

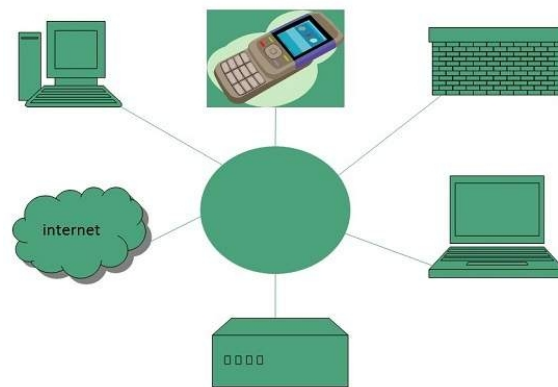
# Overview of World Wide Web

## Essential Computing (Note - 5)

### Internet

Internet is defined as an Information super Highway, to access information over the web. However, It can be defined in many ways as follows:

- Internet is a world-wide global system of interconnected computer networks.
- Internet uses the standard Internet Protocol (TCP/IP).
- Every computer in internet is identified by a unique IP address.
- IP Address is a unique set of numbers (such as 110.22.33.114) which identifies a computer location.
- A special computer DNS (Domain Name Server) is used to give name to the IP Address so that user can locate a computer by a name. For example, a DNS server will resolve a name <https://manarat.ac.bd/> to a particular IP address to uniquely identify the computer on which this website is hosted.



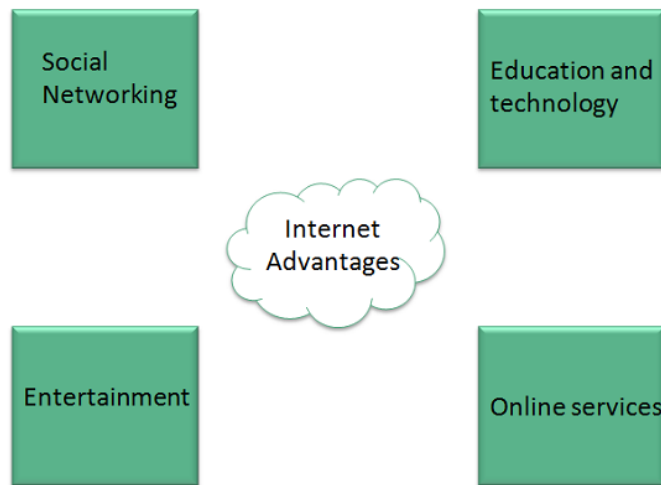
### Evolution

The concept of Internet was originated in 1969 and has undergone several technological & Infrastructural changes as discussed below:

- The origin of Internet devised from the concept of **Advanced Research Project Agency Network (ARPANET)**.
- **ARPANET** was developed by United States Department of Defense.
- Basic purpose of ARPANET was to provide communication among the various bodies of government.
- Initially, there were only four nodes, formally called **Hosts**.
- In 1972, the **ARPANET** spread over the globe with 23 nodes located at different countries and thus became known as Internet.
- By the time, with invention of new technologies such as TCP/IP protocols, DNS, WWW, browsers, scripting languages etc., Internet provided a medium to publish and access information over the web.

## Advantages

Internet covers almost every aspect of life, one can think of. Here, we will discuss some of the advantages of Internet:



- Internet allows us to communicate with the people sitting at remote locations. There are various apps available on the web that uses Internet as a medium for communication. One can find various social networking sites such as:
  - ◆ Facebook, Twitter, Yahoo, Google+, Flickr, Reddit
- One can surf for any kind of information over the internet. Information regarding various topics such as Technology, Health & Science, Social Studies, Geographical Information, Information Technology, Products etc can be surfed with help of a search engine.
- Apart from communication and source of information, internet also serves a medium for entertainment. Following are the various modes for entertainment over internet.
  - ◆ Online Television, Online Games, Music, Videos
- Internet allows us to use many services like:
  - ◆ Internet Banking, Online Shopping, Online Bill Payment, E-mail
- Internet provides concept of **electronic commerce**, that allows the business deals to be conducted on electronic systems

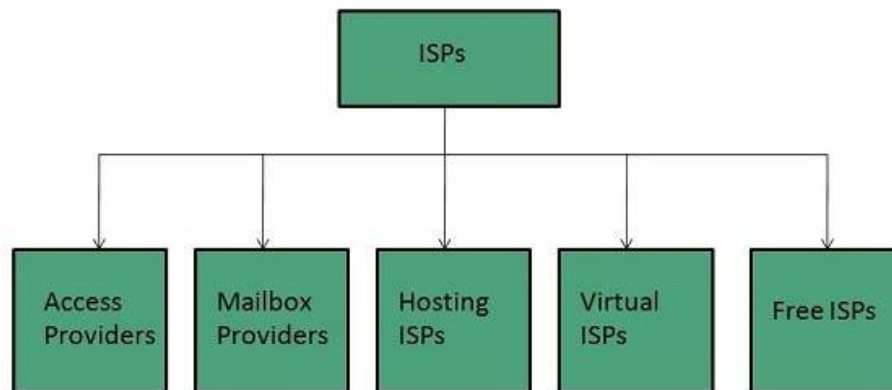
## Internet Service Providers (ISP)

ISP is a company offering access to internet. They offer various services:

- Internet Access
- Domain name registration
- Dial-up access
- Leased line access

## ISP Types

ISPs can broadly be classified into six categories as shown in the following diagram:



### Access providers

They provide access to internet through telephone lines, cable wi-fi or fiber optics.

### Mailbox Provider

Such providers offer mailbox hosting services.

### Hosting ISPs

Hosting ISPs offers e-mail, and other web hosting services such as virtual machines, clouds etc.

### Virtual ISPs

Such ISPs offer internet access via other ISP services.

### Free ISPs

Free ISPs do not charge for internet services.

## Connection Types

There exist several ways to connect to the internet. Following are these connection types available:

1. Dial-up Connection
2. ISDN
3. DSL
4. Cable TV Internet connections
5. Satellite Internet connections
6. Wireless Internet Connections

### Dial-up Connection

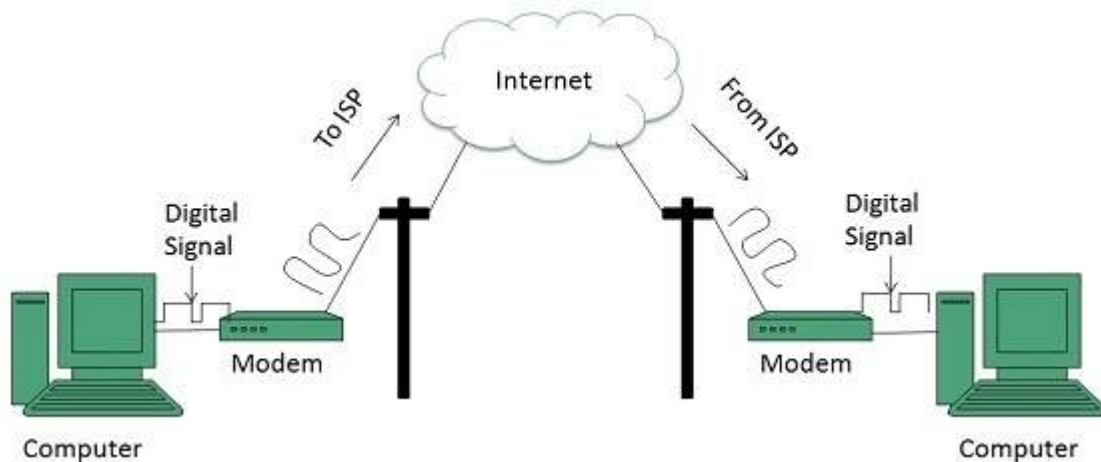
**Dial-up** connection uses telephone line to connect PC to the internet. It requires a modem to setup dial-up connection. This modem works as an interface between PC and the telephone line.

There is also a communication program that instructs the modem to make a call to specific number provided by an ISP.

Dial-up connection uses either of the following protocols:

1. Serial Line Internet Protocol (SLIP)
2. Point to Point Protocol (PPP)

The following diagram shows the accessing internet using modem:



## ISDN

ISDN is acronym of **Integrated Services Digital Network**. It establishes the connection using the phone lines which carry digital signals instead of analog signals.

There are two techniques to deliver ISDN services:

1. Basic Rate Interface (BRI)
2. Primary Rate Interface (PRI)

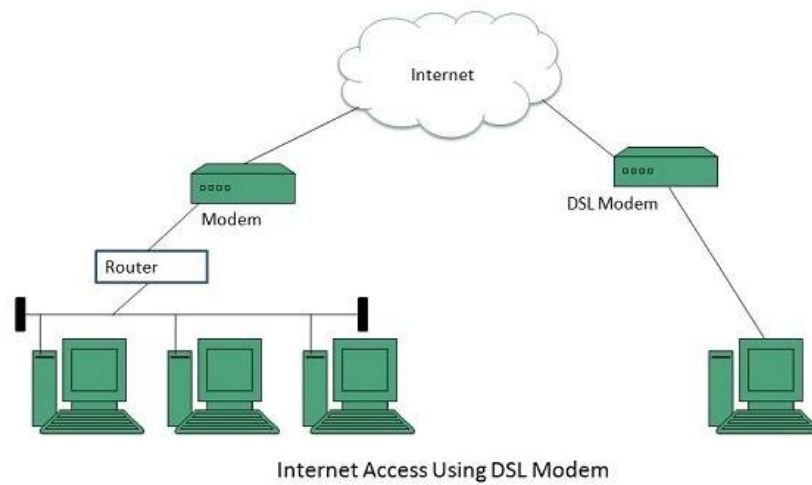
## DSL

DSL is acronym of **Digital Subscriber Line**. It is a form of broadband connection as it provides connection over ordinary telephone lines.

Following are the several versions of DSL technique available today:

1. Asymmetric DSL (ADSL)
2. Symmetric DSL (SDSL)
3. High bit-rate DSL (HDSL)
4. Rate adaptive DSL (RDSL)
5. Very high bit-rate DSL (VDSL)
6. ISDN DSL (IDSL)

All of the above mentioned technologies differ in their upload and download speed, bit transfer rate and level of service. The following diagram shows that how we can connect to internet using DSL technology:



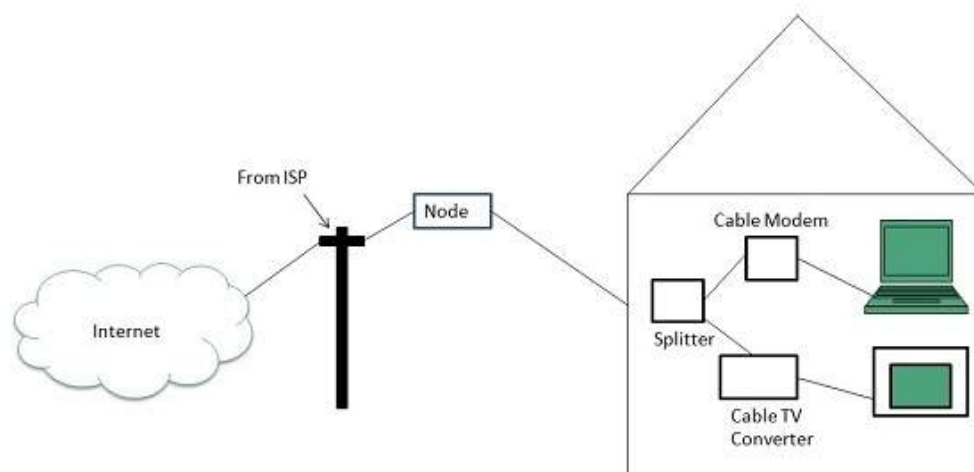
### Cable TV Internet Connection

Cable TV Internet connection is provided through Cable TV lines. It uses coaxial cable which is capable of transferring data at much higher speed than common telephone line.

#### Key Points:

- A cable modem is used to access this service, provided by the cable operator.
- The Cable modem comprises of two connections: one for internet service and other for Cable TV signals.
- Since Cable TV internet connections share a set amount of bandwidth with a group of customers, therefore, data transfer rate also depends on number of customers using the internet at the same time.

The following diagram shows that how internet is accessed using Cable TV connection:



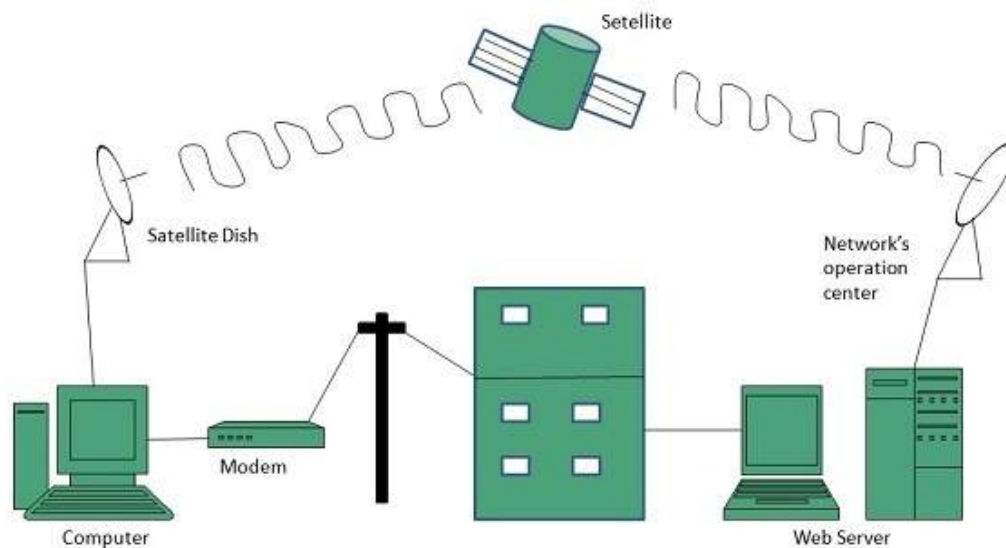
## Satellite Internet Connection

Satellite Internet connection offers high speed connection to the internet. There are two types of satellite internet connection: one way connection or two way connection.

In *one way connection*, we can only download data but if we want to upload, we need a dialup access through ISP over telephone line.

In *two way connection*, we can download and upload the data by the satellite. It does not require any dialup connection.

The following diagram shows how internet is accessed using satellite internet connection:



## Internet Security

Internet security refers to securing communication over the internet. It includes specific security protocols such as:

- Internet Security Protocol (IPSec)
- Secure Socket Layer (SSL)

### Internet Security Protocol (IPSec)

It consists of a set of protocols designed by Internet Engineering Task Force (IETF). It provides security at network level and helps to create authenticated and confidential packets for IP layer.

### Secure Socket Layer (SSL)

It is a security protocol developed by Netscape Communications Corporation. It provides security at transport layer. It addresses the following security issues:

- Privacy
- Integrity
- Authentication

## Threats

Internet security threats impact the network, data security and other internet connected systems. Cyber criminals have evolved several techniques to threat privacy and integrity of bank accounts, businesses, and organizations.

Following are some of the internet security threats:

- Mobile worms
- Malware
- PC and Mobile ransomware
- Large scale attacks like Stuxnet that attempts to destroy infrastructure.
- Hacking as a Service
- Spam
- Phishing

## Internet Domain Name System

When **DNS** was not into existence, one had to download a **Host file** containing host names and their corresponding IP address. But with increase in number of hosts of internet, the size of host file also increased. This resulted in increased traffic on downloading this file. To solve this problem the DNS system was introduced.

**Domain Name System** helps to resolve the host name to an address. It uses a hierarchical naming scheme and distributed database of IP addresses and associated names

## IP Address

IP address is a unique logical address assigned to a machine over the network. An IP address exhibits the following properties:

- IP address is the unique address assigned to each host present on Internet.
- IP address is 32 bits (4 bytes) long.
- IP address consists of two components: **network component** and **host component**.
- Each of the 4 bytes is represented by a number from 0 to 255, separated with dots. For example 137.170.4.124

IP address is 32-bit number while on the other hand domain names are easy to remember names. For example, when we enter an email address we always enter a symbolic string such as *manarat.ac.bd*

## Uniform Resource Locator (URL)

**Uniform Resource Locator (URL)** refers to a web address which uniquely identifies a document over the internet.

This document can be a web page, image, audio, video or anything else present on the web.

## URL Types

There are two forms of URL as listed below:

- Absolute URL
- Relative URL

### Absolute URL

Absolute URL is a complete address of a resource on the web. This completed address comprises of protocol used, server name, path name and file name.

For example <https://manarat.ac.bd/academics/academic-syllabus/bpm/> where:

- **https** is the protocol.
- [manarat.ac.bd/academics/academic-syllabus/bpm/](https://manarat.ac.bd/academics/academic-syllabus/bpm/) is the server name.

The protocol part tells the web browser how to handle the file. Similarly we have some other protocols also that can be used to create URL are:

- FTP, https, Gopher, mailto, news

### Relative URL

Relative URL is a partial address of a web page. Unlike absolute URL, the protocol and server part are omitted from relative URL. Relative URLs are used for internal links i.e. to create links to file that are part of same website as the Web Pages on which you are placing the link.

## Difference between Absolute and Relative URL

Absolute URL	Relative URL
Used to link web pages on different websites	Used to link web pages within the same website.
Difficult to manage.	Easy to Manage
Changes when the server name or directory name changes	Remains same even if we change the server name or directory name.
Take time to access	Comparatively faster to access.

## Domain Name System Architecture

The Domain name system comprises of **Domain Names, Domain Name Space, Name Server** that have been described below:

### Domain Names

Domain Name is a symbolic string associated with an IP address. There are several domain names available; some of them are generic such as **com, edu, gov, net** etc, while some country level domain names such as **au, in, za, us** etc.



The following table shows the **Generic** Top-Level Domain names:

**Domain Name Meaning**

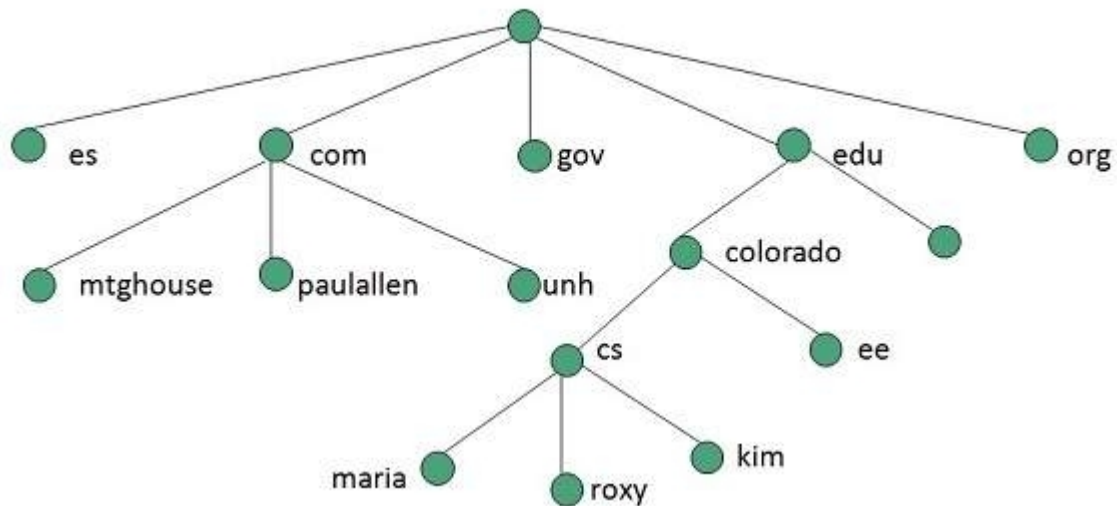
com	Commercial business
edu	Education
gov	U.S. government agency
int	International entity
mil	U.S. military
net	Networking organization
org	Non profit organization

The following table shows the **Country top-level** domain names:

Domain Name	Meaning	Domain Name	Meaning
bd	Bangladesh	fr	France
au	Australia	us	United States
in	India	za	South Africa
cl	Chile	uk	United Kingdom

**Domain Name Space**

The domain name space refers a hierarchy in the internet naming structure. This hierarchy has multiple levels (from 0 to 127), with a root at the top. The following diagram shows the domain name space hierarchy:



In the above diagram each sub-tree represents a domain. Each domain can be partitioned into sub domains and these can be further partitioned and so on.

## Name Server

Name server contains the DNS database. This database comprises of various names and their corresponding IP addresses. Since it is not possible for a single server to maintain entire DNS database, therefore, the information is distributed among many DNS servers.

- Hierarchy of server is same as hierarchy of names.
- The entire name space is divided into the zones

## Zones

Zone is collection of nodes (sub domains) under the main domain. The server maintains a database called zone file for every zone.

## World Wide Web (WWW)

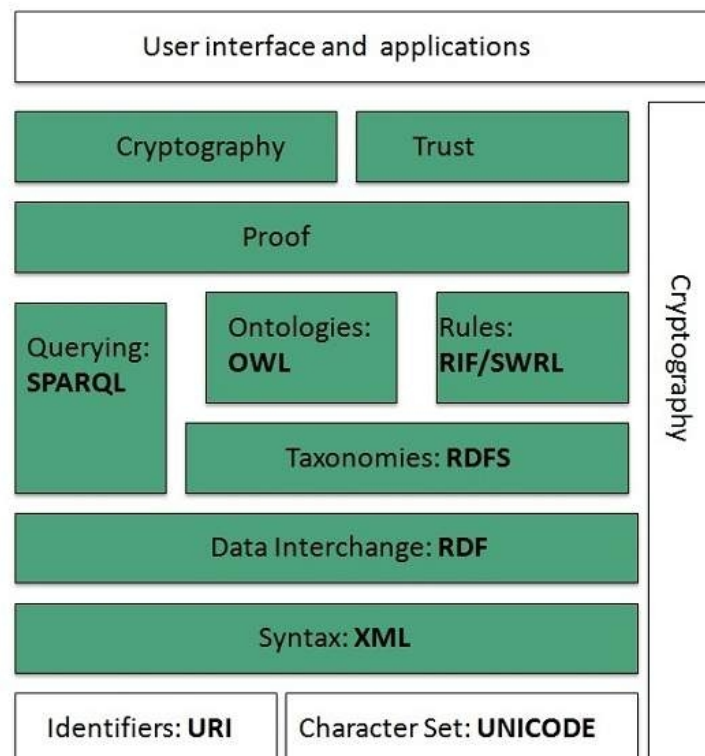
A technical definition of the World Wide Web is : all the resources and users on the Internet that are using the Hypertext Transfer Protocol (HTTP). The World Wide Web is the universe of network-accessible information. In simple terms, It is a way of exchanging information between computers on the Internet, tying them together into a vast collection of interactive multimedia resources.

## Evolution

World Wide Web was created by **Timothy Berners Lee** in 1989 at **CERN** in **Geneva**. World Wide Web came into existence as a proposal by him, to allow researchers to work together effectively and efficiently at **CERN**. Eventually it became World Wide Web.

## 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.

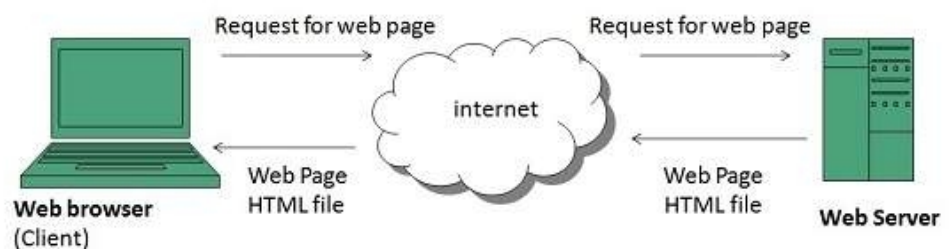
**Cryptography:** Cryptography means such as digital signature for verification of the origin of sources is used.

**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, <https://manarat.ac.bd/>) of the web page in the address bar of web browser. Then the browser requests the Domain Name Server for the IP address corresponding to <https://manarat.ac.bd/>
2. 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.
3. 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.
4. Now the web browser receives the web page, It interprets it and display the contents of web page in web browser's window.



## Types of Name Servers

Following are the three categories of Name Servers that manages the entire Domain Name System:

- Root Server
- Primary Server
- Secondary Server

## Root Server

Root Server is the top level server which consists of the entire DNS tree. It does not contain the information about domains but delegates the authority to the other server

## Primary Servers

Primary Server stores a file about its zone. It has authority to create, maintain, and update the zone file.

## Secondary Server

Secondary Server transfers complete information about a zone from another server which may be primary or secondary server. The secondary server does not have authority to create or update a zone file.

## DNS Working

DNS translates the domain name into IP address automatically. Following steps will take you through the steps included in domain resolution process:

- When we type **manarat.ac.bd** into the browser, it asks the local DNS Server for its IP address.  
Here the local DNS is at ISP end.
- When the local DNS does not find the IP address of requested domain name, it forwards the request to the root DNS server and again enquires about IP address of it.
- The root DNS server replies with delegation that I do not know the IP address of manarat.ac.bd but know the IP address of DNS Server.
- The local DNS server then asks the com DNS Server the same question.
- The **bd** DNS Server replies the same that it does not know the IP address of manarat.ac.bd but knows the address of manarat.ac.bd
- Then the local DNS asks the manarat.ac.bd DNS server the same question.
- Then manarat.ac.bd DNS server replies with IP address of [manarat.ac.bd](http://manarat.ac.bd).
- Now, the local DNS sends the IP address of manarat.ac.bd to the computer that sends the request.

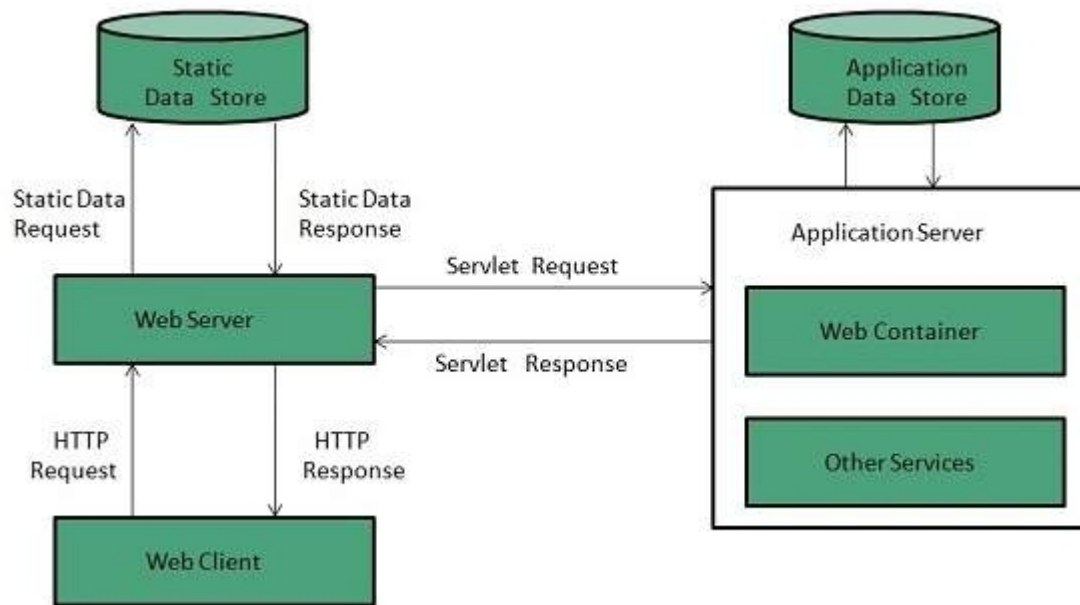
## Web Server

Web server is a computer where the web content is stored. Basically web server is used to host the web sites but there exists other web servers also such as gaming, storage, FTP, email etc.

## Web Server Working

Web server respond to the client request in either of the following two ways:

- Sending the file to the client associated with the requested URL.
- Generating response by invoking a script and communicating with database



### Key Points

- When client sends request for a web page, the web server search for the requested page if requested page is found then it will send it to client with an HTTP response.
- If the requested web page is not found, web server will the send an **HTTP response: Error 404 Not found**.
- If client has requested for some other resources then the web server will contact to the application server and data store to construct the HTTP response.

### Architecture

Web Server Architecture follows the following two approaches:

1. Concurrent Approach
2. Single-Process-Event-Driven Approach.

### Concurrent Approach

Concurrent approach allows the web server to handle multiple client requests at the same time. It can be achieved by following methods:

- Multi-process
- Multi-threaded
- Hybrid method.

### Multi-processing

In this a single process (parent process) initiates several single-threaded child processes and distribute incoming requests to these child processes. Each of the child processes are responsible for handling

single request. It is the responsibility of parent process to monitor the load and decide if processes should be killed or forked.

### **Multi-threaded**

Unlike Multi-process, it creates multiple single-threaded process.

### **Hybrid**

It is combination of above two approaches. In this approach multiple process are created and each process initiates multiple threads. Each of the threads handles one connection. Using multiple threads in single process results in less load on system resources.

### **Websites security**

Websites are always prone to security risks. **Cyber crime** impacts your business by hacking your website. Your website is then used for hacking assaults that install malicious software or malware on your visitor's computer. Hackers may also steal important customer data such as credit card information, destroy your business and propagate illegal content to your users.

### **Security Considerations**

- **Updated Software:** It is mandatory to keep your software updated. It plays a vital role in keeping your website secure.
- **SQL Injection:** It is an attempt by the hackers to manipulate your database. It is easy to insert rogue code into your query that can be used to manipulate your database such as change tables, get information or delete data.
- **Cross Site Scripting (XSS):** It allows the attackers to inject client side script into web pages. Therefore, while creating a form it is good to ensure that you check the data being submitted and encode or strip out any HTML.
- **Error Messages:** You need to be careful about how much information to be given in the error messages. For example, if the user fails to log in the error message should not let the user know which field is incorrect: username or password
- **Validation of Data:** The validation should be performed on both server side and client side.
- **Passwords:** It is good to enforce password requirements such as of minimum of eight characters, including upper case, lower case and special character. It will help to protect user's information in long run.
- **SSL:** It is good practice to use SSL protocol while passing personal information between website and web server or database.