

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



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.

 CONSULTING TECHNOLOGY OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved 3

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

1.2: History of Internet

History of Internet

- Need to share information.
- Advanced Research Projects Agency (ARPA)
 - ARPANET
 - Comprised individual packet switching computers interconnected by leased lines.

Copyright © Capgemini 2015. All Rights Reserved 4

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



Copyright © Capgemini 2015. All Rights Reserved

5

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 x 8), or 4 bytes

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

Copyright © Capgemini 2015. All Rights Reserved 7

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.



Copyright © Capgemini 2015. All Rights Reserved 8

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.

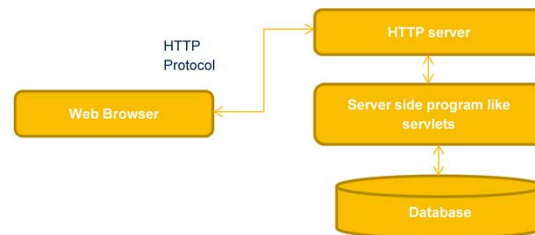
Note: InterNIC (Internet Network Information Center) lets you apply for any domain name you like, regardless of your company name. The only restriction is that the name must be available and not already reserved by someone else.

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.

1.7: HTTP

Basics

- HTTP is the fundamental 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 – Hyper Text Transfer Protocol

A browser works as an HTTP client because it sends requests to an HTTP server which is called Web server. The Web Server then sends responses back to the client. Diagram on above slide shows where HTTP Protocol fits in communication.

Apart from HTTP, other examples of protocols are :

- File Transfer Protocol(FTP)
- Common Internet File System (CIFS)
- Network File System (NFS)
- Simple Network Management Protocol (SNMP): Most widely used protocol for monitoring network devices such as hubs, routers, workstations, and computers. Windows NT supports SNMP.

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.



Copyright © Capgemini 2015. All Rights Reserved 10

HTTP Version:

Earlier versions were 0.9 and 1.0 – Rarely used

Almost all clients and browsers support 1.1 today and it offers:

- Better performance
- Better support for proxies and cache

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



Copyright © Capgemini 2015. All Rights Reserved 11

HTTP- stateless nature

After a request is made, the client disconnects from the server and waits for a response and the server needs to re-establish the connection after it processes the request and hence HTTP is connection less.

In addition to being connectionless and stateless HTTP is also media independent which means that any type of data can be sent by HTTP as long as both the client and server know how to handle the data content. How content is handled is determined by the MIME specification.

Advantage : The stateless design simplifies the server design because there is no need to dynamically allocate storage to deal with conversations in progress. If a client dies in mid-transaction, no part of the system needs to be responsible for cleaning the present state of the server.

Disadvantage: A disadvantage of statelessness is that it may be necessary to include additional information in every request, and this extra information will need to be interpreted by the server

Examples of stateful protocol:

- TCP (Transmission Control Protocol)
- IP (Internet Protocol)
- BGP (Border Gateway Protocol)

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



Copyright © Capgemini 2015. All Rights Reserved 12

Parts of URL:

Consider the given URL as

<http://localhost:8080/Example/test.html?username='test'>

Scheme : http

Host: localhost

Port : 8080

Path : /Example/test.html

Query String: username='test'

In HTTP 1.0, the host was not always include in the HTTP messages

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



Copyright © Capgemini 2015. All Rights Reserved 13

URL parts for example given on slide:

Scheme : http

Host : www.example.com

Port : 80


Path : /path/hello.jsp

Query : k=10&h=40

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.

CONSULTING TECHNOLOGY OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved 14

HTTP Request Methods

GET: The GET method is used to retrieve information from the given server using a given URI. Requests using GET should only retrieve data and should have no other effect on the data.

POST: A POST request is used to send data to the server, for example, customer information, file upload, etc. using HTML forms

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



Copyright © Capgemini 2015. All Rights Reserved 15


Above diagram shows clients communicating with a server via internet.

Servers often provide essential services across a network, either to private users inside a large organization or to public users via the Internet.

1.8: Servers

Servers - Example

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

Capgemini
CONSULTING TECHNOLOGY OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved 16

Examples of Servers

Application server - A server dedicated to running certain software applications

Database server – It provides database services to other computer programs or computers

File server – It provides remote access to files


Print server – It provides printer services

Web server – A server that HTTP clients connect to in order to send commands and receive responses along with data contents

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

Copyright © Capgemini 2015. All Rights Reserved 17

Delivery of web pages over the web means HTML documents and any additional content that may be included by a document, such as images, style sheets and scripts

Web servers are not always used for serving the World Wide Web. They can also be found embedded in devices such as printers, routers, webcams and serving only a local network. The web server may then be used as a part of a system for monitoring and/or administering the device in question. This usually means that no additional software has to be installed on the client computer, since only a web browser is required (which now is included with most operating systems).

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.

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:

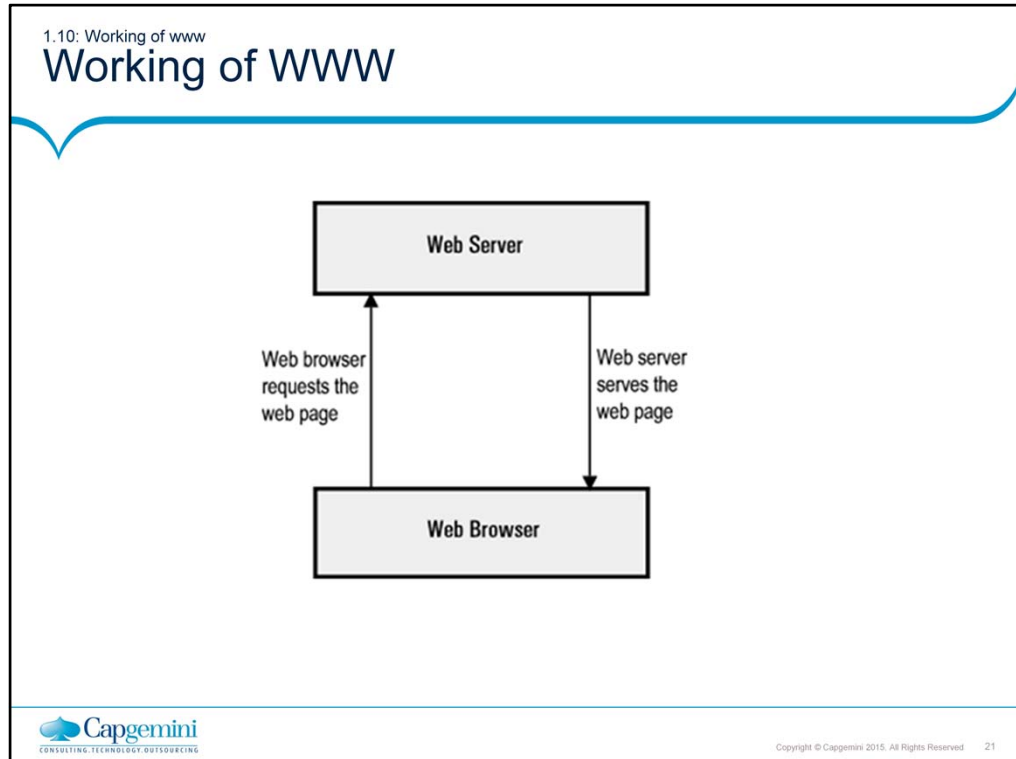


Diagram shown on above slide is simplistic version, Here is the detailed description:

Your web browser first needs to know which IP address the website "www.igate.com" resolves to. If it doesn't already have this information stored in its cache, it requests the information from one or more DNS servers (via the internet). The DNS server tells the browser which IP address the website is located at. Note that the IP address was assigned when the website was first created on the web server.

Now that the web browser knows which IP address the website is located at, it can request the full URL from the web server.

The web server responds by sending back the requested page. If the page doesn't exist (or another error occurs), it will send back the appropriate error message.

Your web browser receives the page and renders it as required.

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.

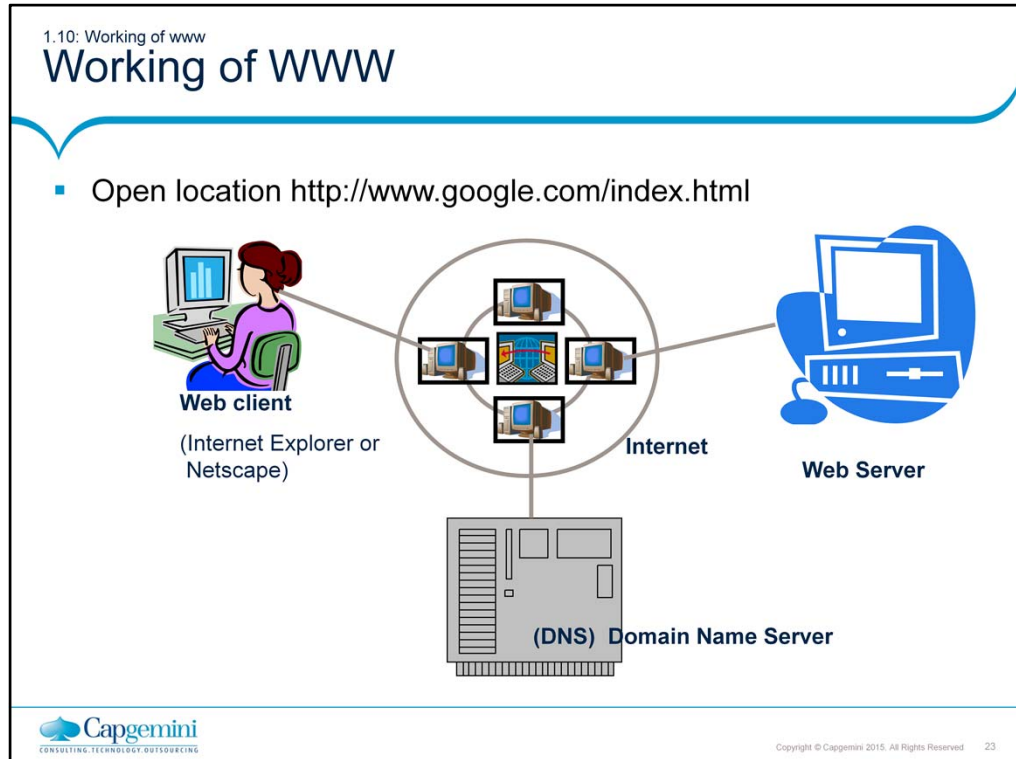
For example, you could configure your web server to use "index.html" in the event no filename has been specified. You could even specify different default documents for different directories if you like.

SSL Certificates

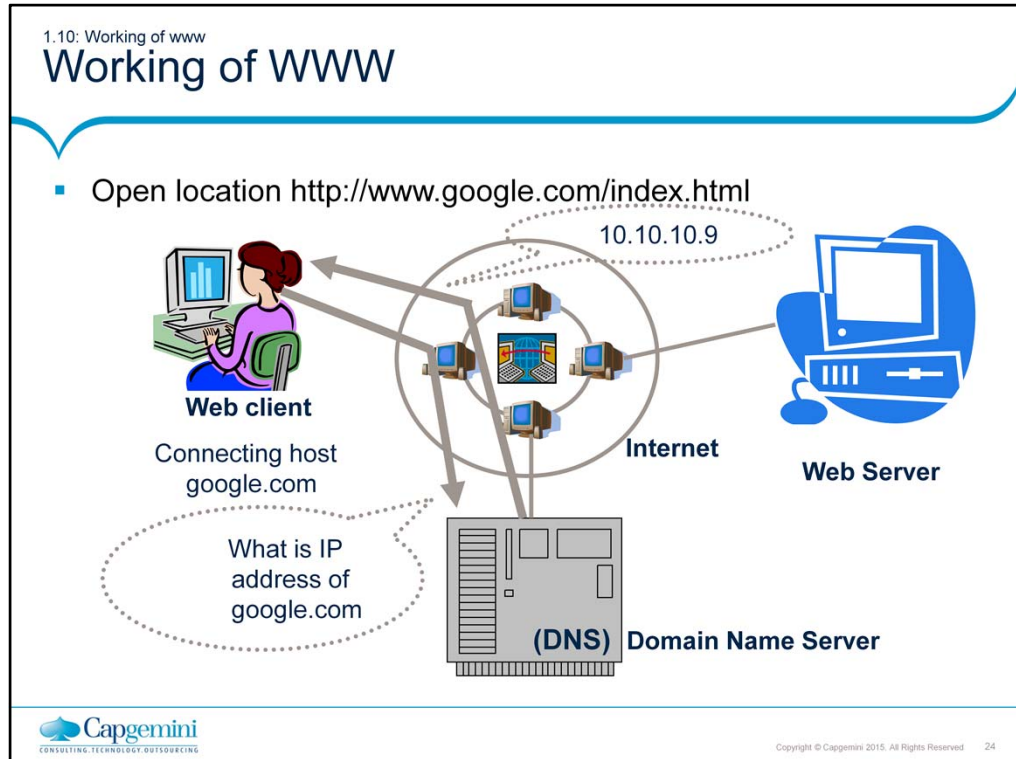
You can apply SSL certificates against a website via the web server. First you need to generate the certificate either by yourself (i.e. using a certificate generator), or by a Certificate Authority (CA). Then, once it has been generated, you apply it to your website via your web server. Applying an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

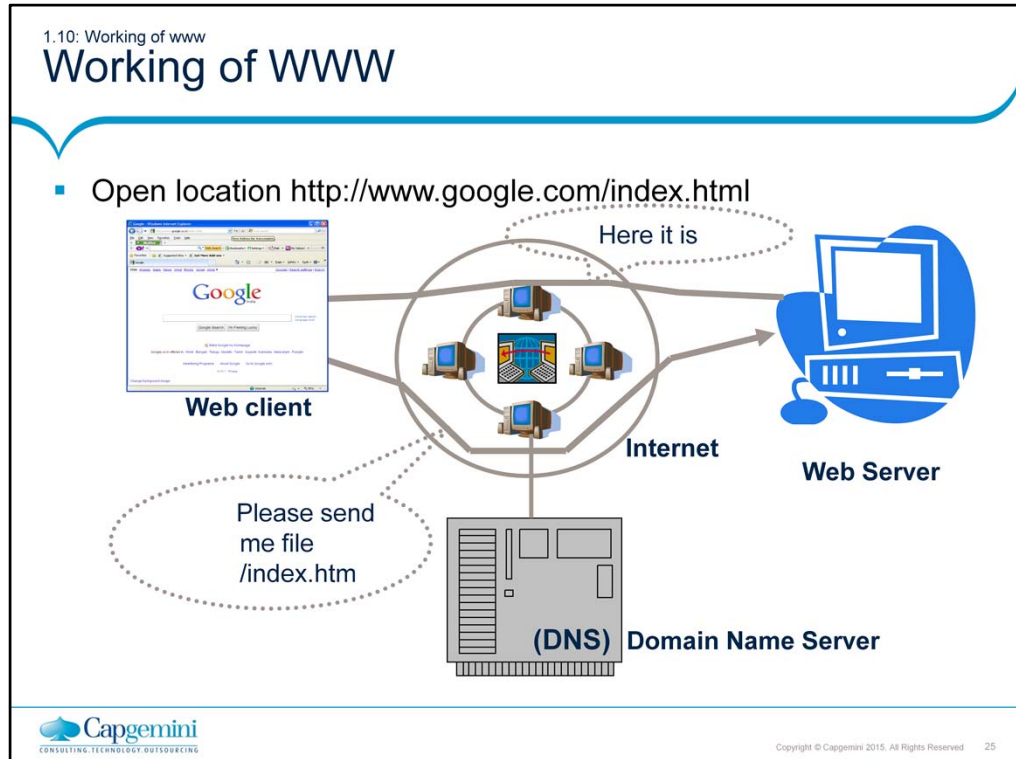
To navigate a website using HTTPS, you simply replace the HTTP with HTTPS at the start of the URL in your browsers' location bar ("`https://www.example.com`")



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



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

HTML tags are keywords (tag names) surrounded by angle brackets like <html>

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.

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.



Copyright © Capgemini 2015. All Rights Reserved 28

In Client side scripting, the dynamic behavior occurs within the presentation. The Client-side content is generated on the user's local computer system.

Such web pages use presentation technology called rich interfaced pages. Client-side scripting languages like JavaScript or ActionScript, used for Dynamic HTML (DHTML) and Flash technologies respectively, are frequently used to orchestrate media types (sound, animations, changing text, etc.) of the presentation.

Web pages that need server side scripting are often created with the help of server-side languages such as ASP, ColdFusion, Perl, PHP, Ruby, WebDNA and other languages. These server-side languages often use the Common Gateway Interface (CGI) to produce dynamic web pages. Three notable exceptions are ASP.NET, JSP, and LSP, which reuse CGI concepts in their APIs but actually dispatch all web requests into a shared virtual machine.

Dynamic web pages are often cached when there are few or no changes expected and the page is anticipated to receive considerable amount of web traffic that would create slow load times for the server if it had to generate the pages on the fly for each request.

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).
 - TCP/IP concepts
 - HTTP concepts
- WWW:
 - Webservers
 - Web Browsers
 - Working of WWW
- HTML
 - Static Web Page
 - Dynamic Web Page

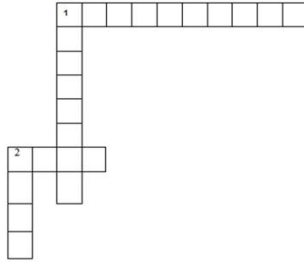


Review Question

- 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



Review Question : Crossword



Clues:

ACROSS

- 1 Client for accessing webpage
- 2 Stateless & connectionless protocol

DOWN

- 1 Set of related web pages
- 2 Markup Language

