

CSE 686 Internet Programming

Week 1: How the Web Works

**Edmund Yu, PhD
Associate Teaching Professor
esyu@syr.edu**

January 17 & 22, 2018

The Internet

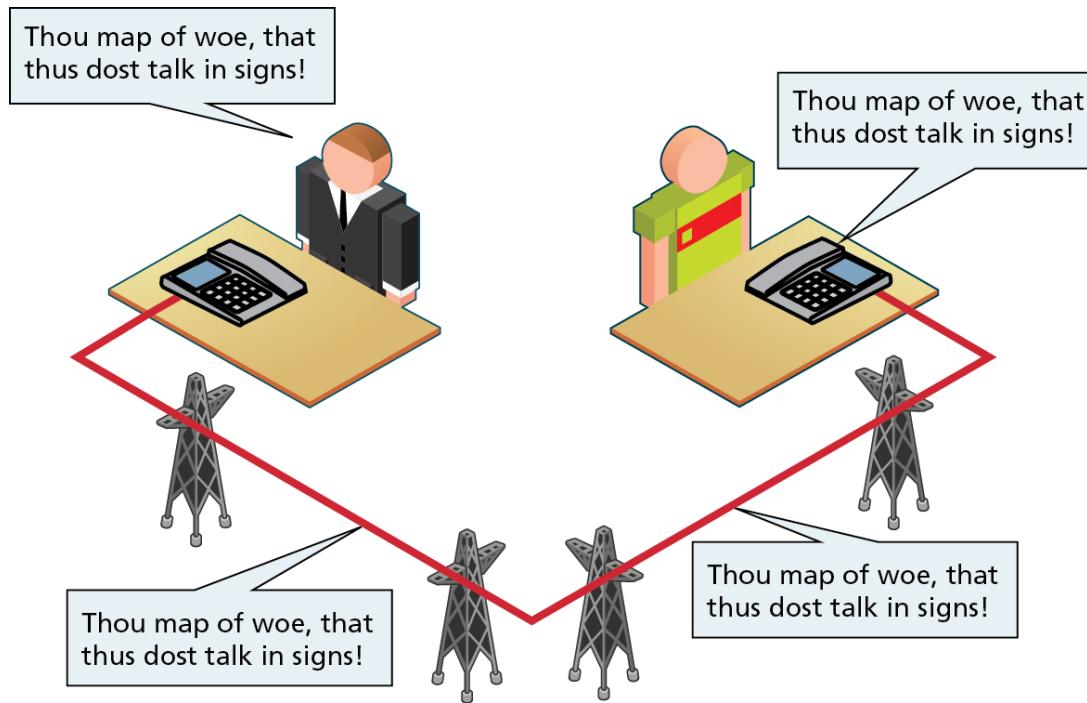
- ❖ Who invented the Internet?

ARPANET: The Beginning of the Internet

- ❖ The research network ARPANET was created in the 1960s
- ❖ ARPANET did not use **circuit switching**, it used **packet switching**
 - ❖ A packet-switched network does not require a continuous connection. Instead it splits the messages into smaller chunks called packets and routes them to the appropriate place based on the destination address.
 - ❖ The packets can take different routes to the destination.

Circuit Switching

- ❖ A **circuit switching** establishes an actual physical connection between two people through a series of physical switches.

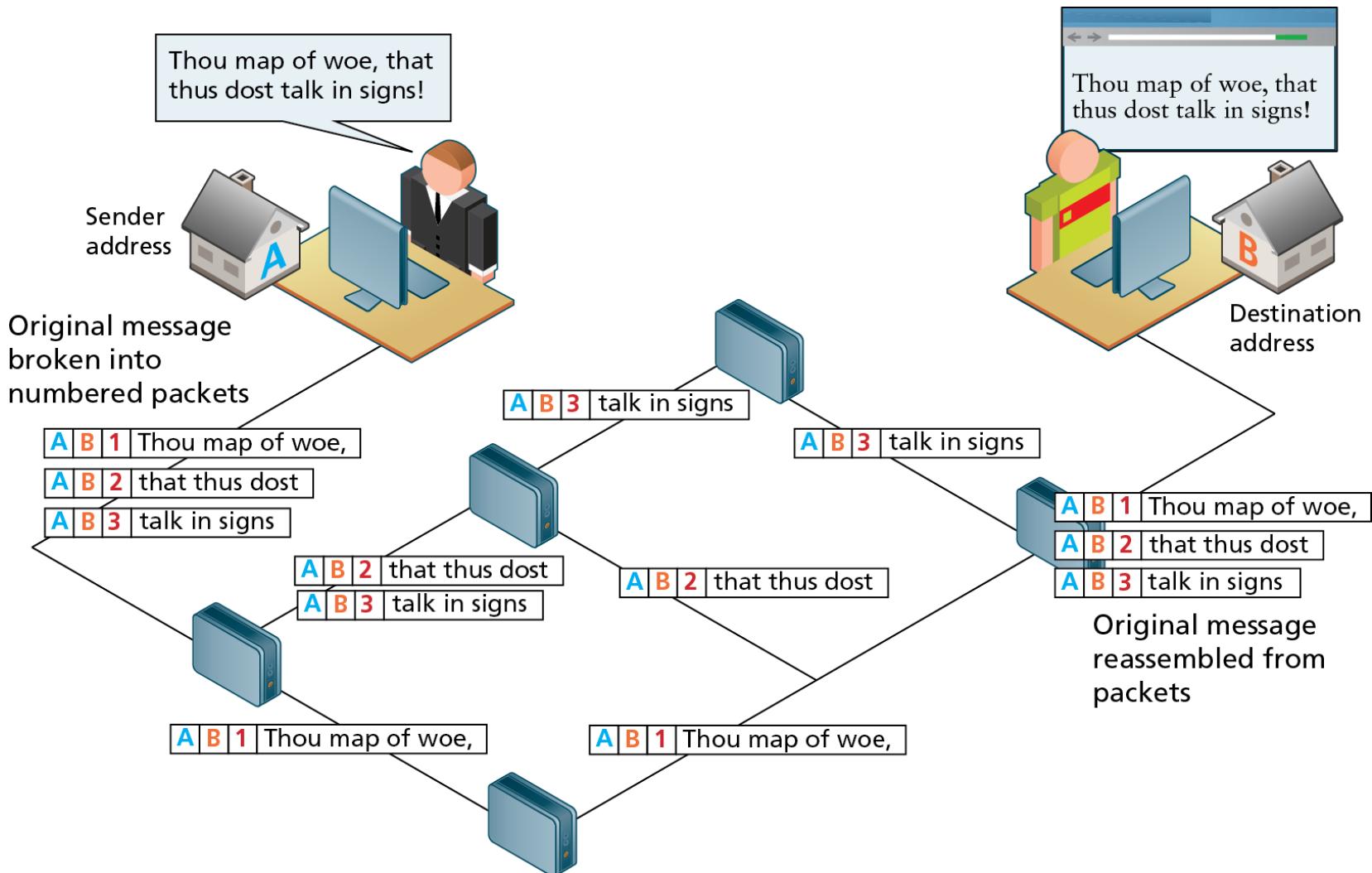


Circuit Switching: Its Limitations

❖ Circuit Switching Weaknesses

1. You must establish a link and maintain a dedicated circuit for the duration of the call
2. Difficult to have multiple conversations simultaneously
3. Wastes bandwidth since even the silences are transmitted

Packet Switching



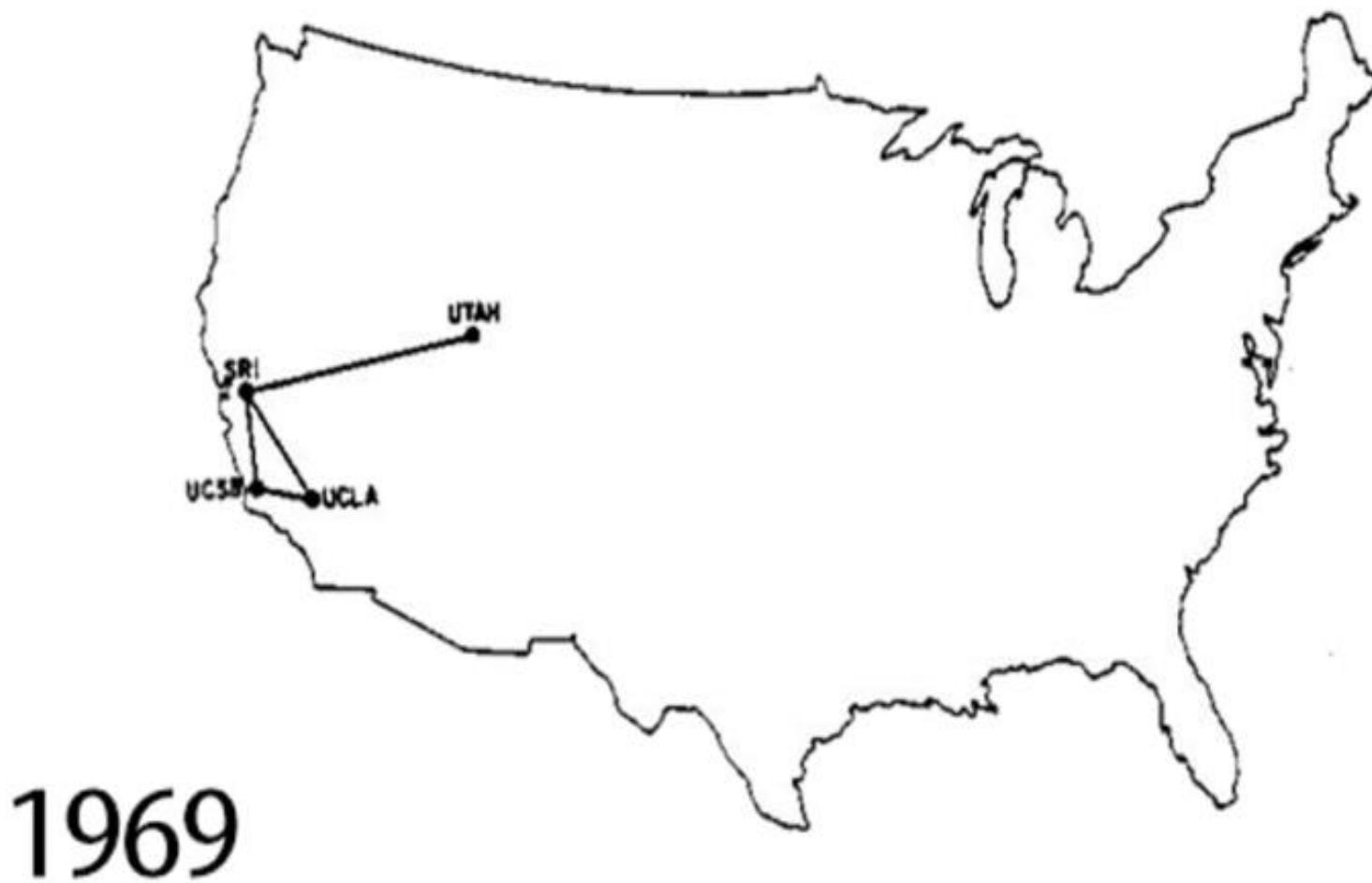
Packet Switching

- ❖ While **packet switching** may seem a more complicated and inefficient approach than **circuit switching**, it is:
 - ❖ more robust (it is not reliant on a single pathway that may fail)
 - ❖ a more efficient use of network resources (since a circuit can communicate multiple connections).

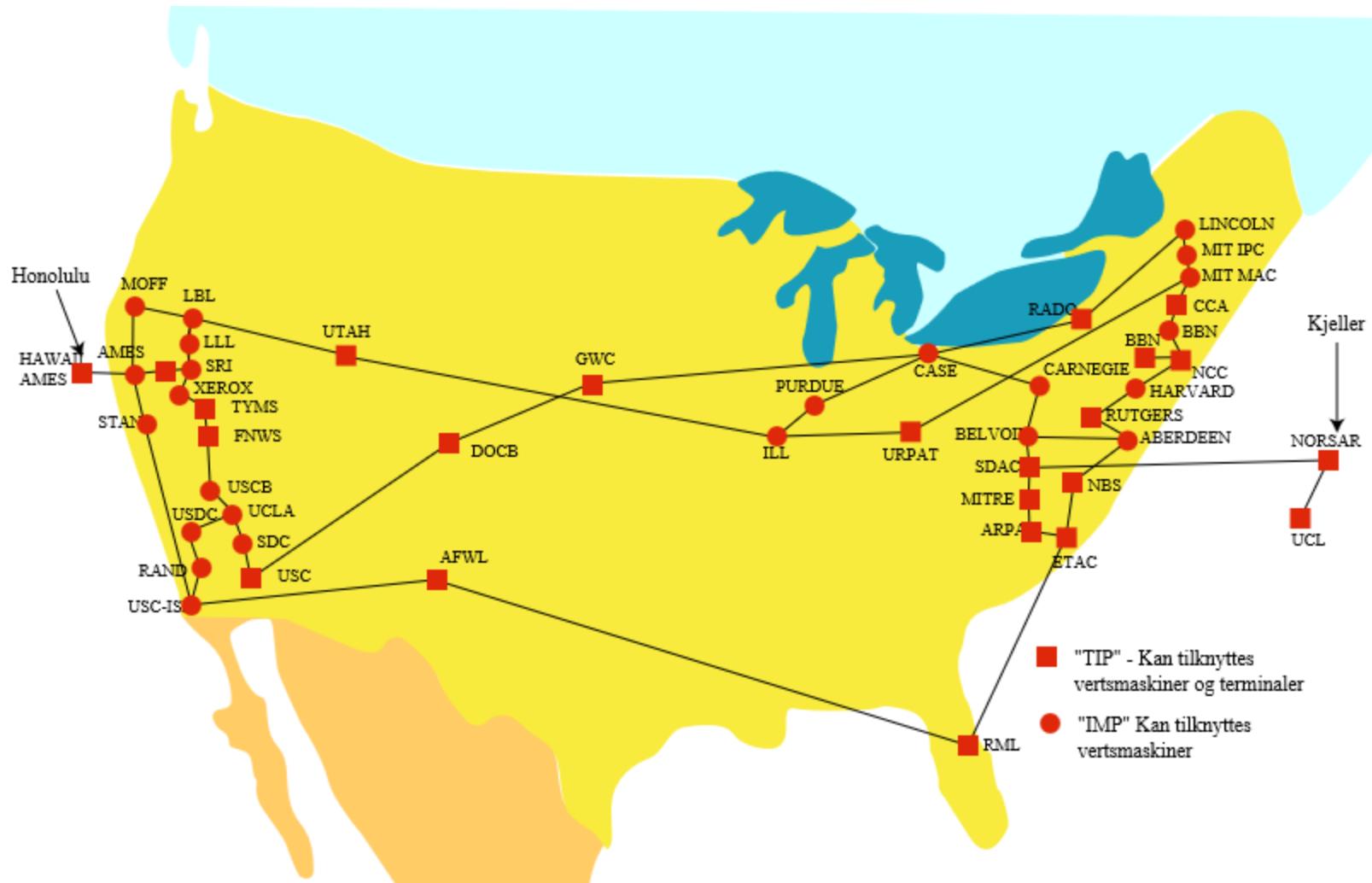
Short History of the Internet

- ❖ The early ARPANET network was funded and controlled by the United States government, and was used exclusively for academic and scientific purposes.
- ❖ The early network started small with just a handful of connected campuses in 1969 and grew to a few hundred by the early 1980s. (See next slide)

ARPANET 1974



ARPANET 1974



Short History of the Internet

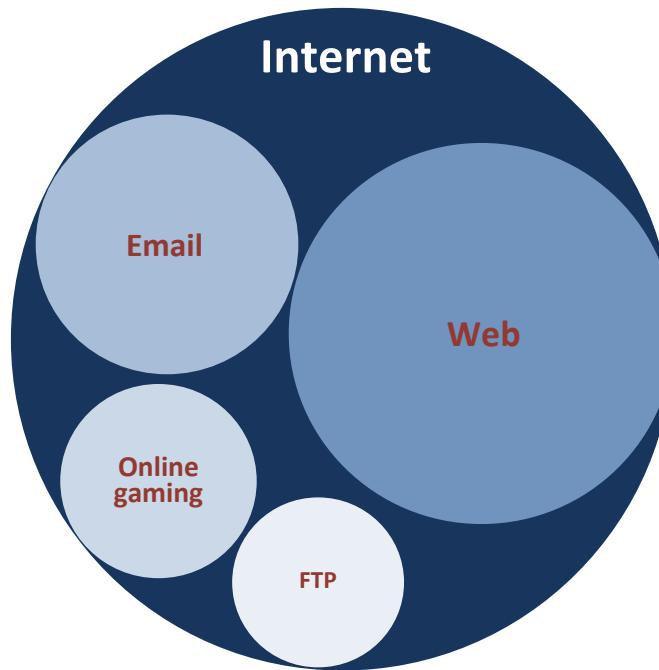
- ❖ At the same time, alternative networks were created like **X.25** in 1974, which allowed (and encouraged) business use.
- ❖ **USENET**, built in 1979, had fewer restrictions still, and as a result grew quickly to 550 hosts by 1981.
- ❖ Although there was growth in these various networks, the inability for them to communicate with each other was a real limitation.
- ❖ To promote the growth and unification of the disparate networks, a suite of communication **protocols** was invented to unify the networks.

Short History of the Internet

- ❖ By **1981** protocols for the Internet were published and ready for use.
 - ❖ New networks built in the United States began to adopt the **TCP/IP** (Transmission Control Protocol/Internet Protocol), while older networks were transitioned over to it.
 - ❖ Any organization, private or public, could potentially connect to this new network so long as they adopted the TCP/IP protocol.
- ❖ On January 1, **1983**, TCP/IP was adopted across all of ARPANET, marking the end of the research network that spawned the Internet.
- ❖ Over the next two decades, TCP/IP networking was adopted across the globe.

Web = Internet?

- ❖ The World-Wide Web (WWW or simply the Web) is certainly what most people think of when they see the word “internet.”
- ❖ But the WWW is only a subset of the Internet.



The Web

- ❖ Who invented the Web?

Web Server

- ❖ In 1989, Sir **Tim Berners-Lee** proposed a new project to his employer **CERN** (European Organization for Nuclear Research) in Geneva that runs the particle accelerators for the physicists studying sub-atomic structure.
(published in 1990)



- ❖ To ease the exchange of information between scientists by using a hypertext system
- ❖ As a result of this project, in 1990, Berners-Lee wrote two programs:
 - ❖ A browser called **WorldWideWeb**, and
 - ❖ The world's first **Web server**, which ran on NeXTSTEP

Hypertext Systems

- ❖ Hypertext systems use **tags** (hypertext links) to link to related documents:

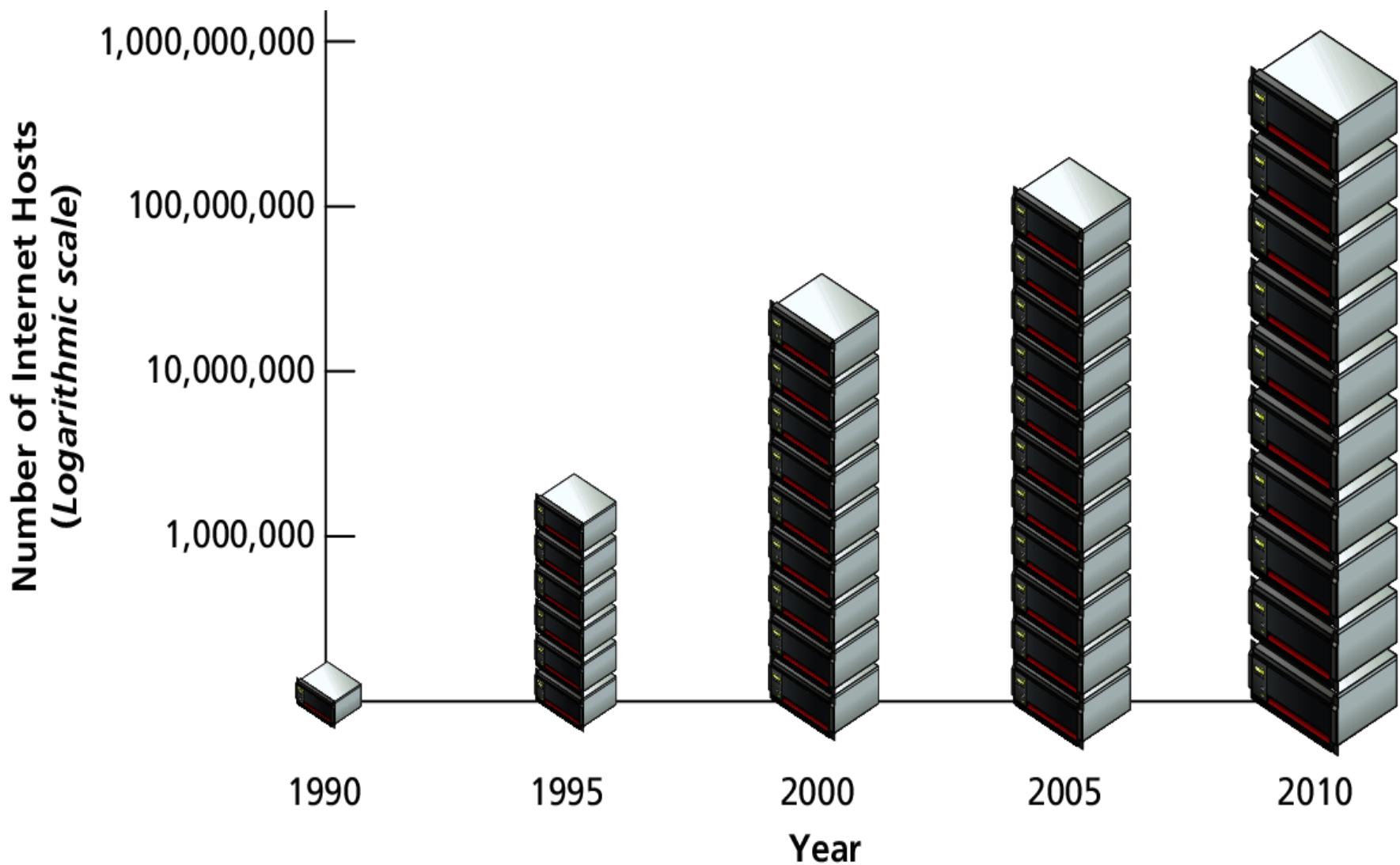
```
<a href=http://news.google.com>Breaking News from Google!</a>
```

- ❖ The program (the **web browser** in modern terminology) that displayed the **hypertext** used some mechanism to highlight the tag and to allow a user to activate this link.
- ❖ Such an action replaced the current document in the browser with the linked document.

The Rise of the Web

- ❖ Between 1991 and 1994, the simplicity and effectiveness of those early technologies used to surf and exchange data through the World Wide Web helped to:
 - ❖ Port them to many different operating systems
 - ❖ Spread to other social groups of people: scientific organizations → universities → the industry
 - ❖ By early 1993, there were around **50** HTTP/web servers on the Internet, and there were about **100 MB** per month of HTML network traffic.
 - ❖ By late 1993, there were more than **600** web servers on the Internet and HTML traffic had grown to **200 GB** per month.
 - ❖ Business interest in the WWW was sparked by reports appearing in the *New York Times* and *The Economist*.
 - ❖ The following years saw an exponential growth of the number of web sites and servers

The Growth of Web Servers



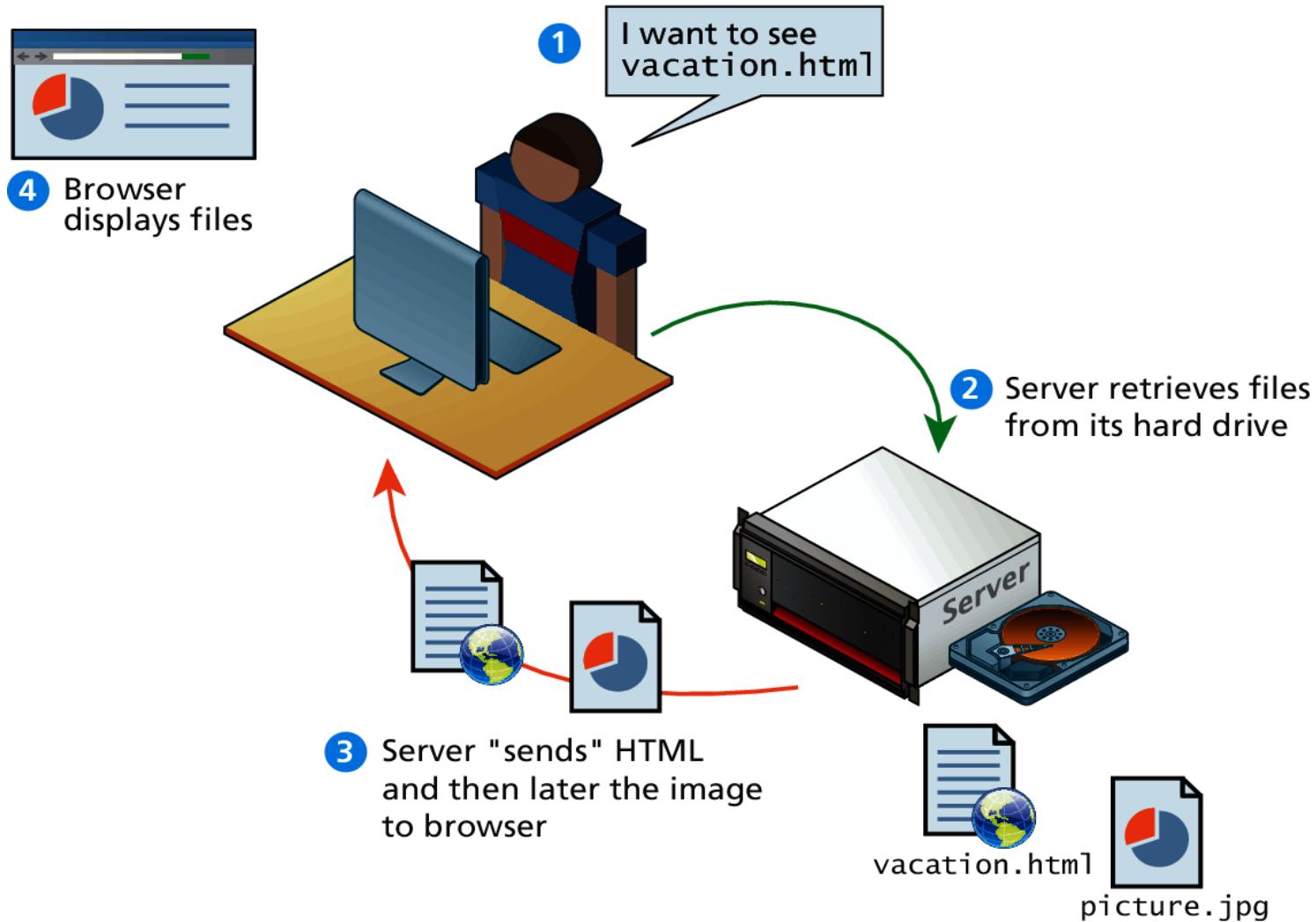
W3C

- ❖ Also in late 1994, Berners-Lee helped found the **World Wide Web Consortium (W3C)**, which would soon become the international standards organization that would oversee the growth of the web.
- ❖ This growth was very much facilitated by the decision of CERN to not patent the work and ideas done by its employee and instead left the web protocols and code-base royalty free.

Static Web Sites

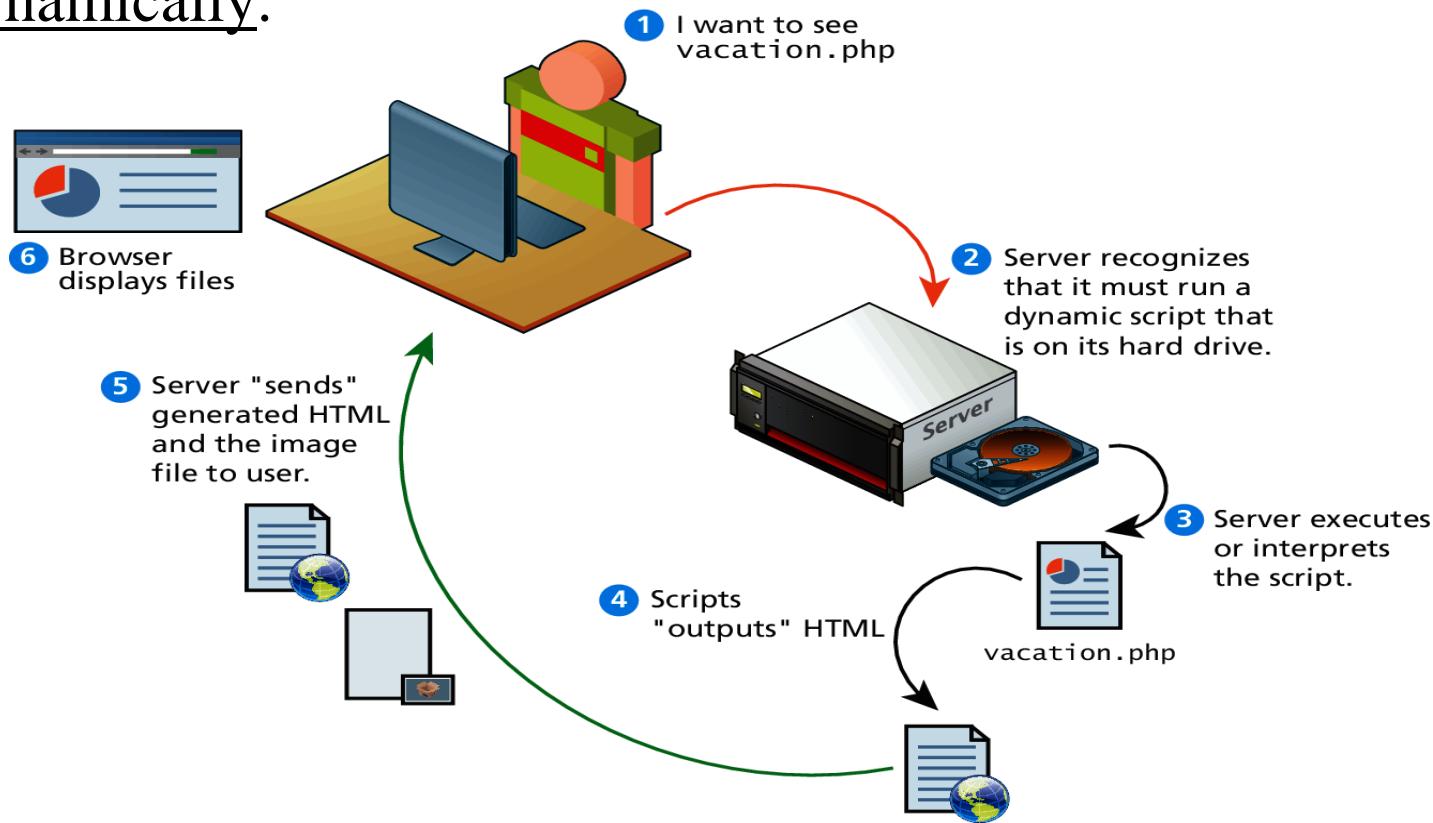
- ❖ In the earliest days of the web, a **webmaster** (the term popular in the 1990s for the person who was responsible for creating and supporting a web site) would publish web pages, and periodically update them.
- ❖ In those early days, the skills needed to create a web site were pretty basic: one needed knowledge of the HTML markup language and perhaps familiarity with editing and creating images.
- ❖ This type of web site is commonly referred to as a **static web site**, in that it consists only of HTML pages that look identical for all users at all times.

Static Web Sites



Dynamic Web Sites

- ❖ Within a few years of the invention of the web, sites began to get more complicated as more and more sites began to use programs running on web servers to generate content dynamically.



Dynamic Web Sites

- ❖ These server-based programs would read content from databases, interface with other computer systems, communicate with financial institutions, and then output HTML that would be sent back to the users' browsers.
- ❖ We call this type of web site **dynamic web site** because the content of the returned page (response) is being created at run-time by a program created by a programmer.

→ **Server Side Scripting**

- ❖ This page content can vary from user to user.

Web 2.0 and Beyond

- ❖ In the mid 2000s, a new buzz-word entered the computer lexicon:
web 2.0.
- ❖ This term had two meanings, one for users and one for developers.
 - ❖ For the users: Web 2.0 referred to an interactive experience where users could contribute and consume web content → **user-generated content**, thus creating a more user-driven web experience. → **social media**
 - ❖ For software developers, Web 2.0 also referred to a change in the paradigm of how dynamic web sites are created.
 - ❖ Programming logic, which previously existed only on the server, began to migrate to the client (browser). → **Client Side Scripting**
 - ❖ This required learning **Javascript**, a programming language that runs in the browser, as well as mastering the rather difficult programming techniques involved in **asynchronous** communication (AJAX).

Core Features of the Web

- ❖ Shortly after that initial proposal, Berners-Lee developed the main features of the web:
 1. **A URL to uniquely identify a resource on the WWW.**
 2. The HTTP protocol to describe how requests and responses operate.
 3. A software program (later called web server software) that can respond to HTTP requests.
 4. HTML to publish documents.
 5. A program (later called a web browser) to make HTTP requests from URLs and that can display the HTML it receives.

URL

- ❖ In order to allow clients to request particular resources from the server, a naming mechanism is required so that the client knows how to ask the server for the file.
- ❖ For the web that naming mechanism is the **Uniform Resource Locator (URL)**.

`http://www.funwebdev.com/index.php?page=17#article`

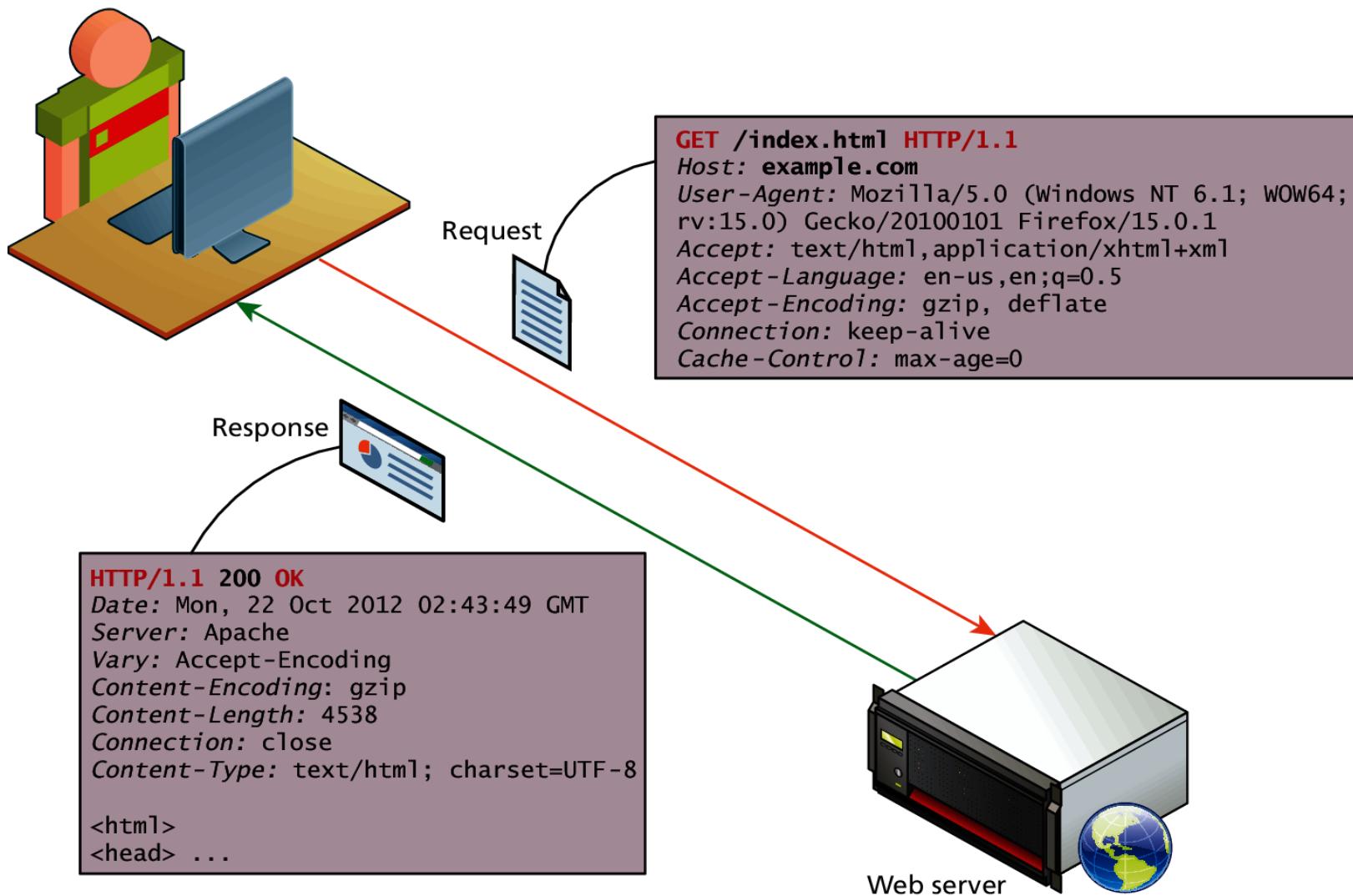
The diagram illustrates the structure of a URL by dividing it into five labeled segments: Protocol, Domain, Path, Query String, and Fragment. Each segment is represented by a blue horizontal line with vertical end caps, positioned below the corresponding part of the URL string. The URL itself is written in red text.

Protocol	Domain	Path	Query String	Fragment
http:	//www.funwebdev.com/	index.php	?page=17	#article

Core Features of the Web

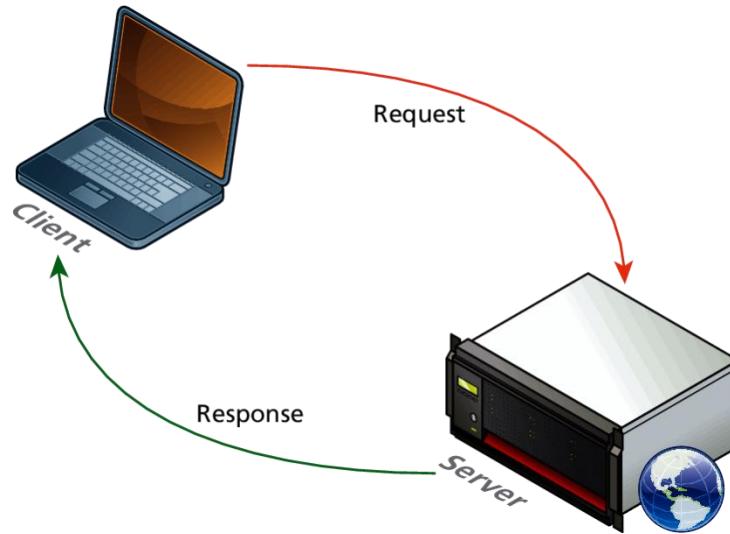
- ❖ Shortly after that initial proposal Berners-Lee developed the main features of the web:
 1. A URL to uniquely identify a resource on the WWW.
 2. **The HTTP protocol to describe how requests and responses operate.**
 3. A software program (later called web server software) that can respond to HTTP requests.
 4. HTML to publish documents.
 5. A program (later called a web browser) to make HTTP requests from URLs and that can display the HTML it receives.

HTTP (HyperText Transfer Protocol)



Request-Response Loop

- ❖ The client initiates a **request** to a server and gets a **response** that could include some resource like an HTML file, an image or some other data.



- ❖ The **request-response loop** is the most basic mechanism on the server for receiving requests and transmitting data in response.

Core Features of the Web

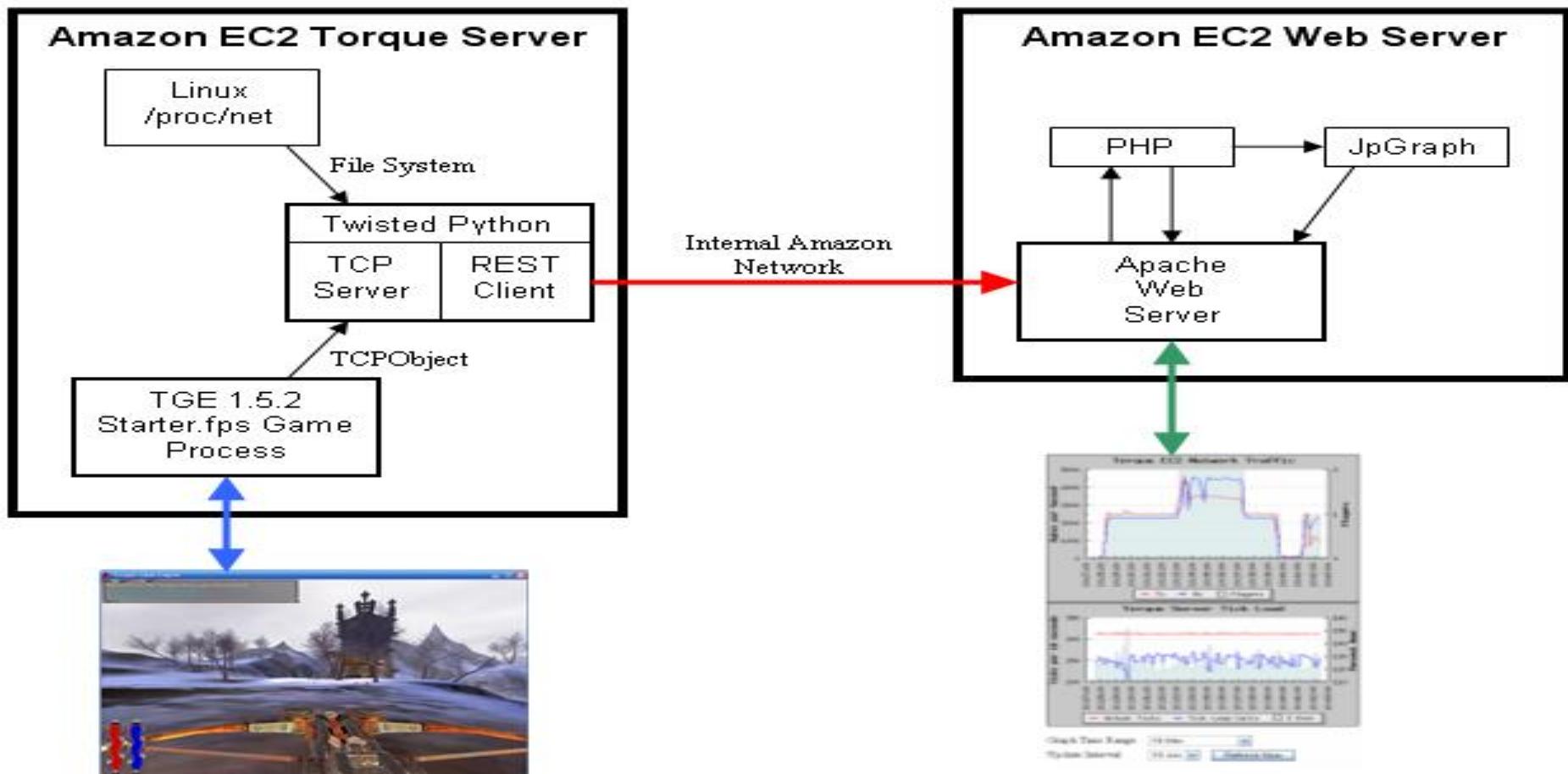
- ❖ Shortly after that initial proposal, Berners-Lee developed the main features of the web:
 1. A URL to uniquely identify a resource on the WWW.
 2. The HTTP protocol to describe how requests and responses operate.
 3. **A software program (later called web server software) that can respond to HTTP requests.**
 4. HTML to publish documents.
 5. A program (later called a web browser) to make HTTP requests from URLs and that can display the HTML it receives.

Web Servers

- ❖ **Web server** can refer to either the hardware (the computer) or the software (the computer application) that helps to deliver Web content that can be accessed through the Internet.
- ❖ A **web server** can be:
 - ❖ A **computer program**
 - ❖ Responsible for accepting HTTP requests from clients (web browsers)
 - ❖ Returns HTTP responses with optional data contents
 - ❖ Usually web pages
 - ❖ HTML documents
 - ❖ Linked objects (images, etc.).
 - ❖ A **computer** that runs a computer program which provides the above functionality

Web Servers

- The most common use of web servers is to host websites, but there are other uses such as gaming, data storage or running enterprise applications.



Core Features of the Web

- ❖ Shortly after that initial proposal Berners-Lee developed the main features of the web:
 1. A URL to uniquely identify a resource on the WWW.
 2. The HTTP protocol to describe how requests and responses operate.
 3. A software program (later called web server software) that can respond to HTTP requests.
 - 4. HTML to publish documents.**
 5. A program (later called a web browser) to make HTTP requests from URLs and that can display the HTML it receives.

HTML

- ❖ A typical HTML document will have the form:

```
<html>
```

```
<head>
```

*meta data, function definitions for client side scripting,
title, ...*

```
</head>
```

```
<body>
```

Content text and associated format tags

Embedded hypertext links

Possibly forms with input tags

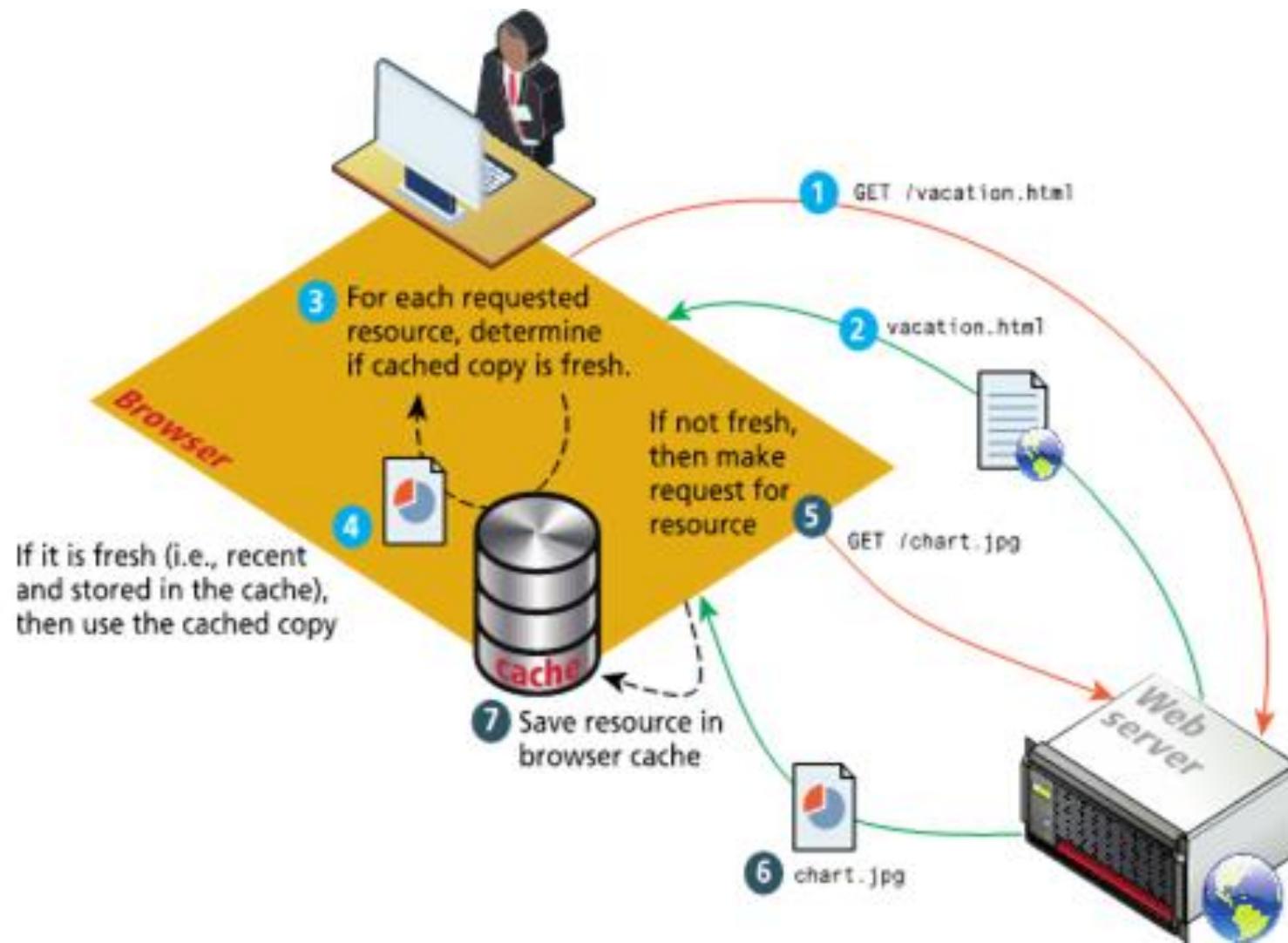
```
</body>
```

```
</html>
```

Core Features of the Web

- ❖ Shortly after that initial proposal Berners-Lee developed the main features of the web:
 1. A URL to uniquely identify a resource on the WWW.
 2. The HTTP protocol to describe how requests and responses operate.
 3. A software program (later called web server software) that can respond to HTTP requests.
 4. HTML to publish documents.
 5. **A program (later called a web browser) to make HTTP requests from URLs and that can display the HTML it receives.**

Web Browsers

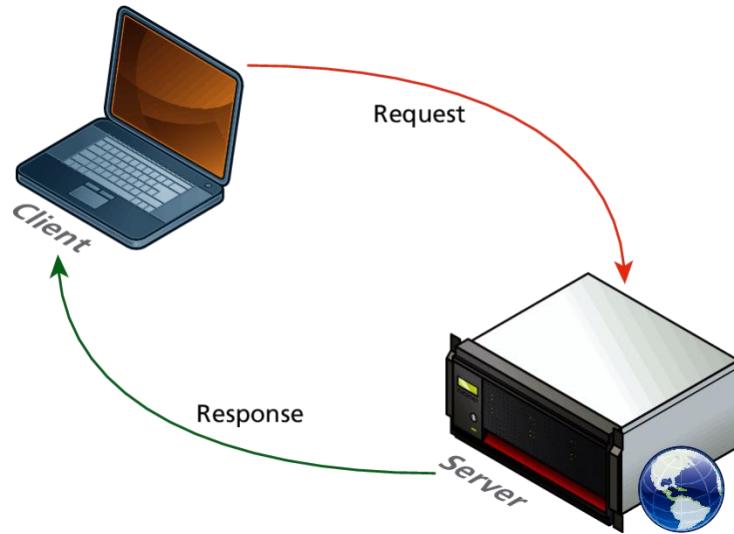


Client-Server Model

- ❖ The web is sometimes referred to as a client-server model of communications. A client-server model is a system model which shows how data and processing is distributed across a range of components.
- ❖ In the **client-server model**, there are two types of components:
 - ❖ **Servers** provide specific services such as access to Web pages, data management, etc. They are normally active 24 hours a day, 7 days a week, listening for messages from any client who makes a request.
 - ❖ **Clients** use these services. They make requests and receive responses from the server, in the form of text files, images, response codes, etc. They access servers over a network, such as the Internet.
- ❖ The clients (browsers) and servers may be running on a single (local) machine or separate machines.

Client-Server Computing

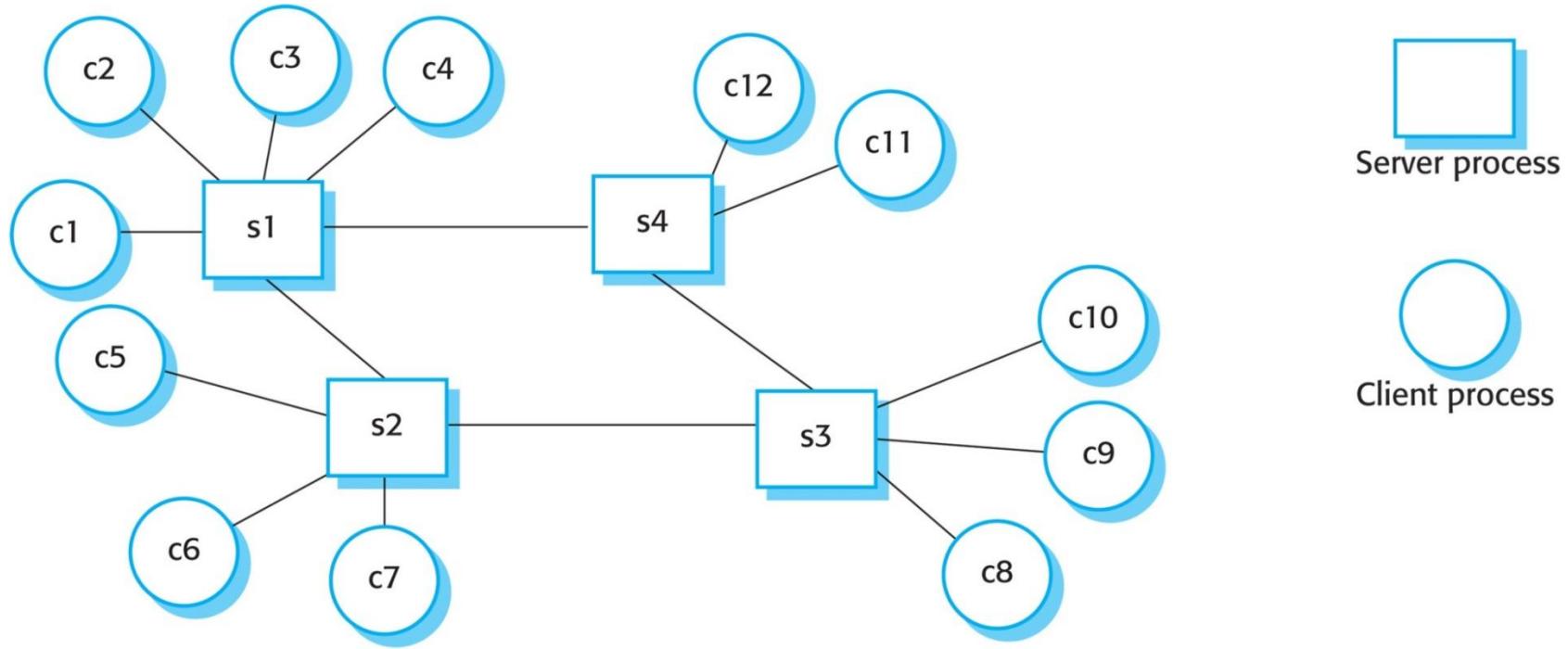
- ❖ More formally, the client initiates a **request** to a server and gets a **response** that could include some resource like an HTML file, an image or some other data.



- ❖ The **request-response loop** is the most basic mechanism on the server for receiving requests and transmitting data in response.

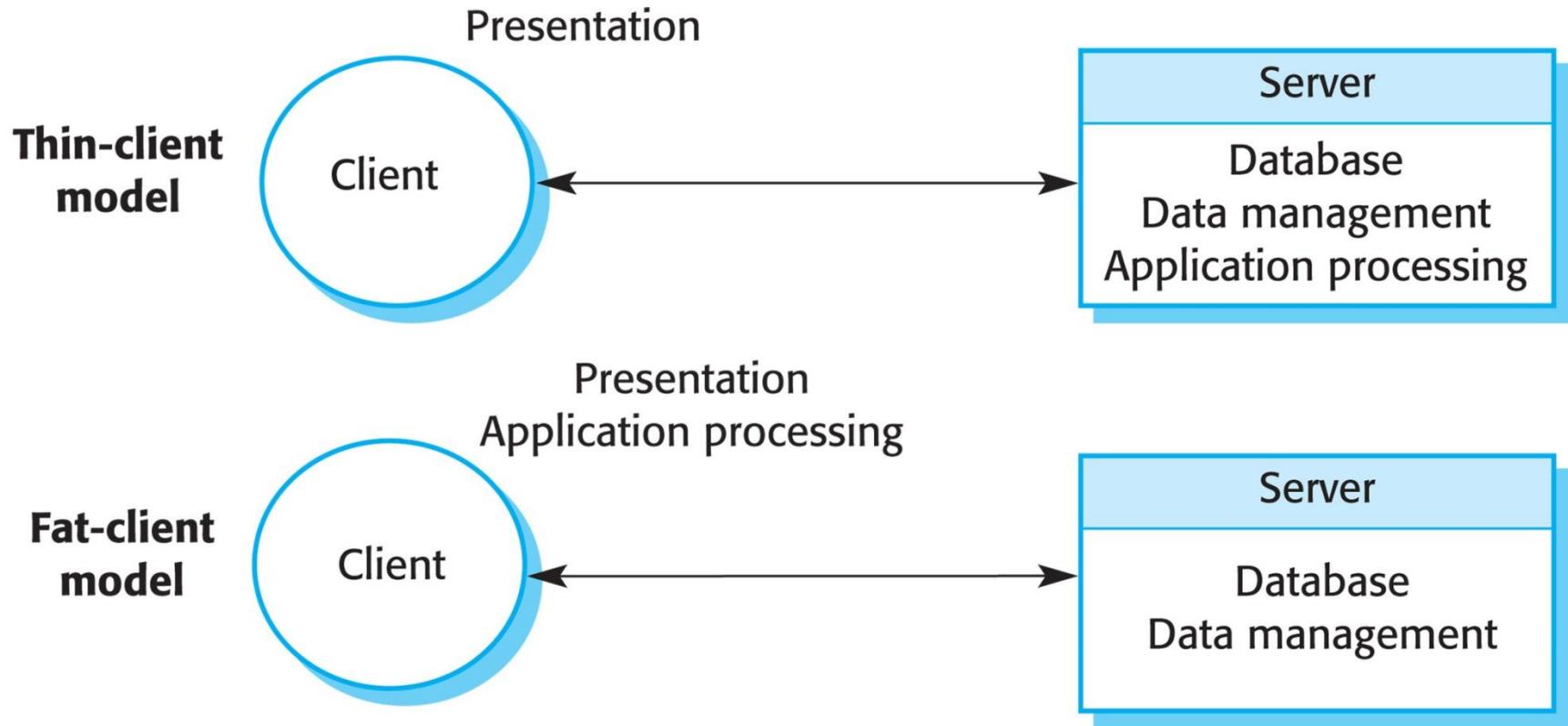
2-tier Client Server Systems

- ❖ In a two-tier client-server system, the system is implemented as a single logical server plus an indefinite number of clients that use that server.



Copyright ©2016 Pearson Education, All Rights Reserved

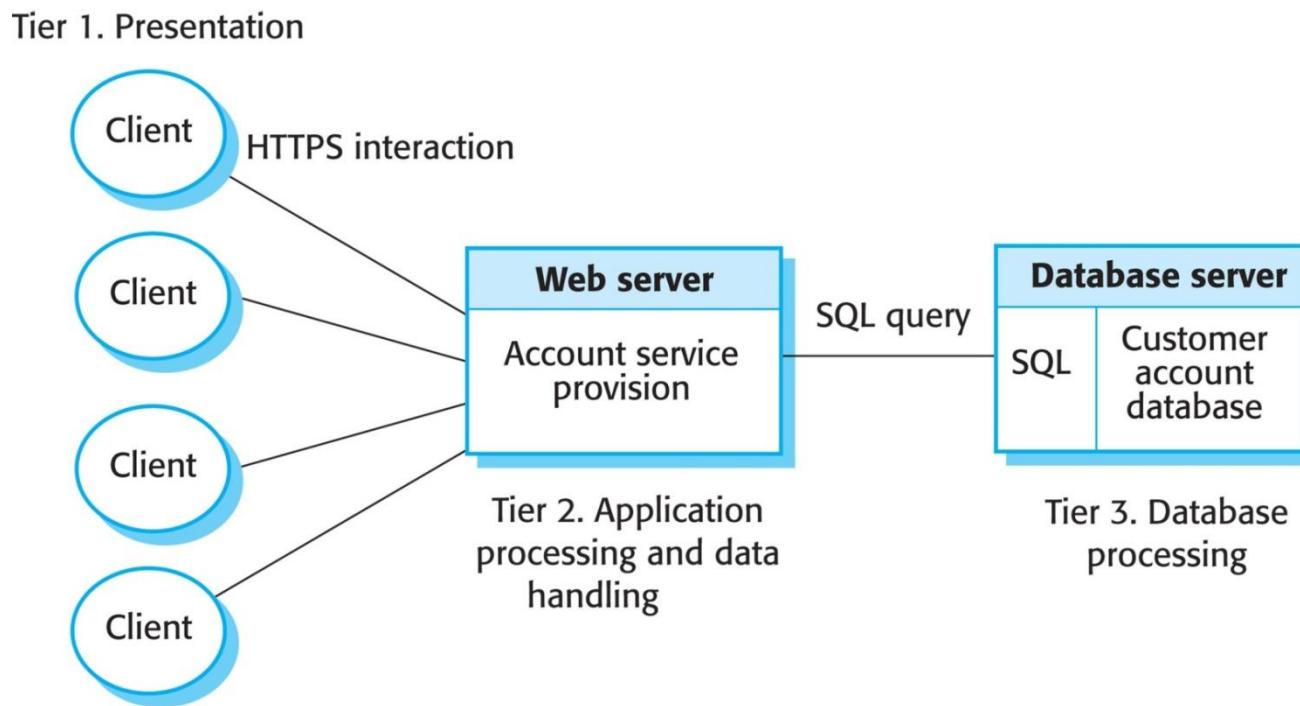
Thin- & Fat-client Architectural Models



Copyright ©2016 Pearson Education, All Rights Reserved

Multi-tier CS Architectures

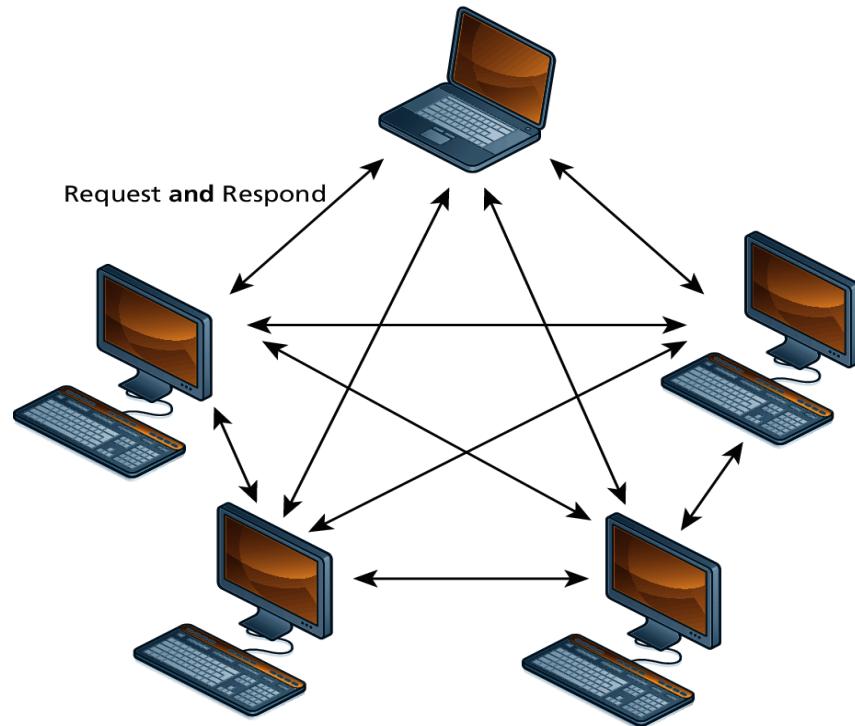
- ❖ In a ‘multi-tier client–server’ architecture, the different layers of the system, namely presentation, application processing, data management,, and database, are separate processes that may execute on different processors.



Copyright ©2016 Pearson Education, All Rights Reserved

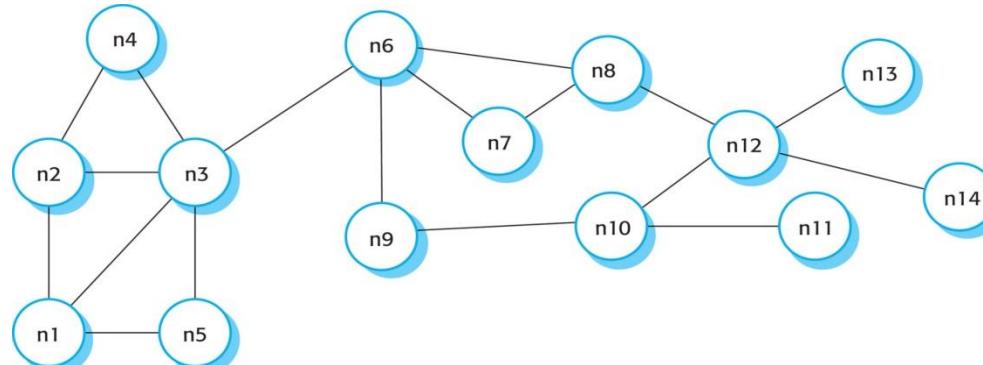
The Peer-to-Peer Alternative

- ❖ In the **peer-to-peer model** where each computer is functionally identical, each node is able to send and receive directly with one another.
- ❖ In such a model each peer acts as both a client and server able to upload and download information.



Peer-to-Peer Systems

- ❖ More formally, peer to peer (p2p) systems are decentralized distributed systems where computations may be carried out by any node (machine) in the network.
- ❖ The overall system is designed to take advantage of the computational power and storage of a large number of networked computers.
- ❖ Most p2p systems have been personal systems but there is increasing business use of this technology.



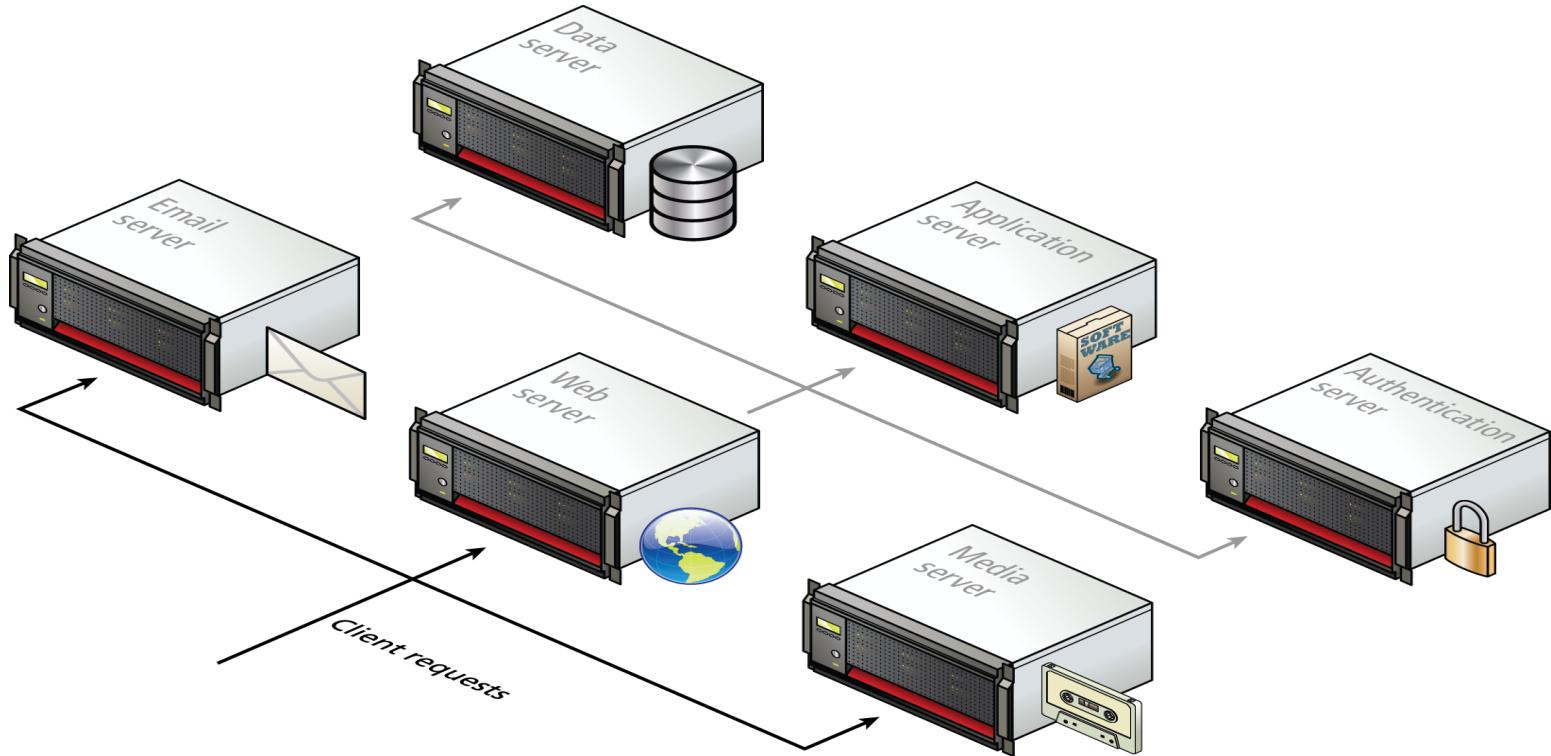
Copyright ©2016 Pearson Education, All Rights Reserved

Peer-to-Peer Systems

- ❖ It is appropriate to use a peer-to-peer architectural model for a system in two circumstances:
 1. The system is **computationally intensive** and it is possible to separate the processing into a large number of independent computations.
 - ❖ E.g. Computational drug discovery system.
 - ❖ Each node checks a separate molecule to see if it has the characteristics required to suppress the growth of cancers
 2. The system primarily involves the **exchange of information** between individual computers/processors.
 - ❖ E.g. file/music/video-sharing systems (gnutella, torrents, etc.), voice/video communication systems (Skype).

Server Types

- ❖ Most real-world web sites are typically not served from a single server machine, but by many servers.
- ❖ It is common to split the functionality of a web site between several different types of server.

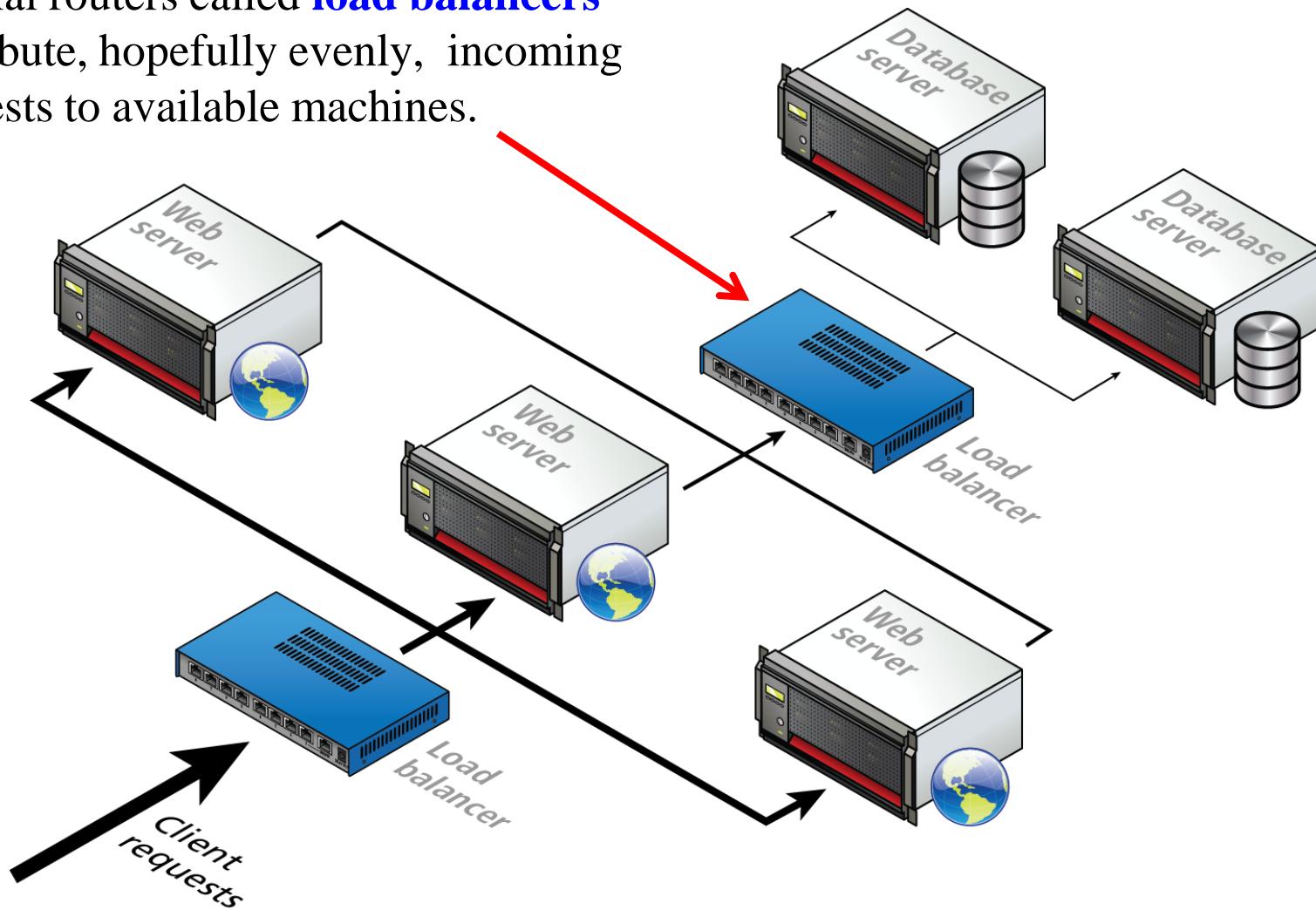


Real-World Server Installations

- ❖ Not only are there different types of servers, there is often replication of each of the different server types.
 - ❖ A busy site can receive thousands or even tens of thousands of requests a second
 - ❖ Globally popular sites such as Facebook receive millions of requests a second.
- ❖ A single web server will be hard-pressed to handle more than a few hundred requests a second, so the usual strategy for busier sites is to use a **server farm**.

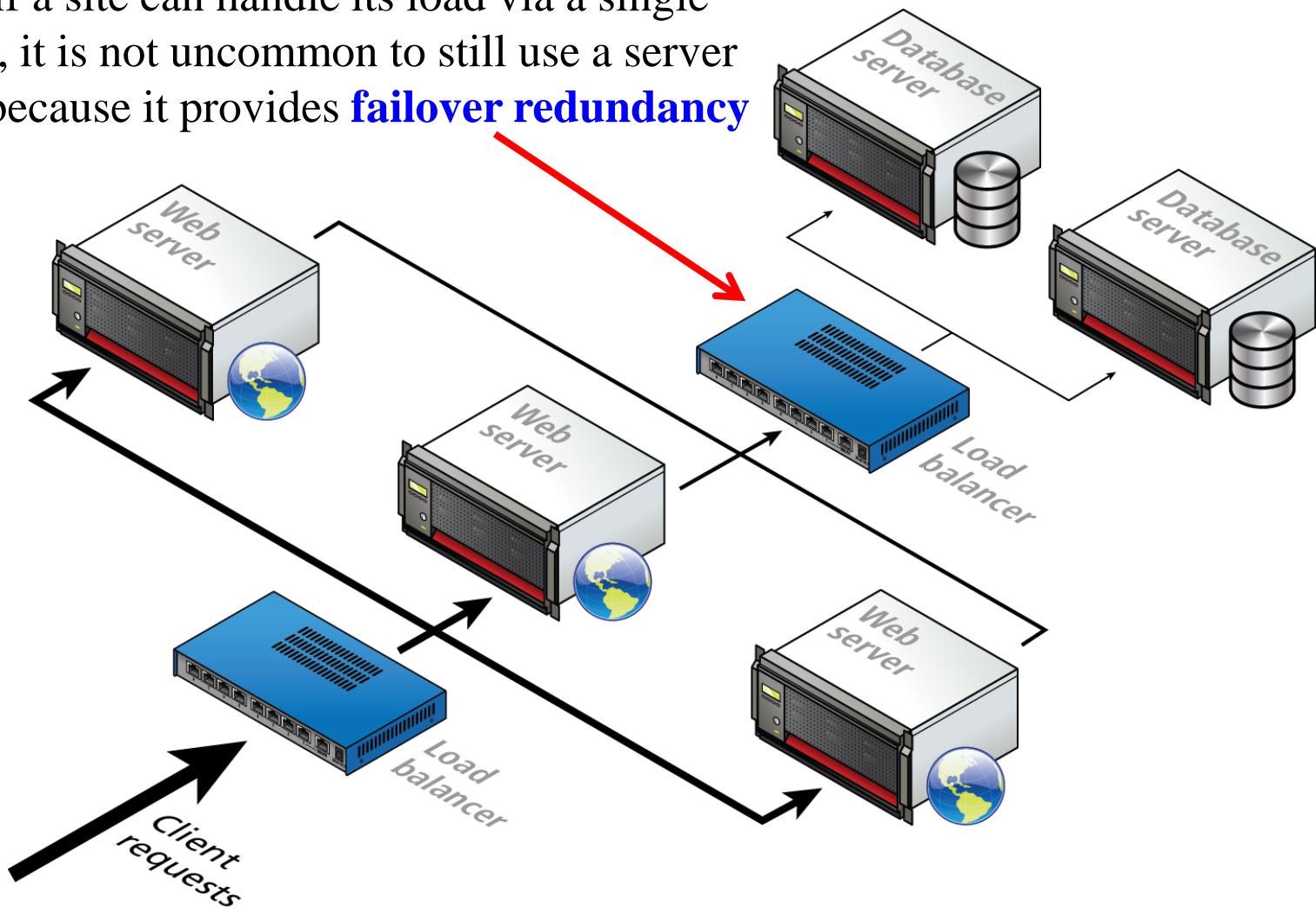
Server Farm

Special routers called **load balancers** distribute, hopefully evenly, incoming requests to available machines.



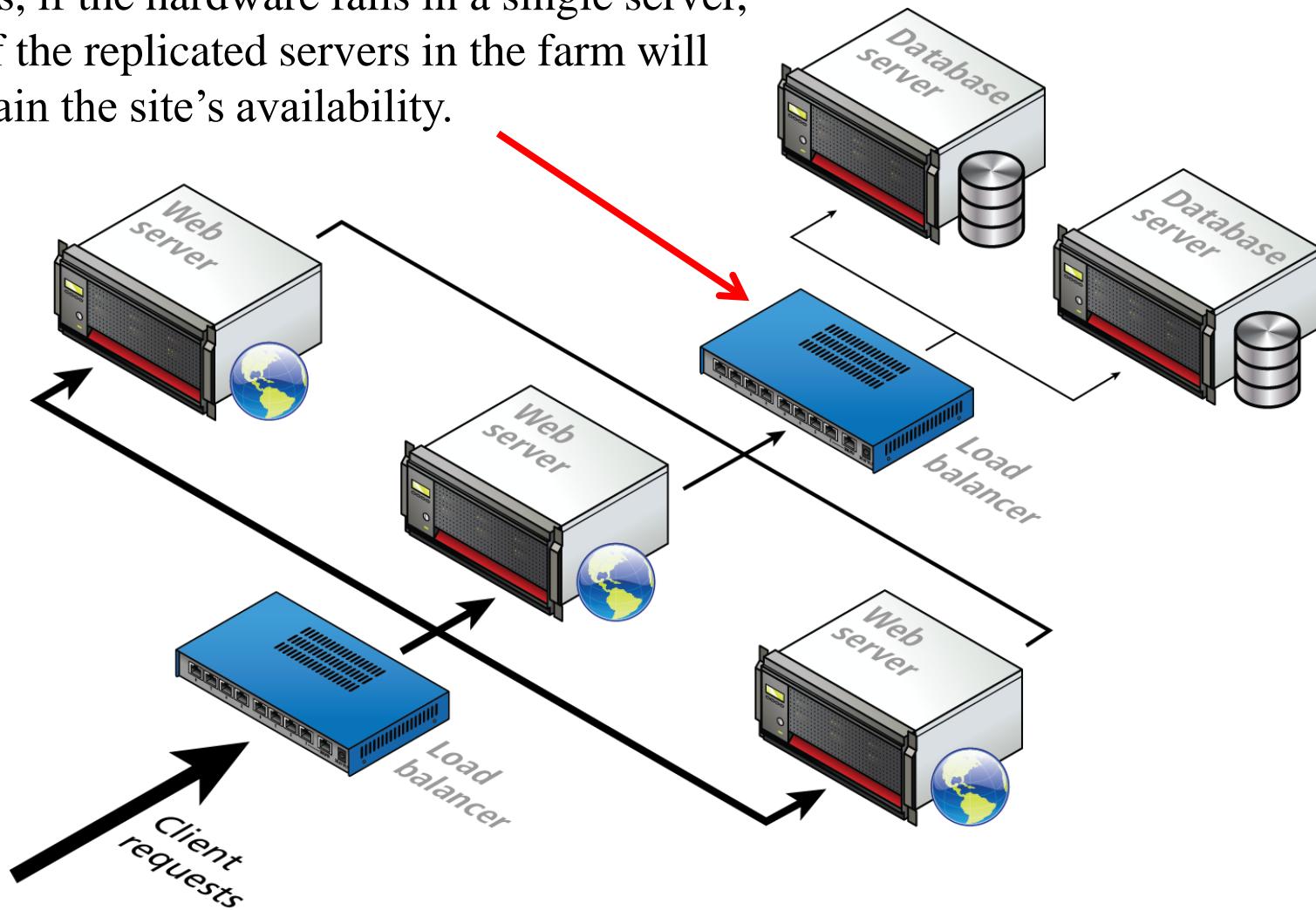
Server Farm

Even if a site can handle its load via a single server, it is not uncommon to still use a server farm because it provides **failover redundancy**



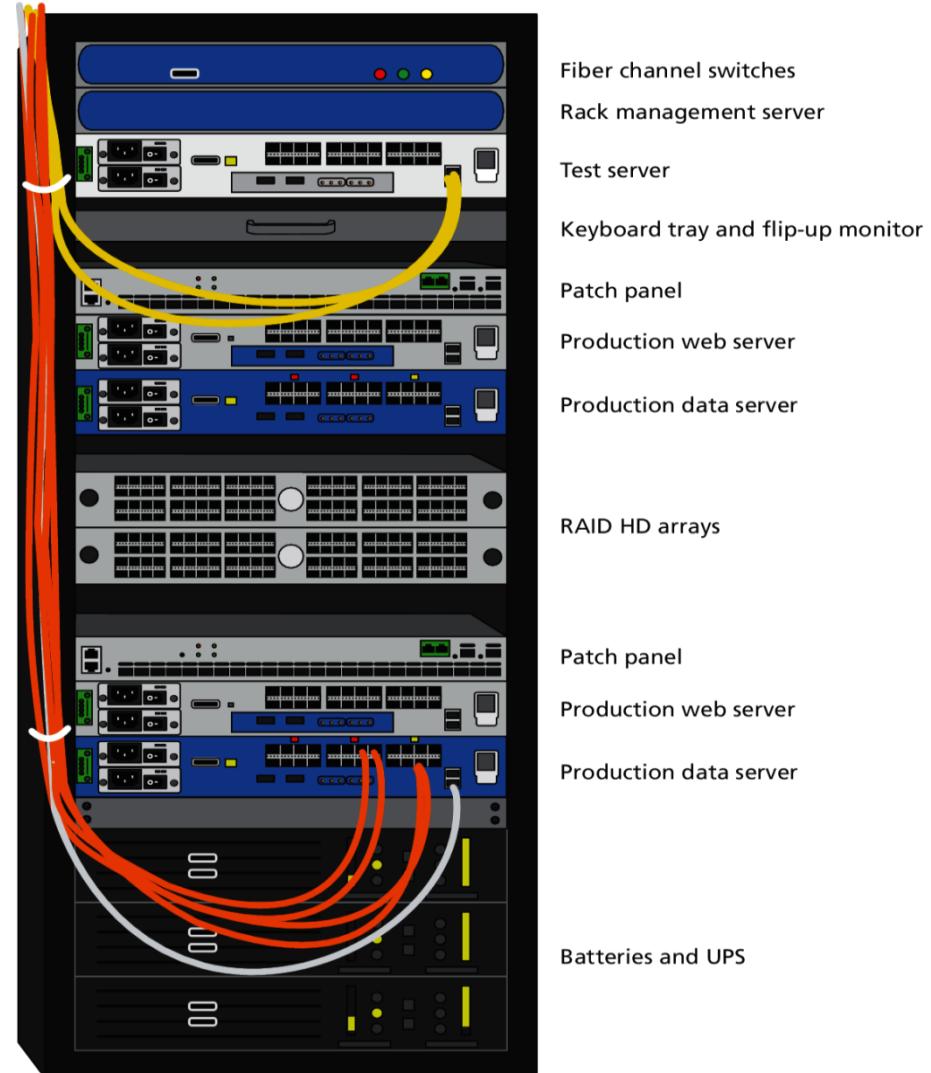
Server Farm

That is, if the hardware fails in a single server, one of the replicated servers in the farm will maintain the site's availability.



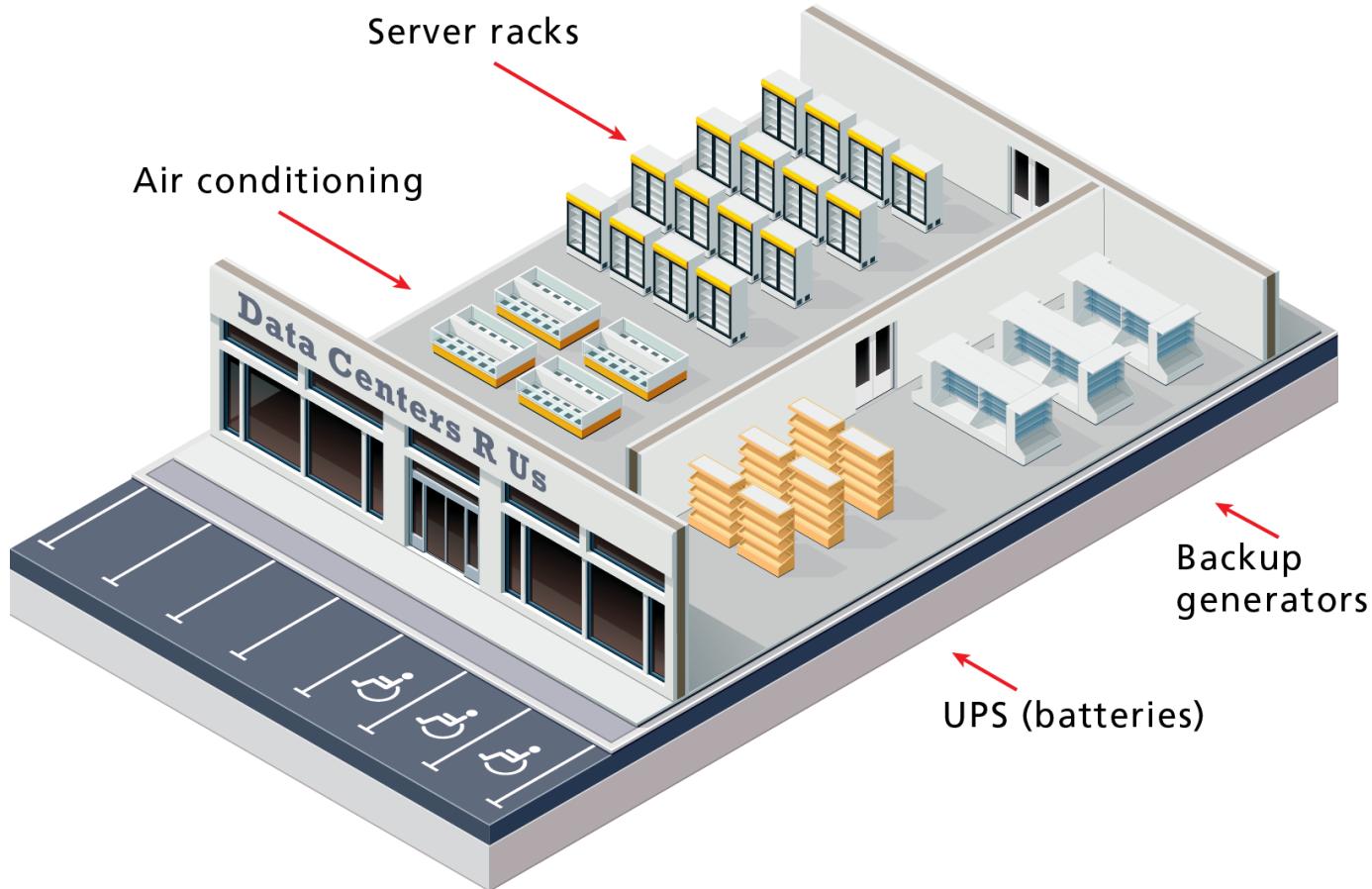
Server Rack

- ❖ In a server farm, the computers do not look like the ones in your house. Instead, these computers are more like the plates stacked in your kitchen cabinets.
- ❖ That is, a farm will have its servers and hard drives stacked on top of each other in **server racks**.
- ❖ A typical server farm will consist of many server racks, each containing many servers.



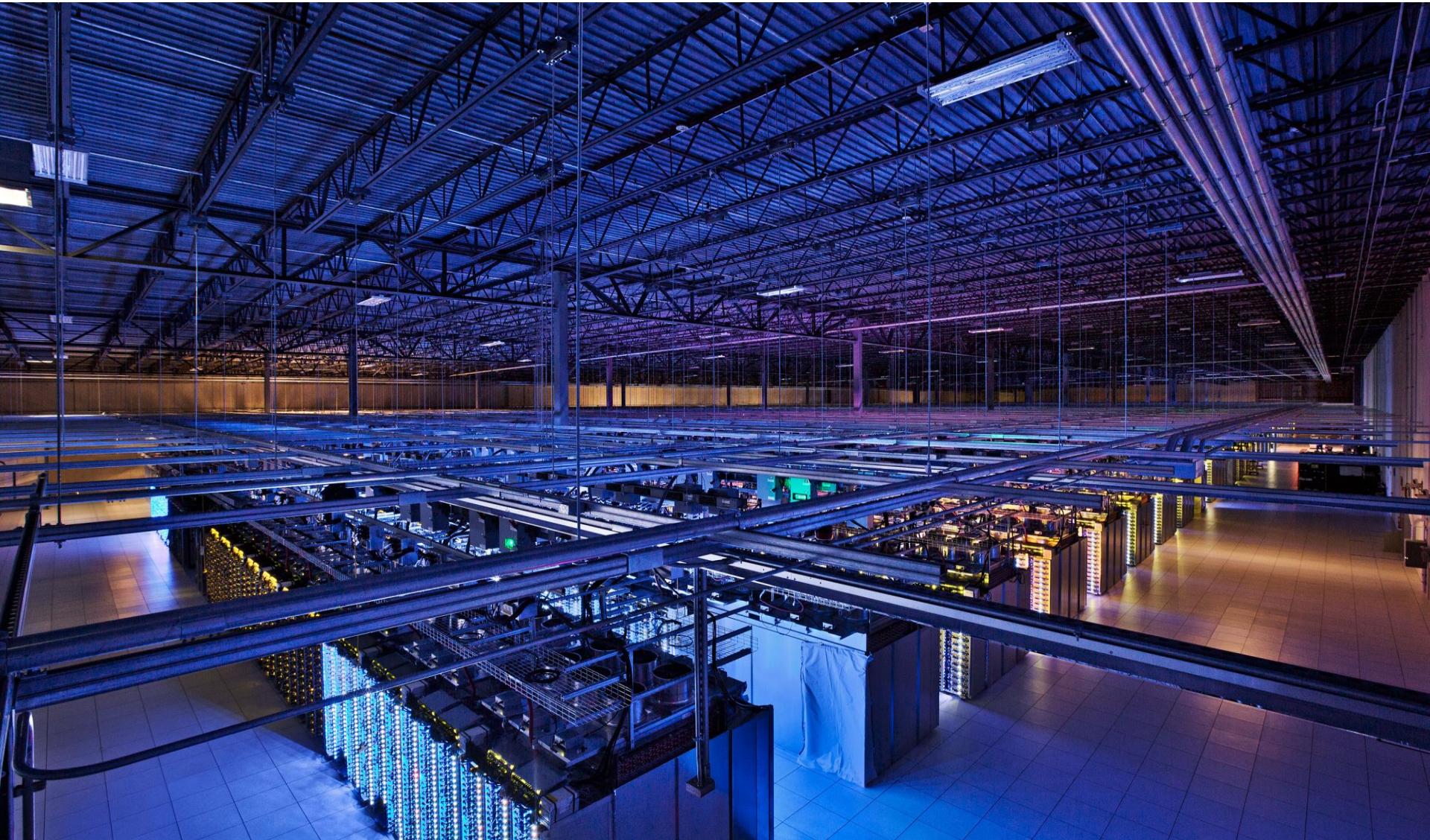
Data Centers

- ❖ Server farms are typically housed in special facilities called **data centers**.



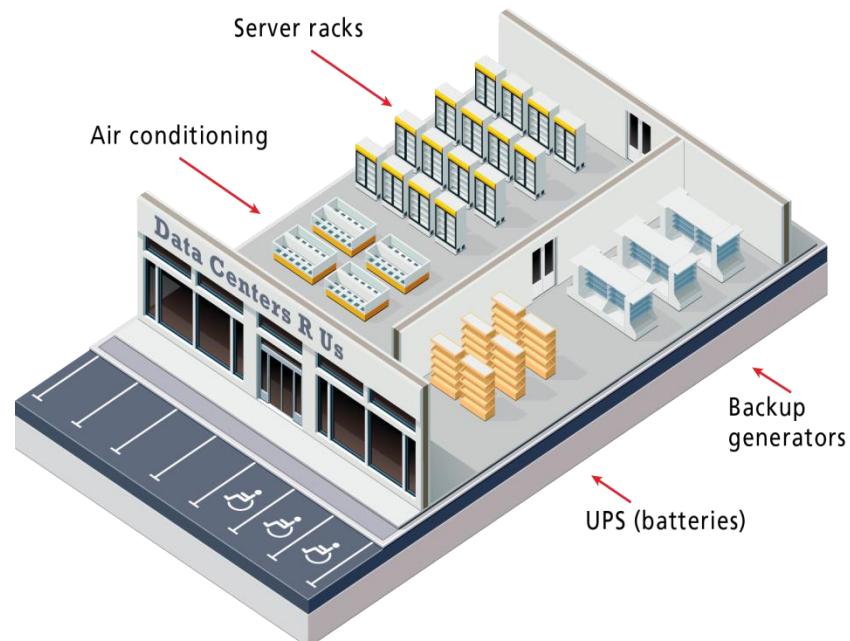
A Google Data Center

Google's data center in Council Bluffs, Iowa, houses servers in over 115,000 square feet of space.



Data Centers

- ❖ To prevent the potential for site down times, most large web sites will exist in mirrored data centers in different parts of the country, or even world. (See next slide)
- ❖ As a consequence, the costs for multiple redundant data centers are quite high, and only larger web companies can afford to create and manage their own.
- ❖ Most web companies will instead lease space from a third-party data center.



Data center locations

We own and operate data centers around the world to keep our products running 24 hours a day, 7 days a week. Find out more about our data center locations, community involvement, and [job opportunities](#) in our locations around the world.

Americas

- Berkeley County, South Carolina
Council Bluffs, Iowa
Douglas County, Georgia
Jackson County, Alabama
Lenoir, North Carolina
Mayes County, Oklahoma
Montgomery County, Tennessee
Quilicura, Chile
The Dalles, Oregon



Asia

- Changhua County, Taiwan
Singapore

Europe

- Dublin, Ireland
Eemshaven, Netherlands
Hamina, Finland
Göteborg, Sweden

Commercial Web Hosting

- ❖ It is also common for the reverse to be true – that is, a single server machine may host multiple sites.
- ❖ Large commercial web hosting companies such as GoDaddy, and others will typically host hundreds or even thousands of sites on a single machine (or mirrored on several servers).