

Web Technology

Introduction to Internet

A global computer network providing a variety of information and communication facilities, consisting of interconnected networks using standardized communication protocols.

The internet is a huge collection of computers connected in a communication network.

History

- The first workable prototype of the Internet came in the late 1960s with the creation of **ARPANET**, or the Advanced Research Projects Agency Network. Originally funded by the U.S. Department of Defense, ARPANET used packet switching to allow multiple computers to communicate on a single network.
- In 1973 global networking becomes a reality as the University College of London (England) and Royal Radar Establishment (Norway) connect to ARPANET. The term **Internet** is born.
- In 1974 the first Internet Service Provider (ISP) is born with the introduction of a commercial version of ARPANET, known as **Telenet**.
- In 1986, the NSF(National Science Foundation) created **NSFNET**, a 56 kbit/s backbone to support the NSF-sponsored supercomputing centers. The NSFNET also provided support for the creation of regional research and education networks in the United States, and for the connection of university and college campus networks to the regional networks.
- By 1990, NSFnet had replaced ARPAnet for most nonmilitary uses, and a wide variety of organizations had established nodes on the new network—by 1992 NSFnet connected more than 1 million computers around the world. In 1995, a small part of NSFnet returned to being a research network. The rest became known as the **Internet**, although this term was used much earlier for both ARPAnet and NSFnet.

World Wide Web

- The World Wide Web (WWW) is a **network of online content** that is formatted in HTML and accessed via HTTP.
- The term refers to all the interlinked HTML pages that can be accessed over the Internet. The World Wide Web was originally designed in 1991 by Tim Berners-Lee.
- It is all the Web pages, pictures, videos and other online content that can be accessed via a Web browser.
- The Internet, in contrast, is the underlying network connection that allows us to access the World Wide Web. (Internet is infrastructure while the Web is service on top of that infrastructure.)

INTERNET	WORLD WIDE WEB
A global system of interconnected computer networks that use the TCP/IP protocol to link devices worldwide	Online content that is formatted in HTML and accessed via HTTP protocol
A massive interconnection of computer networks around the world	Service provided by the internet
Uses Transmission Control Protocol/Internet Protocol (TCP/IP)	Uses Hyper Text Transfer Protocol (HTTP)
	Visit www.PEDIAA.com

Web Browsers

A web browser (commonly referred to as a browser) is a software application for accessing information on the World Wide Web.

A browser is a client on the Web because it initiates the communication with a server.

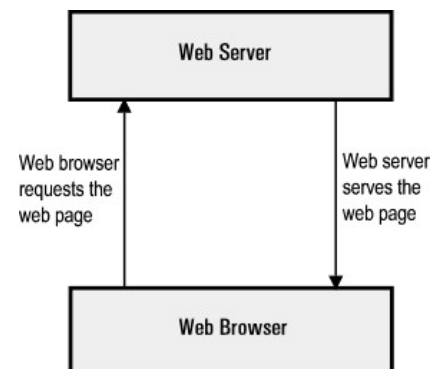
Each individual web page, image, and video is identified by a distinct **Uniform Resource Locator (URL)**, enabling browsers to retrieve these resources from a web server and display them on a user's device.

Web Servers

- Web server is a system that provide documents to requesting browsers.
- A web server consists of a physical server, server operating system (OS) and software used to facilitate HTTP communication.
- Servers are slave programs: They act only when requests are made to them by browsers running on other computers on the Internet.

How server works

1. Browser Resolves the Domain Name to an IP Address.
(from cache or DNS servers)
2. Browser Requests the Full URL (from web server).
3. Web Server sends the Requested Page (or error message).
4. Browser Displays the Webpage.



Popular web servers

Apache, Microsoft IIS, Nginx Web Server, LightSpeed Web Server.

The two most popular ones are-

Apache HTTP Server: It is the most popular web server available and is widely used. It is developed and maintained by Apache Software Foundation. The Software is produced under the Apache License, which makes it free and open source.

It is available for a variety of operating Systems- Windows, Mac OS X, Unix, Linux, Solaris, Novell Netware and FreeBSD.

Microsoft Internet Information Service: Microsoft IIS (Internet Information Service) is the second most popularly used web server, and its market share is increasing quite rapidly with each passing day and is likely to overhaul Apache in the upcoming years.

IIS server has all the features just like Apache, but it is not open source. It is developed, and maintained by Microsoft, thus it works with all the Windows operating system platforms.

IP address

- An Internet Protocol address (IP address) is a numerical label assigned to each device connected to a computer network that uses the Internet Protocol for communication.
- Networks using the TCP/IP protocol route messages based on the IP address of the destination.
- The format of an IP address is a 32-bit numeric address written as four numbers separated by periods. Each number can be zero to 255. For example, 1.160.10.240 could be an IP address.

An example of an IPv6 address(128-bit) is: 2001:0db8:85a3:0000:0000:8a2e:0370:7334.

Domain names

A domain name is simply a human readable form of an IP address.

Examples: www.facebook.com 157.240.25.35 |
www.google.com 74.125.200.100/ 102

```
C:\Users\dileep>nslookup
Default Server: UnKnown
Address: fe80::1

> google.com
Server: UnKnown
Address: fe80::1

Non-authoritative answer:
Name: google.com
Addresses: 2404:6800:4003:c04::65
          74.125.200.100
          74.125.200.102
          74.125.200.113
          74.125.200.138
          74.125.200.139
          74.125.200.101

> facebook.com
Server: UnKnown
Address: fe80::1

Non-authoritative answer:
Name: facebook.com
Addresses: 2a03:2880:f139:183:face:b00c:0:25de
          157.240.25.35
```

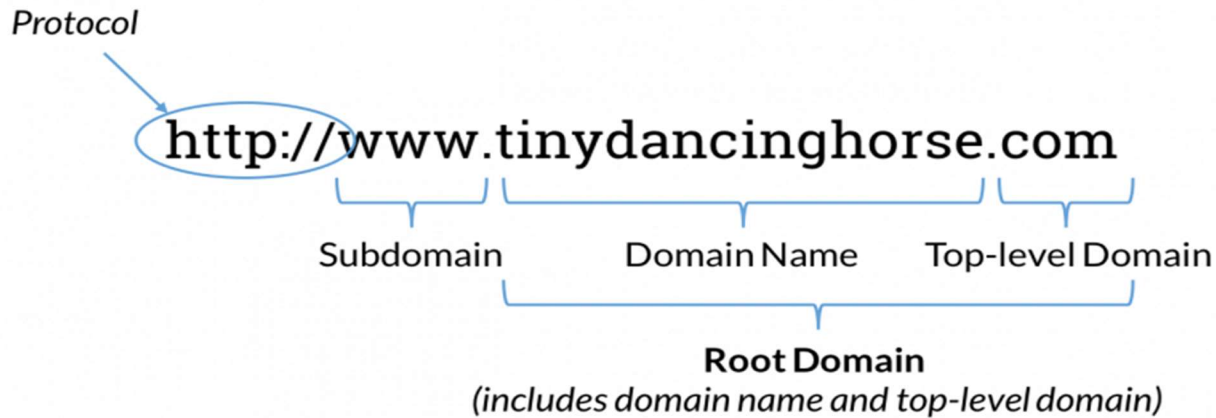
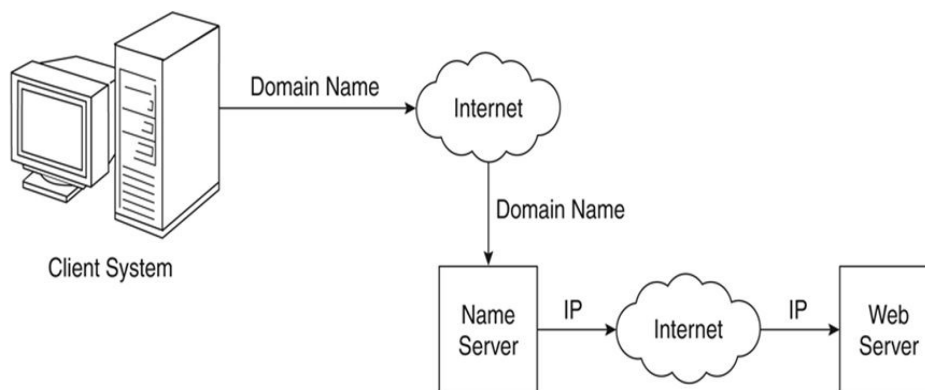


Fig: Domain name description

- The **Domain Name System (DNS)** is the way internet domain names are located and translated into internet protocol addresses.
- The Domain Name Systems (DNS) is the phonebook of the Internet.
- Humans access information online through domain names. Web browsers interact through Internet Protocol (IP) addresses.

DNS translates domain names to IP addresses so browsers can load Internet resources.

Domain name conversion

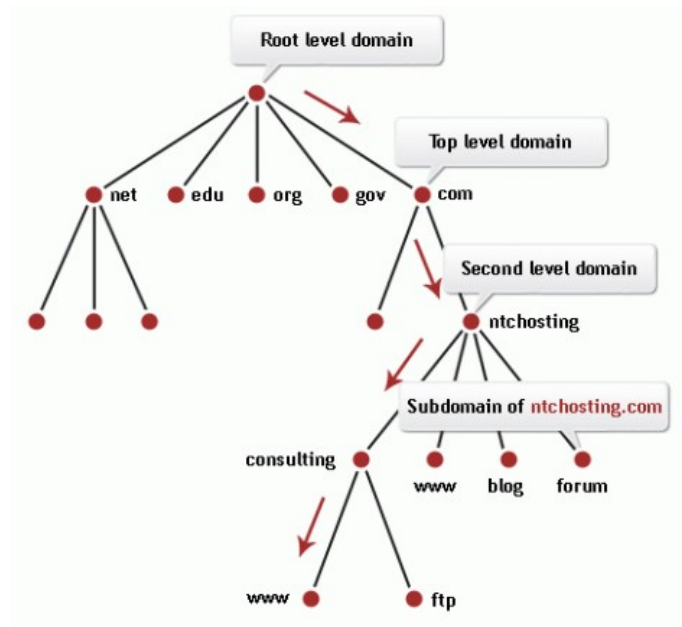
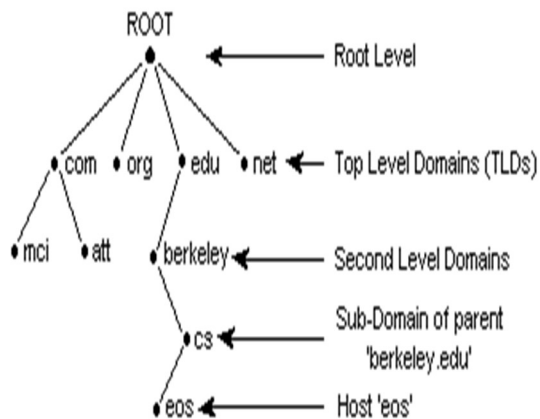


7

- A **DNS server** is a computer used to resolve hostnames to IP addresses.
- **Nameservers** are simply a set of servers (computers) that contain the DNS information.

DNS hierarchy

DNS Hierarchy



The 8 steps in a DNS lookup:

1. A user types 'example.com' into a web browser and the query travels into the Internet and is received by a **DNS recursive resolver**.
2. The resolver then queries a **DNS root nameserver** (.).
3. The root server then responds to the resolver with the address of a **Top Level Domain** (TLD) DNS server (such as .com or .net), which stores the information for its domains. When searching for example.com, our request is pointed toward the .com TLD.
4. The resolver then makes a request to the .com TLD.
5. The TLD server then responds with the IP address of the domain's nameserver, example.com.
6. Lastly, the recursive resolver sends a query to the domain's nameserver.
7. The IP address for example.com is then returned to the resolver from the nameserver.
8. The DNS resolver then responds to the web browser with the IP address of the domain requested initially.

Uniform resource locators

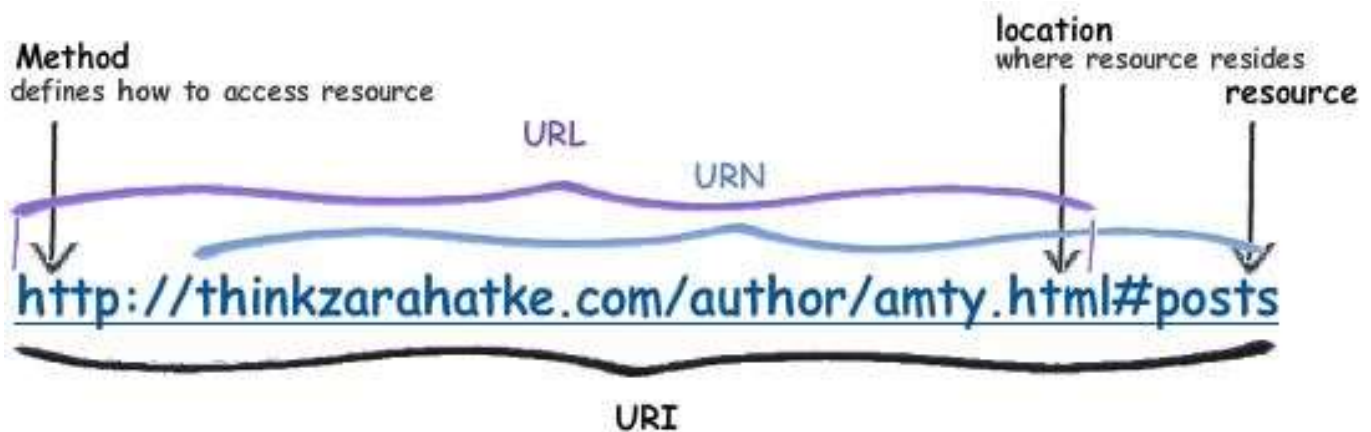
- **Uniform resource locators (URLs)** are used to identify documents (resources) on the Internet.

Following is the general format of a URL:

scheme://host.domain:port/path/filename

Example: `http://www.w3.org:8000/imaginary/test.jpg`

- ❖ The scheme tells web servers which communication protocol to use when it accesses a page on your website. (Examples: http,ftp,file,mailto etc)
 - ❖ The host name is the name of the server computer that stores the document (or provides access to it on some other computer).
 - ❖ Domain names are used in URLs to identify particular Web pages.
 - ❖ A port is a logical construct that identifies a specific process or a type of network service.
 - ❖ Path- Path to the location of the particular file inside the server computer.
 - ❖ File Name - The requested file.
- ✓ **URI** stands for **Uniform Resource Identifier** and it identifies a resource either by location, or a name, or both.
 - ✓ A URI has two specializations: one is **URN** and the other is **URL**.
 - ✓ **URN** stands for **Uniform Resource Name** and it identifies a resource by name in a given namespace.
 - ✓ **URL** stands for Uniform Resource Locator and it is a specialization of URI that defines the network location of a specific resource. Unlike a URN, the URL defines how the resource can be obtained.
 - ✓ **URL+URN = URI**



Stands for Uniform Resource Locator	Stands for Uniform Resource Identifier
A type of URI	Superset of URL
Helps to identify a web resource using the location	Helps to identify a web resource either by name, location or both

- ❖ Protocol specification is provided in URL but not in URI

HTTP

- HTTP means **Hypertext Transfer Protocol**.
- HTTP is the underlying **protocol** used by the World Wide Web and this protocol defines how messages are formatted and transmitted, and what actions Web servers and browsers should take in response to various commands.
- HTTP is called a stateless protocol because each command is executed independently, without any knowledge of the commands that came before it.
- HTTP uses port 80 for communication whereas HTTPS uses port 443.
- HTTPS means Hyper Text Transfer Protocol Secure. Basically, it is the secure version of HTTP. Communications between the browser and website are encrypted by Transport Layer Security (TLS), or its predecessor, Secure Sockets Layer (SSL).

HTTP	HTTPS
An application protocol for distributed, collaborative, and hypermedia information systems	An extension of HTTP protocol for secure communication over a computer network
Stands for Hyper Text Transfer Protocol	Stands for Hyper Text Transfer Protocol Secure
Less secure	More secure
Works in the application layer of the OSI model	Works in the transport layer of the OSI model
Does not use a certificate	Uses SSL certificate
There is no encryption and decryption	There is encryption and decryption
Helps to transfer text, audio, video, images through web pages	Helps to transfer data securely via the network

HTTP request phase

A simple request message from a client computer consists of the following components:

- A request line to get a required resource, for example a request

GET /content/page1.html is requesting a resource called /content/page1.html from the server.

- Header fields (Example – Accept:Language EN).

- An empty line.

- A message body which is optional.

All the lines should end with a carriage return and line feed. The empty line should only contain carriage return and line feed without any spaces.

HTTP response phase

A simple response from the server contains the following components:

- HTTP Status Code (For example HTTP/1.1 301 Moved Permanently, means the requested resource was permanently moved and redirecting to some other resource).

- Response Header fields (Example – Content-Type: html)

- An empty line.

- A response body which is optional.

All the lines in the server response should end with a carriage return and line feed. Similar to request, the empty line in a response also should only have carriage return and line feed without any spaces

[Further reading: <https://code.tutsplus.com/tutorials/http-headers-for-dummies--net-8039>]

HTTP Request Structure

method	path	protocol
GET	/tutorials/other/top-20-mysql-best-practices/	HTTP/1.1
Host: net.tutsplus.com		
User-Agent: Mozilla/5.0 (Windows; U; Windows NT 6.1; en-US; rv:1.9.1		
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=		
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: 300		
Connection: keep-alive		
Cookie: PHPSESSID=r2t5uvjq435r4q7ib3vtdjq120		
Pragma: no-cache		
Cache-Control: no-cache		

protocol	status code
HTTP/1.1	200 OK
Transfer-Encoding: chunked	
Date: Sat, 28 Nov 2009 04:36:25 GMT	
Server: LiteSpeed	
Connection: close	
X-Powered-By: W3 Total Cache/0.8	
Pragma: public	
Expires: Sat, 28 Nov 2009 05:36:25 GMT	
Etag: "pub1259380237;gz"	
Cache-Control: max-age=3600, public	
Content-Type: text/html; charset=UTF-8	
Last-Modified: Sat, 28 Nov 2009 03:50:37 GMT	
X-Pingback: http://net.tutsplus.com/xmlrpc.php	
Content-Encoding: gzip	
Vary: Accept-Encoding, Cookie, User-Agent	

Figure 1: HTTP response structure

MIME (Multipurpose Internet Mail Extensions)

- A Multipurpose Internet Mail Extensions (MIME) type is a standard that **indicates the nature and format of a document**, file, or assortment of bytes.
- Browsers use the MIME type, not the file extension, to determine how to process a URL — it is important that servers send the correct MIME type in the response's Content-Type header.
- General structure is **type/subtype**.
- MIME (Multi-Purpose Internet Mail Extensions) is an extension of the original Internet email protocol that lets people use the protocol to exchange different kinds of data files on the Internet: audio, video, images, application programs, and other kinds, as well as the ASCII text handled in the original protocol, the Simple Mail Transport Protocol (SMTP).
- Servers insert the MIME header at the beginning of any Web transmission. Clients use this header to select an appropriate "player" application for the type of data the header indicates. Some of these players are built into the Web client or browser (for example, all browsers come with GIF and JPEG image players as well as the ability to handle HTML files); other players may need to be downloaded.
- Examples: text/plain
text/html
text/javascript
text/css
image/jpeg

Protocols

1. IMAP

- The **Internet Message Access Protocol (IMAP)** is a mail protocol used for accessing email on a remote web server from a local client.
- IMAP and POP3 are the two most commonly used Internet mail protocols **for retrieving emails**.
- IMAP allows simultaneous access by multiple clients.
- Some of IMAP benefits include the ability to delete messages, search for keywords in the body of emails, create and manage multiple mailboxes or folders, and view the headings for easy visual scans of emails.
- Port 143 - this is the default IMAP non-encrypted port
- Port 993 - this is the port you need to use if you want to connect using IMAP securely

2. POP3

- **Post Office Protocol version 3 (POP3)** is a standard mail protocol used to receive emails from a remote server to a local email client.
- POP3 allows you to download email messages on your local computer and read them even when you are offline.
- POP3 is designed to delete mail on the server as soon as the user has downloaded it.
- POP can be thought of as a "store-and-forward" service.
- With POP3, your messages are stored on your local computer, which reduces the space your email account uses on your web server.
- Port 110 - this is the default POP3 non-encrypted port

- Port 995 - this is the port you need to use if you want to connect using POP3 securely.

3. SMTP

- **Simple Mail Transfer Protocol (SMTP)** is the standard protocol for sending and receiving emails across the Internet.
- SMTP is created and maintained by the Internet Engineering Task Force (IETF).
- SMTP is generally integrated within an email client application and is composed of four key components:
 - Local user or client-end utility known as the mail user agent (MUA)
 - Server known as mail submission agent (MSA)
 - Mail transfer agent (MTA)
 - Mail delivery agent (MDA)
- SMTP works by initiating a session between the user and server, whereas MTA and MDA provide domain searching and local delivery services.
- Port 25 - this is the default SMTP non-encrypted port.
- Port 465 - this is the port used if you want to send messages using SMTP securely.

4. WAP

- **Wireless Application Protocol (WAP)** is a technical standard for accessing information over a mobile wireless network.
- A WAP browser is a web browser for mobile devices such as mobile phones that uses the protocol.
- WAP Protocol Suite consists of
 - Wireless Application Environment (WAE)
 - Wireless Session Protocol (WSP)
 - Wireless Transaction Protocol (WTP)
 - Wireless Transport Layer Security (WTLS)
 - Wireless Datagram Protocol (WDP)
- WAP Push was incorporated into the specification to allow WAP content to be pushed to the mobile handset with minimum user intervention.

5. FTP

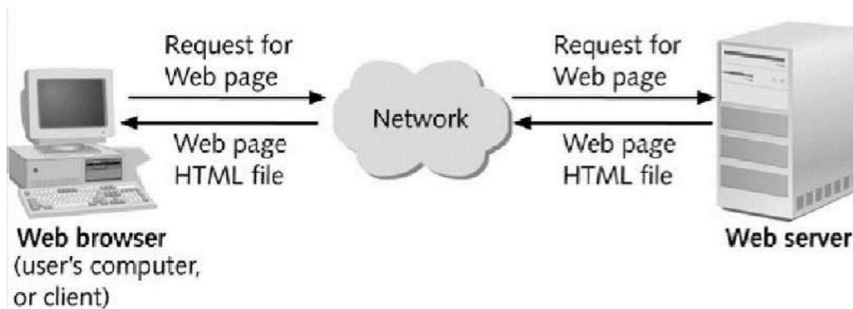
- The **File Transfer Protocol (FTP)** is a standard network protocol used for the transfer of computer files between a client and server on a computer network.
- Transferring files from a client computer to a server computer is called "uploading" and transferring from a server to a client is "downloading".
- FTP is the file transfer protocol in the Internet's TCP/IP protocol suite's Application Layer.
- FTP operates in a client/server environment.
- An FTP Client is software(installed in client's computer) that is designed to move files back-and-forth between two computers over the Internet.
- FTP allows files to have ownership and access restrictions.
- FTP systems generally encode and transmit their data in binary sets which allow for faster data transfer
- FTP utilizes two ports, a 'data' port and a 'command' port (also known as the control port). Traditionally these are port 21 for the command port and port 20 for the data port.

WEB ARCHITECTURE

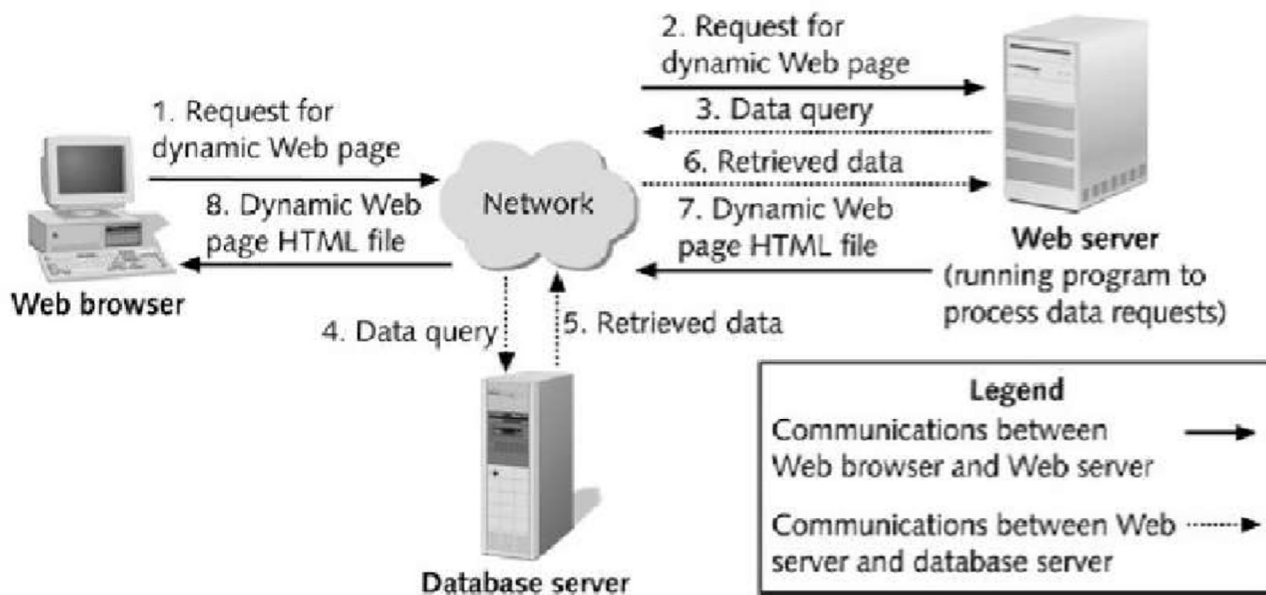
- Web Architecture can be defined as the conceptual structure of the internet.
- Types of web architecture include the client-server model and three-tier model.

Client-server (two tier) architecture

- Web browser display content that return from Web Server
- Web server provide resource for client



Three tier web architecture



Web Standards

- Web standards are rules and guidelines established by the World Wide Web Consortium (W3C) developed to promote consistency in the design code which makes up a web page.
- simply it's the guideline for the mark-up language which determines how a web page displays in a visitor's browser window.
- The World Wide Web Consortium (W3C) is an international community where Member organizations, full-time staff and the general public work together to develop Web standards.

- The best known and widely used Web standards are: HTML (HyperText Markup Language), XML (eXtensible Markup Language) and CSS (Cascading Style Sheets).

Domain name registration process

- First, come up with an idea for your **website name**(domain name).
- After choosing the perfect name, **check its availability**. You can do it via **name generators** or on different **registrar's sites**.
- Register the domain name via chosen **registrar or a reseller**. You'll need to use a bank card for this, so make sure you have it.
- After the registration, the **registry** creates a **WHOIS record**. The registrant may choose to remain anonymous, for example, using WhoIsGuard services. Generic information then will be displayed instead of private registrant's data. The registry also sends domain zone files to master servers.

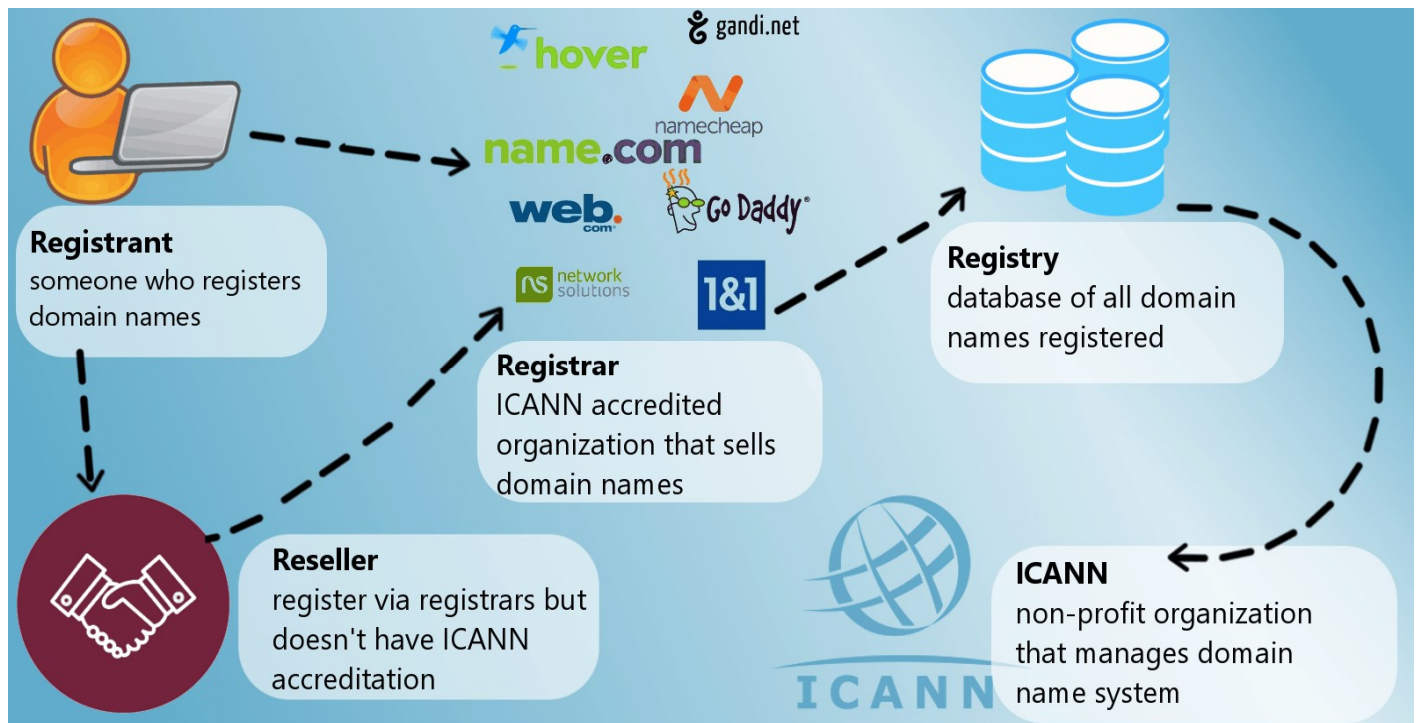


Figure 2:A simple explanation on how domain name registration works.

(source: <https://www.rankapage.com/what-is-domain-name-registrar/>)