

# Module - 1

## Internet Standards

Internet Standards refer to all the documented requirements both in technology as well as methodology pertaining to the Internet.

### The organizations of Internet Standards are

**Internet Engineering Task Force (IETF)** IETF formulates, publishes and regulates Internet Standards, particularly those related to TCP/IP. The organization is open standard, with no formal memberships. Development of IETF standards is open to all. Any interested person can participate for their development. IETF documents are free and easily available over the Internet. IETF specifications are on individual protocols that may be used in different systems.

### WWW

World Wide Web, which is also known as a Web, is a collection of websites or web pages stored in web servers and connected to local computers through the internet. These websites contain text pages, digital images, audios, videos, etc. Users can access the content of these sites from any part of the world over the internet using their devices such as computers, laptops, cell phones, etc. The WWW, along with internet, enables the retrieval and display of text and media to your device.



The building blocks of the Web are web pages which are formatted in HTML and connected by links called "hypertext" or hyperlinks and accessed by HTTP. These links are electronic connections that link related pieces of information so that users can access the desired information quickly. Hypertext offers the advantage to select a word or phrase from text and thus to access other pages that provide additional information related to that word or phrase.

A web page is given an online address called a Uniform Resource Locator (URL). A particular collection of web pages that belong to a specific URL is called a website, e.g., *www.facebook.com*, *www.google.com*, etc. So, the World Wide Web is like a huge electronic book whose pages are stored on multiple servers across the world.

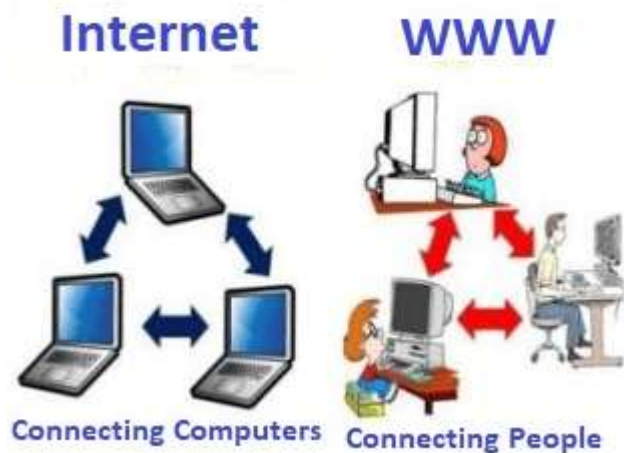
Small websites store all of their WebPages on a single server, but big websites or organizations place their WebPages on different servers in different countries so that when users of a country search their site they could get the information quickly from the nearest server.

So, the web provides a communication platform for users to retrieve and exchange information over the internet. Unlike a book, where we move from one page to another in a sequence, on World Wide Web we follow a web of hypertext links to visit a web page and from that web

page to move to other web pages. You need a browser, which is installed on your computer, to access the Web.

### **Difference between World Wide Web and Internet:**

Some people use the terms 'internet' and 'World Wide Web' interchangeably. They think they are the same thing, but it is not so. Internet is entirely different from WWW. It is a worldwide network of devices like computers, laptops, tablets, etc. It enables users to send emails to other users and chat with them online. For example, when you send an email or chatting with someone online, you are using the internet.



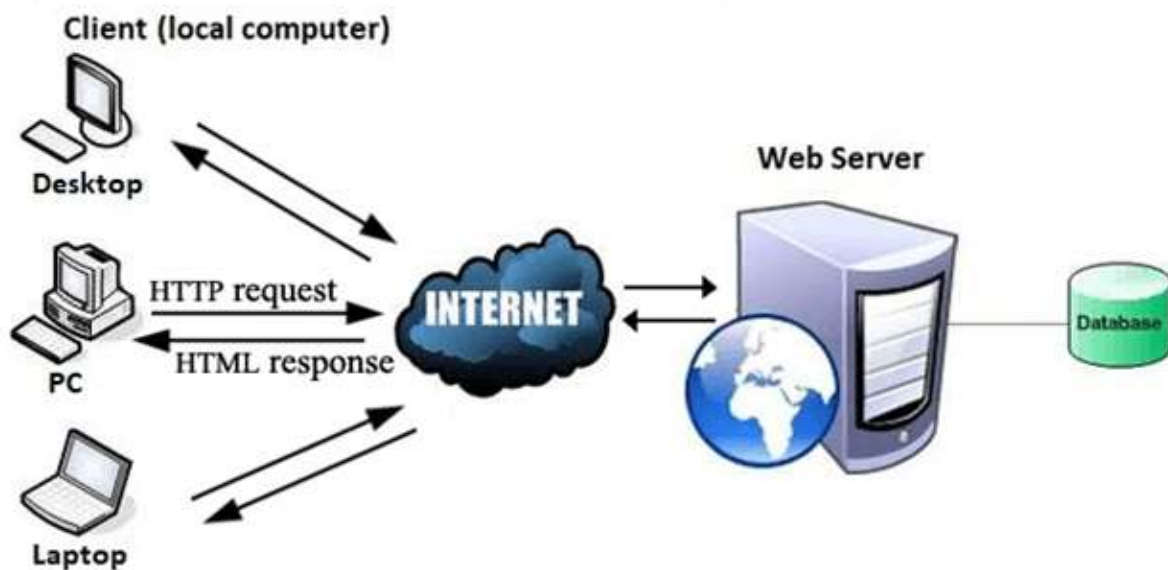
But, when you have opened a website like google.com for information, you are using the World Wide Web; a network of servers over the internet. You request a webpage from your computer using a browser, and the server renders that page to your browser. Your computer is called a client who runs a program (web browser), and asks the other computer (server) for the information it needs.

WWW is a collection of websites connected to the internet so that people can search and share information

The World Wide Web was invented by a British scientist, Tim Berners-Lee in 1989

### **How the World Wide Web Works?**

The Web works as per the internet's basic client-server format as shown in the following image. The servers store and transfer web pages or information to user's computers on the network when requested by the users. A web server is a software program which serves the web pages requested by web users using a browser. The computer of a user who requests documents from a server is known as a client. Browser, which is installed on the user's computer, allows users to view the retrieved documents.



The moment you open the browser and type a URL in the address bar or search something on Google, the WWW starts working. There are three main technologies involved in transferring information (web pages) from servers to clients (computers of users). These technologies include Hypertext Markup Language (HTML), Hypertext Transfer Protocol (HTTP) and Web browsers.

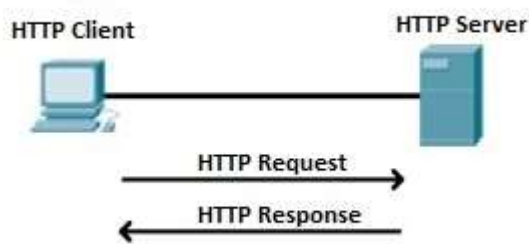
### **Hypertext Markup Language (HTML):**



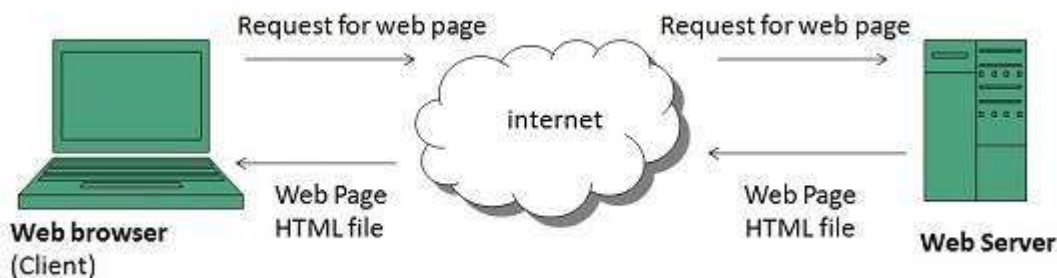
HTML is a standard markup language which is used for creating web pages. It describes the structure of web pages through HTML elements or tags. These tags are used to organize the pieces of content such as 'heading,' 'paragraph,' 'table,' 'Image,' and more. You don't see HTML tags when you open a webpage as browsers don't display the tags and use them only to render the content of a web page. In simple words, HTML is used to display text, images, and other resources through a Web browser.

### **Hypertext Transfer Protocol (HTTP):**

Hyper Text Transfer Protocol (HTTP) is an application layer protocol which enables WWW to work smoothly and effectively. It is based on a client-server model. The client is a web browser which communicates with the web server which hosts the website. This protocol defines how messages are formatted and transmitted and what actions the Web Server and browser should take in response to different commands. When you enter a URL in the browser, an HTTP command is sent to the Web server, and it transmits the requested Web Page.



When we open a website using a browser, a connection to the web server is opened, and the browser communicates with the server through HTTP and sends a request. HTTP is carried over TCP/IP to communicate with the server. The server processes the browser's request and sends a response, and then the connection is closed. Thus, the browser retrieves content from the server for the user.



**HTML (HyperText Markup Language)**—a special type of computer language called a markup language designed to specify the content and structure of web pages (also called documents) in a portable manner.

**Tim Berners-Lee** of CERN (the European Organization for Nuclear Research) began to develop a technology for sharing information via hyperlinked text documents. Berners-Lee called his invention the HyperText Markup Language (HTML). He also wrote communication protocols to form the backbone of his new information system, which he called the World Wide Web. In particular, he wrote the **Hypertext Transfer Protocol (HTTP)**—a communications protocol used to send information over the web. The **URL (Uniform Resource Locator)** specifies the address (i.e., location) of the web page displayed in the browser window. Each web page on the Internet is associated with a unique URL. URLs usually begin with `http://`. HTML documents normally contain **hyperlinks**, which, when clicked, load a specified web document. Both images and text may be hyperlinked. When the user clicks a hyperlink, a web server locates the requested web page and sends it to the user's web browser.

URIs (Uniform Resource Identifiers) identify resources on the Internet. URIs that start with `http://` are called URLs (Uniform Resource Locators). Common URLs refer to files, directories or server-side code that performs tasks such as database lookups, Internet searches and business-application processing.

**HTTP get and post Requests**

The two most common HTTP request types (also known as request methods) are get and post. A get request typically gets (or retrieves) information from a server, such as an HTML document, an image or search results based on a user-submitted search term. A post request typically posts (or sends) data to a server. Common uses of post requests are to send form data or documents to a server.

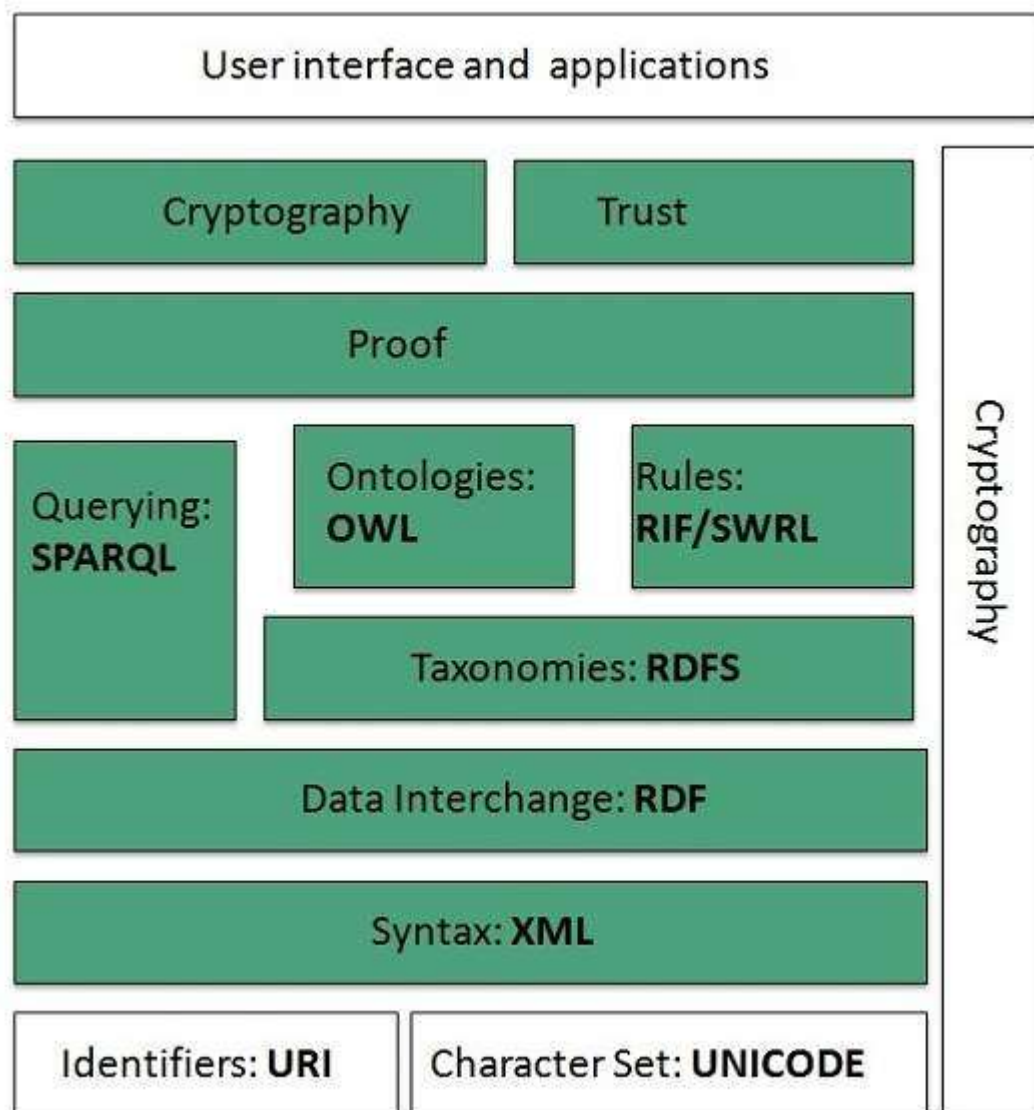
A get request appends data to the URL, e.g., [www.google.com/search?q=deitel](http://www.google.com/search?q=deitel). A post request sends form data as part of the HTTP message, not as part of the URL. The post method is also sometimes preferred because it hides the submitted data from the user by embedding it in an HTTP message.

**Client-side scripting** with JavaScript can be used to validate user input, to interact with the browser, to enhance web pages, and to add client/server communication between a browser and a web server.

Programmers have more flexibility with **server-side scripts**, which often generate custom responses for clients. For example, a client might connect to an airline's web server and request a list of flights from Boston to San Francisco between April 19 and May 5. The server queries the database, dynamically generates an HTML document containing the flight list and sends the document to the client. This technology allows clients to obtain the most current flight information from the database by connecting to an airline's web server.

### WWW Architecture

WWW architecture is divided into several layers as shown in the following diagram:



Identifiers and Character Set

**Uniform Resource Identifier (URI)** is used to uniquely identify resources on the web and **UNICODE** makes it possible to built web pages that can be read and write in human languages.

Syntax

**XML (Extensible Markup Language)** helps to define common syntax in semantic web.

Data Interchange

**Resource Description Framework (RDF)** framework helps in defining core representation of data for web. RDF represents data about resource in graph form.

Taxonomies

**RDF Schema (RDFS)** allows more standardized description of **taxonomies** and other **ontological** constructs.

Ontologies

**Web Ontology Language (OWL)** offers more constructs over RDFS. It comes in following three versions:

- OWL Lite for taxonomies and simple constraints.
- OWL DL for full description logic support.
- OWL for more syntactic freedom of RDF

Rules

**RIF** and **SWRL** offers rules beyond the constructs that are available from **RDFs** and **OWL**. Simple Protocol and **RDF Query Language (SPARQL)** is SQL like language used for querying RDF data and OWL Ontologies.

Proof

All semantic and rules that are executed at layers below Proof and their result will be used to prove deductions.

Cryptography

**Cryptography** means such as digital signature for verification of the origin of sources is used.

User Interface and Applications

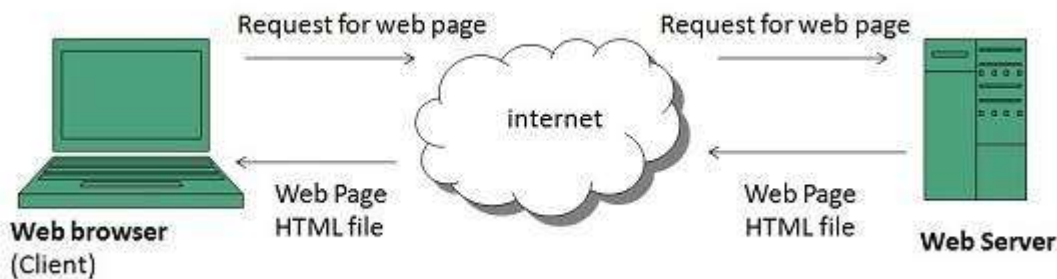
On the top of layer **User interface and Applications** layer is built for user interaction.

## WWW Operation

**WWW** works on client- server approach. Following steps explains how the web works:

1. User enters the URL (say, **http://www.tutorialspoint.com**) of the web page in the address bar of web browser.
2. Then browser requests the Domain Name Server for the IP address corresponding to **www.tutorialspoint.com**.
3. After receiving IP address, browser sends the request for web page to the web server using HTTP protocol which specifies the way the browser and web server communicates.
4. Then web server receives request using HTTP protocol and checks its search for the requested web page. If found it returns it back to the web browser and close the HTTP connection.
5. Now the web browser receives the web page, It interprets it and display the contents of web page in web browser's window.





## SMTP

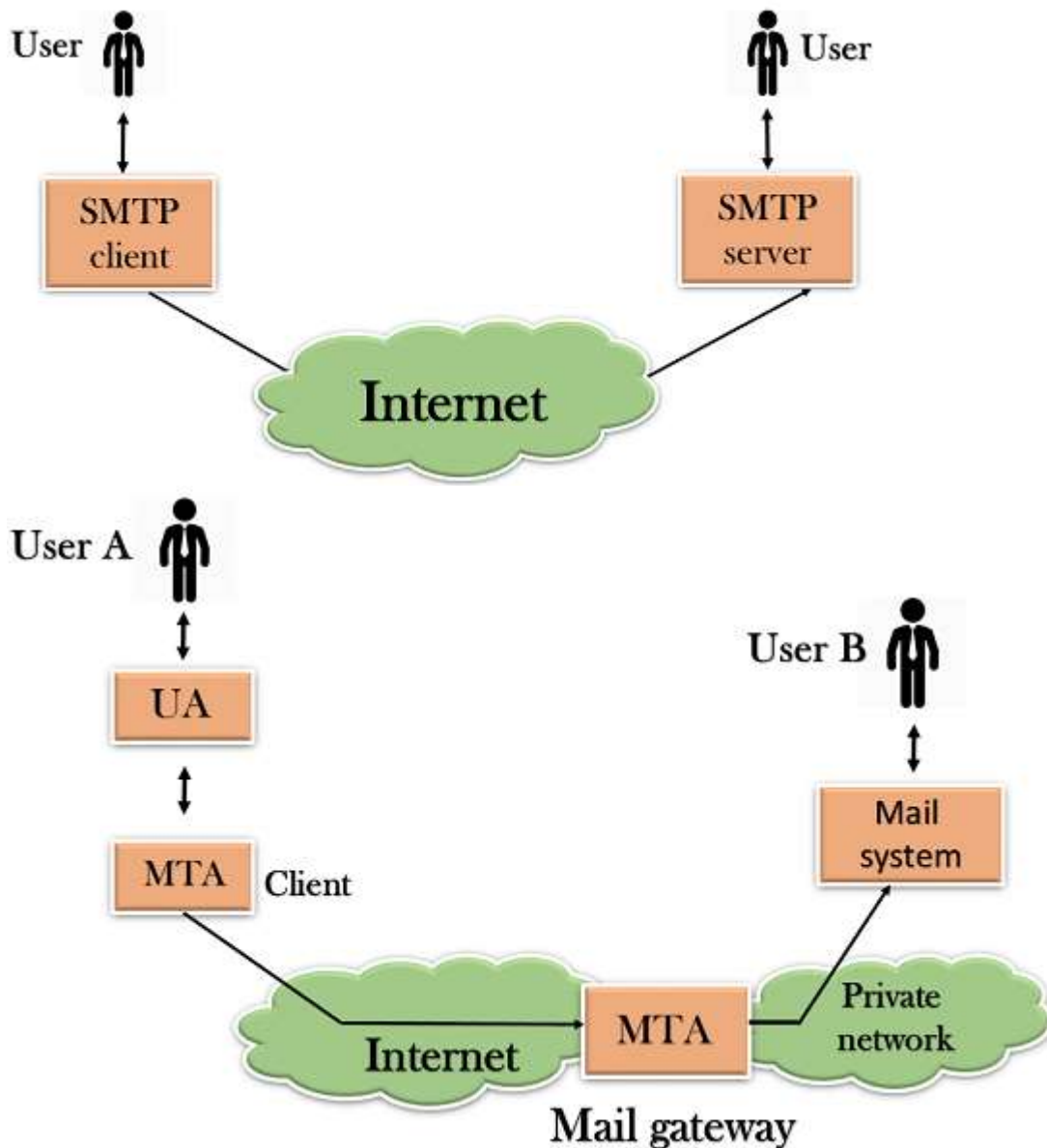
- SMTP stands for Simple Mail Transfer Protocol.
- SMTP is a set of communication guidelines that allow software to transmit an electronic mail over the internet is called **Simple Mail Transfer Protocol**.
- It is a program used for sending messages to other computer users based on e-mail addresses.
- It provides a mail exchange between users on the same or different computers, and it also supports:
  - It can send a single message to one or more recipients.
  - Sending message can include text, voice, video or graphics.
  - It can also send the messages on networks outside the internet.
- The main purpose of SMTP is used to set up communication rules between servers. The servers have a way of identifying themselves and announcing what kind of communication they are trying to perform. They also have a way of handling the errors such as incorrect email address. For example, if the recipient address is wrong, then receiving server reply with an error message of some kind.

### Components of SMTP

First, we will break the SMTP client and SMTP server into two components such as user agent (UA) and mail transfer agent (MTA). The user agent (UA) prepares the message, creates the envelope and then puts the message in the envelope. The mail transfer agent (MTA) transfers this mail across the internet.

SMTP allows a more complex system by adding a relaying system. Instead of just having one MTA at sending side and one at receiving side, more MTAs can be added, acting either as a client or server to relay the email.

The relaying system without TCP/IP protocol can also be used to send the emails to users, and this is achieved by the use of the mail gateway. The mail gateway is a relay MTA that can be used to receive an email.



### Working of SMTP

1. **Composition of Mail:** A user sends an e-mail by composing an electronic mail message using a Mail User Agent (MUA). Mail User Agent is a program which is used to send and receive mail. The message contains two parts: body and header. The body is the main part of the message while the header includes information such as the sender and recipient address. The header also includes descriptive information such as the subject of the message. In this case, the message body is like a letter and header is like an envelope that contains the recipient's address.
2. **Submission of Mail:** After composing an email, the mail client then submits the completed e-mail to the SMTP server by using SMTP on TCP port 25.
3. **Delivery of Mail:** E-mail addresses contain two parts: username of the recipient and domain name. For example, vivek@gmail.com, where "vivek" is the username of the recipient and "gmail.com" is the domain name. If the domain name of the recipient's email address is different from the sender's domain name, then MSA will send the mail to the Mail Transfer Agent (MTA). To relay the email, the MTA will find the target domain. It checks the MX record from Domain Name System to obtain the target domain. The MX record contains the domain name



and IP address of the recipient's domain. Once the record is located, MTA connects to the exchange server to relay the message.

4. **Receipt and Processing of Mail:** Once the incoming message is received, the exchange server delivers it to the incoming server (Mail Delivery Agent) which stores the e-mail where it waits for the user to retrieve it.
5. **Access and Retrieval of Mail:** The stored email in MDA can be retrieved by using MUA (Mail User Agent). MUA can be accessed by using login and password.

What is POP3 (Post Office Protocol 3)?

Post Office Protocol 3, or POP3, is the most commonly used protocol for receiving email over the internet. This standard protocol, which most email servers and their clients support, is used to receive emails from a remote server and send to a local client.

POP3 is a one-way client-server protocol in which email is received and held on the email server. The "3" refers to the third version of the original POP protocol.

A recipient or their email client can download mail periodically from the server using POP3. Thus, POP3 offers a means of downloading email from a server to the client so the recipient can view the email offline. POP3 can be thought of as a "store-and-forward" service.

Once the email is on the client, POP3 then deletes it from the server. With some implementations, users or an administrator can specify that mail be saved for some time, allowing users to download email as many times as they wish within the specified period.

POP3 and email applications

POP3 is built into most popular email clients, including Microsoft Outlook. The protocol will work provided that the email program is configured to host POP3. Each POP3 mail server has a different address that must be entered into the email program for it to connect with the protocol. Users must also enter their username and password to receive email successfully.

Additionally, since POP3 is built into standard internet browsers, including Internet Explorer and Mozilla Thunderbird, users can check their email even without an email client.

How POP3 operates

The server starts POP3 service by listening on TCP port 110. When a client wishes to use POP3 for email retrieval, it establishes a TCP connection with the server host. Once this connection is established, the POP3 server sends a greeting. At this point, the session enters the *authorization* state.

In the *transaction* state that follows, the client and server exchange commands and responses until the connection is either closed or aborted. Commands from the client consist of case-insensitive keywords, possibly followed by arguments. Responses from the server consist of a status indicator and a keyword, which may be followed by additional information.

When the client issues the *quit* command, the session enters the *update* state. The POP3 server releases any resources acquired during the *transaction* state, and says "goodbye," which is when the TCP connection is closed.

After the POP3 session enters the *update* state, the POP3 server deletes the message.

SMTP sends email, a mail handler receives it on the recipient's behalf, and the mail is read using POP3

FTP

- FTP stands for File transfer protocol.
- FTP is a standard internet protocol provided by TCP/IP used for transmitting the files from one host to another.
- It is mainly used for transferring the web page files from their creator to the computer that acts as a server for other computers on the internet.
- It is also used for downloading the files to computer from other servers.

**Objectives of FTP**

- It provides the sharing of files.

- It is used to encourage the use of remote computers.
- It transfers the data more reliably and efficiently.

### Why FTP?

Although transferring files from one system to another is very simple and straightforward, but sometimes it can cause problems. For example, two systems may have different file conventions. Two systems may have different ways to represent text and data. Two systems may have different directory structures. FTP protocol overcomes these problems by establishing two connections between hosts. One connection is used for data transfer, and another connection is used for the control connection.

### FTP Clients

- FTP client is a program that implements a file transfer protocol which allows you to transfer files between two hosts on the internet.
- It allows a user to connect to a remote host and upload or download the files.
- It has a set of commands that we can use to connect to a host, transfer the files between you and your host and close the connection.
- The FTP program is also available as a built-in component in a Web browser. This GUI based FTP client makes the file transfer very easy and also does not require to remember the FTP commands.

### Advantages of FTP:

- **Speed:** One of the biggest advantages of FTP is speed. The FTP is one of the fastest way to transfer the files from one computer to another computer.
- **Efficient:** It is more efficient as we do not need to complete all the operations to get the entire file.
- **Security:** To access the FTP server, we need to login with the username and password. Therefore, we can say that FTP is more secure.
- **Back & forth movement:** FTP allows us to transfer the files back and forth. Suppose you are a manager of the company, you send some information to all the employees, and they all send information back on the same server.

### Disadvantages of FTP:

- The standard requirement of the industry is that all the FTP transmissions should be encrypted. However, not all the FTP providers are equal and not all the providers offer encryption. So, we will have to look out for the FTP providers that provides encryption.
- FTP serves two operations, i.e., to send and receive large files on a network. However, the size limit of the file is 2GB that can be sent. It also doesn't allow you to run simultaneous transfers to multiple receivers.
- Passwords and file contents are sent in clear text that allows unwanted eavesdropping. So, it is quite possible that attackers can carry out the brute force attack by trying to guess the FTP password.
- It is not compatible with every system.