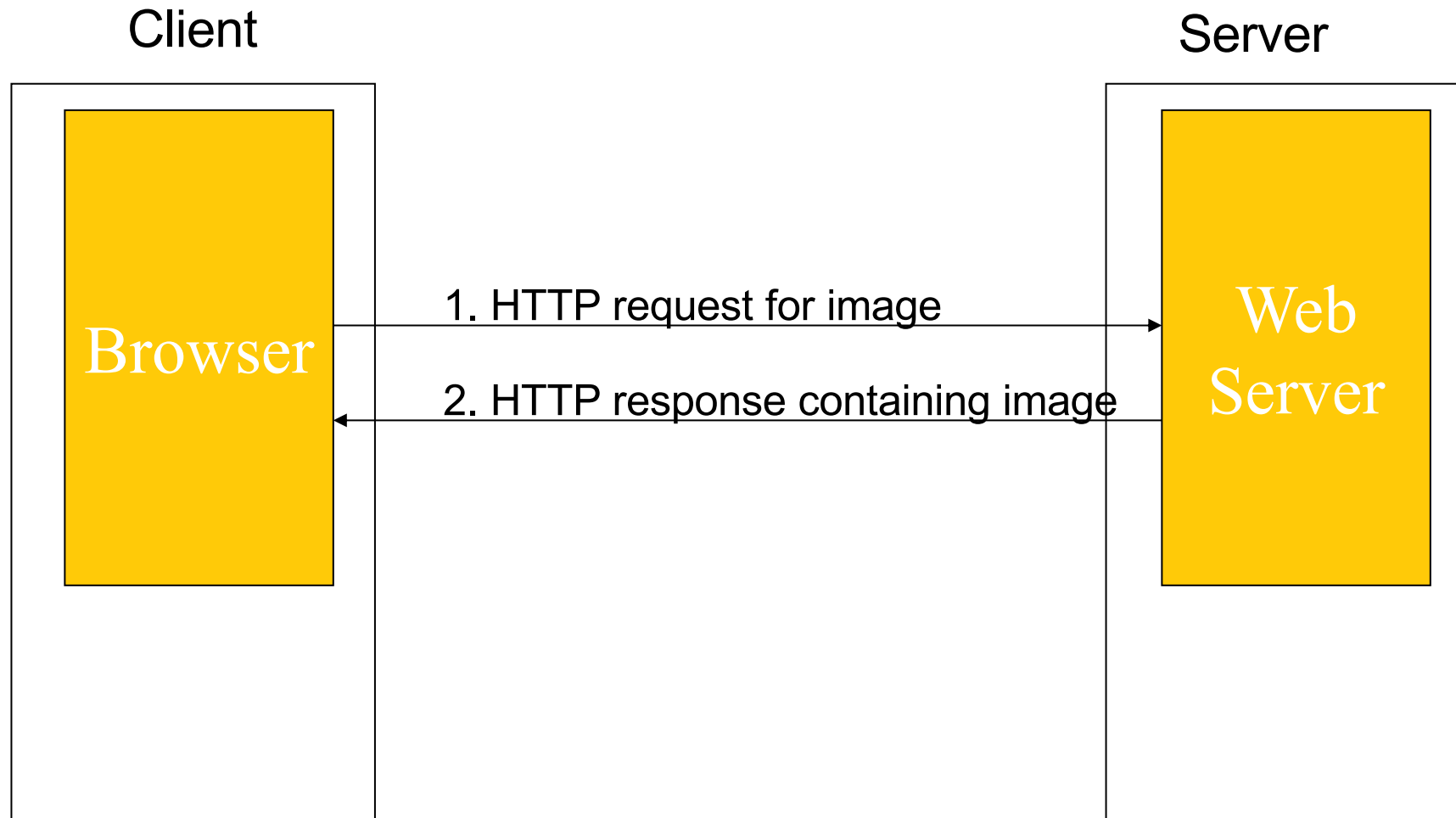
# Web Servers

- httpd: Apache HTTP Server was launched in 1995
- IIS: Microsoft Internet Information Server was released in 1995
- Nginx: open-source software, was released in 2004

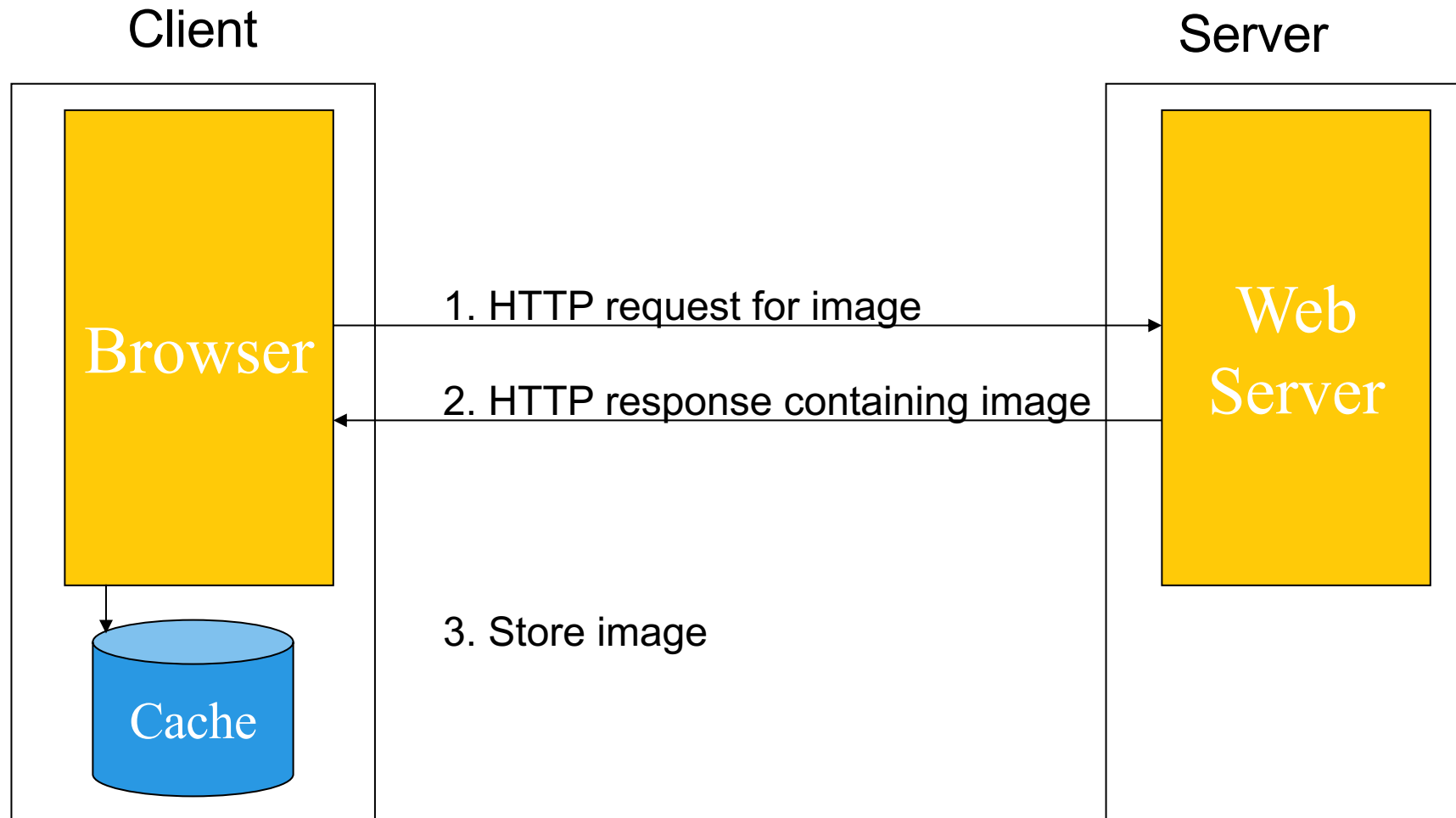


Web server survey, Feb.2022

# Web server/client



# Web server/client

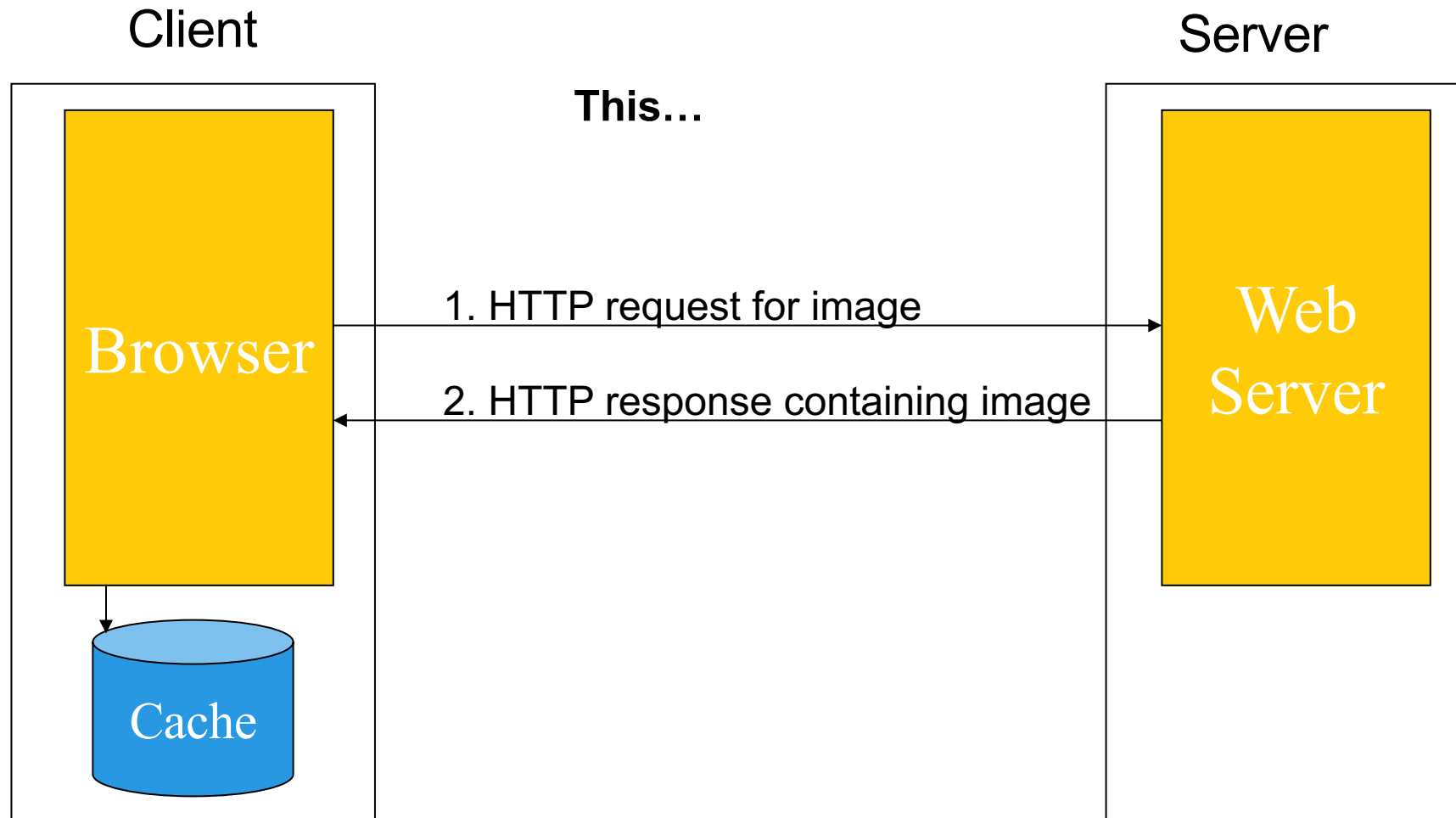




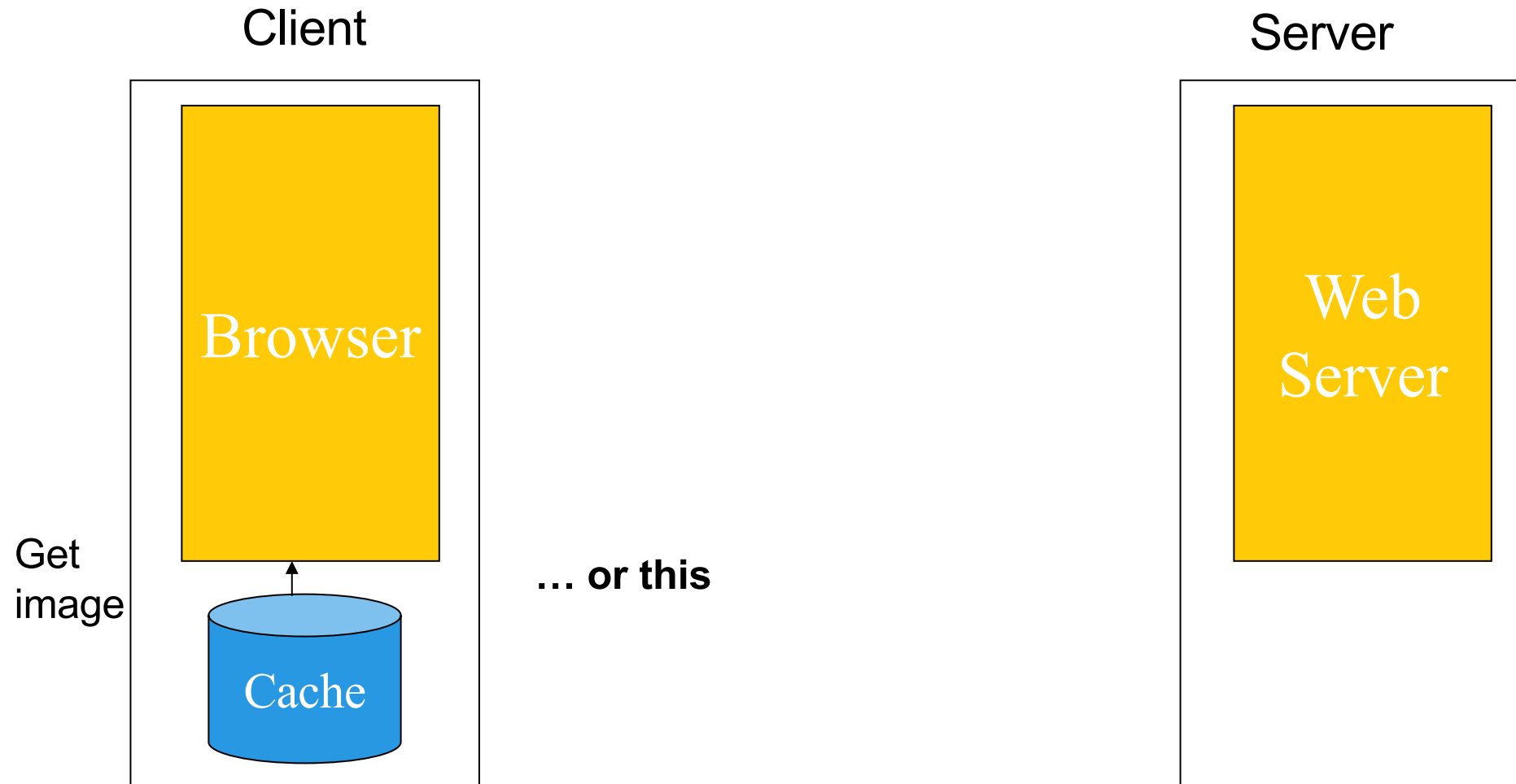
# Web server/client



# Web server/client

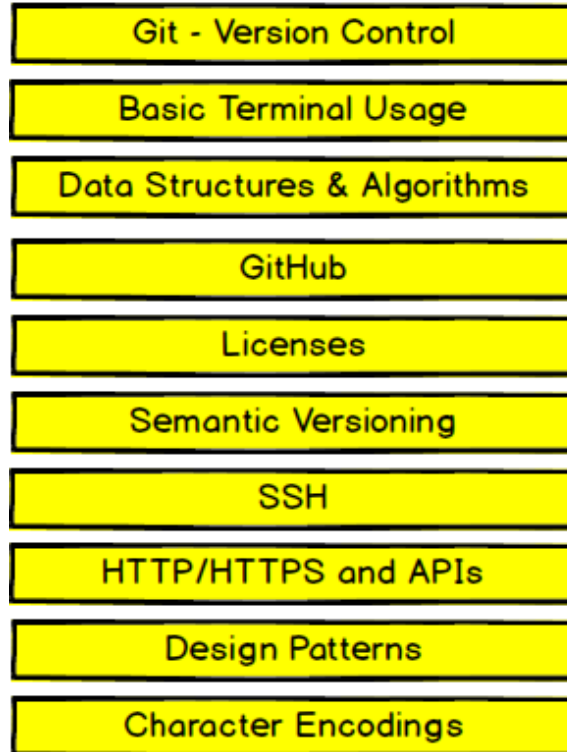


# Web server/client

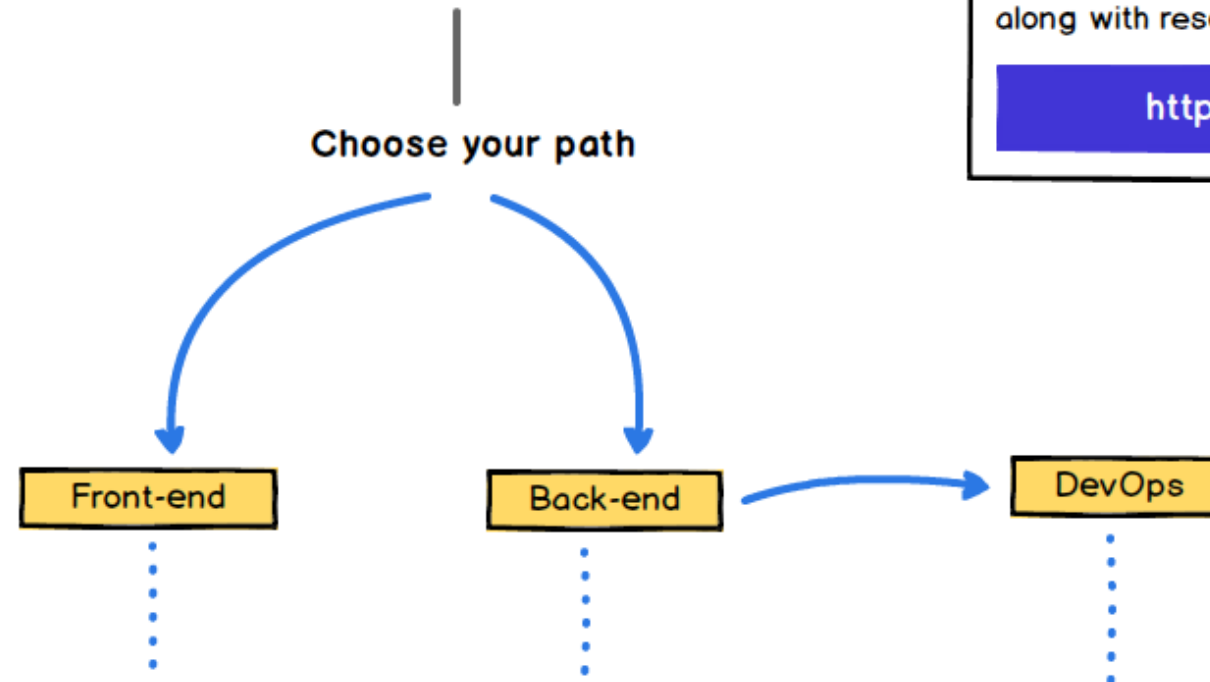


# Web Developer Roadmap 2021

Required for any path



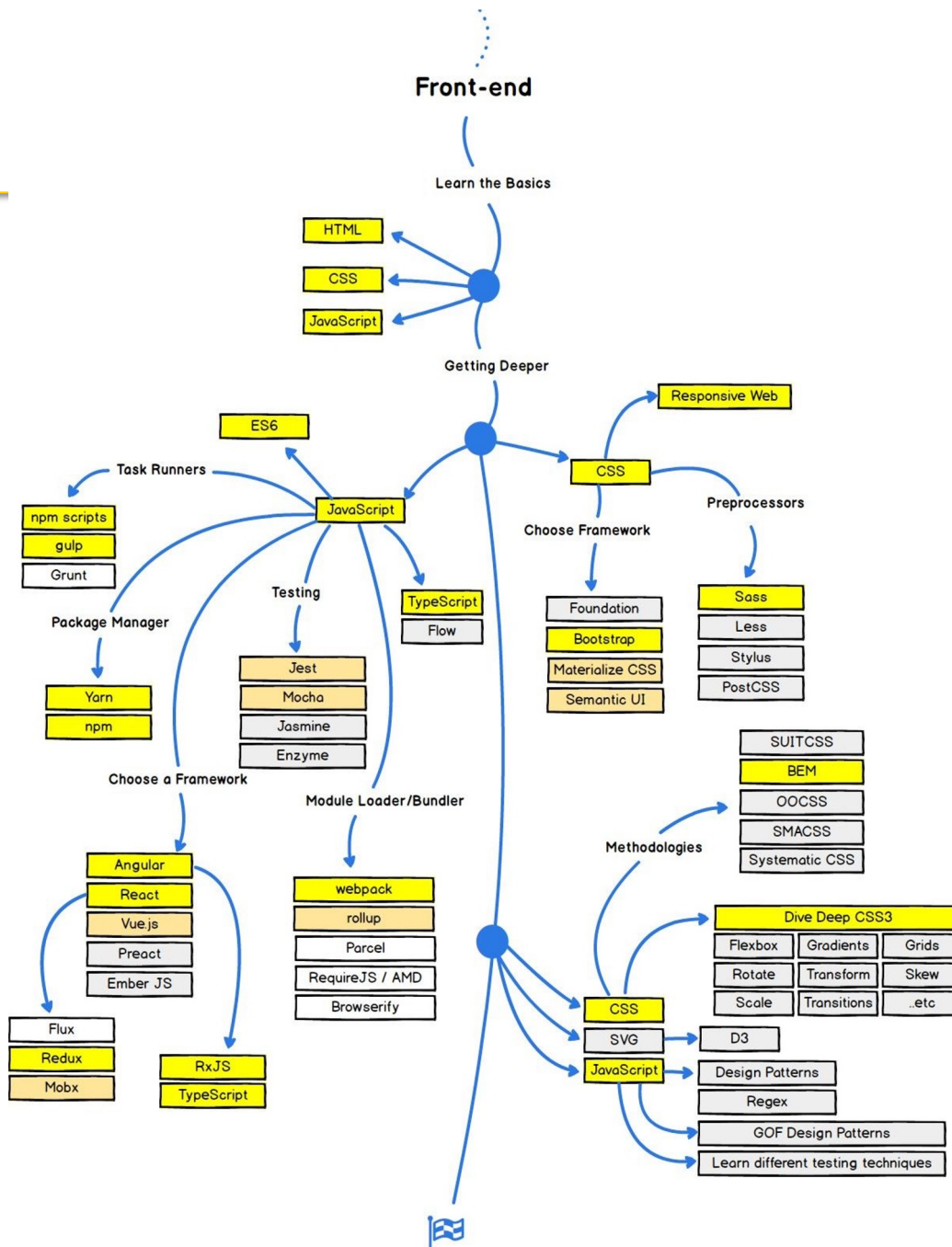
## Web Developer in 2021



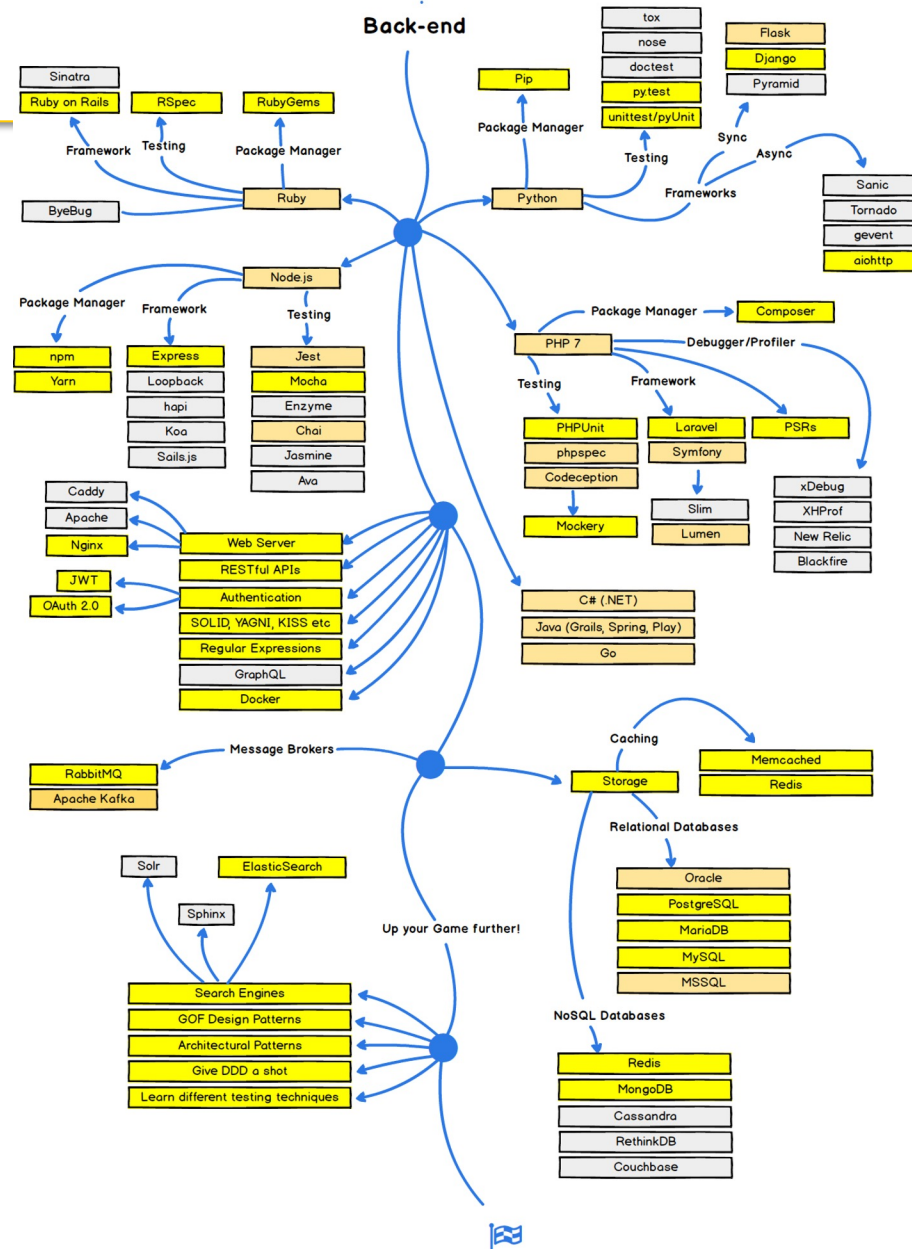
Find the detailed version of this roadmap along with resources and other roadmaps

[http:// roadmap.sh](http://roadmap.sh)

# Front-end



# Back-end



# DevOps

