Stony Brook University

WSE 380: RESEARCH ROTATION

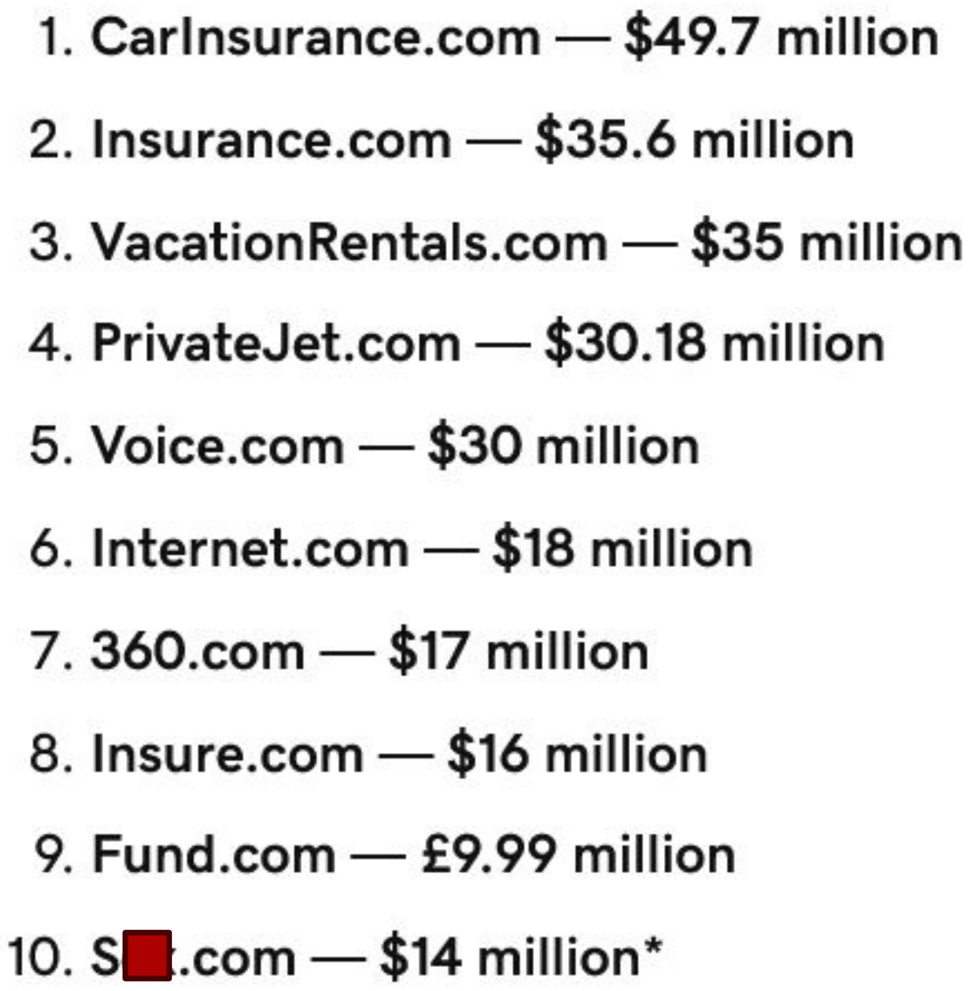
Technical Foundations of Startups

Johnny So
josso@cs.stonybrook.edu

Nick Nikiforakis
nick@cs.stonybrook.edu

What's in a name?

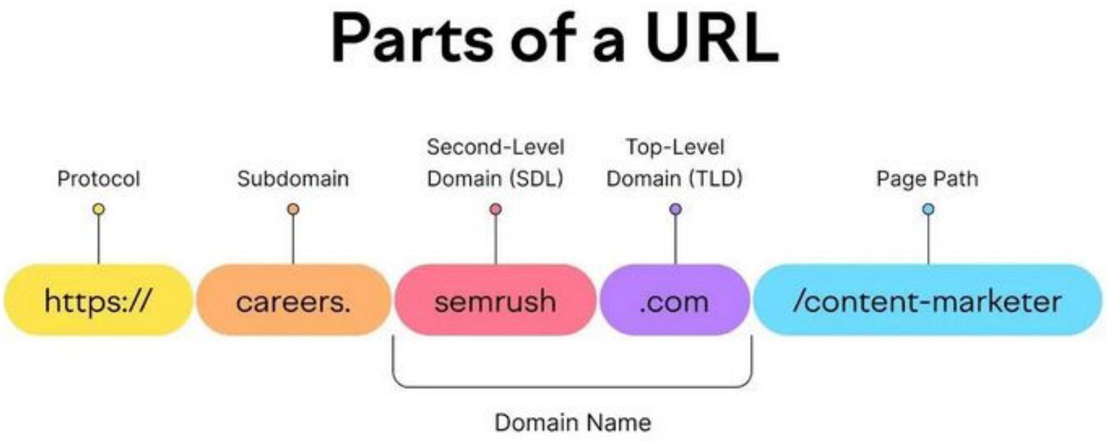
- Names are important
  - And by extension domain names
- Some of the most popular domain names have been sold for millions of dollars
- Domains can either be registered at a registrar (e.g. GoDaddy and Namecheap) or purchased by individuals who already own them

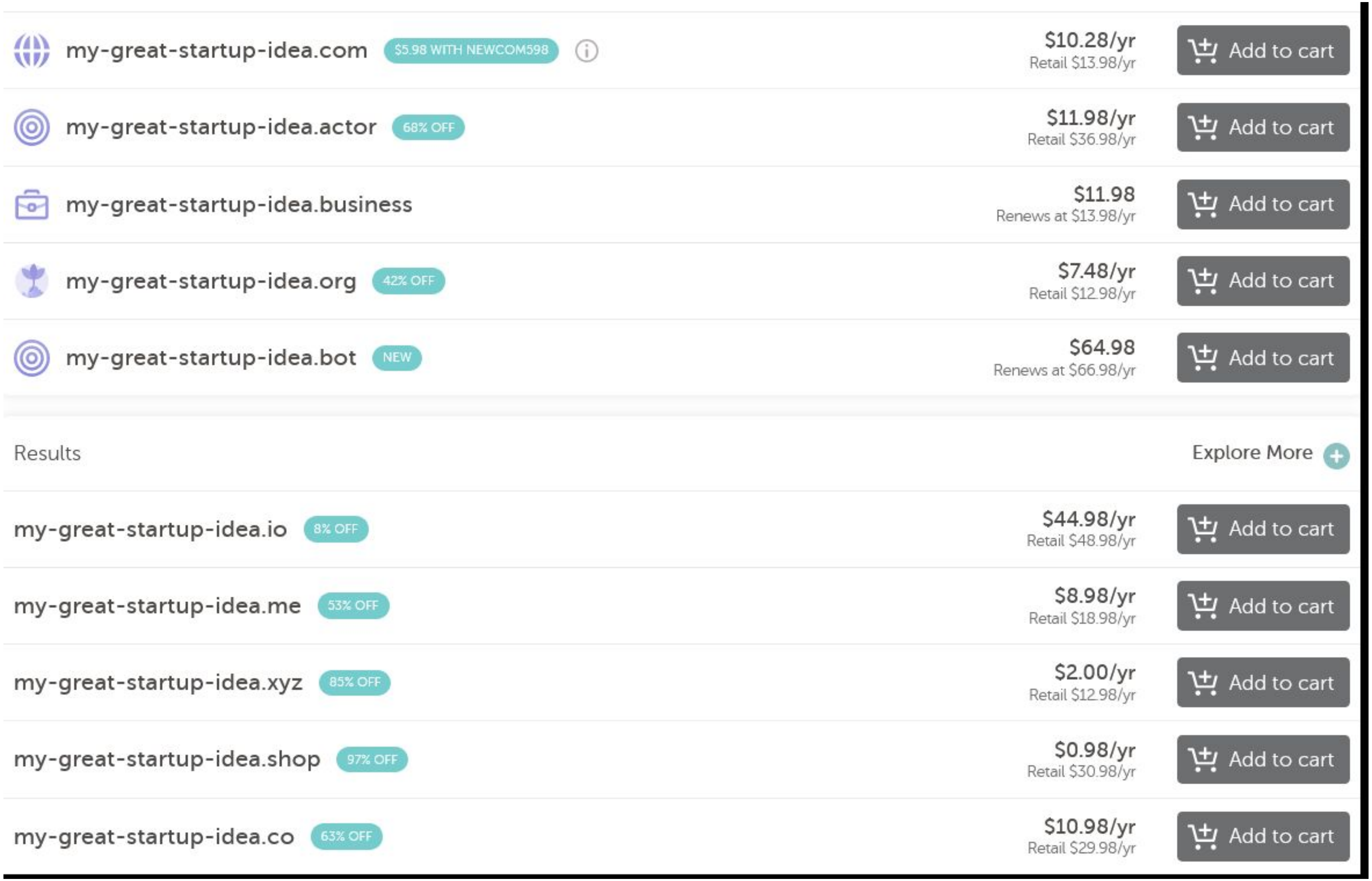


1. CarInsurance.com — $49.7 million
2. Insurance.com — $35.6 million
3. VacationRentals.com — $35 million
4. PrivateJet.com — $30.18 million
5. Voice.com — $30 million
6. Internet.com — $18 million
7. 360.com — $17 million
8. Insure.com — $16 million
9. Fund.com — £9.99 million
10. S[redacted].com — $14 million*

Domain-name parts

- .com, .net, .org are called top-level domains
  - google.com is different than google.net
- Originally meant to indicate what the site was about
  - **Com**mercial
  - **Org**anization
  - **Net**work
- Nowadays there isn't much distinction between TLDs
  - New TLDs are added regularly





my-great-startup-idea.com	$5.98 WITH NEWCOM598	$10.28/yr Retail $13.98/yr	Add to cart
my-great-startup-idea.actor	68% OFF	$11.98/yr Retail $36.98/yr	Add to cart
my-great-startup-idea.business		$11.98 Renews at $13.98/yr	Add to cart
my-great-startup-idea.org	42% OFF	$7.48/yr Retail $12.98/yr	Add to cart
my-great-startup-idea.bot	NEW	$64.98 Renews at $66.98/yr	Add to cart

Results

Explore More

my-great-startup-idea.io	8% OFF	$44.98/yr Retail $48.98/yr	Add to cart
my-great-startup-idea.me	53% OFF	$8.98/yr Retail $18.98/yr	Add to cart
my-great-startup-idea.xyz	85% OFF	$2.00/yr Retail $12.98/yr	Add to cart
my-great-startup-idea.shop	97% OFF	$0.98/yr Retail $30.98/yr	Add to cart
my-great-startup-idea.co	63% OFF	$10.98/yr Retail $29.98/yr	Add to cart

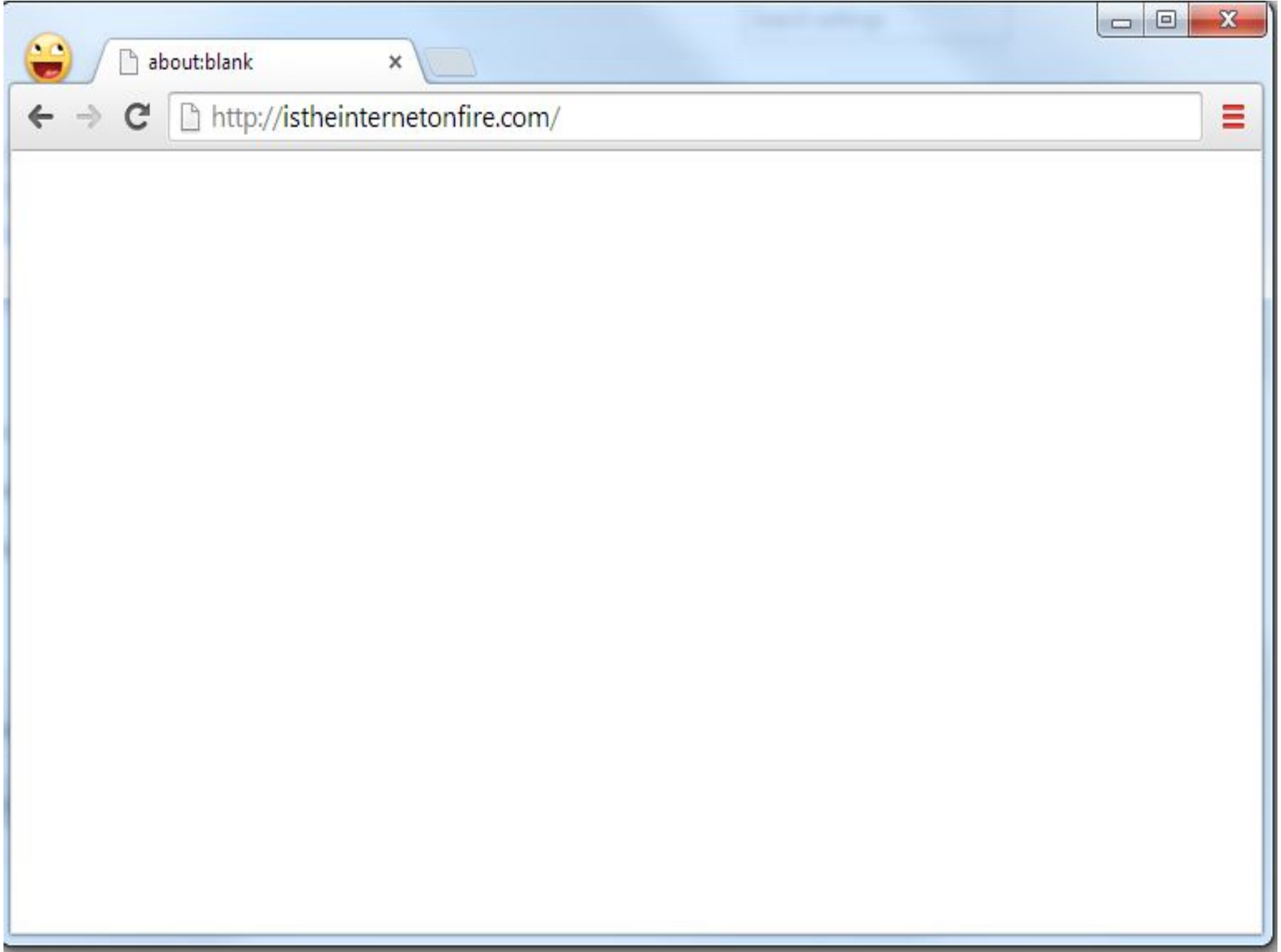
Different TLDs different prices

I am in the "phonebook"

- Purchasing a domain name is the first step but by itself it doesn't do anything
  - Buying a domain gives you the permission to point it to an IP address of your choice
  - Computers and routers don't deal with names, they deal with IP addresses
- The translation process is called name resolution and is done by DNS clients and DNS servers

Person	Phone number
Alice	631-444-2290
Bob	631-123-4567

Domain name	IP address
google.com	142.251.16.139
stonybrook.edu	104.18.6.126
cs.stonybrook.edu	23.185.0.4



http://istheinternetonfire.com/

DNS

- istheinternetonfire.com does not mean anything to a computer
  - So first your browser needs to find the IP address belonging to that domain name

```
nslookup istheinternetonfire.com
Server:         97.107.133.4
Address:         97.107.133.4#53

Non-authoritative answer:
Name:   istheinternetonfire.com
Address: 166.84.7.99
```

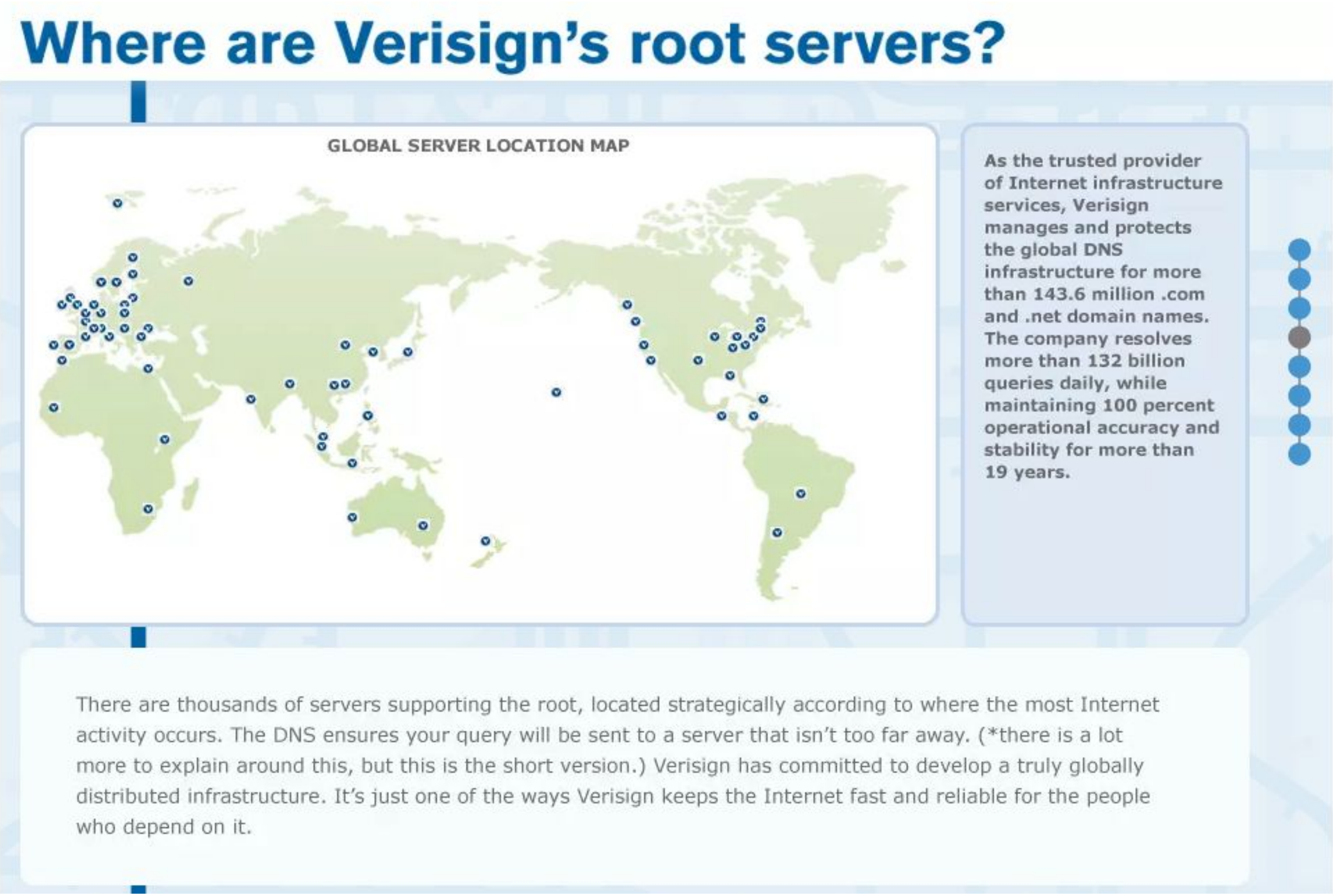
How does that work?

- DNS (Domain Name System) works through distributed hierarchical database of DNS servers

- Your computer has what is called a “stub resolver”.
  - This stub resolver does two things:
    - 1. Ask your recursive resolver (typically provided to you by your ISP) to resolve domains for it
    - 2. Remember (cache) the answer of recent queries

How does that work?

- Given that this is the first time you tried to go to this website, your stub resolver asks your network’s recursive resolver the same question
  - If another user asked that question recently, your recursive resolver (like your stub resolver) remembers the answer and provides it immediately
  - If not then the recursive resolver ask the root servers
    - Root server == “Gate keepers of worldwide DNS”
    - 13 Root servers distributed across the world managed by various entities
      - E.g. Verisign operates 2 out of the 13 servers



Where are Verisign's root servers?

As the trusted provider of Internet infrastructure services, Verisign manages and protects the global DNS infrastructure for more than 143.6 million .com and .net domain names. The company resolves more than 132 billion queries daily, while maintaining 100 percent operational accuracy and stability for more than 19 years.

There are thousands of servers supporting the root, located strategically according to where the most Internet activity occurs. The DNS ensures your query will be sent to a server that isn't too far away. (*there is a lot more to explain around this, but this is the short version.) Verisign has committed to develop a truly globally distributed infrastructure. It's just one of the ways Verisign keeps the Internet fast and reliable for the people who depend on it.

Note: 2 root servers DOES NOT mean two physical machines

Root servers

- The only thing that root servers know, is where the TLD name servers are
  - Servers for .com, .net, .org, etc.
- When your ISP’s recursive resolver asks a root server for the address of istheinternetonfire.com the answer is:
  - I don’t know, but here is a list of .com nameservers that will probably know

TLD Nameserver

- Q: Hey .com Nameserver, what is the IP address of istheinternetonfire.com ?

- A: I don’t know, but go ask the nameservers that are responsible for resolving it, a.dns.gandi.net, b.dns.gandi.net, c.dns.gandi.net
  - Notice that the NS server is located on the .net TLD
  - To save us the trip up to the root and down the .net server, the .com nameserver provides the IP address of the nameserver in its response
    - This is possible because .com and .net are both operated by Verisign

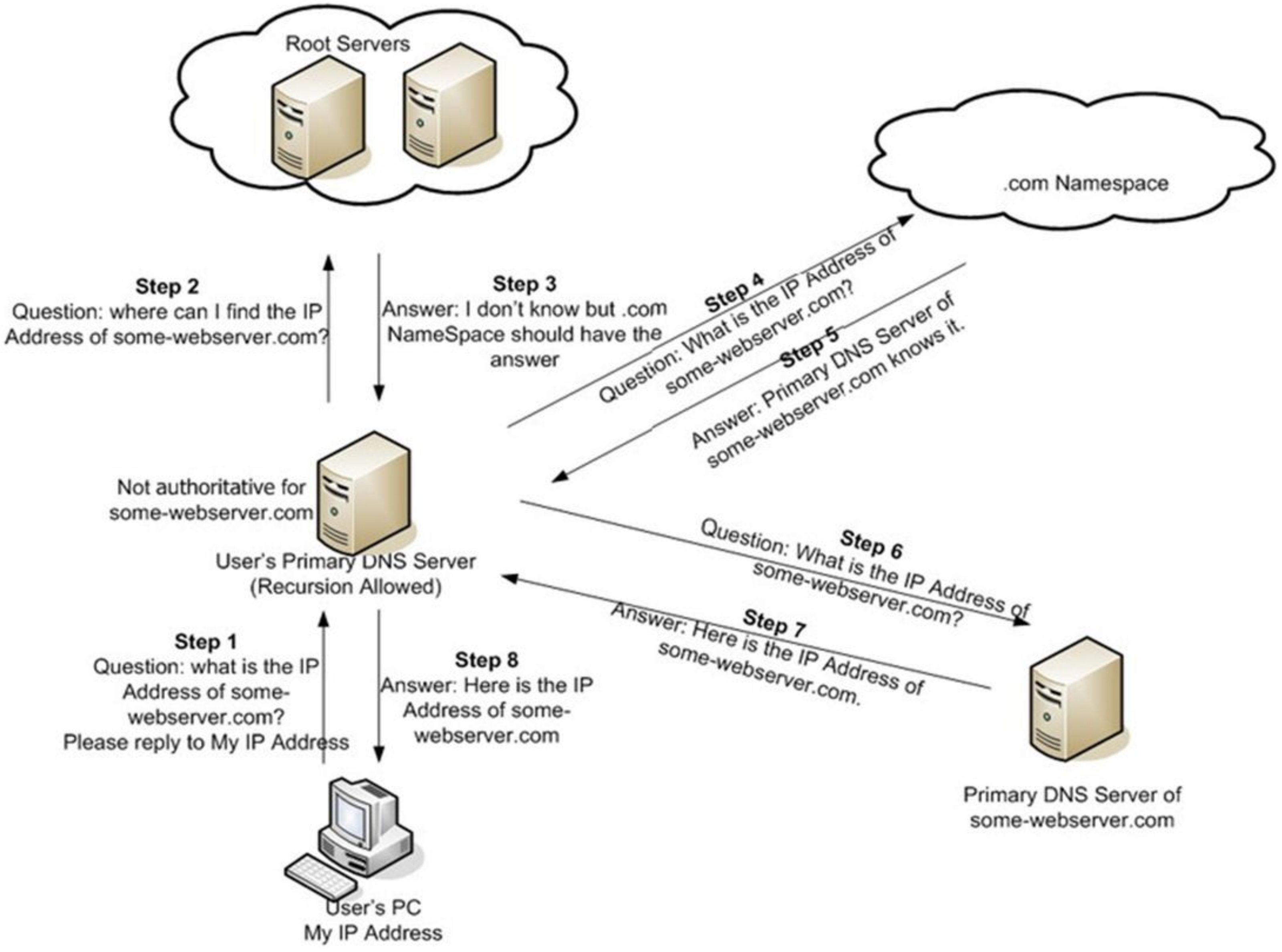
Authoritative Nameserver

- Q: Hey b.dns.gandi.net what is the IP address of istheinternetonfire.com ?

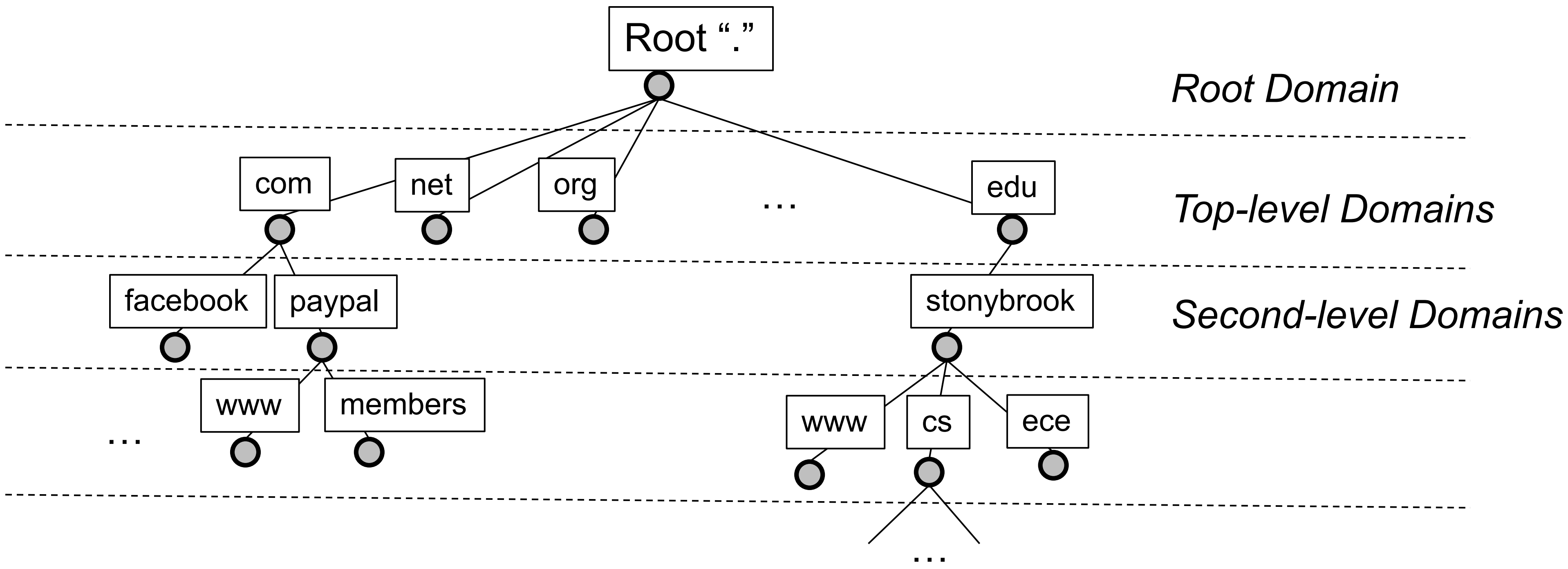
- A:The IP address of istheinternetonfire.com is 166.84.7.99

Now the recursive resolver caches the result and returns the address to your stub resolver running in your operating system

Visually


Root Servers
.com Namespace
Step 2
Question: where can I find the IP Address of some-webserver.com?
Step 3
Answer: I don't know but .com NameSpace should have the answer
Step 4
Question: What is the IP Address of some-webserver.com?
Step 5
Answer: Primary DNS Server of some-webserver.com knows it.
Not authoritative for some-webserver.com
User's Primary DNS Server (Recursion Allowed)
Step 6
Question: What is the IP Address of some-webserver.com?
Step 7
Answer: Here is the IP Address of some-webserver.com.
Step 1
Question: what is the IP Address of some-webserver.com? Please reply to My IP Address
Step 8
Answer: Here is the IP Address of some-webserver.com
Primary DNS Server of some-webserver.com
User's PC
My IP Address

DNS Hierarchy (it’s a tree!)



Where do we get IP addresses?

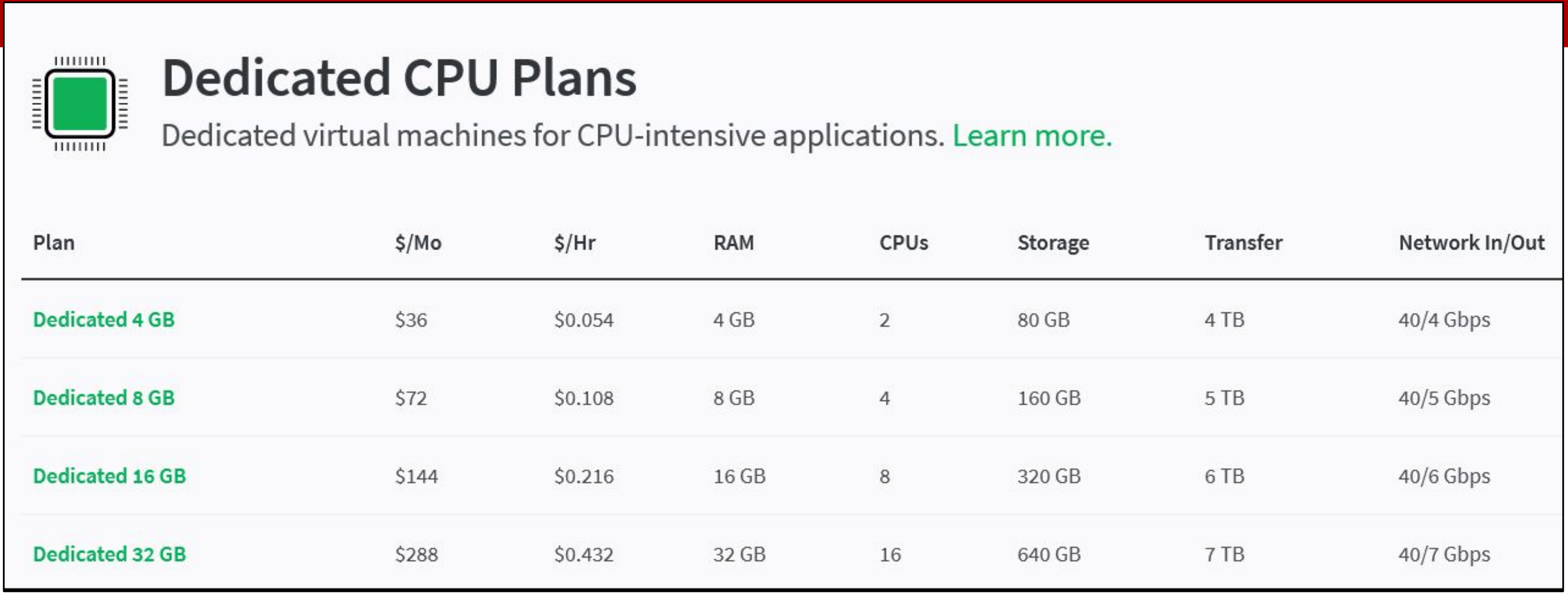
- We need a hosting company that will rent us a server with a public IP address
  - These are called "Hosting providers"
- Servers are just general-purpose computers
  - Typically, with better hardware and uptime than your laptop
- Many available options at different pricing tiers
  - From $5/month to thousands of dollars per month (possibly more)
  - The pricing has to do with:
    - How powerful the server is
    - How much traffic can we send to it and from it

linode

amazon web services

Google Cloud

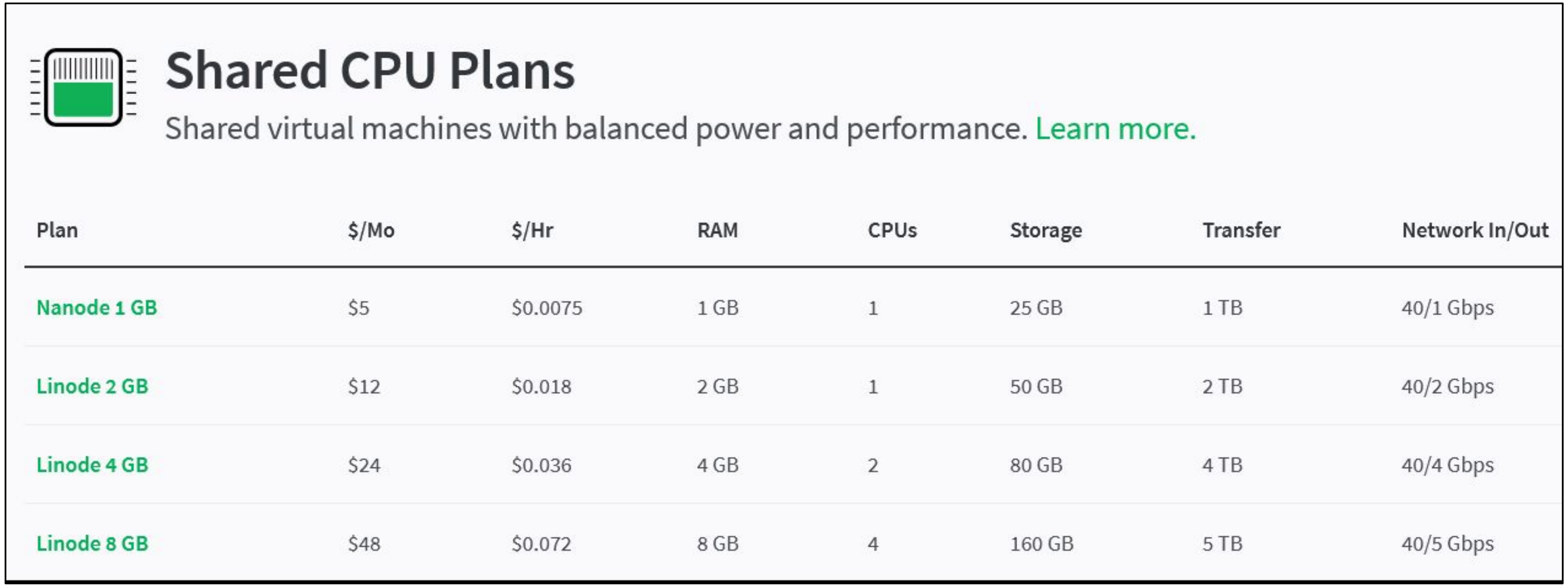
HOSTINGER



Dedicated CPU Plans

Dedicated virtual machines for CPU-intensive applications. Learn more.

Plan	$/Mo	$/Hr	RAM	CPUs	Storage	Transfer	Network In/Out
Dedicated 4 GB	$36	$0.054	4 GB	2	80 GB	4 TB	40/4 Gbps
Dedicated 8 GB	$72	$0.108	8 GB	4	160 GB	5 TB	40/5 Gbps
Dedicated 16 GB	$144	$0.216	16 GB	8	320 GB	6 TB	40/6 Gbps
Dedicated 32 GB	$288	$0.432	32 GB	16	640 GB	7 TB	40/7 Gbps



Shared CPU Plans

Shared virtual machines with balanced power and performance. Learn more.

Plan	$/Mo	$/Hr	RAM	CPUs	Storage	Transfer	Network In/Out
Nanode 1 GB	$5	$0.0075	1 GB	1	25 GB	1 TB	40/1 Gbps
Linode 2 GB	$12	$0.018	2 GB	1	50 GB	2 TB	40/2 Gbps
Linode 4 GB	$24	$0.036	4 GB	2	80 GB	4 TB	40/4 Gbps
Linode 8 GB	$48	$0.072	8 GB	4	160 GB	5 TB	40/5 Gbps

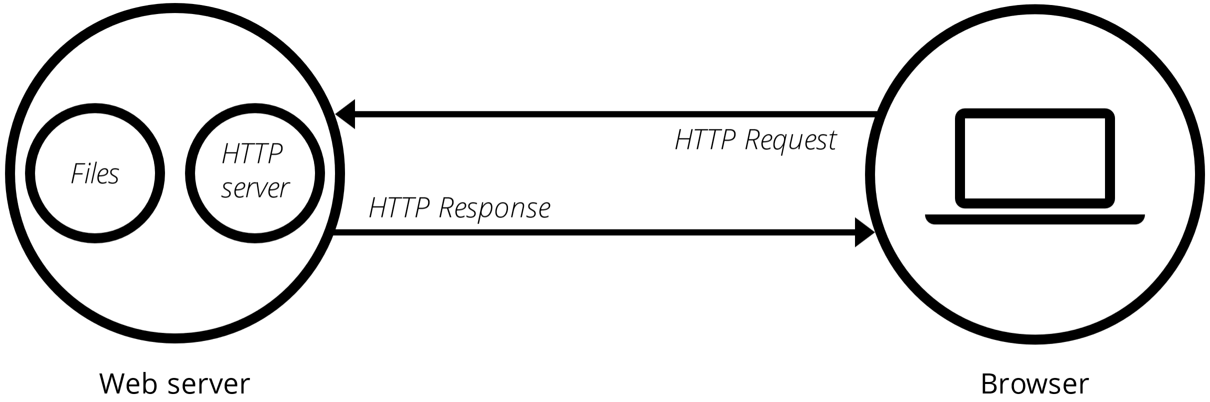
Pricing data from linode.com in 2024

So far so good…

- We now own a cool domain name, and we can point it to an IP address of a server we obtained from a hosting provider
  - When someone types our domain name in their browser, that browser will eventually connect to our server and ask for our website
  - Two ingredients missing
    - Who (what software) will answer the user's browser?
    - What will it send back?

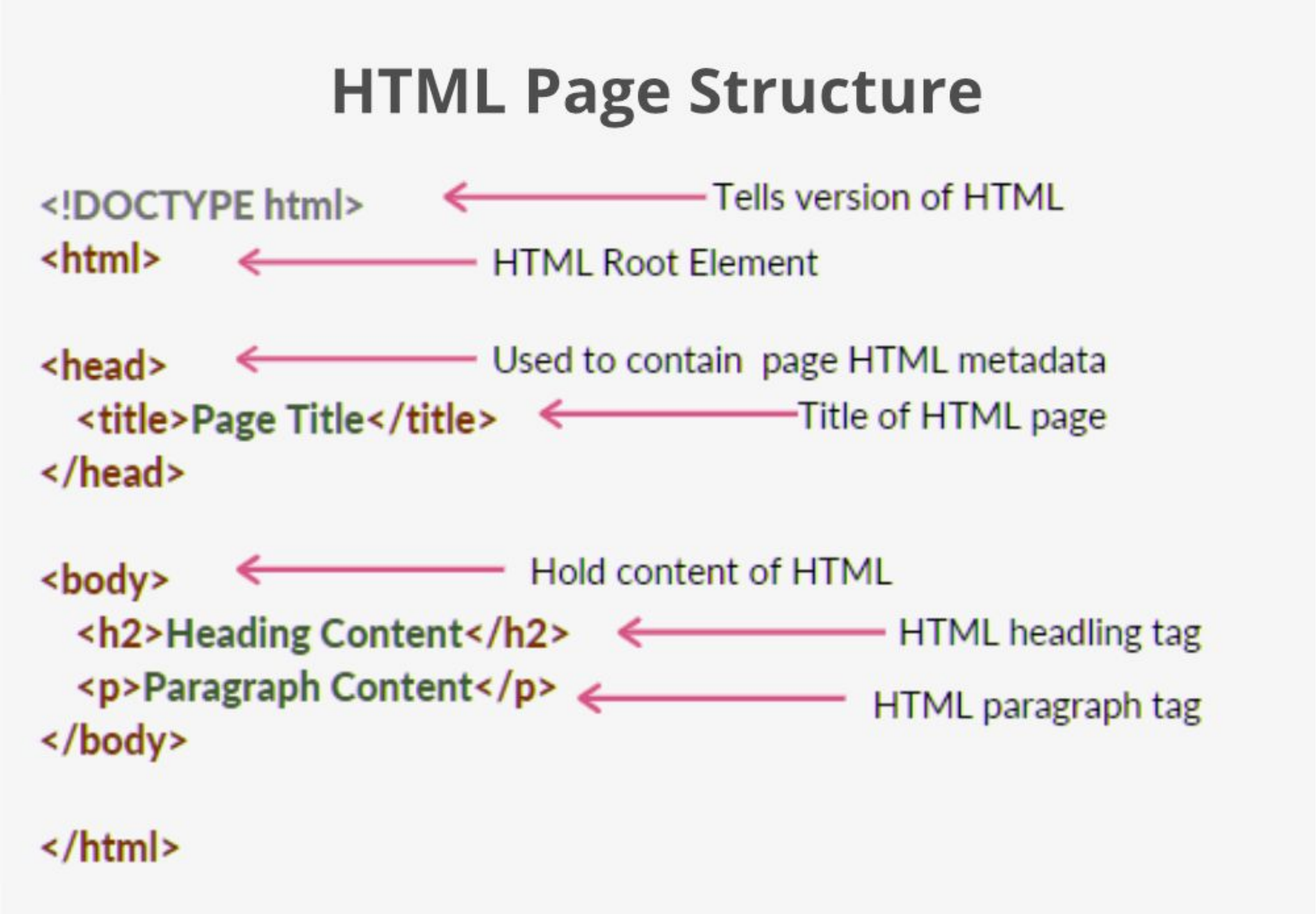
Web server

- A web server (or HTTP server) is a piece of software that listens to traffic and can speak the HyperText Transport Protocol (HTTP)
  - It receives HTTP requests from clients and sends responses
  - Requests asks for specific resources (pages, images, JavaScript code)
  - Responses include the requested resources along with metadata



HTML

- In its simplest form, a website is just a single static page
  - At the server side, this is stored as a single HTML file
  - HTML: Hyper Text Markup Language
- This page can be self-contained or reference external resources
  - Images, other pages, Cascading Style Sheets, etc.



HTML Page Structure

```
<!DOCTYPE html>              ← Tells version of HTML
<html>                       ← HTML Root Element

<head>                       ← Used to contain  page HTML metadata
  <title>Page Title</title>  ← Title of HTML page
</head>

<body>                       ← Hold content of HTML
  <h2>Heading Content</h2>   ← HTML headling tag
  <p>Paragraph Content</p>   ← HTML paragraph tag
</body>

</html>
```

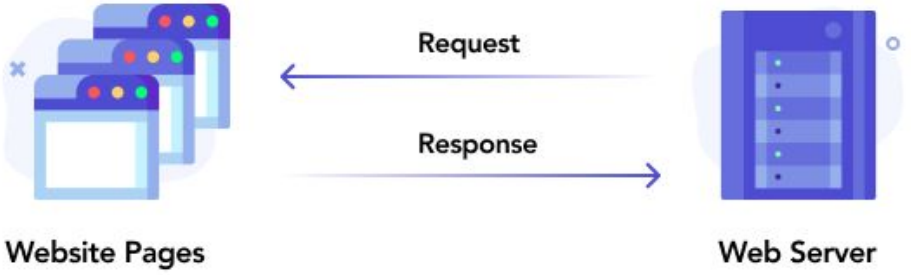
Static content

- When a web server receives an HTTP request, it attempts to locate the file that that request corresponds to
  - When asking for the main page of example.com, it will find the HTML file corresponding to that and send it back to the browser
  - The browser will "draw" the page according to the HTML instructions
    - Images, sections, text, bullet points, etc.

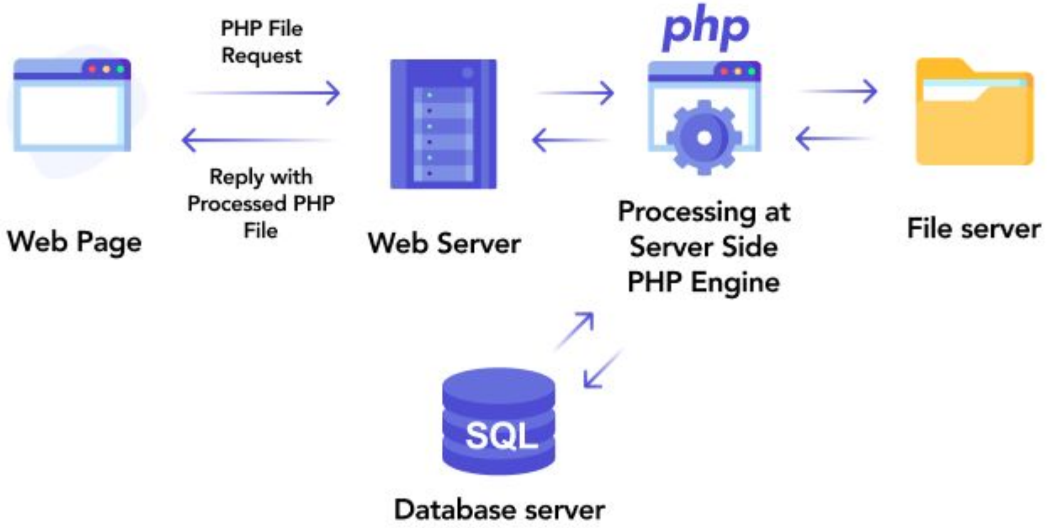
- This would classify as a "static" website
  - The web server just finds files on the server and sends them back to the user
  - 100 users asking for the main page of example.com will receive 100 identical responses

Static vs. Dynamic Content

- Static content is great for simple websites
  - Wikipedia
  - "Calling card" websites of brick-and-mortar businesses
  - Personal hobby sites

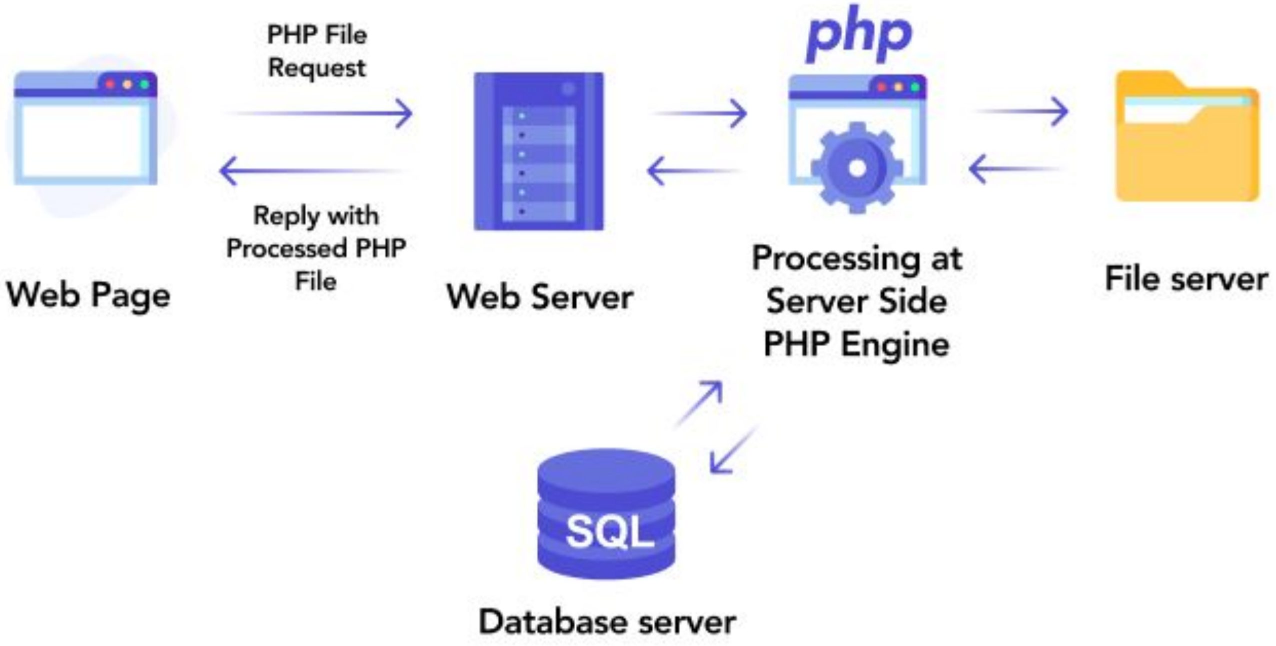


- Anything complicated requires dynamic content
  - Anything that requires handling user input requires a dynamic website
  - All your favorite web applications
    - Social media, YouTube



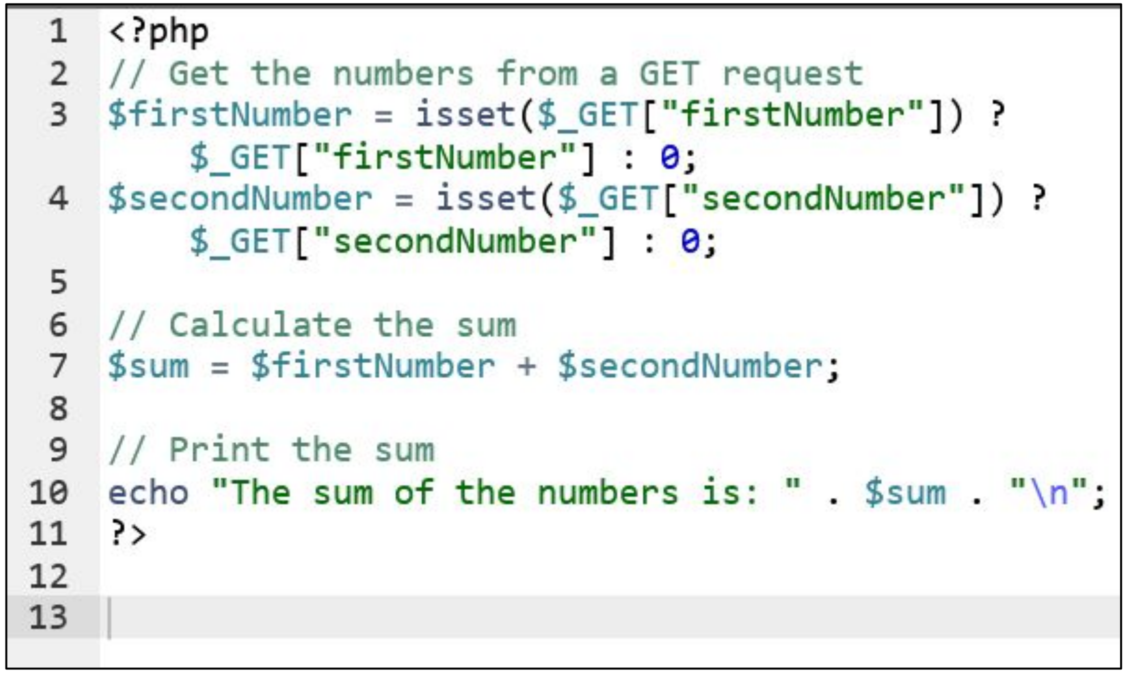
Dynamic websites

- Dynamic websites run full programs at the server, handling user input and generating the response sent back to users
  - Databases are commonly involved in these websites
- Common server-side programming languages
  - PHP
  - Python
  - NodeJS



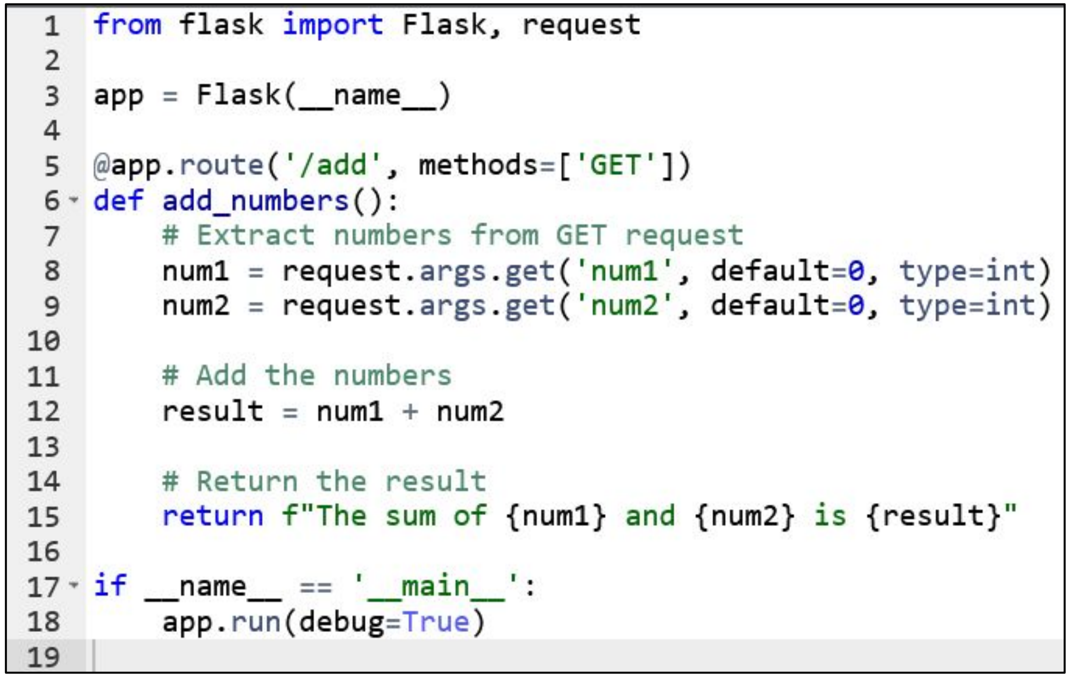
Adding two numbers on a dynamic website

PHP



```php
<?php
// Get the numbers from a GET request
$firstNumber = isset($_GET["firstNumber"]) ?
    $_GET["firstNumber"] : 0;
$secondNumber = isset($_GET["secondNumber"]) ?
    $_GET["secondNumber"] : 0;

// Calculate the sum
$sum = $firstNumber + $secondNumber;

// Print the sum
echo "The sum of the numbers is: " . $sum . "\n";
?>

```

Python



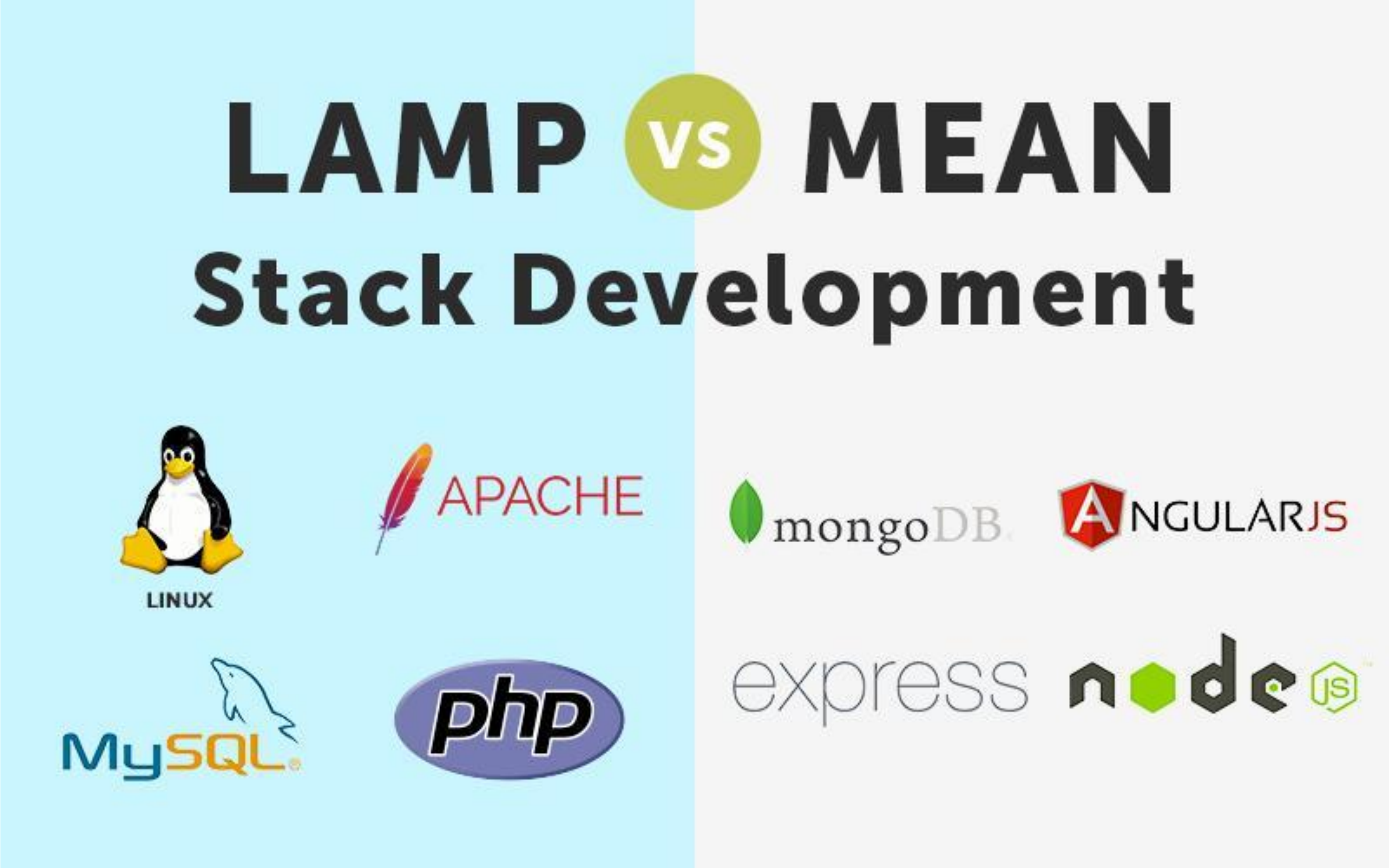
```python
from flask import Flask, request

app = Flask(__name__)

@app.route('/add', methods=['GET'])
def add_numbers():
    # Extract numbers from GET request
    num1 = request.args.get('num1', default=0, type=int)
    num2 = request.args.get('num2', default=0, type=int)

    # Add the numbers
    result = num1 + num2

    # Return the result
    return f"The sum of {num1} and {num2} is {result}"

if __name__ == '__main__':
    app.run(debug=True)

```

Database servers

- Database servers are used to keep data in structured ways
  - Authentication data
  - Order data
  - Item inventory and pricing

- Structured vs. unstructured databases
  - **Structured:** Schemas have to be defined ahead of time
    - E.g. MySQL and PostgreSQL
  - **Unstructured**: Loose collection of properties
    - E.g. MongoDB and Cassandra

Combining different technologies together



Securing your web application

- Now you have a dynamic web application that can power your next startup
  - How do we secure it against attacks?

- Common attacks
  - **Credential stuffing attacks on login forms**
  - Vulnerabilities in your written code
    - Or in the ready-made code you have deployed
  - Vulnerabilities in the underlying operating system

How do attackers use passwords?

- Once a database of credentials is leaked, attackers can use them in multiple ways
  - Extract emails and usernames
    - Chances are that users are reusing the same username/email address in other unrelated services
  - Learn what are the most common passwords that most users use
  - Learn what are the passwords that specific users use

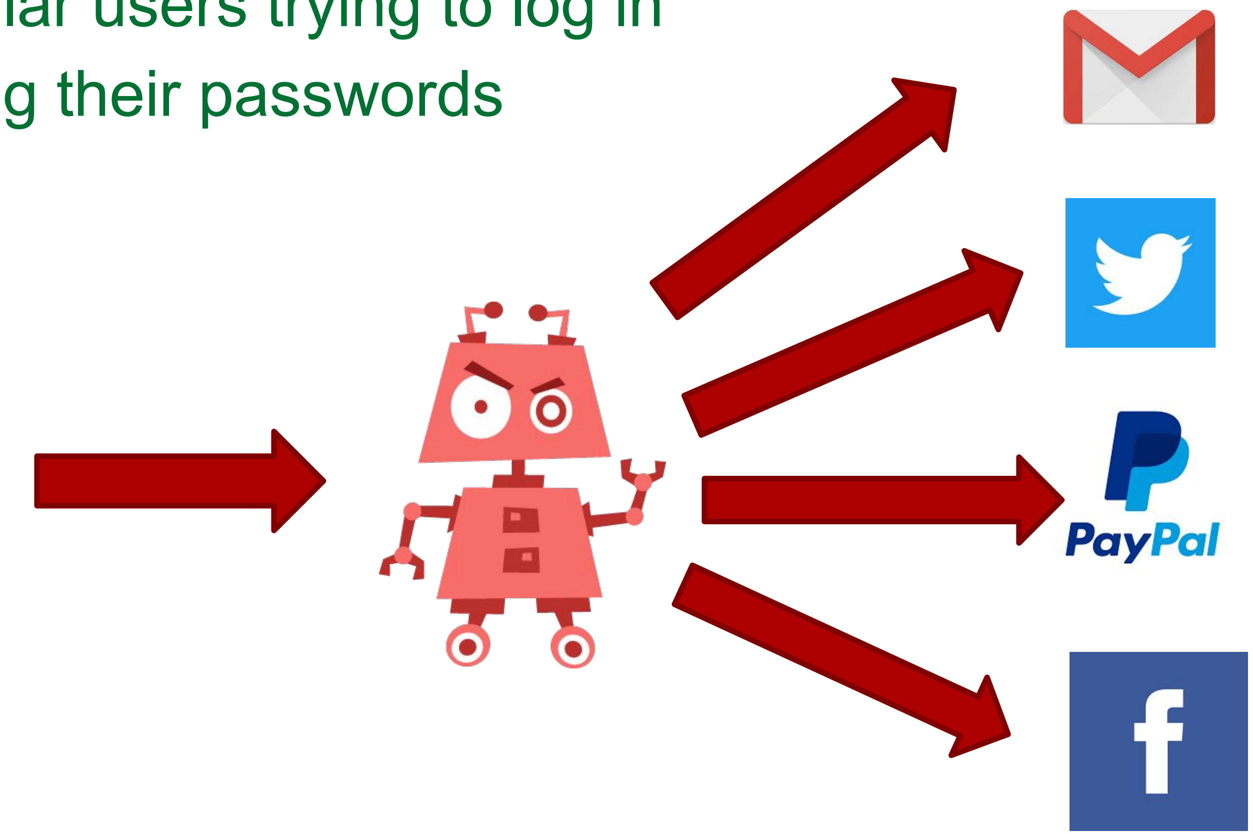
Username	Password
alice@gmail.com	ilovedogs
bob@yahoo.com	Password!
eve@outlook.com	1q2w3e4r
john@stonybrook.edu	g@rfield1

Credential stuffing

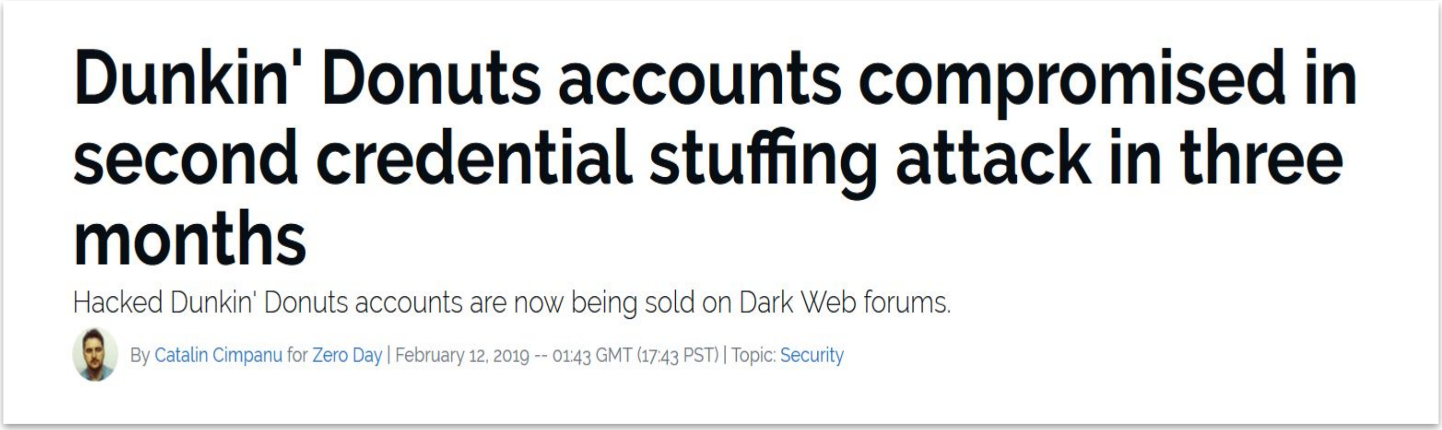
- Attackers build programs that try these credentials against other services
  - These programs act like regular users trying to log in
  - Attackers bet on users reusing their passwords

Username	Password
alice@gmail.com	ilovedogs
bob@yahoo.com	Password!
eve@outlook.com	1q2w3e4r
john@stonybrook.edu	g@rfield1

supercutecats.com



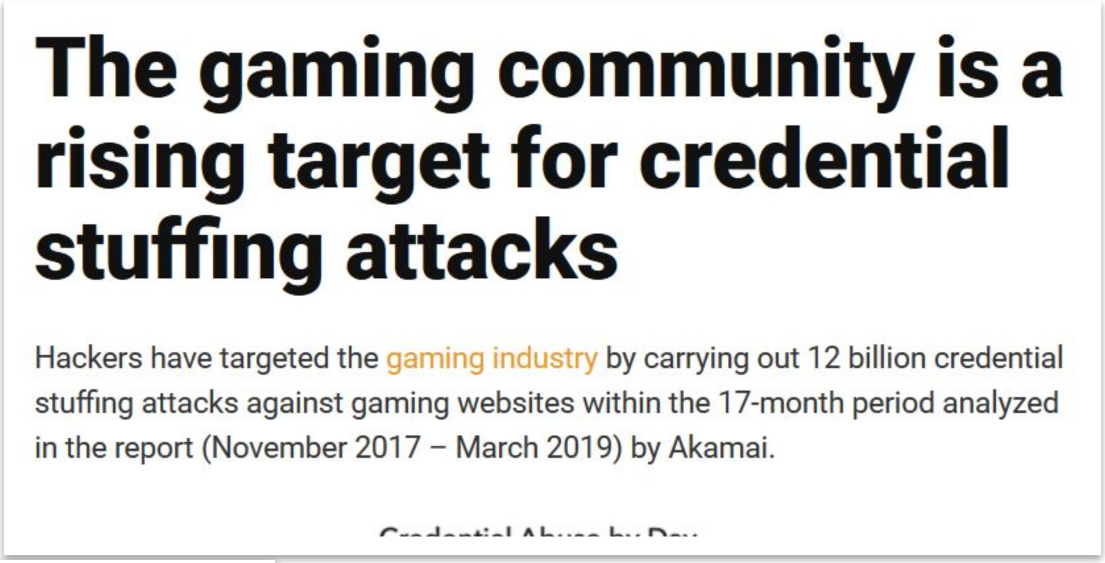
Credential stuffing is a real and growing problem



Dunkin' Donuts accounts compromised in second credential stuffing attack in three months

Hacked Dunkin' Donuts accounts are now being sold on Dark Web forums.

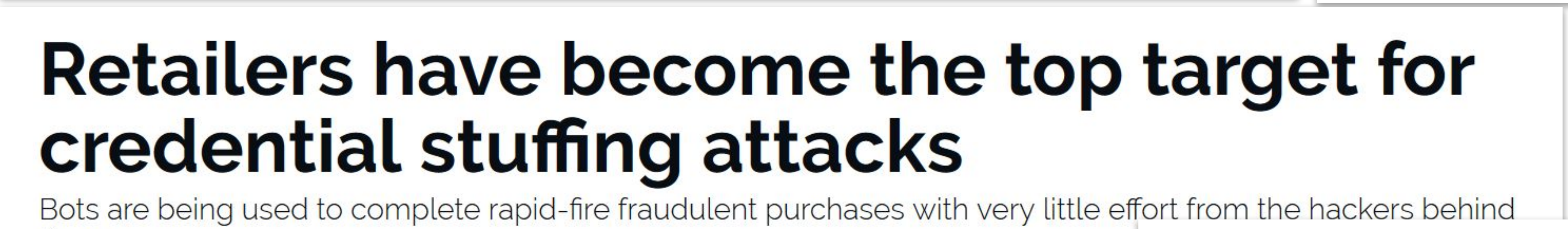
By Catalin Cimpanu for Zero Day | February 12, 2019 -- 01:43 GMT (17:43 PST) | Topic: Security



The gaming community is a rising target for credential stuffing attacks

Hackers have targeted the gaming industry by carrying out 12 billion credential stuffing attacks against gaming websites within the 17-month period analyzed in the report (November 2017 – March 2019) by Akamai.

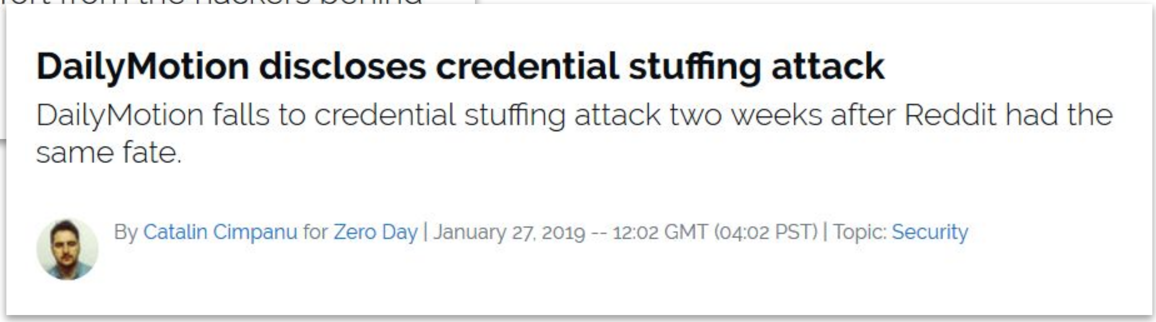
Credential Abuse by Day



Retailers have become the top target for credential stuffing attacks

Bots are being used to complete rapid-fire fraudulent purchases with very little effort from the hackers behind them.

By Charlie Osborne for Zero Day | February 27, 2019 -- 11:00 GMT (03:00 PST) | Topic: Security



DailyMotion discloses credential stuffing attack

DailyMotion falls to credential stuffing attack two weeks after Reddit had the same fate.

By Catalin Cimpanu for Zero Day | January 27, 2019 -- 12:02 GMT (04:02 PST) | Topic: Security

Vision for this rotation

- **High-level**: Help you deploy a real web application on a custom domain of your choosing, while learning multiple underlying technologies (Linux, DNS, web servers, Docker, etc.)
- Details:
  - Understand the basics of setting up a fully featured web application
  - Learn basic Linux usage so that you can deploy your website on a real server
  - Deploy a real website and develop it according to your interests
  - Progressively secure the website against common attacks by deploying freely-available security tools