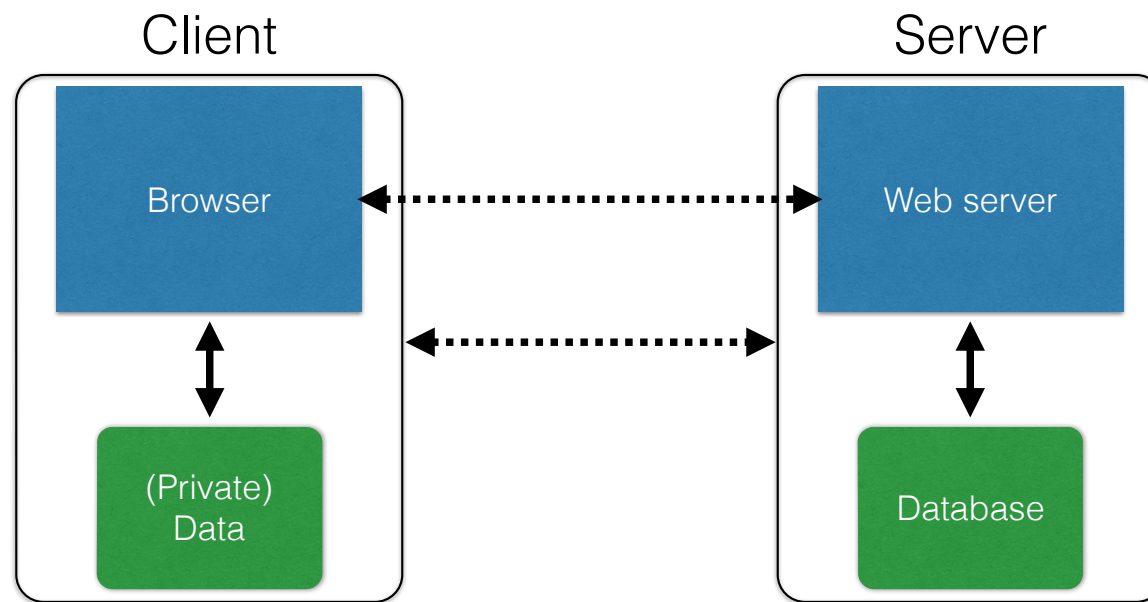


# Web Basics

# The Web, Basically



**(Much) user data is part of the browser**

**DB is a separate entity, logically (and often physically)**

# Interacting with web servers

## **Resources which are identified by a URL**

(Universal Resource Locator)

`http://www.cs.umd.edu/~mwh/index.html`

### **Protocol      Hostname/server      Path to a resource**

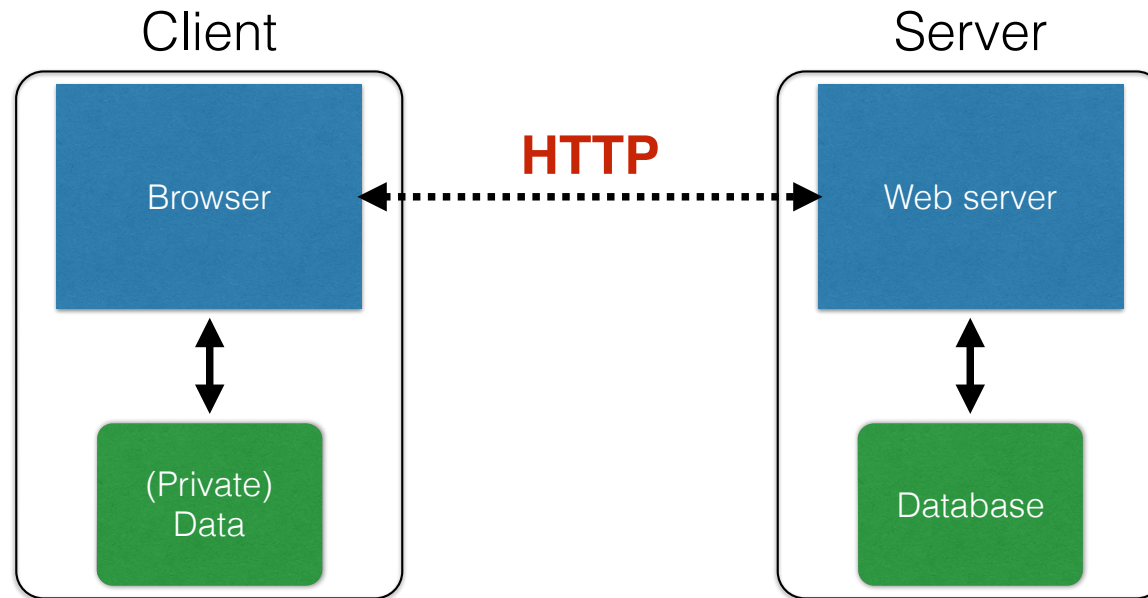
ftp	Translated to an IP address by DNS	
https	Here, the file <code>index.html</code> is <b>static content</b>	
tor	i.e., (e.g. <code>128.8.127.3</code> ) a fixed file returned by the server	

`http://facebook.com/delete.php?f=joe123&w=16`

### **Path to a resource Arguments**

Here, the file `delete.php` is **dynamic content**  
i.e., the server generates the content on the fly

# *Basic* structure of web traffic



- HyperText Transfer Protocol (**HTTP**)
  - An “application-layer” protocol for exchanging collections of data

# *Basic* structure of web traffic



- **Requests contain:**
  - The **URL** of the resource the client wishes to obtain
  - **Headers** describing what the browser can do
- **Request types** can be **GET** or **POST**
  - **GET**: all data is in the URL itself (no server side effects)
  - **POST**: includes the data as separate fields (can have side effects)

# HTTP GET requests

<http://www.reddit.com/r/security>

## HTTP Headers

http://www.reddit.com/r/security

GET /r/security HTTP/1.1

Host: www.reddit.com

User-Agent: Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.9.2.11) Gecko/20101013 Ubuntu/9.04 (jaunty) Firefox/3.6.11

Accept: text/html,application/xhtml+xml,application/xml;q=0.9,\*/\*;q=0.8

Accept-Language: en-us,en;q=0.5

Accept-Encoding: gzip,deflate

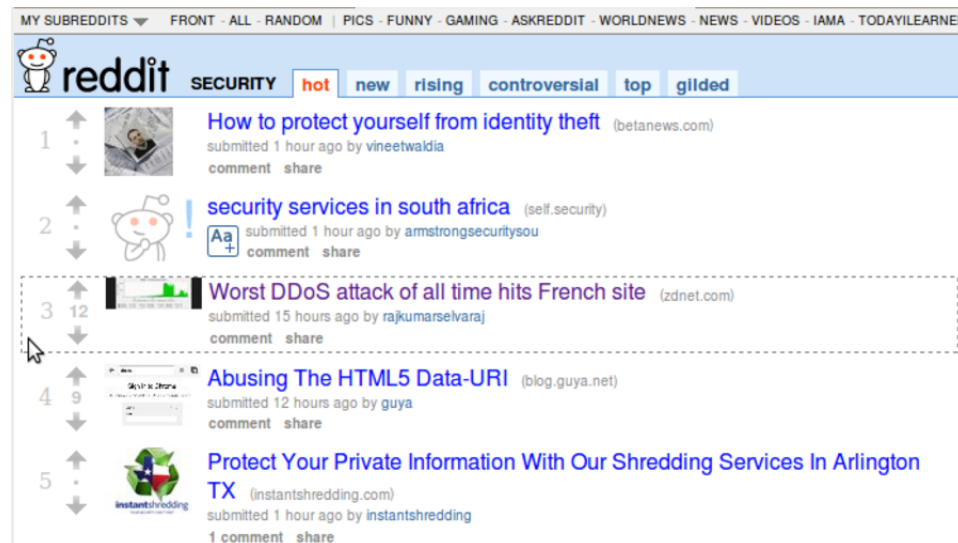
Accept-Charset: ISO-8859-1,utf-8;q=0.7,\*;q=0.7

Keep-Alive: 115

Connection: keep-alive

\_\_utmc=55650...

**User-Agent** is typically a **browser**  
but it can be `wget`, `JDK`, etc.



## HTTP Headers

<http://www.zdnet.com/worst-ddos-attack-of-all-time-hits-french-site-7000026330/>

GET /worst-ddos-attack-of-all-time-hits-french-site-7000026330/ HTTP/1.1

Host: www.zdnet.com

User-Agent: Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.9.2.11) Gecko/20101013 Ubuntu/9.04 (jaunty) Firefox/3.6.11

Accept: text/html,application/xhtml+xml,application/xml;q=0.9,\*/\*;q=0.8

Accept-Language: en-us,en;q=0.5

Accept-Encoding: gzip,deflate

Accept-Charset: ISO-8859-1,utf-8;q=0.7,\*;q=0.7

Keep-Alive: 115

Connection: keep-alive

Referer: <http://www.reddit.com/r/security>

**Referrer URL: the site from which  
this request was issued.**

# HTTP POST requests

## Posting on Piazza

### HTTP Headers

https://piazza.com/logic/api?method=content.create&aid=hrteve7t83et

POST /logic/api?method=content.create&aid=hrteve7t83et HTTP/1.1

Host: piazza.com

User-Agent: Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.9.2.11) Gecko/20101013 Ubuntu/9.04 (jaunty) Firefox/3.6.11

Accept: application/json, text/javascript, \*/\*; q=0.01

Accept-Language: en-us,en;q=0.5

Accept-Encoding: gzip,deflate

Accept-Charset: ISO-8859-1,utf-8;q=0.7,\*;q=0.7

Keep-Alive: 115

Connection: keep-alive

Content-Type: application/x-www-form-urlencoded; charset=UTF-8

X-Requested-With: XMLHttpRequest

Referer: https://piazza.com/class

Content-Length: 339

Cookie: piazza\_session="DFwuCEFIGvEGwwHLJyuCvHIGTHKECKL.5%25x+x+ux%255M5%22%215%3F5%26x%26%26%7C%22%21r...

Pragma: no-cache

Cache-Control: no-cache

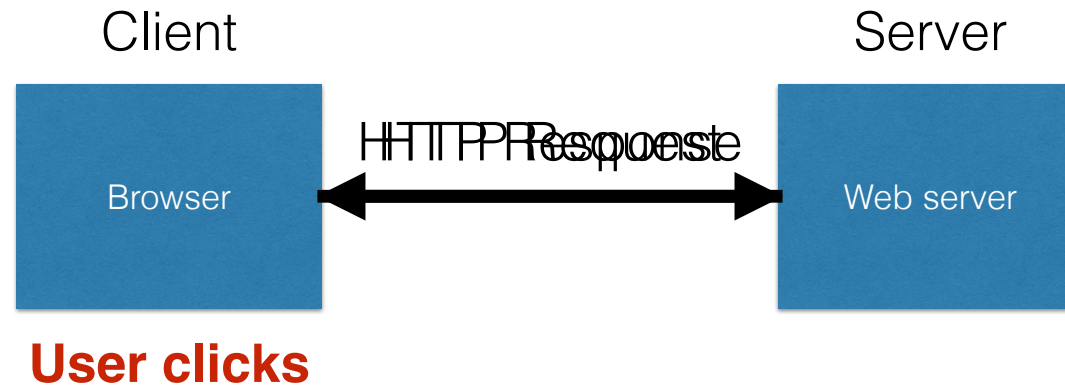
{"method":"content.create","params":{"cid":"hrpng9q2ndos","subject":"<p>Interesting.. perhaps it has to do with a change to the ...

Implicitly includes data  
as a part of the URL

Explicitly includes data as a part of the request's content



# Basic structure of web traffic



- **Responses** contain:
  - **Status** code
  - **Headers** describing what the server provides
  - **Data**
  - **Cookies** (much more on these later)
    - Represent *state* the server would like the browser to store on its behalf

# HTTP responses

**HTTP version**   **Status code**   **Reason phrase**

**Headers**

**Data**

```
HTTP/1.1 200 OK
Date: Tue, 18 Feb 2014 08:20:34 GMT
Server: Apache
Set-Cookie: session-zdnet-production=6bhqca1i0cbciagu11sisac2p3; path=/; domain=zdnet.com
Set-Cookie: zdregion=MTI5LjluMTI5LjE1Mzp1czp1czpjZDJmNWY5YTdkODU1N2Q2YzM5NGU3M2Y1ZTRmN0
Set-Cookie: zdregion=MTI5LjluMTI5LjE1Mzp1czp1czpjZDJmNWY5YTdkODU1N2Q2YzM5NGU3M2Y1ZTRmN0
Set-Cookie: edition=us; expires=Wed, 18-Feb-2015 08:20:34 GMT; path=/; domain=.zdnet.com
Set-Cookie: session-zdnet-production=59ob97fpinqe4bg6lde4dvvq11; path=/; domain=zdnet.com
Set-Cookie: user_agent=desktop
Set-Cookie: zdnet_ad_session=f
Set-Cookie: firstpg=0
Expires: Thu, 19 Nov 1981 08:52:00 GMT
Cache-Control: no-store, no-cache, must-revalidate, post-check=0, pre-check=0
Pragma: no-cache
X-UA-Compatible: IE=edge,chrome=1
Vary: Accept-Encoding
Content-Encoding: gzip
Content-Length: 18922
Keep-Alive: timeout=70, max=146
Connection: Keep-Alive
Content-Type: text/html; charset=UTF-8

<html> ..... </html>
```