Web Basics – HTML5

Lesson 1. Introduction to the Internet

Lesson Objectives

- After completing this module you will be able to:
  - Understand the history of Internet.
  - Understand Web terminology.
  - Understand IP addresses
  - TCP/IP Protocol
  - Domain Name System
  - HTTP Protocol
  - Servers – Web Servers
  - Web Browsers
  - Working of WWW
  - HTML – Static and Dynamic Web Pages

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

1.1: Internet
What is Internet?

- Internet:
  - 'Network of networks' or "world's largest network".
  - A concept, like the economy.
  - Collection of inter-networked regional networks.
  - Not owned by anyone.
  - Based on TCP/IP.

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Internet:

Internet is a network of networks in which many machines are interconnected with each other using TCP/IP protocol suite.

No one person runs the Internet, and no single organization pays all the costs; there is no Internet Corporation. It exists as a result of the cooperation from people all over the world, who work in various types of organizational and computing environments. Internet never closes down, mostly because of its decentralized structure.   Today, it is reliable and predictable. Individual servers may close for upgrades or hardware replacement, but the network is always available without interruption.

Services offered by internet are:

Communication : Users can easily communicate with each other at any time from anywhere. For an Example, Email Communication, GTALK, etc..

Data Transfer : Users can transfer file in the format of picture, word document, pdf, etc.. as an attachment in the E-Mail. Also files are possible to be shared through FTP servers

History of Internet

Internet was established more than 25 years ago to meet research needs of the U. S. defence industry. However, it has grown into a huge global network serving universities, academic researchers, government agencies, and commercial interests, both in the United States and in more than 100 other countries.

In 1969, Advanced Research Projects Agency (ARPA) of the U. S. Department of defense established ARPAnet, an experimental four-computer network, so that research-scientists could communicate among themselves. By 1971, ARPAnet comprised almost two dozen sites. By 1974, that number grew to 62, and by 1981, it comprised more than 200 sites.

As more and more computers using different operating systems were connected, the need for a common communications protocol became apparent. Theory required that any computer on the network should be able to talk to any other computer, as a peer.

1.3: Internet Basic Definition
Internet – Basic Definitions

- Internet Service Provider (ISP)
- Network Information Center (NIC)
- Internet Address
- Internet Domain Name
- Routing
- Gateways
- Protocols
  - TCP/IP
  - HTTP

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Internet Service Provider (ISP): Internet Service Providers (ISPs) are companies that enable user to connect to the Internet . They offer an account on their systems and access to the Internet.

Network Information Center (NIC): Network Information Center (NIC) assigns and regulates IP addresses on the Internet. You can get one directly from the NIC, or you can ask your ISP to secure an IP address on your behalf.

Internet Address: TCP/IP requires each host on a TCP/IP network have their own unique IP address.

Internet Domain Name: A domain name maps or translates the actual numeric IP address used for your Web server into an easy-to-remember alphanumeric name. Domain refers to a collection of network host computers, known by the same name. Your domain name should reflect your organization or corporation, for example, .com , .edu , .gov , .int, .mil , .net, and so on.

Routing: Process of getting your data from point A to point B.

Gateway: Forwards datagrams to a destination if it knows where the destination is.

Protocol: specify interactions between the communicating entities. Example for protocol are TCP/IP, HTTP, telnet, etc..

1.4: Internet Address

Internet Address

- Every device (eg: computer, printer) that participates in a computer network is assigned a numeric label called as Internet Protocol address (IP address).

- The designers of the Internet Protocol defined an IP address as a 32-bit number and this system is known as Internet Protocol Version 4 (IPv4).

- IP addresses are binary numbers, but they are usually stored in text files and displayed in human-readable notations, such as 172.16.254.1

  - IP Address is a 32-bit address, in the form of x.x.x.x.

An IPv4 address (dotted-decimal notation)

172 . 16 . 254 . 1

10101100 .00010000 .11111110 .00000001

One byte =Eight bits

Thirty-two bits (4 × 8), or 4 bytes

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

1.4: Internet Address
Internet Address (Contd…)

- Internet address classification
  - Class A   N.H.H.H Used for very large networks
  - Class B   N.N.H.H Used for medium sized networks
  - Class C   N.N.N.H Used for smaller networks
  - Class D   Multicast Address
  - Class E   Reserved for future use.

- Addresses that start with a value between:
  - 1 and 126   Class A (First bit value is 0)
  - 128 and 191   Class B (First two bit values are 10)
  - 192 and 223   Class C (First three bit values are 110)
  - 224 and 239   Class D (First four bit values are 1110)
  - 240 and 255   Class E (First five bit values are 11110)

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Internet Address Classification:

   Internet address is classified into 5 classes such as Class A, B, C, D and E. IP address start and end value will vary based on classification. IP address for each classification is described using N and H.

        N stands for Network (Assigned by the NIC)
        H stands for Host (Assigned by the local administrator)

IP address classification : start and finish address

Class	Start address	Finish address
A	0.0.0.0	127.255.255.255
B	128.0.0.0	191.255.255.255
C	192.0.0.0	223.255.255.255
D	224.0.0.0	239.255.255.255
E	240.0.0.0	255.255.255.255

Depends on classification, IPAddress can be in the range value between start address and finish address.

1.5: TCP/IP Basics
TCP/IP Basics

- Transmission Control Protocol (TCP):
  - Connection-oriented transport layer protocol.
  - Sets up a connection between the sender and receiver.
  - Uses the services of IP to send and receive data.
  - Re-orders received information.

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

TCP (Transmission Control Protocol) is a connection-oriented protocol. TCP ensures that data arrives and that it arrives in the correct order. It reorders information that is received out of order and requests the information that is not received is to be resent.

Internet uses TCP/IP to link computers. TCP/IP stands for Transmission Control Protocol/Internet Protocol, which are two significant parts of what is now also known as the Internet Protocol Suit.

The feature that makes TCP/IP different from many other networking protocols is that it was designed to link networks instead of simply linking computers in a network.

1.6: Domain Name System
Domain Name System

- Computers work best with numbers.
  - Synonymous to identifying people with names.
- Domain Name:
  - Maps or translates the actual numeric IP address into an easy-to-remember alphanumeric name.
- Internet Network Information Center (InterNIC) Registration Service:
  - Manages IP addresses and domain name assignment to internet users.

Domain Name System (DNS)

It is as easy to identify computers with numbers, as it is to identify people with names. To bridge this dichotomy, the Domain Name System (DNS), a distributed database, was invented. Domain name maps or translates an actual, numeric IP address into an easy-to-remember alphanumeric name that your Web server uses.

Before DNS can do this for you, you must register any name you want to use.

Domain refers to a collection of network host computers, known by the same name. Your domain name should reflect your organization or corporation.

InterNIC (Internet Network Information Center) Registration Service, manages the task of assigning IP addresses and domain names to Internet users.

Note: InterNIC (Internet Network Information Center) lets you apply for any domain name you like, regardless of your company name. The only restriction is that the name must be available and not already reserved by someone else.

Rightmost part of a name is called its zone. The next part is the name of the company. The part to the left of the company name is the particular machine within the company. Seven domains were established originally.

1.6: Domain Name System

Domain Name System (Contd…)

▪ Few domain types are as listed below:

Domain type name	Description
.com	Commercial organization. Most companies end up as a part of this domain.
.edu	Educational establishment such as university
.gov	Branch of the U. S. government.
.int	International organization, such as NATO or the United Nations.
.mil	*Mil* is a branch of the U. S. military.
.net	Network organization.
.org	Non-profit organization.

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

1.7: HTTP

Basics

- HTTP is the fundamentals means of communication used by WWW.
- It defines formal syntax that allows user agents, such as browsers, to interact with web servers.
- It is one of the many protocols designed to allow clients to store and retrieve files from servers.
- HTTP requests can specify the language the browser would like to see in a page as well as information about how the data is encoded.

HTTP Protocol

Web Browser

HTTP server

Server side program like servlets

Database

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING
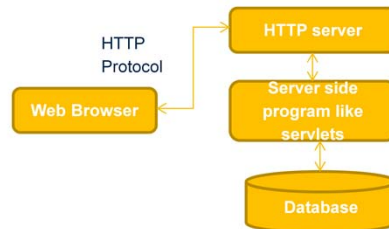
1.7: HTTP

Client Server Interaction in HTTP

- Client Server interaction in HTTP includes following four basic steps:
  - The client opens Transmission Control Protocol (TCP) connection.
  - Client then sends a HTTP request such as 'GET index.HTML'
  - Then, the server sends an HTTP response, including a status and a requested object.
  - Finally, a TCP connection is ended.

1.7: HTTP

HTTP – Stateless Nature

- HTTP is connection less.
- HTTP being stateless is direct implication of HTTP being connectionless.
  - The server and client are aware of each other only during a request. Afterwards, each forgets the other. For this reason neither the client nor the browser can retain information between different request across the web pages.

- Advantage
  - Simple design
  - If transaction fails, then server state is not required to be cleaned.

- Disadvantage
  - Additional information need to be sent in every request.
  - Information need to be interpreted

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

1.7: HTTP
URL and Parts of URL

- HTTP clients use Uniform Resource Locator (URL) to interact with resources
- For HTTP, URL is composed of:
  - Scheme
  - It is implied by the fact that it is HTTP message.
  - Host
  - In HTTP 1.1, it is included in HOST header.
  - Port
  - Port is used by TCP and not HTTP. By default it is 80.
  - Path
  - Relative path of the requested resource
  - Query
  - Both Path and Query are contained in request start line.
  - Scheme and Host are case insensitive
  - Path and query are case sensitive

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

1.7: HTTP
URL and Parts of URL

- Consider following example:
  - http://www.example.com:80/path/hello.jsp?k=10&h=40

- See below the detailed description of the given URL:

URL Parts name	Example	Comparison
Scheme	http	Equals to HTTP
Host Name	www.example.com	Equals to WWW.EXAMPLE.COM
Path	/hello.jsp	Not equals to /HELLO.jsp
Query	k=10&h=40	Not equals to K=10&H=40

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

1.7: HTTP

HTTP Request Methods

- GET
  - It is used to retrieve a resource from server
- POST
  - POST is used to pass information to the server.
  - POST allows clients to send messages to forums or update databases.

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

1.8: **Servers**

Servers

- A server is a system (software and suitable computer hardware) that responds to requests across a computer network to provide, or help to provide, a network service.

- Servers operate within a client-server architecture, servers are computer programs running to serve the requests of other programs, the clients.

- Thus, the server performs some task on behalf of clients. The clients typically connect to the server through the network but may run on the same computer

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

1.8: Servers

Servers - Example

- Application Server
- Database Server
- File Server
- Print Server
- Web Server

1.8: Servers
Web Servers

- Web server can refer to either the hardware (the computer) or the software (the computer application) that helps to deliver web content that can be accessed through the Internet.

- The primary function of a web server is to deliver web pages on the request of clients using the Hypertext Transfer Protocol (HTTP).

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

1.8: Servers
Web Servers - Features

- Virtual Hosting
  - To serve many websites using one IP address.
- Large File Support
  - To be able to serve files whose size is greater than 2GB or 32 bit OS.
- Bandwidth throttling
  - To limit the speed of responses in order to not saturate the network and to be able to serve more clients.
- Server-side scripting
  - To generate dynamic web pages, still keeping web server and website implementations separate from each other.

1.8: Servers
Web Servers – Path Translation

- Web servers are able to map the path component of URL into:
  - A local file system resource (for static requests)
  - An internal or external program name (for dynamic requests)
- Consider URL :
  - http://www.exam.com/path/file.html is requested by client.
- The client's user agent will translate it into a connection to www.example.com with the following HTTP 1.1 request:
  - GET /path/file.html HTTP/1.1
  - Host: www.example.com
- The web server on www.example.com will append the given path to the path of its root directory.
  - On an Apache server, this is commonly /home/www
  - On Unix machines, usually /var/www
  - The result is the local file system resource: /home/www/path/file.html
- The Web Server then reads the file, if it exists and sends a response to the client's web browser.

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

1.8: Servers

Web Servers – Examples

- Examples of web servers
  - Apache *by Apache*
  - IIS *by Microsoft*
  - nginx *by NGINX Inc*
  - GWS *by Google*

Developer	December 2014	Percent	January 2015	Percent
Apache	358,159,405	39.11%	348,460,753	39.74%
Microsoft	273,967,294	29.81%	241,276,347	27.52%
nginx	132,467,763	14.47%	128,083,920	14.61%
Google	20,011,260	2.19%	20,209,649	2.30%

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

1.9: Web Browser

Web Browser

- A web browser (commonly referred to as a browser) is a software application for retrieving, presenting and traversing information resources on the World Wide Web.

- An information resource is identified by a Uniform Resource Identifier (URI) and may be a web page, image, video or other piece of content

- A web browser can also be defined as an application software or program designed to enable users to access, retrieve and view documents and other resources on the Internet.

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

1.9: Web Browser
Web Browser

- Available web browsers range in features from minimal, text-based user interfaces with bare-bones support for HTML to rich user interfaces supporting a wide variety of file formats and protocols.
- Browsers which include additional components to support e-mail, Usenet news, and Internet Relay Chat (IRC), are sometimes referred to as "Internet suites" rather than merely "web browsers".
- Most browsers can be extended via plug-ins, downloadable components that provide additional features.

Capgemini
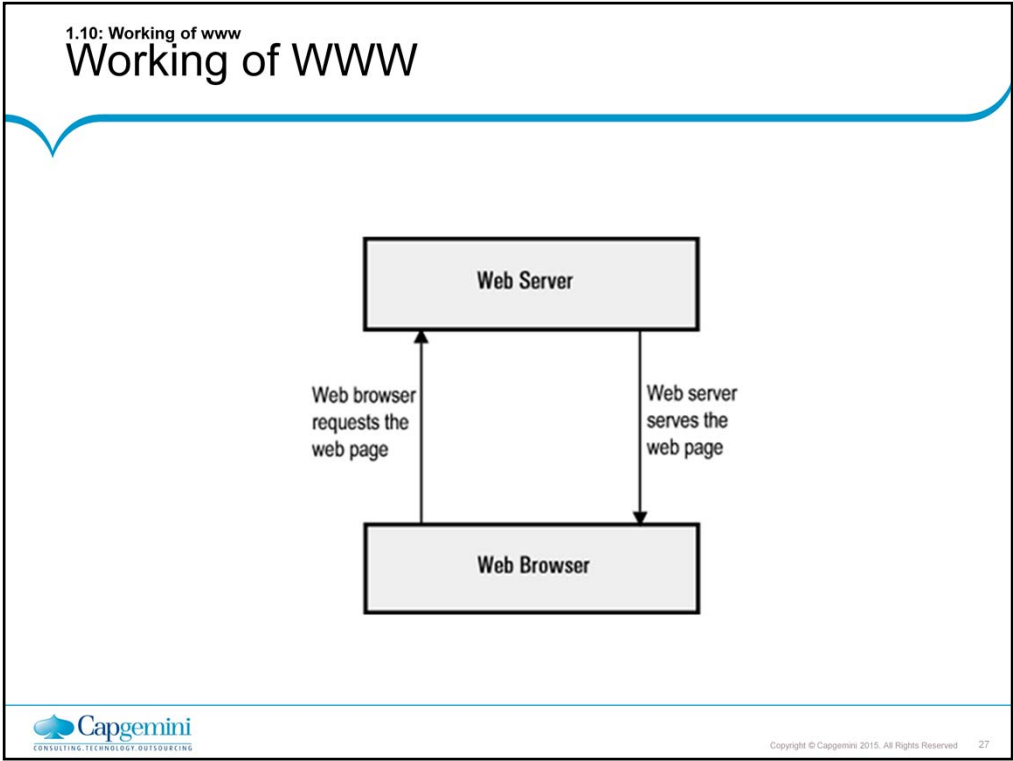CONSULTING.TECHNOLOGY.OUTSOURCING

1.9: Web Browser

Web Browser – Examples

- Google Chrome
- Mozilla Firefox
- Microsoft Internet Explorer
- Opera by Opera Software
- Apple Safari

1.10: Working of www
Working of WWW

- WWW (World Wide Web) refers to all of the publicly accessible web sites in the world, in addition to other information sources that web browsers can access.
  - These other sources include FTP sites, USENET newsgroups, and a few surviving Gopher sites.
- Typically Internet follows client/server model where:
  - Web-browsers acts as client software on the remote machine.
  - The server software is hosted on the webserver which acts as host.
- Whenever you view a web page on the internet, you are requesting that page from a web server. When you type a URL into your browser (for example, "http://www.igate.com/igate-profile.aspx"), your browser requests the page from the web server and the web server sends the page back:

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

1.10: Working of www
Working of WWW

Web Server

Web browser requests the web page

Web server serves the web page

Web Browser

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Multiple Websites : A web server can contain more than one website. In fact, many hosting companies host hundreds, or even thousands of websites on a single web server. Each website is usually assigned a unique IP address which distinguishes it from other websites on the same machine. This IP address is also what the DNS server uses to resolve the domain name.

It is also possible to configure multiple websites without using different IP addresses using host headers and/or different ports.

Page Not Found : If the requested page isn't found, the web server sends the appropriate error code/message back to the client. You can create user friendly error messages, then configure your web server to display that page instead of the usual error page. This can add a nice touch to your website.

Default Documents: If you've ever created a website, you may have found that if you have an "index" file (index.html for example), you don't need to specify the name of the file. For example, the following URLs both load the same page:

http://www.example.com/html/tutorial

http://www.example.com/html/tutorial/index.html

In this example, "index.html" is the *default document.* You can configure your web server so that any file name can be the default document.
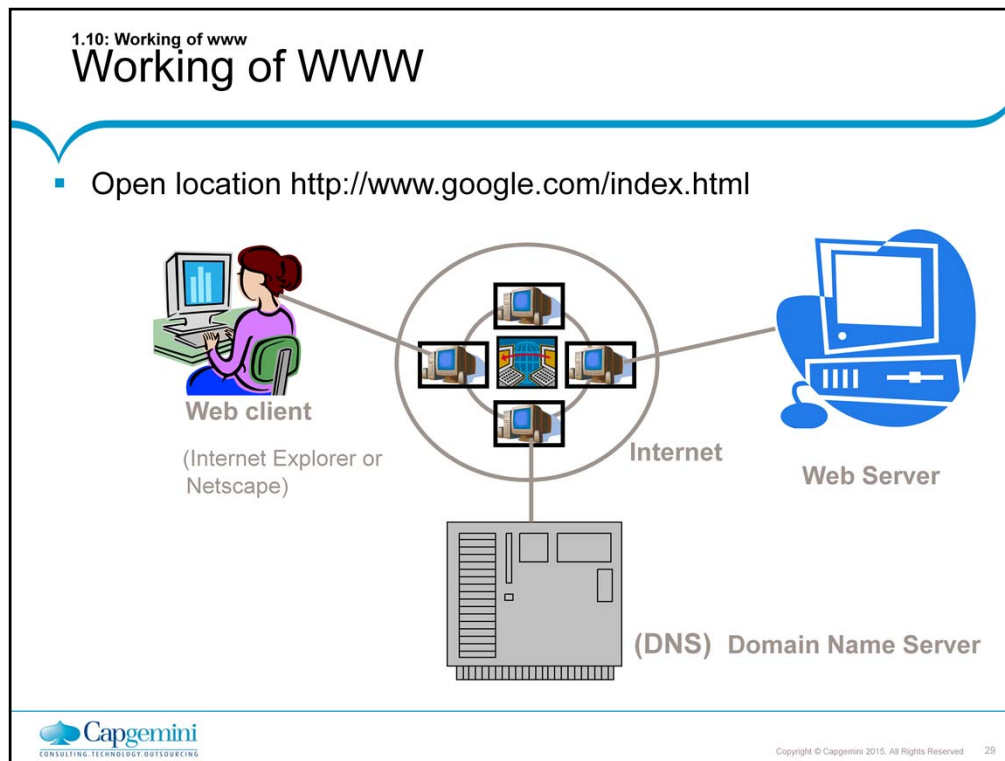
For example, you could configure your web server to use "index.html" in the event no filename has been specified. You could even specify different default documents for different directories if you like.
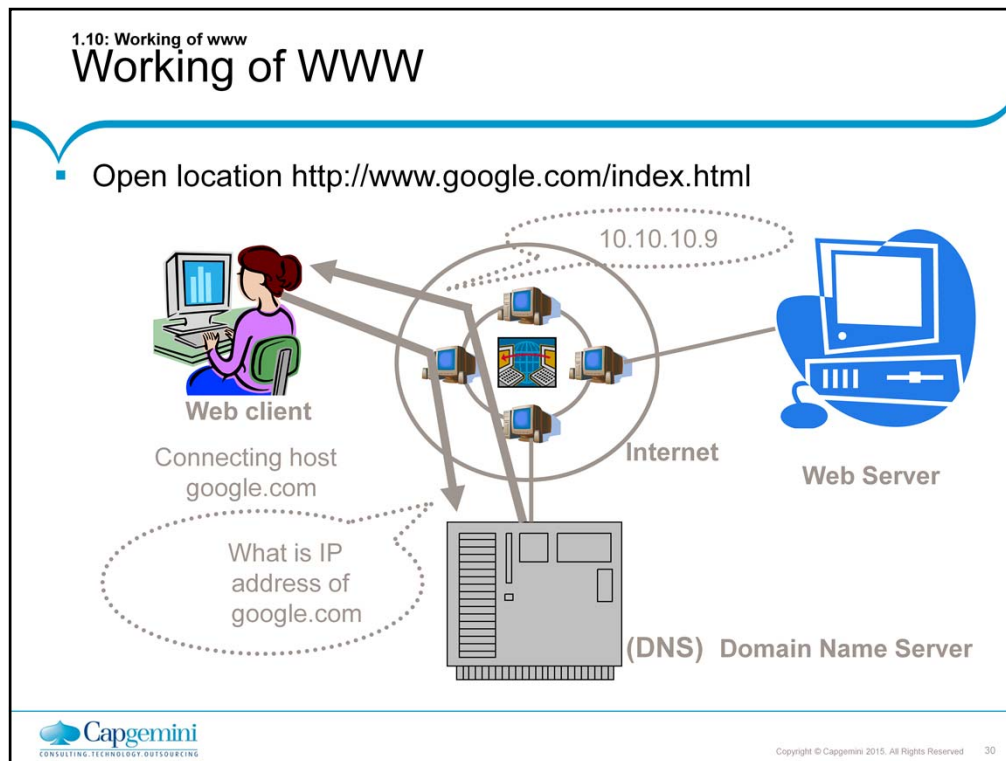
SSL Certificates

You can apply SSL certificates against a website via the web server. First you need to generate the certificate either by yourself (i.e. using a certificate generator), or by a Certificate Authority (CA). Then, once it has been generated, you apply it to your website via your web server. Applying an SSL certificate to a website is a straight forward task.

Once you've applied an SSL certificate against a website, you can navigate it using HTTPS (as opposed to HTTP). HTTPS encrypts any data that is transferred over the internet. This reduces the possibility of some malicious person being able to read your users' sensitive information.

To navigate a website using HTTPS, you simply replace the HTTP with HTTPS at the start of the URL in your browsers' location bar ("*https*://www.example.com")

1.10: Working of www
Working of WWW

- Open location http://www.google.com/index.html

Web client
(Internet Explorer or Netscape)

Internet

Web Server

(DNS) Domain Name Server

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved   29

Consider http://www.google.com/index.html is typed in web client by the user.

1.10: Working of www
Working of WWW

- Open location http://www.google.com/index.html

After the request is made, Domain Name server will be connected via Internet to find the ipaddress (for the domain name) of the web server which need to be connected for processing the request.
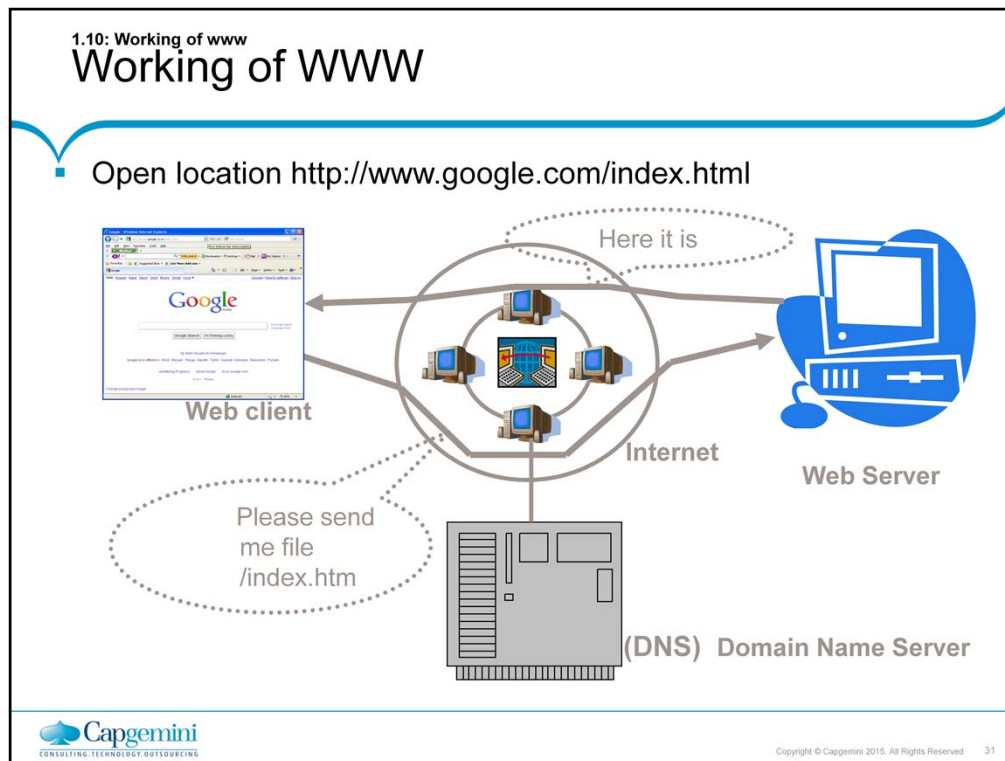
Once Ipaddress of the webserver is found, request from the webclient will be sent to the web server. In the webserver , request will be processed and response will be sent back to the web client.

1.10: HTML

HTML and WWW

- What is HTML?
  - HTML stands for Hyper Text Markup Language
  - HTML is a markup language used to design web pages.
  - A markup language is a set of markup tags
  - The tags describe document content
  - HTML documents contain HTML tags and plain text
  - HTML documents are also called web pages

- Web Pages are of two types:
  - Static Web Page
  - Dynamic Web Page

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

1.10: HTML
Static Web Page

- A static web page is a web page that is delivered to the user exactly as stored, in contrast to dynamic web pages which are generated by a web application.
- A static web page displays the same information for all users, from all contexts, subject to modern capabilities of a web server to negotiate content-type or language of the document where such versions are available and the server is configured to do so.
- Static web pages are often HTML documents stored as files in the file system and made available by the web server over HTTP.
- Disadvantages:
  - Any personalization or interactivity has to run client-side, which is restricting.
  - Maintaining large numbers of static pages as files can be impractical without automated tools.

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

1.10: HTML
Dynamic Web Page

- A dynamic web page is a web page with web content that varies based on parameters provided by a user or a computer program.
- For dynamic behavior, client side scripting and server side scripting are used.
- Client-side scripting is changing interface behaviors within a specific web page in response to mouse or keyboard actions, or at specified timing events.
- Server side scripting involves program running on a web server and is used to change the web content on various web pages, or to adjust the sequence of or reload of the web pages. Server responses may be determined by such conditions as data in a posted HTML form, parameters in the URL, the type of browser being used, the passage of time, or a database or server state.

Lesson Summary

- In this lesson, you have learnt about:
- Internet: Connection of interrelated networks.
  - Protocols: Developed to maintain communication standard across:
    - Different computers and operating systems (Platforms).
    - TCIP/IP concepts
    - HTTP concepts
- WWW:
  - Webservers
  - Web Browsers
  - Working of WWW
- HTML
  - Static Web Page
  - Dynamic Web Page

Summary

Review Question

- Question 1: URL is a networked extension of the standard filename concept.
  - True/False

- Question 2: HTTP is a Stateless Protocol
  - True/False

- Question 3: Which of the following Webservers host maximum number of websites?
  - Apache
  - IIS
  - Nginx
  - GWS

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Review Question : Crossword
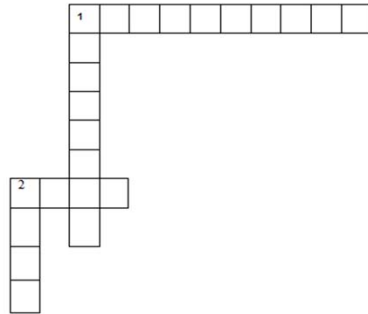
Clues:

ACROSS
1 Client for accessing webpage
2 Stateless & connectionless protocol

DOWN
1 Set of related web pages
2 Markup Language