

CHAPTER 1

Fundamentals of Web Technology

What is web technology?

- We probably know that computers don't communicate with each other the way that people do.
- That means computers require codes, or directions. These binary codes and commands allow computers to process needed information.
- So to establish communication between computers Web Technology was invented.
- Web Technology refers to the methods that communicate computers with each other through the use of markup language and multimedia packages.**

Markup Language:

- A markup language is a computer language that uses tags to define elements within document.**
- It is human-readable and markup files contain standard words , rather than typical programming syntax.
- Among several markup languages HTML and XML are most popular.
- HTML is used for creating webpages.
- XML is used for storing structured data , rather than formatting information on a page.

Multimedia Packages:

forms such as text, audio, images, animations, video and interactive content.

- Multimedia contrasts with media that use only rudimentary computer displays such as text-only or traditional forms of printed or hand-produced material.
- Multimedia devices are electronic media devices used to store and experience multimedia content.

Internet:

- The concept of a new large-scale computer network was first developed by the U.S Department of Defense (DoD), in the 1960s.
- The main purpose of this network were communications , program sharing , and remote computer access for researchers working on defense-related contracts.

- NSFNET was decommissioned in 1995, removing the last restrictions on the use of the Internet to carry commercial traffic.
- That means, internet is a globally connected network system that uses TCP/IP to transmit data via various types of media.**
- The internet is a network of global exchanges – including private, public, business, academic and government networks – connected by guided, wireless and fiber-optic technologies.
- The terms **internet** and **World Wide Web** are often used interchangeably, but they are not exactly the same thing; the internet refers to the global communication system, including hardware and infrastructure, while the web is one of the services communicated over the internet.
- Internet Protocol (IP) is the internet's primary component and communications backbone.
- IP is used to allocate the devices that are connected to the network.

WWW(World Wide Web):

- The World Wide Web ("WWW" or simply the "Web") is a global information medium which users can read and write via computers connected to the Internet.
- The **World Wide Web (WWW)**, is an information space where documents and other web resources are identified by Uniform Resource Locators (URLs), interlinked by hypertext links, and accessible via the Internet.
- An information space is a type of information design in which representations of information objects are situated in a principled space.
- A new protocol for the Internet , as well as a system of document access to use it was proposed by a small group of people led by Tim Berners-Lee at CERN(Conseil Europeen pour la Recherche Nucleaire) , In 1989.
- The World Wide Web has been central to the development of the Information Age and is the primary tool billions of people use to interact on the Internet.
- Hypertext is the text that is specially coded using a standard coding language known as HyperText Markup Language.

Web Browsers:

- In some network , when two computers communicate , in many cases one acts as client and the other as a server.
- First the communication initiates by client , which is often a request for information stored on the server which sends information back to the clients.

- Browsers are programs running on client machines where documents are requested to server.
- A **web browser** (commonly referred to as a **browser**) is a software application for accessing information on the World Wide Web.
- They are called browsers because they allow the user to browse the resources available on the servers.
- The first browsers were text based – they were not capable of displaying graphic information nor did they have a graphical user interface.

Web Servers:

- Web servers are programs that provide documents to requesting browsers identified using URLs.
- Servers are slave Programs : They act only when requests are made to them by browser running on others computers on the Internet.
- The most commonly used Web servers are Apache , which runs under Windows operating systems.

1. Web Server Operation:

- First, Web browsers initiate network communications with servers by sending URLs.
- One of two different things can be specified by URL: the address of data file stored on the server or a stored program on the server that clients wants executed and the output of the program returned to the client.
- The standard Web protocol HTTP(Hypertext Transfer Protocol), is responsible for all the communication between Web client and Web servers.
- When Web server begins execution , it informs the operating system under which it is running that it is now ready to accept incoming network connections through a specific port on the machine.
- In this running state , the server runs as a background process in the operating system.
- Then a network connection is established between Web client or browser and Web server.
- And sends information requests and possibly data to the server , receives information from the server and close the connection.

2. General Server Characteristics:

- Document root
The file structure of the serer has two separate directories. The root of one of this is document root.
- Server root
This is another root directory of web server file structure.
- Virtual document tree

Many servers allow a part of the servable document stored outside of the document root this is called virtual document tree.

Virtual hosts

Many servers can support more than one site on a computer potentially reducing the cost of each site and makes their maintenance more convenient. such secondary hosts are called virtual hosts.

Proxy server

Some servers can serve documents that are in the document root of the other machines are on the web are called proxy servers.

Web servers not only support HTTP but also ftp, Gopher, News and mail.

URL(Uniform Resource Locators)

- Uniform (or Universal) Resource Locators (URLs) are used to identify document (resources) on the Internet.
- Due to the URL address, the user gets information about where the needed information is located.
- It can refer to the website, some particular document, or an image.
- It is also referred to as a web address.
- URLs consist of multiple parts -- including a protocol and domain name -- that tell a web browser how and where to retrieve a resource.
- End users use URLs by typing them directly into the address bar of a browser or by clicking a hyperlink found on a webpage, bookmark list, in an email or from another application.
- URLs are strings of characters used to identify a resource over a network.

1. URL structure

- The URL contains the name of the protocol needed to access a resource, as well as a resource name.
 - The first part of a URL identifies what protocol to use as the primary access medium.
 - The second part identifies the IP address or domain name -- and possibly subdomain -- where the resource is located.

DOMAIN

QUERY STRING

<https://sitechecker.pro/knowledge-base/?name=article&topic=seo#top>

PROTOCOL

PATH

FRAGMENT

figure: Example of URLs

Optionally, after the domain, a URL can also specify:

- a path to a specific page or file within a domain;
- a network port to use to make the connection;
- a specific reference point within a file, such as a named anchor in an HTML file; and
- a query or search parameters used -- commonly found in URLs for search results.

MIME(Multipurpose Internet Mail Extensions)

- A browser needs some way of determining the formats of document it receives from a Web servers.
- It is not possible to render documents without knowing the formats of the document because different document formats require different rendering software.
- MIME** was developed to specify the format of different kinds of documents to be sent via Internet mail.
- MIME (Multi-Purpose Internet Mail Extensions)** is an extension of the original Internet e-mail 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.
- A MIME type consists of a **type** and a **subtype** — two strings separated by /.
- The *type* represents the category and can be a *discrete* or a *multipart* type. The *subtype* is specific to each type.

Discrete types indicate the category of the document. They can be one of the following table:

Type	Description	Example of typical subtypes
text	Any document that contains text and is theoretically human readable	text/plain, text/markdown
image	Any kind of image. Videos are not included, though animated images (like animated GIF) are described with an image type.	image/gif, image/jpeg, image/bmp, image/vnd.microsoft.icon
audio	Any kind of audio file	audio/midi, audio/webm, audio/wav
video	Any kind of video file	video/ogg, video/webm

application	Any kind of binary data, especially data that will be executed or interpreted somehow.	application/javascript, application/octet-stream, application/pkcs12, application/vnd.ms-powerpoint, application/xml,application/pdf, application/xml, application/xhtml+xml
-------------	--	--

HTTP(Hypertext Transfer Protocol)

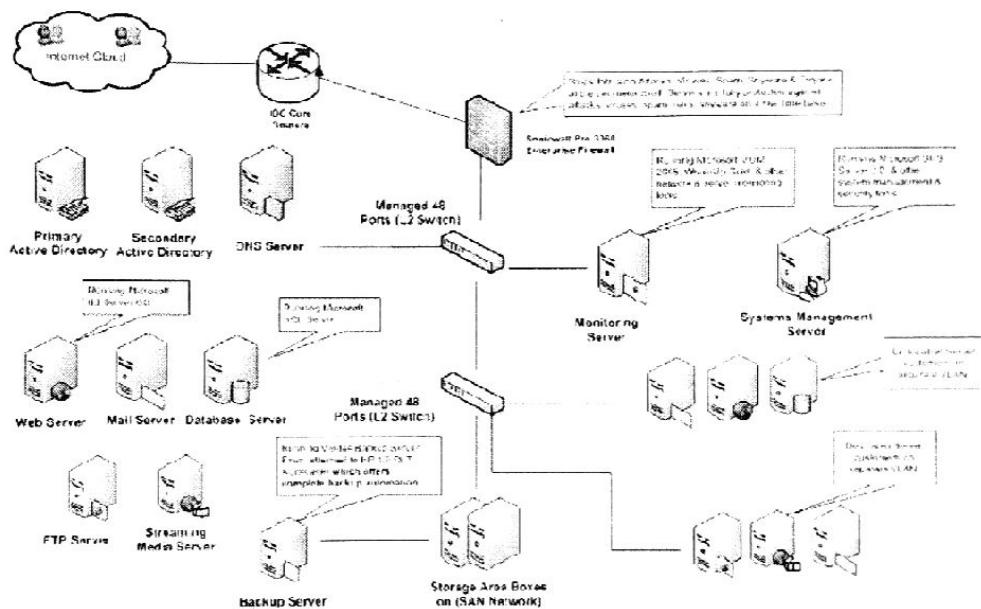
- HTTP is the protocol that is used for transferring hypertext (i.e text ,graphic , audio, video ,animation etc) between two computers and is particularly used on the World Wide Web.
- It is a TCP/IP based communication protocol and provides a standard for Web browsers and servers to communicate.
- Where as Hypertext is the text that is specially coded using a standard coding language called Hypertext Markup Language(HTML) which basically creates hyperlinks and thereby controls how the World Wide Web works and how Webpages are formatted and displayed.
- These hyperlinks can be in the form of text, graphic , image ,sound or video and are used to “link” the user to some other file.
- HTTP defines how message formatted and transmitted and what actions Web servers and browsers should take in response to various commands .
- For example , when you enter a URL in your browser , this actually sends an HTTP command to the Web server directing it to fetch and transmit the requested Webpage.
- HTTP is based on Client/Server principle .
- Communication between the host and the client occurs through a request/response pair.
 - A connection is established between two computers- out of which one is client (generally the browser) that initiates the request and the other is the server that responds to the request.
- Also, HTTP identifies the resource that the client has requested for and informs the server about the action to be taken.
- When the user clicks on the hypertext link ,the client program on their computer uses HTTP to contact the server , identify the resource and ask the server to respond with an action.

HTTP has three important features.

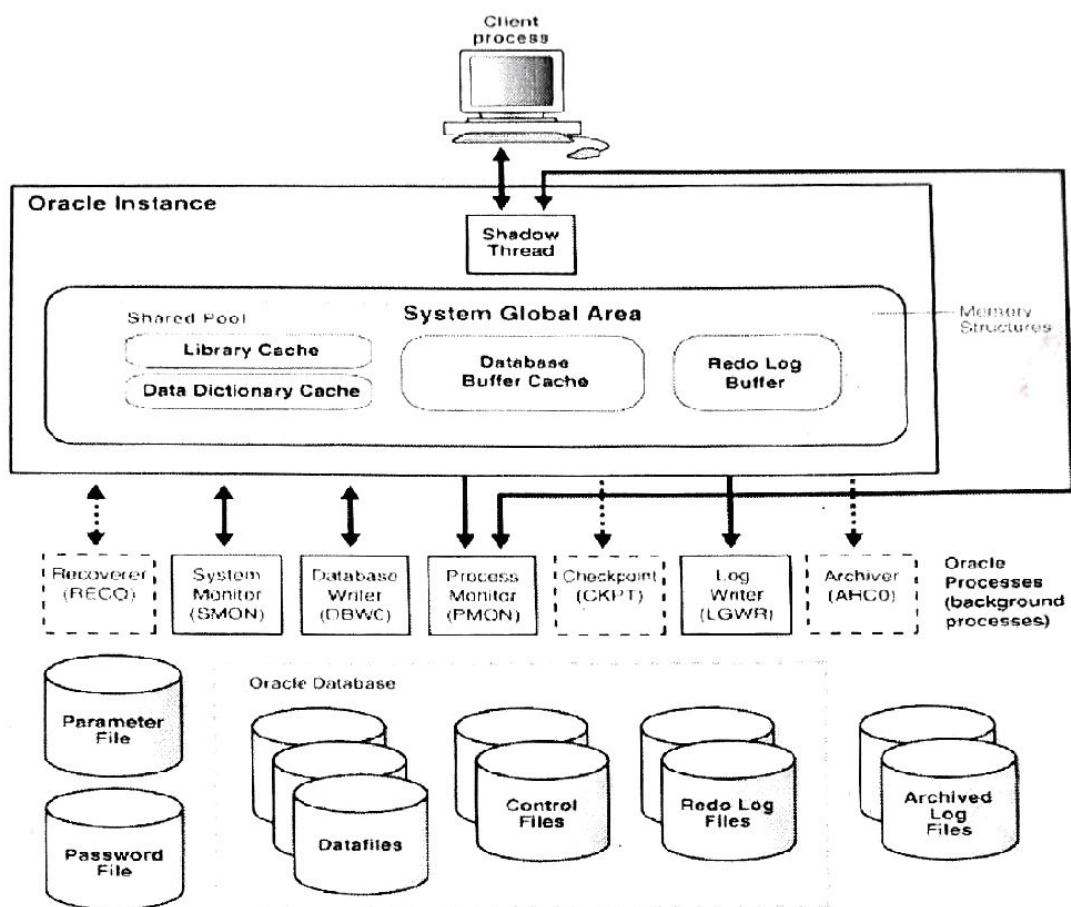
- **Firstly**, it is connectionless .After a request is made , the client disconnects from the server and waits for a response. To process the request , the server has to re-establish the connection with the client .
- **Secondly**, HTTP is media independent. This means any type of data(text , audio, video, etc) can be sent by HTTP as long as both the client and server know how to handle the data content.
- **Thirdly**, HTTP is stateless. This is because the server and the client are aware of each other only during request. Afterwards , they get disconnected.

Web Architecture

- Before we can get to define what **web architecture** is, first it is necessary to frame it properly.
- There are some roles related to architecture within the Information Technology industry, but basically we can do the following division:
 - **System architects:** combine different hardware components (computers and other devices) with software elements (operating systems) to build systems capable of providing the resources needed by applications or services which run on top of them.



- **Data architects:** design how information handled by applications is structured using databases.



- **Storage Architects:** design storage area networks (SAN) that allow to store all the information generated by applications running on different information systems.
- **Networking Architects:** plan and design communication networks that allow data exchange between different information systems.
- **Software Architects:** more typically known as **software engineers**, these professionals design and build applications that provide services needed by users of information systems or other applications.

Web architecture determines how software logic is divided between a server and a client, as well as how these two components communicate with each other. The ways web architecture types work are quite different and depend on a client model and requests a client sends/receives to/from a server.

Depending on the needs of your web site or a web app, you choose the web architecture that would work best for it. Some most conceived part of web architect are:

- **Designing the user interface** of web applications. In the case of a website, it would be the design of the website itself: look & feel

(colors, images, fonts used, etc.), page layout, menu structure, visual elements, etc.

- **Design and implementation of application logic**, i.e. the set of capabilities which it provides, such as user-entered **data processing, results calculation** from different input data, **algorithms designing and implementation, manipulation of information** stored in databases, **executing different actions** as a result of compliance with some given conditions or event triggering, etc. i.e. planning and designing of what later is implemented using one or more **programming languages**.
- **Designing of information architecture**, i.e. determine the real-world information that the application will be able to handle, design a **conceptual model** which is an accurate reflection of the real world with its different **entities and relationships** between those entities, determine the **data model** which best suits that conceptual model, implement the data model on a particular **database engine** and move the necessary information to it for the application proper functioning.

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 (Wireless Application Protocol) is a specification for a set of communication protocols to standardize the way that wireless devices, such as cellular telephones and radio transceivers, can be used for Internet access, including e-mail, the World Wide Web, newsgroups, and instant messaging.
- Instant messaging , is the exchange of near real-time messages through a stand-alone application or embedded software.
- WAP only access text rather than images or video or audio etc, that means WAP can provides us graphic interface.

The WAP layers are:

- Wireless Application Environment (WAE)
- Wireless Session Layer (WSL)
- Wireless Transport Layer Security (WTLS)
- Wireless Transport Layer (WTP)

The WAP was conceived by four companies: Ericsson, Motorola, Nokia, and Unwired Planet (now Phone.com). The Wireless Markup Language (WML) is used to create pages that can be delivered using WAP.

Web standards

- Web standards are the formal, non-proprietary standards and other technical specifications that define and describe aspects of the WWW.
- 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.

The advantages in adhering to these standards are many:

- Web pages will display in a wide variety of browsers and computers, including new technology.
- W3C standards promote the use of “Cascading Style Sheets”(CSS) or design code which is attached to the web page. The use of style sheets significantly reduces the page file size.
- Design features such as colors and fonts can be easily changed by just modifying one style sheet instead of editing every individual page in a website, reducing the costs to modify site.
- Search engines are able to access and index pages designed to web standards with greater efficiency.

Email Protocols

SMTP(Simple Mail Transfer Protocol):

- Simple Mail Transfer Protocol (SMTP) is the standard protocol for **sending emails** across the Internet.
 - Most e-mail systems that send mail over the Internet use SMTP to send messages from one server to another; the messages can then be retrieved with an e-mail client using either POP or IMAP.
 - Most e-mail systems that send mail over the Internet use SMTP to send messages from one server to another; the messages can then be retrieved with an e-mail client using either POP or IMAP.

By default, the SMTP protocol works on three ports:

- Port 25** - this is the default SMTP non-encrypted port

- Port 2525** - this port is opened on all Site Ground servers in case port 25 is filtered (by your ISP for example) and you want to send non-encrypted emails with SMTP
- Port 465** - this is the port used if you want to send messages using SMTP securely

POP(Post Office Protocol)

- POP is a protocol that is used to retrieve mail from mail server.
- POP access mails that are only located in inbox.
- 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 and removed from the email server.
- This means if you want to access your mail from multiple location , you can't.
- That means POP do not allow multiple user.

By default, the POP3 protocol works on two ports:

- 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

IMAP(Internet Message Access Protocol)

- The **Internet Message Access Protocol (IMAP)** is an Internet standard protocol used by email clients to retrieve email messages from a mail server over a TCP/IP connection.
- Stores email messages on a mail server, but allows the end user to view and manipulate the messages as though they were stored locally on the end user's computing device(s).
- This allows users to organize messages into folders, have multiple client applications know which messages have been read, flag messages for urgency or follow-up and save draft messages on the server.
- Most implementations of IMAP support multiple logins; this allows the end user to simultaneously connect to the email server with different devices.

By default, the IMAP protocol works on two ports:

- 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

FTP (File Transfer Protocol)

- 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.
- FTP is built on a client-server model architecture using separate control and data connections between the client and the server.
- FTP uses the Internet's TCP/IP protocols to enable data transfer.
- FTP promotes sharing of files via remote computers with reliable and efficient data transfer.
- Files can be transferred between two computers using FTP software.
- The user's computer is called the local host machine and is connected to the Internet.
- The second machine, called the remote host, is also running FTP software and connected to the Internet.
- A user typically needs to log on to the FTP server, although some servers make some or all of their content available without login, also known as anonymous FTP.
- FTP is less secure protocol because it uses plain text format when transferring files.

Web Hosting

- Web hosting is a service that allows organizations and individuals to post a website or web page onto the Internet.
- Web hosts are companies that provide space on a server owned or leased for use by clients, as well as providing Internet connectivity, typically in a data center.
- Websites are hosted, or stored, on special computers called servers.

Types of hosting

- **Shared web hosting service** refers to a web hosting service where many websites reside on one web server connected to the Internet.
- **Reseller hosting** is a form of web hosting wherein the account owner has the ability to use his or her allotted hard drive space and bandwidth to host websites on behalf of third parties.
- **A virtual private server (VPS)** is a virtual machine sold as a service by an Internet hosting service.
- **A dedicated hosting service, dedicated server, or managed hosting service** is a type of Internet hosting in which the client leases an entire server not shared with anyone else.
- **A colocation centre** (also spelled co-location, or **colo**) or "**carrier hotel**", is a type of data centre where equipment, space, and bandwidth are available for rental to retail customers.

- **Cloud computing** is shared pools of configurable computer system resources and higher-level services that can be rapidly provisioned with minimal management effort, often over the Internet.

Assignment 1:

1. Write short note about HTTP.
2. Explain Domain name and its hierarchy.
3. Write about domain name registration process.

Additional Topics:

Free source software:

- The Free Software Definition written by Richard Stallman and published by Free Software Foundation (FSF), defines free software as being software that ensures that the end users have freedom in using, studying, sharing and modifying that software.
- Free and open-source software (FOSS) is software that can be classified as both free software and open-source software.
- That is, anyone is freely licensed to use, copy, study, and change the software in any way, and the source code is openly shared so that people are encouraged to voluntarily improve the design of the software.

Unlike the Open Source term, Free Software only has 4 “Freedoms” with its definition and are numbered 0-3:

1. The freedom to run the program for any purpose (Freedom 0)
2. The freedom to study how the program works and adapt it to your needs (Freedom 1)
3. The freedom of redistribution of software (Freedom 2)
4. The freedom to improve the program and release your improvements to the public to benefit the while community. (Freedom 3)

Open source software:

- Open-source software (OSS) is computer software distributed with its source code available for modification.
- The software usually includes a license for programmers to change the software in any way they choose.
- They can fix bugs, improve functions, or adapt the software to suit their own needs.
- The Open Source Initiative (OSI) is a leading authority on OSS
- For software to be considered “Open Source” it must meet ten conditions. Among these ten conditions the first three that are really at the core of Open Source and differentiates it from other software.
 - Free Redistribution: the software can be freely given away or sold.
 - Source Code: the source code must either be included or freely obtainable.
 - Derived Works: redistribution of modifications must be allowed.

OSI: Open Source Initiative (OSI) is a non-profit organization dedicated to promoting open-source software.

Advantages of Open-Source Software

While cost is a driving factor, OSS has several additional benefits:

- High-quality results when the source code is passed around, tested and fixed.
- It is a valuable learning opportunity for programmers. They can learn and apply skills to the most popular programs available today.
- Many consider open-source software more secure than proprietary software because bugs are identified and fixed quickly.

Most of the software is free. Costs may arise later, however, such as subscriptions or support fees.

Popular Types of Open-Source Software

- Mozilla's Firefox web browser
- Thunderbird email client
- PHP scripting language
- Python programming language
- Apache HTTP web server
- database system

Proprietary Software:

- Proprietary software is software that is owned by an individual or a company (usually the one that developed it).
- There are almost always major restrictions on its use, and its source code is almost always kept secret.
- Source code is the form in which a program is originally written by a human using a programming language and prior to being converted to machine code which is directly readable by a computer's CPU (central processing unit).
- It is necessary to have the source code in order to be able to modify or improve a program.
- Virtually all Microsoft software is proprietary, including the Windows family of operating systems and Microsoft Office.
- This includes software that is given away at no charge, such as Internet Explorer.
- Other major producers of proprietary software include Adobe, Borland, IBM, Macromedia, Sun Microsystems and Oracle.

Advantages:

- One advantage to using a proprietary-software system is that you will generally be able to take advantage of the software company's customer service department for troubleshooting and setup purposes.

Disadvantage:

- These types of software are costly and not always needed.